CWI
P.O. Box 94079, 1090 GB Amsterdam, The Netherlands
Telephone + 31 - 20 592 9333
Telefax + 31 - 20 592 4199
WWW page http://www.cwi.nl/publications_bibl/

CWI is the nationally funded Dutch institute for research in Mathematics and Computer Science.

Minimax estimation in regression and random censorship models

E.N. Belitser

1991 Mathematics Subject Classification: 62G07 (Curve estimation: nonparametric regression, density estimation, etc.), 62G05 (Estimation), 62G20 (Asymptotic properties).

ISBN 90 6196 488 1
NUGI-code: 811

Contents

Chapter 1

Preliminaries

1.1 Nonparametric minimax estimation

Fisher's works in the 1920s laid the foundations for statistics to become a separate discipline of mathematics. During the last 50 years, a large part of statistics has finally been incorporated into the rigid framework of theoretical mathematics, primarily through the elegant use of measure theoretic concepts. At the same time *estimation theory* has grown, from a mathematical technique established in 1806 with the first publication on the least squares estimator by Legendre, to become an independent topic in statistics.

In probability theory a random phenomenon is described by a probability space $(\Omega, \mathcal{A}, P)$. The measurable space $(\Omega, \mathcal{A})$ gives its qualitative and the measure P its quantitative descriptions respectively. In the theory of probability, the underlying probability space $(\Omega, \mathcal{A}, P)$ is assumed to be predetermined and one studies its properties. In statistics one deals with the converse situation. That is to say, one tries to retrieve certain characteristics of the unknown probability space on the basis of some observed properties.

Observation is one of the fundamental notions in statistics. The observations may be a sample of real valued random variables, a stochastic process or of some other nature obtained as a result of a statistical experiment. The general statistical estimation problem is to gain information about some features of the underlying probability measure, using the observed data.

Mathematically, the observations are usually interpreted as a sample of random elements $X_1, X_2, \cdots, X_n$ from the probability distribution P

on a measurable space $(\mathcal{X}, \mathcal{B})$. Let $\mathcal{X}$ be some metric space and $\mathcal{B}$ be its Borel σ-algebra. Another ingredient of a statistical estimation problem is the following formalization of prior knowledge about the distribution P – one thinks of P as ranging over $\mathcal{P}$, a class of distributions on $(\mathcal{X}, \mathcal{B})$. The class $\mathcal{P}$ is assumed to be known to the statistician and is, in fact, the statistical model. Depending on how "big" the class $\mathcal{P}$ is, one can speak of *parametric* or *nonparametric models*. Recently, a class of *semiparametric models* was recognized as intermediate between parametric and nonparametric models (see van der Vaart (1988), Bickel et al. (1993)).

For a long time parametric modeling has been a subject of investigation. The results that are developed are applied to the problem of fitting probability laws to data. A parametric model is usually described by assuming that the family of distributions $\mathcal{P}$ can be parameterized and represented in the form $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$, where Θ is a subset of a Euclidean space. Thus, the problem of retrieving information about P_θ is equivalent to the problem of retrieving information about parameter θ.

A disadvantage of parametric modeling is that prior information about the underlying probability law is often more vague than any parametric family would allow – parametric families are too specific, or "narrow". Therefore, parametric modeling may in general not be robust in the sense that a slight contamination of the data might lead to wrong conclusions. Moreover, the data might be of such a type that there is no suitable parametric family that gives a good fit. Under these circumstances, nonparametric modeling can serve as a good alternative. Broadly speaking, nonparametric models are those that are characterized only by a qualitative description of the class $\mathcal{P}$. A way to describe a nonparametric model is to assume that $\mathcal{P} = \{\mathcal{P}_\theta, \theta \in \Theta\}$, where Θ is a subset of an infinite dimensional space.

The first paper in the area of nonparametric density estimation is due to Rosenblatt (1956). Since then, a large amount of literature dedicated to methods for estimating infinite dimensional objects – densities, regression functions, spectral densities, distribution functions, failure rates, images etc. – has appeared. It should be mentioned though that the change-over from parametric to nonparametric modeling has produced a side effect – the theory of nonparametric estimation still lacks coherence and generality: "...instead of a single, natural minimax theorem, there is a whole forest of results growing in various and sometimes conflicting directions..." (Donoho et al. (1995)).

Traditionally, two types of statistical estimation problem are recognized: so called *regular* problems and *irregular* (or *nonregular*). By regular problems one understands conventionally problems when one wants to estimate a "smooth" functional of the underlying distribution. Standard examples of regular estimation problems include estimation of the distribution function, mean and median. For other applications, one can refer to, among others, van der Vaart (1988), (1991), Groeneboom and Wellner (1992), Bickel et al. (1993), Groeneboom (1996). A typical feature of regular estimation problems is that $\sqrt{n}$-consistent estimation is possible, where n denotes sample size. Usually the notion of a regular estimation problem is associated with differentiability of the functional that is to be estimated (see Koshevnik and Levit (1976), Levit (1978), Pfanzagl (1982), van der Vaart (1991)). Another beneficial feature of regular models is that a unified and relatively simple treatment of asymptotic lower bounds in estimating a differentiable functional is possible. The construction of asymptotically exact lower bounds for various risks is essentially implemented by classical methods.

By nonregular problems one understands usually all problems of estimating a functional (of the underlying distribution) of interest which is not differentiable. Common examples of nonregular estimation problems are density estimation and regression estimation problems. In contrast to regular models, estimation theory in nonregular models is more complicated and more varied. There is no single general theorem describing lower bounds for the minimax risk (a measure of the complexity of the estimation problem) and optimal estimators. The behaviour of the minimax risk depends strongly on the model and sometimes on the definition of the minimax risk itself. In this book we concentrate on nonregular estimation problems, among which density estimation and regression estimation problems have a significant place in recent research in nonparametric statistics.

The great majority of the estimation problems considered in the literature fits in the following general model of a statistical experiment. Suppose a random element, viewed as an "observation", $X^{(\epsilon)}$ and a family of probability distributions $\{P_\theta^{(\epsilon)}\}$, indexed by a positive number ϵ and the unknown parameter θ belonging in general to an infinite dimensional set Θ, are given. Suppose $X^{(\epsilon)}$ takes its value in a measurable space $(\mathcal{X}^{(\epsilon)}, \mathcal{U}^{(\epsilon)})$. We wish to estimate the value of a functional $f(\theta)$ on the basis of the observed data, where $f : \Theta \to B$, B is a metric space with the distance function $d(\cdot, \cdot)$. In the most general setup one would do this by taking a mapping $f_\epsilon(X^{(\epsilon)})$, $f_\epsilon : \mathcal{X}^{(\epsilon)} \to B$. We are going to study this problem within the minimax framework and to compare estimators on the basis of their risk functions, which implies taking certain expectations. For that, we will need the measurability of $d(f_\epsilon, f)$ as a function of the observations. Let $\mathcal{A}$ be the usual Borel σ-field on R. Any mapping $f_\epsilon : \mathcal{X}^{(\epsilon)} \to B$ such that $d(f_\epsilon, f)$ is $(\mathcal{U}^{(\epsilon)}, \mathcal{A})$-measurable is called an *estimator* of f.

"Solutions of nonasymptotic estimation problems, although an important task for its own sake, can not serve as a subject of sufficiently general mathematical theory" (Ibragimov and Hasminskii (1981)) and we study the estimation problem in the asymptotic setup as $\epsilon \to 0$. In the case of a sample of n independent observations one can assume, for example, $\epsilon = n^{-1/2}$.

The purpose of the rest of this section is to give a simple introduction to the nonparametric minimax estimation problem in a nonregular model. We will not pursue generality anymore and restrict ourselves to the problems of estimating a regression function and a probability density. One can define all the notions below in a general setup if need be.

First we state the problem of density estimation. Given a sequence of independent real valued random variables $Y_1, Y_2, \ldots, Y_n$ identically distributed with common distribution function F such that F has a density f with respect to the Lebesgue measure on the real line, we want to recover f, using the observed data.

Now consider the problem of regression function estimation. Let

$$(X_1, Y_1),\ (X_2, Y_2), \ldots,\ (X_n, Y_n)$$

be a sample of n independent identically distributed pairs of real valued random variables having an unknown distribution. Let a random pair (X, Y) have the same distribution as each of sample pairs and one is interested in the dependence structure between X and Y. The problem of nonparametric regression estimation is to recover the function $f(x) = \mathbf{E}[Y|X = x]$ on the basis of observations

$$(X_1, Y_1),\ (X_2, Y_2), \ldots,\ (X_n, Y_n).$$

Collection $(X_1, \ldots, X_n)$ is called the design. Note that

$$Y_i = f(X_i) + \xi_i, \qquad i = 1, 2, \ldots, n, \tag{1.1}$$

where the ξ_i's are independent random variables such that $\mathbf{E}[\xi_i|X_i] = 0$. If we begin with the model (1.1) and assume further that the joint distribution of the ξ_i's is independent of the joint distribution of the X_i's, the X_i's are independent identically distributed random variables, the ξ_i's are independent identically distributed random variables with zero mean, then this model is called the *additive regression model.* Clearly, this is a particular case of the general regression model above and $f(x)$ is the regression function. Very often a nonrandom design (i.e. $(X_1, \ldots, X_n)$ is a nonrandom vector) on a bounded interval is considered in such models. A common example of nonrandom design is the equidistant one, i.e. the X_i's are nonrandom and equispaced on a bounded interval.

From now on, by the term *curve* we mean either a regression function or a density. The curve f is assumed to belong to a subset Θ of some linear space equipped with a metric $d(\cdot,\cdot)$. Often, Θ is a subset of some normed space. We will use the usual notation $\|\cdot\|$ for this norm. In this case $d(f,g) = \|f - g\|$.

Now we will assign a specific meaning to the rather loosely formulated "to recover the function". Let $w(u)$, $u \in R$, be a *loss function,* i.e. a nonnegative symmetric function nondecreasing on the positive semi-axis, satisfying $\operatorname{ess\,inf} w < \operatorname{ess\,sup} w$. The class of loss functions contains many desirable functions, for instance, bounded loss

$$w(u) = I\{u \geq C\}$$

for some positive constant C, where $I\{S\}$ denotes the indicator function of set S; squared loss

$$w(u) = u^2.$$

As a measure of quality of an estimator $\hat{f}_n$ introduce the *risk function:*

$$R_n(\hat{f}_n, f) = R_n(\hat{f}_n, f, \psi_n, w) = \mathbf{E}_f \left(w(\psi_n^{-1} d(\hat{f}_n, f)) \right).$$

We will call it also just the *risk.* Here $\mathbf{E}_f$ denotes the expectation with respect to the distribution of the observations given that the true curve is $f(\cdot)$ and the positive sequence ψ_n, whose meaning will be clarified later, is a normalizing factor.

If $R_n(\hat{f}_n, f) \leq R_n(\tilde{f}_n, f)$ for all $f \in \Theta$ and $R_n(\hat{f}_n, f) < R_n(\tilde{f}_n, f)$ for at least one $f \in \Theta$, then, as is intuitively clear, we say that the estimator $\hat{f}_n$ is better than $\tilde{f}_n$. Unfortunately, common situation is that $\hat{f}_n$ is better than $\tilde{f}_n$ for some curves $f \in \Theta$ and worse for others, and typically there is no uniformly best estimator in this sense. This suggests finding a scalar characteristic of the risk that depends only on $\hat{f}_n$. It has already become a tradition in statistics to evaluate performance of an estimator by the worst case principle: the quantity

$$\sup_{f \in \Theta} R_n(\hat{f}_n, f)$$

is called the *maximal risk* of the estimator $\hat{f}_n$ over Θ. At this point we incorporated prior information about f $(f \in \Theta)$ into our consideration. The best performance is then the *minimax risk*:

$$r_n(\Theta) = r_n(\Theta, \psi_n, w) = \inf_{\hat{f}_n} \sup_{f \in \Theta} R_n(\hat{f}_n, f),$$

where the infimum is taken over all possible estimators. The minimax risk expresses the least possible mean loss when the worst case happens and, in a way, reflects the complexity of the estimation problem over the class Θ.

There is another point in support of this setting – one would like to derive a nonvoid theory of lower bounds in estimation theory. Indeed, suppose we wish to estimate a density function f at a point x_0, using the quadratic loss function $(\hat{f}_n(x_0) - f(x_0))^2$. Taking the infimum of the risk over all estimators, we have trivially that $\inf_{\hat{f}_n} R_n(\hat{f}_n, f) \geq 0$. On the other hand, this bound is attained by a dummy estimator $\hat{f}_n(x_0) = a$, $a = f(x_0)$, which is absolutely unacceptable for estimating any other $f_1(x_0) \neq a$. Taking the supremum over the class Θ before infimum over all estimators makes the lower bound nontrivial (of course, if the class Θ contains at least two different curves and the loss function is reasonably chosen) and, certainly, estimators as above will not attain this bound. Another way to build up a nonvoid theory of lower bounds is to restrict the class of estimators. For instance, restricting to unbiased estimators in the theory of parametric estimation leads to, under some regularity conditions, the Cramér-Rao bound for the quadratic risk.

Note that the minimax risk still depends on the loss function w. Specific estimation problems are sometimes best treated with different choices of loss function, and the selection of the loss function in such cases is governed by mathematical convenience rather than by any other arguments.

The minimax approach provides a solid basis for nonparametric estimation theory. In the last two decades, Wald's concept of minimaxity

(Wald (1950)) for nonparametric models was developed to fruition by Ibragimov and Hasminskii (1980)–(1991), Stone (1980), (1982), (1984), (1985), Bretagnolle and Huber (1979), Birgé (1983), (1985), (1986), (1987) and many others. In a long series of works by many researchers worldwide, nonparametric minimax estimation theory has reached maturity and a developed machinery for optimality considerations within the minimax framework was built up. The minimax framework consists basically in specifying its three ingredients:

- the model of observations;
- the functional class Θ;
- the loss function and then the minimax risk.

A positive sequence $\psi_n = \psi_n(\Theta)$ is called the *minimax rate of convergence of estimation* if there exist positive constants $C_1 = C_1(\Theta)$, $C_2 = C_2(\Theta)$ and an estimator $\hat{f}_n$ such that

$$\liminf_{n\to\infty} \inf_{\tilde{f}_n} \sup_{f\in\Theta} R_n(\tilde{f}_n, f, C_1\psi_n, w) > \operatorname{ess\,inf} w \tag{1.2}$$

and

$$\limsup_{n\to\infty} \sup_{f\in\Theta} R_n(\hat{f}_n, f, C_2\psi_n, w) < \operatorname{ess\,sup} w\,. \tag{1.3}$$

An estimator $\hat{f}_n$ satisfying (1.3) is called *asymptotically minimax up to rate of convergence*. Frequently we will say just *rate of convergence* or *optimal rate of convergence* instead of "minimax rate of convergence".

As is easy to see, the rate of convergence is not uniquely defined: if ψ_n is a rate of convergence, then $\phi_n\psi_n$ is also a rate of convergence, where ϕ_n is any positive sequence such that

$$0 < \liminf_{n\to\infty} \phi_n \le \limsup_{n\to\infty} \phi_n < \infty.$$

To avoid unnecessary technicalities in definitions, from now on we restrict our consideration to loss functions of the form: $w(x) = |x|^q$, $q > 0$. In such cases, it is convenient to define the risk function as

$$R_n(\hat{f}_n, f) = \mathbf{E}_f \left(d(\hat{f}_n, f)\right)^q.$$

Then a sequence $\psi_n = \psi_n(\Theta)$ is the rate of convergence if there exist positive constants $C_1 = C_1(\Theta)$, $C_2 = C_2(\Theta)$ and an estimator $\hat{f}_n$ such that

$$\liminf_{n\to\infty} \inf_{\tilde{f}_n} \sup_{f\in\Theta} \psi_n^{-q} R_n(\tilde{f}_n, f) \ge C_1 \tag{1.4}$$

and

$$\limsup_{n\to\infty} \sup_{f\in\Theta} \psi_n^{-q} R_n(\hat{f}_n, f) \le C_2 \,. \tag{1.5}$$

The problem of best possible estimation in terms of optimal rates of convergence has been extensively studied (Stone (1980), (1982), Ibragimov and Hasminskii (1980), (1981), (1982), (1983), (1984a), (1984b) (1990), Birgé (1983), (1985) Hall (1989), Donoho and Liu (1991), and others). It is well to bear in mind that the optimal rates of convergence may depend on the risk function.

Different estimators turn out to be optimal (minimax) in the sense of the best rate of convergence in different estimation problems. We mention only some of them: kernel estimators (Härdle and Marron (1985), Ibragimov and Hasminskii (1984a), Korostelev (1994)); projection estimators (Ibragimov and Hasminskii (1981)), nearest neighbour estimators (Stone (1980), (1982)); spline estimators (Nussbaum (1985), Speckman (1985)), wavelet estimators (Donoho et al. (1995), Donoho and Johnstone (1994a)). From the practical point of view, estimators based on stochastic approximation procedures are very important (Belitser and Korostelev (1992), Belitser (1993)).

There are also works in which the optimal rates of convergence are established for dependent observations under some conditions, basically for dependence to be decreasing in some way (see Boente and Fraiman (1989), Truong and Stone (1992)).

The possible quality of estimation, in particular the rate of convergence, essentially depends on the nonparametric class Θ. To provide a reasonable quality of estimation, usually one imposes some uniform smoothness condition on functions from the nonparametric class. For example, it is not possible to estimate consistently a density at a point if this density has a jump at this point. Typically, the definition of class involves a parameter which represents some kind of "smoothness" amount for the functions from this class. The best possible rate of convergence within which one can estimate the unknown curve depends on the "smoothness" in a reasonable way: the "smoother" the class, the better the rate one can estimate with. Usually one considers sufficiently smooth classes so that the minimax rate converges to zero. Standard examples of nonparametric classes described by smoothness conditions include the Sobolev class: for some positive Q and integer $m \ge 1$,

$$W(m,Q) = \left\{ f \in L_2([0,1]) : \int_0^1 (f^{(m)}(x))^2 dx \le Q \right\};$$

the Hölder class: for some positive L, integer $k \geq 0$ and $0 < \beta \leq 1$,

$$\Sigma(\alpha, L) = \left\{ f : |f^{(k)}(x_1) - f^{(k)}(x_2)| \leq L|x_1 - x_2|^{\beta},\ x_1, x_2 \in R \right\},$$

where m and $\alpha = k + \beta$ are the smoothness parameters in the Sobolev and Hölder classes respectively.

So far we have restricted our consideration to definitions and general ideas. To elucidate the introduced notions, we give now two examples of estimating a regression function.

Example 1.1. Consider the additive regression model

$$Y_i = f(x_i) + \xi_i, \quad x_i \in [0,1], \quad i = 1, \ldots, n,$$

where observations are taken at points

$$x_i = x_{in} = \frac{i-1}{n-1}, \qquad i = 1, \ldots, n,$$

$\xi_1, \ldots, \xi_n$ are independent standard (with zero mean and unit variance) Gaussian random variables, the unknown regression function f is assumed to belong to a mth-order Sobolev class

$$W(m, Q) = \left\{ f \in L_2([0,1]) : \int_0^1 (f^{(m)}(x))^2 dx \leq Q \right\}$$

for some positive Q and integer $m \geq 1$. Let $\|\cdot\|$ denote the usual norm in $L_2([0,1])$. As a risk function, we take the so called quadratic risk

$$R_n(\tilde{f}_n, f) = \mathbf{E}_f \|\hat{f}_n - f\|^2.$$

As is shown in Nussbaum (1985), the minimax rate of convergence in this problem is the following:

$$\psi_n = n^{-\frac{m}{2m+1}}.$$

Example 1.2. Suppose we want to estimate a nonparametric regression function $f(x)$, $0 \leq x \leq 1$, on the basis of the observations

$$Y_i = f(i/n) + \xi_i, \qquad i = 1, \ldots, n,$$

where $\xi_1, \ldots, \xi_n$ are independent Gaussian random variables with zero means and variances σ^2. The regression function is assumed a priori to belong to the Hölder class

$$\Sigma(\alpha, L) = \left\{ f : |f^{(k)}(x_1) - f^{(k)}(x_2)| \leq L|x_1 - x_2|^{\beta},\ x_1, x_2 \in [0,1] \right\},$$

$\alpha = k + \beta$, for some $0 < \beta \leq 1$, $L > 0$, integer $k \geq 0$. The estimation quality is measured by the normalized sup-norm risk

$$R_n(\tilde{f}_n, f, \psi_n, w) = \mathbf{E}_f w \left(\psi_n^{-1} \sup_{0 \leq x \leq 1} |\hat{f}_n(x) - f(x)| \right),$$

where ψ_n is a normalizing factor, the loss function w is continuous and satisfies

$$w(u) \leq w_0(1 + |x|^q)$$

for some positive constants w_0 and q. Then the minimax rate of convergence in this problem is shown (see Korostelev (1994), Donoho (1994)) to be of the form

$$\psi_n = \left(\frac{\log n}{n} \right)^{\frac{\alpha}{2\alpha+1}}.$$

The common optimal estimators are based on the fact that we know the smoothness: the bandwidth for the kernel methods, the number of terms for the orthogonal series method etc. In practice, however, the amount of smoothness (and other parameters describing the class) is never known. So, the problem of finding a data dependent method of choosing unknown parameters – a so called *adaptive* method – is an important task. There are several methods: cross-validation, generalized cross-validation, plug-in. The problem of minimax adaptivity, within rate of convergence, with respect to the degree of smoothness was raised by Stone (1982). Efromovich and Pinsker (1984) were first to solve this problem for Sobolev classes and L_2-norm in the problem of filtering a signal against the background of a Gaussian white noise process. The method is based on adaptive determining optimal damping coefficients in an orthogonal series estimator. Similar technique was applied to adaptive estimating a square integrable probability density in Efromovich (1985). Later their results were improved in Golubev (1987), Golubev and Nussbaum (1992), Oudshoorn (1996). In the problem of nonparametric regression function estimation a procedure adaptive within rate of convergence was proposed in Härdle and Marron (1985); for a survey see Marron (1988), (1989). An interesting and very general adaptation method was developed in Lepski (1990), (1991), (1992), further advances can be found in Lepski and Spokoiny (1996).

Another fruitful approach was recently proposed in Donoho and Johnstone (1994a), (1995), Donoho et al. (1995) (see further references therein). A wavelet shrinkage method is shown to perform well when estimating a function with inhomogeneous smoothness properties. Namely, a $\log n$ loss in the rate of convergence yields a "nearly minimax" adaptive procedure for the whole scale of Besov classes.

Up to this point, we have been concerning asymptotic optimality only in terms of optimal rates of convergence. On the other hand, in many regular estimation problems much stronger results are available: not only is shown that $\sqrt{n}$ is an optimal rate of convergence, but optimal constants are derived.

Recall the definition of the optimal rate of convergence (1.4) and (1.5). Take now one representative ψ_n of possible minimax rates of convergence. For this fixed convergence rate, we can define the asymptotically best constants in (1.4) and (1.5), i.e. such $C_l = C_l(\Theta)$ and $C_u = C_u(\Theta)$ that

$$\liminf_{n\to\infty} \psi_n^{-q} r_n(\Theta) = C_l \tag{1.6}$$

and

$$\limsup_{n\to\infty} \psi_n^{-q} r_n(\Theta) = C_u, \tag{1.7}$$

where $r_n(\Theta)$ is the minimax risk over the class Θ. The case $C_l = C_u = C_o$ is of a particular interest. The constant $C_o = C_o(\Theta)$ is optimal in the minimax sense for the given nonparametric class Θ. The corresponding estimator $\hat{f}_n$ from (1.5) attaining C_o should be naturally called *asymptotically minimax.* Then the minimax risk has the following simple asymptotic expression:

$$r_n(\Theta) = C_o(\Theta)\psi_n^q(\Theta)(1 + o(1)) \qquad \text{as} \quad n \to \infty.$$

Clearly, one would like to strengthen the optimal rate results by finding the optimal constants when they exist. A further challenging problem is to describe the asymptotics of the normalized minimax risk

$$\psi_n^{-q} r_n(\Theta) = \inf_{\hat{f}_n} \sup_{f\in\Theta} \psi_n^{-q} R_n(\hat{f}_n, f)$$

when $C_l < C_u$. The following general definition of the asymptotically minimax estimator covers this case as well. An estimator $\hat{f}_n$ is called asymptotically minimax (or just *minimax*) if

$$\limsup_{n\to\infty} \psi_n^{-q} \left(\sup_{f\in\Theta} R_n(\hat{f}_n, f) - r_n(\Theta) \right) = 0 \qquad \text{as} \quad n \to \infty,$$

where ψ_n is the minimax rate of convergence.

In the following two examples it has been possible to describe the exact asymptotics of the minimax risk.

Example 1.1 (continuation). In fact, the result of Nussbaum (1985) describes also the optimal constant:

$$\lim_{n\to\infty} \inf_{\tilde{f}_n} \sup_{f\in W(m,Q)} n^{\frac{2m}{2m+1}} \mathbf{E}_f \|\hat{f}_n - f\|^2 = \gamma(m,Q),$$

where

$$\gamma(m,Q) = (Q(2m+1))^{\frac{1}{2m+1}} \left(\frac{m}{\pi(m+1)}\right)^{\frac{2m}{2m+1}}$$

is Pinsker's constant.

Example 1.2 (continuation). For simplicity sake, consider $k = 0$ in the definition of the class $\Sigma(\alpha, L)$, i.e. $\alpha = \beta$. The following result is due to Korostelev (1994):

$$\lim_{n\to\infty} \inf_{\tilde{f}_n} \sup_{f\in\Sigma(\alpha,L)} \mathbf{E}_f w\left(\left(\frac{n}{\log n}\right)^{\frac{\alpha}{2\alpha+1}} \sup_{0\le x\le 1} |\hat{f}_n(x) - f(x)|\right) = w(C_0),$$

where the optimal constant

$$C_0 = \left(L\sigma^{2\alpha}\left(\frac{\alpha+1}{2\alpha^2}\right)^{\alpha}\right)^{\frac{1}{2\alpha+1}}.$$

Recall our observation that the minimax risk can serve, in some sense, as a measure of complexity of estimation problem for the whole class Θ. In practice, however, only one curve is in the background. This gives rise to the question how to characterize the difficulty of estimation problem contributed solely by this particular curve. A natural way to do this is to localize the risk function. To be more precise, let $\mathcal{T}$ be a topology on the class Θ. Introduce the *local minimax risk*: for a neighbourhood $V \in \mathcal{T}$,

$$r_n(V) = \inf_{\hat{f}_n} \sup_{f\in V} R_n(\hat{f}_n, f).$$

Suppose now that ψ_n is the rate of convergence, $V = V(f_0)$ is a "sufficiently small" neighbourhood of some curve f_0. Then one is interested in the asymptotic behaviour of $\psi_n^{-q} r_n(V)$ and finding an estimator achieving this behaviour asymptotically. Precisely, an estimator $\hat{f}_n$ is called *locally asymptotically minimax* (or just locally minimax) if for each curve $f \in \Theta$ there exists a neighbourhood $V_0 \ni f$ such that for any neighbourhood V, $f \in V \subseteq V_0$ (this is a formal way to say: for any sufficiently small neighbourhood of f),

$$\limsup_{n\to\infty} \psi_n^{-q} \left(\sup_{f\in V} R_n(\hat{f}_n, f) - r_n(V) \right) = 0.$$

Sometimes we will call such an estimator *efficient.* If the limit

$$\lim_{V\downarrow f_0} \lim_{n\to\infty} \psi_n^{-q} r_n(V) = C_o(f_0)$$

exists, then, the constant $C_o(f_0)$, together with the rate of convergence ψ_n, describes the exact behavior of the local minimax risk and represents the difficulty of the estimation problem at the point f_0.

1.2 Minimax estimation: a brief survey

In this section we give a brief review on results in nonparametric statistics concerning the exact asymptotics of the minimax risk. In a number of nonparametric estimation problems, recently it has been possible to improve the results on best obtainable rates of convergence by finding the exact asymptotic value of the minimax risk in the class of all estimators. Presently, the problem of finding the minimax optimal constants is of a great interest.

As far as we know results of such type were obtained only in a limited number of works. Nevertheless, this does not pretend to be an exhaustive account of all works concerning the exact asymptotics of the minimax risk, but rather a collection of observations with emphasis on some relevant aspects of nonparametric estimation problems.

Until recently, this kind of minimax estimation problem seemed remote. However, a solution was found by Pinsker (1980) for a filtering problem over ellipsoids in Hilbert space. The essence of Pinsker's method consists in showing that minimax linear estimators over an ellipsoid are asymptotically minimax in the class of all estimators. This technique was extended to estimation of a square integrable probability density in Efromovich and Pinsker (1982). For the Gaussian white noise model, the result of Pinsker was generalized to the case of regression functions with singularities in Oudshoorn (1996). Namely, the unknown regression function from a Sobolev-type ellipsoid was allowed to have a finite but unknown number of jumps. It was shown that properly normalized minimax quadratic risk attains asymptotically Pinsker's constant. A generalization of Pinsker's result to general losses can be found in Tsybakov (1997).

The works of Nussbaum (1985), Speckman (1985), Golubev (1991), Golubev (1992), Golubev and Nussbaum (1990), Golubev and Nussbaum (1992) are dedicated to the regression estimation problem and concern minimax risk with squared error loss and smoothness assumed in an L_2-Sobolev sense.

In the paper of Speckman (1985) the class of estimators is restricted to the linear ones. The best linear estimator is derived and the exact minimax rate of convergence is obtained. This minimax estimator is a variant of spline smoothing. Some practical aspects have been considered. For instance, in case the variance of the errors σ^2 and the constant Q specifying the Sobolev class are unknown, generalized cross-validation is shown to give an adaptive estimator which achieves the minimax optimal rate under the additional assumption of normality.

The assumption of normality of the errors was essential in the paper of Nussbaum (1985) (see Example 1.1) where the best possible minimax constant is obtained. The method of estimation is based on the fact that under normality of the errors the minimax linear estimator is asymptotically minimax in the class of all estimators. The proposed optimal estimator is a smoothing spline. The multidimensional case of this problem for the equidistant design has been studied by Golubev (1992).

The problem of adaptation with respect to the variance σ^2, the constant Q and smoothness α was solved by Golubev and Nussbaum (1992). The minimax adaptive estimator in this paper is no longer a smoothing spline and no longer even linear.

With regard to the lower asymptotic risk bound, the result of Nussbaum was extended to the nonnormal case in the papers of Golubev (1991), Golubev and Nussbaum (1990). In Golubev and Nussbaum (1990) a nonequidistant design was considered and the noise distribution was assumed to be unknown and varying in a shrinking Hellinger neighbourhood of some central measure. The pertaining lower asymptotic risk bound is established, based on an analogy with a location model in the independent identically distributed case.

Schipper (1996) considered the minimax estimation problem of a density on the real line, using the mean integrated square error as a risk function. The Sobolev and analytic classes were studied. For both classes kernel estimators proved to be minimax. The L_2-structure of the problem allowed to employ a characteristic function technique in the derivation of the upper bound and the van Trees inequality in establishing the lower bound for the minimax risk.

The results of Korostelev (1994), Donoho (1994), Korostelev and Nussbaum (1995) and Schipper (1997) concern L_∞-loss function and L_∞-smoothness instead of L_2. Hence these results can be viewed as the L_∞-analog of the L_2-results just mentioned.

Korostelev (1994) (see Example 1.2) considered a Gaussian regression model with equidistant design. The unknown regression function was from the Hölder class on the unit interval with the smoothness parameter α, $0 < \alpha \leq 1$. It was shown that the exact asymptotic minimax risk is attained by a certain kernel estimator and minimaxity for the risk $\mathbf{E}w(\psi_n^{-1}\|\hat{f}_n - f\|_\infty)$ does not depend on (reasonably chosen) loss function $w(\cdot)$.

Donoho (1994) generalized the result of Korostelev to the estimation of kth derivatives, $k \geq 0$, of α-Lipschitz regression functions for $\alpha > 1$. Also, it is shown that the constants in the asymptotics of the minimax risk are the same as the constants arising in certain problems of optimal recovery. The paper makes heavy use of ideas of renormalization and optimal recovery. As is mentioned in Donoho (1994), L_∞-loss has special importance in connection with setting fixed-width simultaneous confidence bands for an unknown regression function.

Korostelev and Nussbaum (1995) derived the exact asymptotics of the minimax risk in the density estimation problem, by using an interesting approach based on asymptotic equivalence of their model with the Gaussian white noise model, for the smoothness parameter $\alpha > 1/2$. By using direct methods, the result of Korostelev and Nussbaum (1995) was generalized by Schipper (1997) in the following respects: the loss function is allowed to grow exponentially fast, densities from the class are allowed to be supported on the whole real line and need not to be bounded away from zero, localized version of main result is obtained (the exact asymptotic of the local minimax risk is derived and a locally minimax estimator is proposed).

Another type of results was recently initiated by Golubev and Levit (1996a) where the problem of estimating values of an analytic density or any of its derivatives at a given point is studied. This problem, being in essence nonregular, exhibits a close resemblance with the problem of estimating a smooth functional. Namely, locally minimax estimation of a density and any of its derivatives at a given point is possible. An asymptotically efficient (locally minimax) estimator is proposed. The existence of an efficient density estimator leads to a second order efficient estimator of the distribution function in a related class (cf. also Golubev and Levit (1996b)). This approach has been also extended to the nonparametric regression model, with an equidistant design, in Golubev et al. (1996). In this paper, the unknown regression function is again assumed to belong to a class of functions analytic in a strip of the complex plane around the real axis, and two different types of results are given. Firstly, an asymptotically minimax estimator of the regression function is presented such that its mth derivative is an asymptotically minimax estimator of the mth derivative of the regression function, for a broad class of loss functions. Secondly, the same problem is considered for L_∞-norm on a bounded interval.

1.3 Scope

In the first chapter we have already presented an elementary introduction to nonparametric curve estimation and a short overview of relevant literature. The subsequent chapters constitute the main content. Let us outline the rest of the book.

A regression model with continuous time, the so called Gaussian white noise model, has received much attention in the literature in the last few decades. As well as being of interest on its own, the Gaussian white noise model, under some conditions, can also serve as a prototype for nonparametric regression model and observation model in the problem of density estimation. So, on the one hand the white noise model can be considered as a mathematical idealization, and on the other hand, this model captures the statistical essence of the original model and preserves its traits in a pure form. The problem of signal recovery in Gaussian white noise is recognized to be a "generic" nonparametric curve estimation problem. It turns out namely that many observation models exhibit asymptotic statistical equivalence with this model; for example, in the density estimation (Nussbaum (1996)), regression model (Brown and Low (1996)), in the convolution problem, the second order minimax estimation of the distribution function (Golubev and Levit (1996b)). However, the equivalence notion has mostly been treated informally until recently and the problem of establishing the equivalence in a precise sense is a delicate and in general difficult task. Recently a serious effort was mounted to make the notion of equivalence formal; see Nussbaum (1996), Brown and Low (1996). We will not dwell on this issue, but rather focus on the model itself.

Pinsker (1980) considered the problem of recovering a signal, in Gaussian noise, assuming that the unknown signal belongs to an ellipsoid in a Hilbert space. For a class of ellipsoids satisfying certain regularity conditions, he derived the asymptotic minimax risk and presented a linear estimator which proved to be asymptotically minimax.

In Chapter 2 we consider a generalization of the white noise model, namely, we allow the noise to be not necessarily white. We call such noise "coloured". We impose some regularity conditions on ellipsoids which seem to be more restrictive than those in Pinsker (1980), but our derivation of the main results is shorter and more transparent.

However, the main novelty in that chapter is the derivation of the exact asymptotic behaviour of the second order term of the minimax risk. It is an interesting and challenging task to study the second order asymptotics of the minimax risk in nonparametric nonregular estimation problems, and we are not aware of any results of this kind. The problem of deriving the second order behaviour of the minimax risk is not just of pure mathematical interest – there are estimation problems in which it is the second order behaviour that should be studied. For instance, it arises naturally in the technically involved problem of the second order minimax estimation of the distribution function (cf. Golubev and Levit (1996b)). Indeed, if the series of the coloured noise variances converges (the noise is "small"), then the first order asymptotic behaviour is trivial and corresponds to the parametric (or regular) situation. In such cases, only the asymptotic behaviour of the second order term of the minimax risk reveals the nonparametric nature of the problem. In any case, studying the second order asymptotics enables one to improve the accuracy of the estimation and to make the structure of the optimal estimator more precise when, for example, there is a class of first order optimal estimators, but only a particular choice of the bandwidth for the kernel method or the number of terms for the orthogonal series method makes an estimator second order optimal.

We propose a linear estimator and show that this estimator is asymptotically minimax up to second order term of the minimax risk. To illustrate the application of the main results, we give a list of examples, including the well known Sobolev and analytic classes for different levels of coloured noise.

The next chapter, where we consider the classical additive regression model with equispaced design points, is closely related to the previous one. This problem has been actively investigated by Nussbaum (1985), Speckman (1985), Golubev (1992), Golubev and Nussbaum (1992) for the Sobolev type classes with the smoothness parameter β. In those results the exact asymptotics of the minimax risk was derived in the form $C_o\psi_n^2$, with the optimal rate of convergence, as a rule, $\psi_n = n^{\beta/(2\beta+1)}$ and the optimal constant C_o.

In Chapter 3, we suppose that the unknown regression curve lies in an ellipsoid from $L_2([0,1])$. A distinguishing feature of our main result is that it covers a rather general class of ellipsoids. So, the periodic Sobolev class and the class of periodic analytic functions can be described as certain examples of such ellipsoids. We describe the exact asymptotic behaviour of the minimax risk and derive the asymptotically minimax estimator. The estimator proves to be a windowed Fourier projection estimator, although it can also be represented as a kernel estimator. In addition, we discuss the questions of consistency, local minimaxity, robust estimation and nonnormality of the noises.

Our method relies on approximation of the initial nonparametric model by a sequence of linear models of dimensions increasing with the number of observations. These approximating linear models are intimately related to the observation model from the previous chapter. In regard to the methods for deriving lower bounds, we propose a new approach based on the elementary but rather powerful van Trees inequality. This method allows the exact evaluation of the quadratic minimax risk. Some practical aspects are considered – in several examples, we give the exact formulae of the optimal kernels.

In the last chapter we are concerned with one of the basic problem in the theory of nonparametric estimation, the density estimation problem, when the observation model is complicated by the presence of censoring. A rich literature is devoted to the random censorship model for various estimation problems. The problem of density estimation under censoring has long been treated in the literature; see a short overview in the last section of Chapter 4. The majority of the proposed density estimators in the random censorship model is based on the well known Kaplan-Meier estimator, an estimator of the distribution function. Many interesting aspects of the problem were investigated, (among them, the rates of convergence of certain estimators) and all these studies led to a better understanding of risk computations for Kaplan-Meier based estimators. The issue of optimality of considered estimators with respect to the rate of convergence remained open. In a recent paper of Liu (1996), a Kaplan-Meier based kernel estimator was shown, under some conditions, to be of optimal rate. In Chapter 4 we explore, in the density estimation problem, the innovative combination of two concepts: the optimality considerations (the minimax approach) and the random censorship model.

In recent work, Golubev and Levit (1996a) considered the problem of estimating an analytic density in the model of independent identically distributed observations and discovered an interesting phenomenon. It turned out that this problem lies on the boundary between regular and nonregular problems (a so called *regularizable* problem) and one can construct asymptotically unbiased and asymptotically efficient estimators of the density at a point, with a convergence rate only slightly worse than $\sqrt{n}$. So far, there has been no results, to the best of our knowledge, in models with a loss of information (or any other than the model of independent identically distributed observations), where a certain density estimator is found to be minimax.

Instead of the global minimax risk, we employ the local minimax risk as a measure of the quality of an estimator, which yields more exact results. We consider both an analytic and a more general class of C^∞ densities. We call this class "infinitely smooth". We describe the asymptotic behaviour of the local minimax risk and propose an efficient (locally asymptotically minimax) estimator – an integral of a properly chosen kernel with respect to the Kaplan-Meier estimator. The methods for the two classes we consider proved to be essentially different. Namely, there are no efficient estimators with compactly supported kernels in the case of analytic functions while in the case of C^∞ densities such estimators do exist. This makes possible applying strong approximation results for the Kaplan-Meier estimator in the latter case. For the class of analytic functions, the risk computations are more cumbersome and relies heavily on the martingale technique.

We propose a wide class of kernels on which the estimator can be based, which turns out to be important in the estimation problem with censored observations. Certainly, it is desirable to have kernels with the lightest possible tails to reduce the influence of the censoring. It is shown that one can take a compactly supported kernel for the infinitely smooth class and a kernel with tails decreasing as exponent of any polynomial for the class of analytic functions.

A certain useful tool, the Approximation Lemma (see Section 4.5), plays an important part in the proofs of the main results. It reflects the fact that any density from either of the two considered nonparametric classes can be approximated by a sequence of smooth functionals with a negligible approximation error, thus linking our problem with a regular estimation problem. The treatment of the lower bound based on the van Trees inequality is in essence similar to that in Golubev and Levit (1996a), but is more involved since one has to take into account the censoring mechanism.

Chapter 2

Minimax filtering over ellipsoids

Suppose a signal $f(t)$ is transmitted over a communication channel with Gaussian white noise of intensity ϵ^2 during the time interval $[0, T]$. The statistical estimation problem is to recover the signal $f(t)$, based on the observation $X_\epsilon(t)$, $0 \le t \le T$:

$$dX_\epsilon(t) = f(t)dt + \epsilon\, dW(t), \quad 0 \le t \le T, \tag{2.1}$$

where $f(t)$ is an unknown function and assumed to belong to a known set $\Theta \subset L_2([0,T])$, $W(t)$ is a standard Wiener process. In this chapter we will measure the quality of an estimator $\hat{f}(t)$ by its global squared L_2-norm risk:

$$\mathbf{E}\|\hat{f} - f\|^2,$$

where $\|\cdot\|$ denotes the usual $L_2([0,T])$-norm throughout this chapter.

Under the assumption that the unknown signal belongs to the Hilbert space $L_2([0,T])$, we can reduce this problem to the problem of estimating an infinite dimensional parameter $\theta \in l_2$. Indeed, let $\{\phi_k,\ k = 1, 2, \ldots\}$ be an orthonormal basis in $L_2([0,T])$. Then

$$Y_k = \theta_k + \epsilon\, \xi_k, \quad k = 1, 2, \ldots,$$

where

$$Y_k = \int_0^T \phi_k(t) dX_\epsilon(t), \tag{2.2}$$

$$\theta_k = \int_0^T \phi_k(t) f(t) dt, \tag{2.3}$$

$$\xi_k = \int_0^T \phi_k(t)dW(t).$$

So, the ξ_k's are independent standard Gaussian random variables.

In this chapter we consider a generalization of the model (2.1): noise is not necessarily white. We will call such noise "coloured".

2.1 "Coloured" Gaussian noise model

First of all, we assume without loss of generality that $T = 1$. Consider now the following generalization of the model (2.1):

$$dX_\epsilon(t) = f(t)dt + \epsilon\, g(t)dW(t), \tag{2.4}$$

where $g(t)$ is some known bounded function. In the same way as in the previous section we write the equivalent model:

$$Y_k = \theta_k + \epsilon\, \sigma_k \xi_k \quad k = 1, 2, \dots\ , \tag{2.5}$$

where the Y_k's, θ_k's are defined by (2.2) and (2.3) respectively, $\{\phi_k\}$ is an orthonormal basis with the weight function $g^2(t)$, the ξ_k's are standard Gaussian random variables and

$$\sigma_k = \left(\int_0^1 \phi_k^2(t) g^2(t) dt \right)^{1/2}.$$

Another generalization of the model (2.1):

$$dX_\epsilon(t) = f(t)dt + \epsilon\, dW(g(t)), \tag{2.6}$$

where $g(t)$ is assumed to be nonnegative, nondecreasing and differentiable. The model (2.5) holds again with Y_k's, θ_k's defined by (2.2) and (2.3) respectively, $\{\phi_k\}$ is an orthonormal basis with the weight function $g'(t)$, ξ_k's are standard Gaussian random variables and

$$\sigma_k = \left(\int_0^1 \phi_k^2(t) dg(t) \right)^{1/2}.$$

Let us consider one more generalization of the model (2.1):

$$dX_\epsilon(t) = (f * g)(t)dt + \epsilon\, dW(t), \tag{2.7}$$

where $*$ denotes the convolution operation, $g(t)$ is some known function such that all its Fourier coefficients with respect to the trigonometric basis are nonzero. Then:

$$Y_k = \theta_k g_k + \epsilon\, \xi_k, \quad k = 1, 2, \dots ,$$

where Y_k's, θ_k's are defined by (2.2) and (2.3) respectively, g_k's are the Fourier coefficients of function g and ξ_k's are independent standard Gaussian random variables. Since coefficients g_k's are assumed to be nonzero, we can reduce this model to model (2.5) by dividing the equality above by g_k.

We could also consider the usual trigonometric basis in the first two examples, but then we would have nonzero covariances between the Gaussian random variables ξ_k's. If the covariance Σ is an operator in l_2 with bounded inverse, then we can transform the observations via $Y' = \Sigma^{-1/2} Y$, $Y = (Y_1, Y_2, \dots)$, giving new data $Y'_k = \theta'_k + \epsilon \xi'_k$, where now ξ'_k's are independent.

From now on we study the observation model (2.5). Here σ_k 's, $\sigma_k \geq 0$, $k = 1, 2, \dots$, are given, ξ_k's are independent standard Gaussian random variables, $\epsilon > 0$ is a small parameter. The unknown infinite-dimensional parameter of interest $\theta = (\theta_1, \theta_2, \dots)$ is assumed to lie in an l_2-ellipsoid Θ:

$$\Theta = \Theta(Q) = \{\theta : \sum_{k=1}^{\infty} a_k^2 \theta_k^2 \leq Q\}, \tag{2.8}$$

where $(a_k,\ k = 1, 2, \dots)$ is a nonnegative sequence converging to infinity. Without loss of generality we let the sequence $(a_k,\ k = 1, 2, \dots)$ be strictly positive and monotone.

The model (2.5), (2.8) was first studied by Pinsker. He considered the problem of recovering a signal from an ellipsoid in Gaussian noise with spectral density ϵ^2, as $\epsilon \to 0$, which can be described, in equivalent terms, by (2.5), (2.8). In this chapter, developing further the approach of Pinsker (1980), we describe the second order behaviour of the minimax estimators and the quadratic minimax risk for the model (2.5), (2.8). These results are illustrated by a number of examples.

2.2 Minimax linear estimation

Let the model of observations be given by (2.5).

In this section we investigate the minimax linear risk which will be shown later to be asymptotically equal, under some conditions, to the minimax risk. Denote $x = (x_1, x_2, \dots)$ and introduce the class of linear estimators:

$$\hat{\theta} = \hat{\theta}(x) = (\hat{\theta}_1, \hat{\theta}_2, \dots), \quad \hat{\theta}_k = x_k Y_k, \quad k = 1, 2, \dots . \tag{2.9}$$

Define the risk of a linear estimator

$$R_\epsilon(x, \theta) = \mathbf{E}_\theta \|\hat{\theta}(x) - \theta\|^2 \tag{2.10}$$

and the minimax linear risk

$$r_\epsilon^l = r_\epsilon^l(\Theta) = \inf_x \sup_{\theta \in \Theta} R_\epsilon(x, \theta), \tag{2.11}$$

where $\|\theta\|^2 = \sum_{k=1}^\infty \theta_k^2$.

To formulate the result about the minimax linear risk, we introduce some notations. Let c_ϵ be a solution of the equation

$$\epsilon^2 \sum_{k=1}^\infty \sigma_k^2 a_k (1 - c a_k)_+ = cQ \tag{2.12}$$

and

$$d_\epsilon = d_\epsilon(\Theta) = \epsilon^2 \sum_{k=1}^\infty \sigma_k^2 (1 - c_\epsilon a_k)_+ . \tag{2.13}$$

Here b_+ denotes nonnegative part of b.

The following theorem describes the minimax linear risk.

Theorem 2.1. *Let c_ϵ and d_ϵ be defined by (2.12) and (2.13). Then*

$$\inf_x \sup_{\theta \in \Theta} R_\epsilon(x, \theta) = \sup_{\theta \in \Theta} \inf_x R_\epsilon(x, \theta); \tag{2.14}$$

the saddle point $(\tilde{x}, \tilde{\theta})$ for the problem (2.11) is given by

$$\tilde{x}_k = (1 - c_\epsilon a_k)_+ , \tag{2.15}$$

$$\tilde{\theta}_k^2 = \frac{\epsilon^2 \sigma_k^2 (1 - c_\epsilon a_k)_+}{c_\epsilon a_k} \tag{2.16}$$

and the linear minimax risk satisfies the following equations:

$$r_\epsilon^l = d_\epsilon = \sup_{\theta \in \Theta} \epsilon^2 \sum_{k=1}^\infty \frac{\sigma_k^2 \theta_k^2}{\theta_k^2 + \epsilon^2 \sigma_k^2}. \tag{2.17}$$

Proof. First note that the risk of a linear estimator has the form

$$R_\epsilon(x,\theta) = \sum_{k=1}^{\infty} \left(\epsilon^2\sigma_k^2 x_k^2 + (1-x_k)^2\theta_k^2\right). \tag{2.18}$$

Since, according to (2.12),

$$Qc_\epsilon^2 = \epsilon^2 \sum_{k=1}^{\infty} \sigma_k^2 c_\epsilon a_k (1-c_\epsilon a_k)_+,$$

by (2.18) we have

$$\begin{aligned}
\inf_x \sup_{\theta\in\Theta} R_\epsilon(x,\theta) &\le \sup_{\theta\in\Theta} R_\epsilon(\tilde{x},\theta) \\
&\le Q \sup_{k\ge 1} (1-\tilde{x}_k)^2/a_k^2 + \sum_{k=1}^{\infty} \epsilon^2\sigma_k^2\tilde{x}_k^2 \\
&\le Qc_\epsilon^2 + \epsilon^2 \sum_{k=1}^{\infty} \sigma_k^2 (1-c_\epsilon a_k)_+^2 \\
&= \epsilon^2 \sum_{k=1}^{\infty} \sigma_k^2((c_\epsilon a_k(1-c_\epsilon a_k)_+ + (1-c_\epsilon a_k)_+^2)) \\
&= \epsilon^2 \sum_{k=1}^{\infty} \sigma_k^2 (1-c_\epsilon a_k)_+ = d_\epsilon .
\end{aligned} \tag{2.19}$$

Note now that the equation (2.12) can be also rewritten as

$$\sum_{k=1}^{\infty} a_k^2\tilde{\theta}_k^2 = Q,$$

so that $\tilde{\theta} \in \Theta$. Taking into account this and (2.19), we obtain

$$\begin{aligned}
d_\epsilon &\ge \sup_{\theta\in\Theta} R_\epsilon(\tilde{x},\theta) \ge \inf_x \sup_{\theta\in\Theta} R_\epsilon(x,\theta) \\
&\ge \sup_{\theta\in\Theta} \inf_x R_\epsilon(x,\theta) = \sup_{\theta\in\Theta} \epsilon^2 \sum_{k=1}^{\infty} \frac{\sigma_k^2\theta_k^2}{\theta_k^2+\epsilon^2\sigma_k^2} \\
&\ge \epsilon^2 \sum_{k=1}^{\infty} \frac{\sigma_k^2\tilde{\theta}_k^2}{\tilde{\theta}_k^2+\epsilon^2\sigma_k^2} = d_\epsilon ,
\end{aligned}$$

which completes the proof of the theorem. □

Remark 2.1. As is indicated in the proof of the theorem, the equation (2.12) can be also rewritten as

$$\sum_{k=1}^{\infty} a_k^2 \tilde{\theta}_k^2 = Q. \tag{2.20}$$

Remark 2.2. Due to monotonicity of $(a_k,\ k = 1, 2, \ldots)$,

$$d_\epsilon = \epsilon^2 \sum_{k=1}^{N} \sigma_k^2 (1 - c_\epsilon a_k),$$

where

$$N = N_\epsilon(\Theta) = \max\{k\colon\ a_k \le c_\epsilon^{-1}\}\,. \tag{2.21}$$

One can easily derive explicit formulae for c_ϵ and N:

$$c_\epsilon = c_\epsilon(\Theta) = \frac{\sum_{k=1}^{N} \sigma_k^2 a_k}{Q\epsilon^{-2} + \sum_{k=1}^{N} \sigma_k^2 a_k^2}\,,$$

$$N = \max\left\{ l:\ \epsilon^2 \sum_{k=1}^{l} \sigma_k^2 a_k (a_l - a_k) \le Q \right\}.$$

Note that (2.21) entails that

$$0 \le c_\epsilon a_k \le 1, \quad k = 1, 2, \ldots, N\ . \tag{2.22}$$

Remark 2.3. The assumption of strict positivity of the sequence $\{a_k\}$ does not restrict the generality. Indeed, suppose $a_k = 0$ for some k then the relations (2.14) and (2.17) still hold, whereas the saddle point does not exist: corresponding $\tilde{x}_k = 1$ and $\tilde{\theta}_k^2 = \infty$. In this case (2.14) and (2.17) would follow from the inequalities:

$$\begin{aligned} d_\epsilon &\ge \sup_{\theta\in\Theta} R_\epsilon(\tilde{x}, \theta) \ge \inf_x \sup_{\theta\in\Theta} R_\epsilon(x, \theta) \\ &\ge \sup_{\theta\in\Theta} \inf_x R_\epsilon(x, \theta) = d_\epsilon\,, \end{aligned}$$

where $\tilde{x}_k$'s are defined by (2.15).

Remark 2.4. The proof of Theorem 2.1 is not constructive. A constructive version of this assertion is given in Section A.1, where we elucidate how one can derive the saddle point $(\tilde{x}, \tilde{\theta})$. The equations (2.12), (2.15) and (2.16) can also be obtained by the Lagrange multiplier method for a problem of maximizing a functional

$$\sum_{k=1}^{\infty} \frac{\epsilon^2 \sigma_k^2 \theta_k^2}{\theta_k^2 + \epsilon^2 \sigma_k^2}$$

subject to the convex constraint (2.8).

2.3 Asymptotically minimax estimation

In this section we investigate the asymptotic behaviour of the minimax risk with respect to all possible estimators.

We define the minimax risk:

$$r_\epsilon = r_\epsilon(\Theta) = \inf_{\hat{\theta}} \sup_{\theta \in \Theta} \mathbf{E}_\theta \|\hat{\theta} - \theta\|^2, \tag{2.23}$$

where $\hat{\theta}$ is an arbitrary estimator based on $Y = (Y_1, Y_2, \ldots)$.

In the proofs of lower bounds we use the van Trees inequality (see van Trees (1968), p. 72). Now we describe the version of this inequality which we use below. Let $dP_\theta(y)$, $y = (y_1, y_2, \ldots)$, denote the distribution of the vector of observations $Y = (Y_1, Y_2, \ldots)$ in (2.5) and $\varphi(y_k, \theta_k)$ be the marginal (Gaussian) density of Y_k. Assume that a prior distribution $d\Lambda(\theta)$, $\theta = (\theta_1, \theta_2, \ldots)$, is defined, according to which the θ_k are independent random variables, with corresponding densities $\nu_k(x)$. Let, for all k, $\nu_k(x)$ be absolutely continuous, with finite Fisher information

$$I(\nu_k) = \int \left(\frac{\partial \log \nu_k(x)}{\partial x} \right)^2 \nu_k(x)\, dx.$$

Assume also that $\nu_k(x)$ is positive inside a bounded interval of the real line and zero outside it.

We write $\mathbf{E}$ for the expectation with respect to the joint distribution of Y and θ. Then, according to the van Trees inequality (cf. Gill and Levit (1995) and Golubev and Levit (1996b)), the Bayes risk $\mathbf{E}(\hat{\theta}_k - \theta_k)^2$ admits a lower bound:

$$\mathbf{E}(\hat{\theta}_k - \theta_k)^2 \geq \frac{1}{I_k + I(\nu_k)}, \tag{2.24}$$

where $I_k = \epsilon^{-2} \sigma_k^{-2}$ is the Fisher information about θ_k contained in the observation Y_k and $\hat{\theta}_k = \hat{\theta}_k(Y)$.

Since our setup here is slightly different from those of Theorem A.1, Gill and Levit (1995) and Golubev and Levit (1996b), below we sketch a short proof of (2.24). Let

$$\begin{aligned} B &= \frac{\partial}{\partial \theta_k} \log \left(\varphi(Y_k, \theta_k)\nu_k(\theta_k)\right) , \\ A &= \hat{\theta}_k - \theta_k . \end{aligned}$$

Denote

$$\begin{aligned} Y^{(k)} &= (Y_1, \ldots, Y_{k-1}, Y_{k+1}, \ldots) , \\ \theta^{(k)} &= (\theta_1, \ldots, \theta_{k-1}, \theta_{k+1}, \ldots) \end{aligned}$$

and let $dP_\theta^{(k)}(y^{(k)})$ and $d\Lambda^{(k)}(\theta^{(k)})$ respectively be their distributions.

Use the Cauchy-Schwarz inequality $\mathbf{E}A^2 \geq (\mathbf{E}AB)^2/\mathbf{E}B^2$. One can assume, without loss of generality, that $\mathbf{E}A^2 < \infty$. Our assumptions permit integration by parts:

$$\begin{aligned} \int \theta_k \left(\varphi(y_k, \theta_k)\nu_k(\theta_k)\right)' d\theta_k &= \theta_k^r \varphi(y_k, \theta_k^r)\nu_k(\theta_k^r) - \theta_k^l \varphi(y_k, \theta_k^l)\nu_k(\theta_k^l) \\ &\quad - \int \varphi(y_k, \theta_k)\nu_k(\theta_k)\, d\theta_k \\ &= - \int \varphi(y_k, \theta_k)\nu_k(\theta_k)\, d\theta_k , \end{aligned}$$

where θ_k^r and θ_k^l are the right and left endpoints of the support of ν_k, respectively. Moreover,

$$\int \left(\varphi(y_k, \theta_k)\nu_k(\theta_k)\right)' \, d\theta_k = 0 .$$

Therefore, interchanging the order of integration in the following integral yields

$$\begin{aligned} \mathbf{E}AB &= \int (\hat{\theta}_k - \theta_k) \frac{\partial}{\partial \theta_k} \log \left(\varphi(y_k, \theta_k)\nu_k(\theta_k)\right) \, dP_\theta(y) \, d\Lambda(\theta) \\ &= \int (\hat{\theta}_k - \theta_k) \left(\varphi(y_k, \theta_k)\nu_k(\theta_k)\right)' \, dy_k \, d\theta_k dP_\theta^{(k)}(y^{(k)}) \, d\Lambda^{(k)}(\theta^{(k)}) \\ &= - \int \theta_k \left(\varphi(y_k, \theta_k)\nu_k(\theta_k)\right)' \, d\theta_k \, dy_k dP_\theta^{(k)}(y^{(k)}) \, d\Lambda^{(k)}(\theta^{(k)}) \\ &= \int dP_\theta(y) \, d\Lambda(\theta) = 1. \end{aligned}$$

It remains to note that $\mathbf{E}B^2 = I_k + I(\nu_k)$.

The next theorem describes the lower bound for the minimax risk. The proof of this and the following results of this section will be given in the last section of this chapter.

Theorem 2.2. *Let* $(m_k,\ k = 1, 2, \ldots)$, $m_k = m_k(\epsilon)$, *be a nonnegative sequence satisfying the condition*

$$\sum_{k=1}^{\infty} a_k^2 m_k^2 + \left(\beta_\epsilon \left(\log \gamma_\epsilon^{-1}\right) \sum_{k=1}^{\infty} a_k^4 m_k^4\right)^{1/2} \leq Q\,, \tag{2.25}$$

for some positive functions γ_ϵ *and* β_ϵ *such that*

$$\lim_{\epsilon \to 0} \beta_\epsilon = \infty\,, \qquad \lim_{\epsilon \to 0} \gamma_\epsilon = 0\,.$$

Then the following asymptotic lower bound holds: for any positive fixed constant α,

$$r_\epsilon \geq \sum_{k=1}^{\infty} \frac{\epsilon^2 \sigma_k^2 m_k^2}{m_k^2 + \epsilon^2 \sigma_k^2}(1 + o(1)) + O\left(\gamma_\epsilon^\alpha\right)\,, \qquad \epsilon \to 0\,, \tag{2.26}$$

where the minimax risk r_ϵ *is defined by (2.23).*

To derive a good lower bound, one should, in principle, maximize the functional appearing in (2.26) under the restriction (2.25). One can show that, under a rather mild condition, this problem is asymptotically equivalent to the maximization problem (2.17) which has already been solved by Theorem 2.1. This implies, in particular, the asymptotic equivalence of the minimax risk and the minimax linear risk. However, this does not give the second order behaviour of the minimax risk since it is not specified how $o(1) \to 0$ as $\epsilon \to 0$ (see Remark 2.11). The following theorem is a version of the previous one, refined in the sense of the second order behaviour of the minimax risk.

Theorem 2.3. *Let* c_ϵ *and* $N = N_\epsilon$ *be defined by (2.12) and (2.21). If condition*

$$\lim_{\epsilon \to 0} \epsilon^4 c_\epsilon^{-2} \log \epsilon^{-1} \sum_{k=1}^{N} a_k^2 \sigma_k^4 (1 - c_\epsilon a_k)^2 = 0\,, \qquad \epsilon \to 0\,,$$

holds, then

$$r_\epsilon = \epsilon^2 \sum_{k=1}^{N} \sigma_k^2 - \epsilon^2 \left(c_\epsilon \sum_{k=1}^{N} \sigma_k^2 a_k\right)(1 + o(1))\,, \qquad \epsilon \to 0\,,$$

where the minimax risk r_ϵ is defined by (2.23).

The next theorem, although looking quite general, provides exact asymptotics of the minimax risk for a more restricted class of ellipsoids. In particular, it is convenient in applications where the sequence $(a_k,\ k = 1, 2, \ldots)$ is increasing faster than k^m for any $m > 0$. Since in such cases the limiting behaviour of the minimax linear risk d_ϵ typically does not depend on Q (cf. Examples 2.4-2.5 in the next section), this theorem also provides the exact asymptotics of the minimax risk r_ϵ. In the context of curve estimation this corresponds to estimating "very smooth" functions, with rapidly decreasing Fourier coefficients (cf. Golubev and Levit (1996b).)

Theorem 2.4. *Let d_ϵ and r_ϵ be defined by (2.13) and (2.23) respectively. Then*

$$d_\epsilon(\Theta(Q/\pi^2)) \le r_\epsilon \le d_\epsilon(\Theta(Q))\,. \tag{2.27}$$

Corollary 2.1. *Let c_ϵ and N be defined by (2.12) and (2.21). If*

$$\sum_{k=1}^{\infty} \sigma_k^2 = \tau < \infty$$

and

$$c_\epsilon \sum_{k=1}^{N} \sigma_k^2 a_k = o\left(\sum_{k=N+1}^{\infty} \sigma_k^2 \right), \qquad \epsilon \to 0\,,$$

then the following asymptotic expansion of the minimax risk holds:

$$r_\epsilon = \epsilon^2 \tau - \epsilon^2 \left(\sum_{k=N+1}^{\infty} \sigma_k^2 \right)(1 + o(1))\,, \qquad \epsilon \to 0\,.$$

Remark 2.5. There are two terms in the asymptotic expression of the minimax risk in Theorem 2.3. They can be either of the same order or the second term can be of smaller order than the first. In the latter case Theorem 2.3 provides at least two terms of the asymptotic expansion of the minimax risk (cf. Example 2.1 below). As is easy to see, in Corollary 2.1 the asymptotic expansion of the minimax risk always gives both the first and the second order terms.

Remark 2.6. Note that the lower and upper bounds for the minimax risk in Theorem 2.4 are nonasymptotic.

Remark 2.7. If the series of the coloured noise variances converges, as in Corollary 2.1, then the first order term of the minimax risk is trivially $\epsilon^2 \sum_{k=1}^{\infty} \sigma_k^2$, i.e. it is the same as in the finite dimensional counterparts. In such cases, which might be called "subparametric", only the second order term of the minimax risk reveals the true nonparametric nature of the problem.

Remark 2.8. Recall that the sequence $(a_k,\ k = 1, 2, \ldots)$ was assumed positive. The results remain valid under the weaker assumption: $a_k \geq 0$, $k = 1, 2, \ldots$.

2.4 Examples

The results presented below illustrate the assertions in the previous section. Denote $\sum_{k=1}^{\infty} \sigma_k^2 = \tau$ when this series is convergent.

Example 2.1. Consider model (2.5), (2.8) with $a_k = k^\alpha$, $\alpha > 0$, $\sigma_k^2 = k^{\delta-1}$, $\alpha + \delta > 0$. In this case it is easy to prove that $c_\epsilon N^\alpha \to 1$ as $\epsilon \to 0$. Using this and (2.12), one can calculate

$$N = \left((2\alpha+\delta)(\alpha+\delta)Q/(\alpha\epsilon^2)\right)^{\frac{1}{2\alpha+\delta}} (1+o(1)),$$

$$c_\epsilon = \left(\alpha\epsilon^2/((2\alpha+\delta)(\alpha+\delta)Q)\right)^{\frac{\alpha}{2\alpha+\delta}} (1+o(1)).$$

Here we make use of the asymptotic relation

$$\sum_{m=1}^{M} m^\kappa = \frac{M^{\kappa+1}}{(\kappa+1)}(1+o(1)) \quad \text{as} \quad M \to \infty, \quad \kappa > -1. \tag{2.28}$$

Now one can easily verify the condition of Theorem 2.3. By applying Theorem 2.3, we derive the asymptotics of the minimax risk.

- Case $\delta > 0$. The asymptotics (2.28) and relations for N and c_ϵ yield (cf. Pinsker (1980) for $\delta = 1$ corresponding to a periodic Sobolev function space)

  $$r_\epsilon = \epsilon^{4\alpha/(2\alpha+\delta)} \delta^{-1} (Q(2\alpha+\delta))^{\delta/(2\alpha+\delta)} (\alpha/(\alpha+\delta))^{2\alpha/(2\alpha+\delta)} (1+o(1)).$$

  In this case Theorem 2.3 gives only the first-order term of the minimax risk. Note also that although $(\sigma_k^2,\ k = 1, 2, \ldots)$ can be increasing to infinity, the minimax risk still converges to zero.

- Case $\delta = 0$. By using (2.28) and the asymptotics

$$\sum_{k=1}^{M} k^{-1} = \log M + C_e + o(1) \quad \text{as} \quad M \to \infty\,, \tag{2.29}$$

one obtains

$$r_\epsilon = \alpha^{-1}\epsilon^2 \log\epsilon^{-1} + \epsilon^2(C_e + (2\alpha)^{-1}\log(2\alpha Q) - \alpha^{-1})(1+o(1))\,,$$

where $C_e = 0.5772156\ldots$ is the Euler constant (see, for example, Gradshtein and Ryzhik (1980), equation 6.360.2).

- Case $\delta < 0$. Using the asymptotic relation

$$\sum_{m=M}^{\infty} m^{-\kappa} = \frac{M^{1-\kappa}}{\kappa - 1}(1+o(1)) \quad \text{as} \quad M \to \infty\,, \quad \kappa > 1\,,$$

we calculate

$$r_\epsilon = \epsilon^2\tau + \epsilon^{\frac{4\alpha}{2\alpha+\delta}}\delta^{-1}(Q(2\alpha+\delta))^{\frac{\delta}{2\alpha+\delta}}(\alpha/(\alpha+\delta))^{\frac{2\alpha}{2\alpha+\delta}}(1+o(1))\,.$$

Example 2.2. $a_k = k^\alpha$, $\alpha > 0$, $\sigma_k^2 = k^{-(1+\alpha)}$. In this case the condition of Theorem 2.3 is again satisfied, and

$$\begin{aligned} N &= (Q\alpha\epsilon^{-2}/\log\epsilon^{-2})^{1/\alpha}(1+o(1)), \\ c_\epsilon &= \epsilon^2\log\epsilon^{-2}(Q\alpha)^{-1}(1+o(1))\,. \end{aligned}$$

Then by Theorem 2.3,

$$r_\epsilon = \epsilon^2\tau - \epsilon^4(\log\epsilon^{-2})^2\alpha^{-2}Q^{-1}(1+o(1))\,.$$

Example 2.3. $a_k = k^\alpha$, $\alpha > 0$, $\sigma_k^2 = k^{-(1+\delta)}$, $\delta > \alpha$. One calculates

$$\begin{aligned} N &= \left((\delta-\alpha)Q\epsilon^{-2}\right)^{1/\alpha}(1+o(1)), \\ c_\epsilon &= \epsilon^2((\delta-\alpha)Q)^{-1}(1+o(1)). \end{aligned}$$

With these asymptotic relations, one can show that

$$d_\epsilon(\Theta(Q)) = \epsilon^2\tau - \epsilon^4 Q^{-1}(\delta-\alpha)^{-2}(1+o(1))\,.$$

By applying Theorem 2.4, we can obtain only the rate of the second-order term of the minimax risk:

$$r_\epsilon = \epsilon^2\tau - \epsilon^4 Q^{-1}(\delta-\alpha)^{-2}\phi_\epsilon\,,$$

where

$$\liminf_{\epsilon\to 0}\phi_\epsilon \ge 1\,, \qquad \limsup_{\epsilon\to 0}\phi_\epsilon \le \pi^2\,.$$

Example 2.4. $a_k = e^{\beta k}$, $\beta > 0$, $\sigma_k^2 = k^{\delta-1}$. From (2.21) one can see that

$$e^{-\beta} \le c_\epsilon e^{N\beta} \le 1\,. \tag{2.30}$$

Using (2.12), (2.30) and the asymptotics

$$\sum_{m=1}^{M} m^\kappa e^{\beta m} = \frac{M^\kappa e^{\beta(M+1)}}{(e^\beta - 1)}(1+o(1)) \quad \text{as} \quad M \to \infty$$

gives

$$N = \beta^{-1}\log\epsilon^{-1} + (2\beta)^{-1}(1-\delta)\log\log\epsilon^{-1} + O(1)\,.$$

By the last two relations and (2.30), we have

$$\begin{aligned} c_\epsilon \sum_{k=1}^{N} \sigma_k^2 a_k &= \frac{N^{\delta-1} c_\epsilon e^{\beta(N+1)}}{e^\beta - 1}(1+o(1)) \\ &\le \frac{N^{\delta-1} e^\beta}{e^\beta - 1}(1+o(1)) \\ &= O\left((\log\epsilon^{-1})^{\delta-1}\right). \end{aligned}$$

We apply Theorem 2.4 to this example.

- Case $\delta > 1$. Since, according to Gradshtein and Ryzhik (1980), equation 0.121,

$$\sum_{m=1}^{M} m^\kappa = \frac{M^{\kappa+1}}{(\kappa+1)} + \frac{M^\kappa}{2}(1+o(1)) \quad \text{as} \quad M \to \infty\,, \quad \kappa > 0\,,$$

  we calculate

$$\begin{aligned} \sum_{k=1}^{N} \sigma_k^2 &= \frac{N^\delta}{\delta} + \frac{N^{\delta-1}}{2}(1+o(1)) \\ &= \frac{(\log\epsilon^{-1})^\delta}{\delta\beta^\delta} + \frac{(1-\delta)(\log\epsilon^{-1})^{\delta-1}\log\log\epsilon^{-1}}{2\beta^\delta}(1+o(1))\,, \end{aligned}$$

  and obtain that

$$r_\epsilon = \frac{\epsilon^2(\log\epsilon^{-1})^\delta}{\delta\beta^\delta} + \frac{\epsilon^2(1-\delta)(\log\epsilon^{-1})^{\delta-1}\log\log\epsilon^{-1}}{2\beta^\delta}(1+o(1))\,.$$

- Case $\delta = 1$. In this case we have that $c_\epsilon \sum_{k=1}^{N} \sigma_k^2 a_k = O(1)$,

$$\sum_{k=1}^{N} \sigma_k^2 = N = \beta^{-1} \log \epsilon^{-1} + O(1),$$

and therefore,

$$r_\epsilon = \beta^{-1} \epsilon^2 \log \epsilon^{-1} + O(\epsilon^2).$$

- Case $0 < \delta < 1$. One can show that

$$\begin{aligned}
\sum_{m=1}^{M} m^{-\kappa} &= \frac{2^\kappa}{2 - 2^\kappa} \left(\sum_{m=1}^{M} (M+m)^{-\kappa} - \sum_{m=1}^{2M} (-1)^{m+1} m^{-\kappa} \right) \\
&= \frac{M^{1-\kappa}}{(1-\kappa)} + \zeta(\kappa) + o(1) \quad \text{as} \quad M \to \infty, \quad 0 < \kappa < 1,
\end{aligned}$$

where

$$\zeta(\kappa) = \frac{2^\kappa}{2^\kappa - 2} \sum_{m=1}^{\infty} (-1)^{m+1} m^{-\kappa}$$

is the Riemann zeta function (Gradshtein and Ryzhik (1980), equation 7.422.2). Using these asymptotics, we obtain

$$\sum_{k=1}^{N} \sigma_k^2 = \frac{(\log \epsilon^{-1})^\delta}{\delta \beta^\delta} + \zeta(1-\delta) + o(1).$$

Consequently,

$$r_\epsilon = \epsilon^2 (\log \epsilon^{-1})^\delta \delta^{-1} \beta^{-\delta} + \epsilon^2 \zeta(1-\delta)(1 + o(1)).$$

- Case $\delta = 0$. Since, by (2.29),

$$\sum_{k=1}^{N} \sigma_k^2 = \log N + C_e + o(1),$$

we get

$$r_\epsilon = \epsilon^2 \log\log \epsilon^{-1} + \epsilon^2 (C_e + \log \beta^{-1})(1 + o(1)).$$

- Case $\delta < 0$. In this case one can verify that

$$\sum_{k=N+1}^{\infty} \sigma_k^2 = -\frac{N^\delta}{\delta}(1+o(1)) = -\frac{(\log\epsilon^{-1})^\delta}{\delta\beta^\delta}(1+o(1)).$$

Therefore, by Corollary 2.1 we have

$$r_\epsilon = \epsilon^2\tau + \epsilon^2(\log\epsilon^{-1})^\delta\beta^{-\delta}\delta^{-1}(1+o(1)).$$

Example 2.5. $a_k = e^{\beta k^r}$, $\beta > 0$, $0 < r < 1$, $\sigma_k^2 = k^{\delta-1}$. With the asymptotics

$$\sum_{m=1}^{M} m^\kappa e^{\beta m^r} = C_r e^{\beta M^r} M^{\kappa+(1-r)_+}(1+o(1)) \quad \text{as} \quad M \to \infty,$$

where

$$C_r = \begin{cases} (r\beta)^{-1}, & 0 < r < 1 \\ e^\beta/(e^\beta - 1), & r = 1 \\ 1, & r > 1, \end{cases}$$

one can obtain

$$N = \left(\beta^{-1}\log\epsilon^{-1}\right)^{1/r}(1+o(1)).$$

By definition of N, we evaluate

$$\begin{aligned} c_\epsilon \sum_{k=1}^{N} \sigma_k^2 a_k &= C_r N^{\delta-1+(1-r)_+} c_\epsilon e^{\beta N^r}(1+o(1)) \\ &\le C_r N^{\delta-1+(1-r)_+}(1+o(1)) \\ &= O\left(\left(\log\epsilon^{-1}\right)^{(\delta-1+(1-r)_+)/r}\right)(1+o(1)). \end{aligned}$$

Now the asymptotics of the minimax risk may be obtained in the same way as in Example 2.4.

- Case $\delta > 0$.

$$r_\epsilon = \epsilon^2(\log\epsilon^{-1})^{\delta/r}\beta^{-\delta/r}\delta^{-1}(1+o(1)).$$

- Case $\delta = 0$.

$$r_\epsilon = r^{-1}\epsilon^2\log\log\epsilon^{-1} + \epsilon^2(C_e + r^{-1}\log\beta^{-1})(1+o(1)).$$

- Case $\delta < 0$.

$$r_\epsilon = \epsilon^2\tau + \epsilon^2(\log\epsilon^{-1})^{\delta/r}\beta^{-\delta/r}\delta^{-1}(1+o(1))\,.$$

Example 2.6. $a_k = k^\alpha$, $\sigma_k^2 = e^{\beta k^r}$, $\alpha, \beta, r > 0$. Let us establish first an upper bound for the minimax risk $r_\epsilon(\Theta)$ (see (2.23)). Such a bound is provided by the minimax linear risk which, according to Theorem 2.1, equals d_ϵ (see (2.12)–(2.13)). Using the asymptotic expansions (as $M \to \infty$): for $0 < r < 1$

$$\begin{aligned}
\sum_{m=1}^{M} m^\alpha e^{\beta m^r} &= M^\alpha e^{\beta M^r}\left(\frac{M^{1-r}}{\beta r} + \frac{r-1-\alpha}{(\beta r)^2}M^{1-2r}(1+o(1))\right);\\
\sum_{m=1}^{M} m^\alpha e^{\beta m} &= M^\alpha e^{\beta M}\left(\frac{e^\beta}{e^\beta - 1} - \frac{\alpha e^\beta}{(e^\beta-1)^2}M^{-1}(1+o(1))\right);\\
\sum_{m=1}^{M} m^\alpha e^{\beta m^r} &= M^\alpha\left(e^{\beta M^r} + e^{\beta(M-1)^r}(1+o(1))\right) \quad \text{for} \quad r > 1;
\end{aligned}$$

one can solve (2.12)–(2.13), thus obtaining

$$c_\epsilon = \left(\beta^{-1}\log\epsilon^{-2}\right)^{-\alpha/r}(1+o(1)),$$

$$d_\epsilon = Qc_\epsilon^2(1+o(1)) = Q\left(\beta^{-1}\log\epsilon^{-2}\right)^{-2\alpha/r}(1+o(1)).$$

The last formula exhibits a distinctive feature of this example, as compared to all previous ones. Indeed, analyzing the proof of Theorem 2.1 (cf. inequality (2.19)), one realizes that the term Qc_ϵ^2, contributing to d_ϵ, arises solely as the squared bias term of the linear minimax estimator. Thus, only the bias of the estimator contributes to its maximal risk, up to the first order.

To show that d_ϵ coincides asymptotically with the minimax risk $r_\epsilon(\Theta)$, we choose a prior distribution Λ on Θ and use the obvious inequality $r_\epsilon(\Theta) \geq \mathcal{R}_\epsilon(\Lambda)$, where $\mathcal{R}_\epsilon(\Lambda)$ denotes the Bayes risk. Let Λ be a distribution on Θ such that

$$\theta_N = \pm\rho, \qquad \text{with probabilities } 1/2$$

and

$$\theta_i = 0, \qquad i \neq N, \qquad \Lambda\text{-almost surely},$$

where $\rho = (Q/a_N)^{1/2}$ and $N = [c_\epsilon^{-\alpha}]$. Clearly $\Lambda(\Theta) = 1$,

$$\rho^2 = Qa_N^{-2} = Qc_\epsilon^2(1 + o(1)) = d_\epsilon(1 + o(1))$$

and $\sigma_N^2 = e^{\beta N^r} = \epsilon^{-2}e^{O(1)}$.

Due to sufficiency considerations, the Bayes risk $\mathcal{R}_\epsilon(\Lambda)$ in estimating θ is equal to the Bayes risk in estimating θ_N, based on the observation Y_N only. Since

$$\lim_{\epsilon \to 0} \frac{\rho^2}{\mathbf{Var}\, Y_N} = 0,$$

it follows (see Ibragimov and Hasminskii (1984c), proof of Lemma 3.2) that

$$r_\epsilon(\Theta) \geq \mathcal{R}_\epsilon(\Lambda) = \rho^2(1 + o(1)) = d_\epsilon(1 + o(1)).$$

Thus

$$r_\epsilon(\Theta) = Q\left(\beta^{-1} \log \epsilon^{-2}\right)^{-2\alpha/r}(1 + o(1)).$$

Remark 2.9. Note that in most cases in Examples 2.4 and 2.5 both the first and the second order terms of the minimax risk do not depend on the "size" Q of the ellipsoid $\Theta(Q)$.

Remark 2.10. Let $a_k = a_k(\beta)$, $k = 1, 2, \ldots$, be as in Example 2.4 or Example 2.5. Define the correspondent hyperrectangle in l_2-space:

$$\mathcal{H}_\beta = \mathcal{H}_\beta(Q) = \{\theta : \ |\theta_k| \leq \sqrt{Q}a_k^{-1}(\beta), \ k = 1, 2, \ldots\}.$$

The assertions of Examples 2.4 and 2.5 concerning the first order behaviour (also the second order behaviour for the cases $\delta = 0$ and $\delta < 0$) of the minimax risk remain valid with $\Theta = \Theta_\beta$ replaced by $\mathcal{H}_\beta$. This is evident from the following easily verified relation:

for any $Q > 0$, $\beta > 0$, $0 < \mu < \beta$ there exists $Q_1 > 0$ such that

$$\Theta_\beta(Q) \subseteq \mathcal{H}_\beta(Q) \subseteq \Theta_{\beta-\mu}(Q_1).$$

2.5 Proofs

Proof of Theorem 2.2. Fix arbitrary δ_1 and δ_2 such that $0 < \delta_1 < 1$, $0 < \delta_2 < 1$. Take a positive number $R = R(\delta_1, \delta_2)$ and a probability density $\nu(x) = \nu(x, \delta_1, \delta_2)$ such that $\nu(x)$ has support $(-R, R)$, is positive and continuously differentiable inside this interval, has finite Fisher information $I(\nu)$ and satisfies the following properties:

$$\mathbf{E}X^2 = 1 - \delta_1$$

and

$$I(\nu) = \int_{-R}^{R} \frac{(\nu'(x))^2}{\nu(x)} dx \le 1 + \delta_2 \,,$$

where X is a random variable with probability density $\nu(x)$. Note that under the imposed conditions on density $\nu(x)$ the relation between $\mathbf{E}X^2$ and $I(\nu)$ is not arbitrary. Indeed, integrating by parts,

$$\begin{aligned} 1 &= \left(\int_{-R}^{R} \nu(x) dx \right)^2 = \left(\int_{-R}^{R} x\nu'(x) dx \right)^2 \\ &\le \int_{-R}^{R} \frac{(\nu'(x))^2}{\nu(x)} dx \int_{-R}^{R} x^2 \nu(x) dx \end{aligned}$$

and therefore the inequality

$$\mathbf{E}X^2 \ge \frac{1}{I(\nu)} \tag{2.31}$$

should hold, which leads to the following relation between δ_1 and δ_2:

$$\delta_2 \ge \frac{\delta_1}{1 - \delta_1} \,.$$

For example, the choice $\delta_1 = \delta/2$ and $\delta_2 = \delta$, with $0 < \delta < 1$, will do. The inequality (2.31) becomes equality for the standard normal density. Thus, for small δ_1, δ_2 we should take R so big that we can find a density ν with support $(-R, R)$ close enough to the standard normal density in order to provide the properties above. Suppose from now on that $R = R(\delta_1, \delta_2)$ is the smallest possible value for which these properties are fulfilled.

Let m_k, $k = 1, 2, \ldots$, be sequence satisfying (2.25). Without loss of generality we assume that $m_k > 0$, $k = 1, 2, \ldots$. Indeed, the zero m_k's give zero contribution to the lower bound (2.26). Introduce

$$\nu_k(x) = \nu_k(x, \delta_1, \delta_2) = m_k^{-1} \nu(m_k^{-1} x), \quad k = 1, 2, \ldots \,.$$

These are probability densities with supports $(-Rm_k, m_k R)$ respectively and if $X_k = m_k X$ then X_k is a random variable with density $\nu_k(x)$. We have

$$\mathbf{E}X_k^2 = m_k^2 (1 - \delta_1) \tag{2.32}$$

$$I(\nu_k) = m_k^{-2} I(\nu) \le m_k^{-2} (1 + \delta_2). \tag{2.33}$$

Let θ be distributed according to a prior measure μ such that θ_k, $k = 1, 2, \dots$, are distributed independently with the densities $\nu_k(x)$, $k = 1, 2, \dots$, respectively. Let $\mathbf{E}$ denote the expectation with respect to the joint distribution of $Y_1, Y_2, \dots$ and $\theta_1, \theta_2, \dots$.

Since Θ is closed and convex, $r_\epsilon = \inf_{\hat\theta\in\Theta} \sup_{\theta\in\Theta} \mathbf{E}_\theta \|\hat\theta - \theta\|^2$. We bound the minimax risk from below as follows:

$$\begin{aligned} r_\epsilon &= \inf_{\hat\theta\in\Theta} \sup_{\theta\in\Theta} \mathbf{E}_\theta \|\hat\theta - \theta\|^2 \\ &\geq \inf_{\hat\theta} \int_\Theta \sum_{k=1}^\infty \mathbf{E}_\theta (\hat\theta_k - \theta_k)^2 d\mu(\theta) \\ &= \inf_{\hat\theta \in supp\, \mu} \int_\Theta \sum_{k=1}^\infty \mathbf{E}_\theta (\hat\theta_k - \theta_k)^2 d\mu(\theta) \\ &\geq \inf_{\hat\theta} \sum_{k=1}^\infty \mathbf{E}(\hat\theta_k - \theta_k)^2 - \sup_{\hat\theta\in supp\,\mu} \int_{\Theta^C} \sum_{k=1}^\infty \mathbf{E}_\theta (\hat\theta_k - \theta_k)^2 d\mu(\theta) \\ &\geq \inf_{\hat\theta} \sum_{k=1}^\infty \mathbf{E}(\hat\theta_k - \theta_k)^2 - 4R^2 \mu(\Theta^C) \sum_{k=1}^\infty m_k^2 . \end{aligned} \tag{2.34}$$

Due to the assumptions on probability density $\nu_k(x)$, we can apply the van Trees inequality (2.24) to the Bayes risk $\mathbf{E}(\hat\theta_k - \theta_k)^2$. Thus, by (2.24) and (2.33), we obtain

$$\inf_{\hat\theta} \sum_{k=1}^\infty \mathbf{E}(\hat\theta_k - \theta_k)^2 \geq \sum_{k=1}^\infty \frac{\epsilon^2 \sigma_k^2 m_k^2 (1+\delta_2)^{-1}}{m_k^2 (1+\delta_2)^{-1} + \epsilon^2 \sigma_k^2} . \tag{2.35}$$

Recall that $|\theta_k| \leq R m_k$ and $\mathbf{E}\theta_k^2 = (1-\delta_1) m_k^2$, $k = 1, 2, \dots$. Therefore, we have

$$|a_k^2(\theta_k^2 - \mathbf{E}\theta_k^2)| \leq a_k^2 m_k^2 |R^2 - 1 + \delta_1| .$$

Using this relation, the condition (2.25) and the Hoeffding's inequality (see Pollard (1984)), we evaluate $\mu(\Theta^C)$:

$$\begin{aligned} \mu(\Theta^C) &= \mu \left\{ \sum_{k=1}^\infty a_k^2 \theta_k^2 > Q \right\} \\ &= \mu \left\{ \sum_{k=1}^\infty a_k^2 (\theta_k^2 - \mathbf{E}\theta_k^2) > Q - \sum_{k=1}^\infty a_k^2 \mathbf{E}\theta_k^2 \right\} \\ &\leq \exp \left\{ - \frac{\left(Q - (1-\delta_1)\sum_{k=1}^\infty a_k^2 m_k^2\right)^2}{2(R^2 - 1 + \delta_1)^2 \sum_{k=1}^\infty a_k^4 m_k^4} \right\} \end{aligned}$$

$$
\begin{aligned}
&\le \exp\left\{-\frac{\beta_\epsilon \log \gamma_\epsilon^{-1}}{2(R^2-1+\delta_1)^2}\right\} \\
&= \gamma_\epsilon^{\beta_\epsilon \tau}
\end{aligned} \tag{2.36}
$$

where $\tau = 1/(2(R^2-1+\delta_1)^2)$. Now let $\delta_1 = \delta_1(\epsilon)$, $\delta_2 = \delta_2(\epsilon)$ depend on ϵ and converge to zero as $\epsilon \to 0$ in such a way (sufficiently slowly) that $R = R_\epsilon = R(\delta_1(\epsilon), \delta_2(\epsilon))$ and $\tau = \tau_\epsilon$ become functions of ϵ satisfying

$$R_\epsilon^2 \gamma_\epsilon^{C_1} \to 0$$

for some $C_1 > 0$ and

$$\beta_\epsilon \tau_\epsilon \to \infty$$

as $\epsilon \to 0$.

It follows from (2.34), (2.35) and (2.36) that

$$
\begin{aligned}
r_\epsilon &\ge \sum_{k=1}^{\infty} \frac{\epsilon^2 \sigma_k^2 m_k^2 (1+\delta_2(\epsilon))^{-1}}{m_k^2(1+\delta_2(\epsilon))^{-1} + \epsilon^2\sigma_k^2} - 4R_\epsilon^2 \gamma_\epsilon^{\beta_\epsilon \tau_\epsilon} \sum_{k=1}^{\infty} m_k^2 \\
&\ge \sum_{k=1}^{\infty} \frac{\epsilon^2 \sigma_k^2 m_k^2 (1+\delta_2(\epsilon))^{-1}}{m_k^2 + \epsilon^2\sigma_k^2} - 4R_\epsilon^2 \gamma_\epsilon^{\beta_\epsilon \tau_\epsilon} Q a_1^{-2} \\
&\ge \frac{1}{1+\delta_2(\epsilon)} \sum_{k=1}^{\infty} \frac{\epsilon^2 \sigma_k^2 m_k^2}{m_k^2 + \epsilon^2\sigma_k^2} - C_2 R_\epsilon^2 \gamma_\epsilon^{C_1+\alpha} \\
&= \sum_{k=1}^{\infty} \frac{\epsilon^2 \sigma_k^2 m_k^2}{m_k^2 + \epsilon^2\sigma_k^2}(1+o(1)) + O\left(\gamma_\epsilon^{\alpha}\right)
\end{aligned}
$$

for any fixed $\alpha > 0$, and the theorem follows. □

Remark 2.11. The most detailed analysis of the proof of the theorem above shows that in fact the following nonasymtotic assertion can be proved.

Let for any a and b, $0 < a < 0$, $0 < b < 1$, $b \ge a(1-a)^{-1}$, $R = R(a,b)$ denote the same function as in the proof of the theorem. Let $\delta' = (\delta_1', \delta_2', \ldots)$ and $\delta'' = (\delta_1'', \delta_2'', \ldots)$ be any sequences such that $0 < \delta_i' < 0$, $0 < \delta_i'' < 1$, $\delta_i'' \ge \delta_i'(1-\delta_i')^{-1}$, $i = 1, 2, \ldots$. Let $(m_k,\ k = 1, 2, \ldots)$, $m_k = m_k(\epsilon)$, be a nonnegative sequence satisfying the condition

$$
\begin{aligned}
&\sum_{k=1}^{\infty} a_k^2 (1-\delta_k') m_k^2 \\
&\qquad + \left(\left(\log \gamma_\epsilon^{-1}\right) 2 \sum_{k=1}^{\infty} (R^2(\delta_k', \delta_k'') - 1 + \delta_k')^2 a_k^4 m_k^4 \right)^{1/2} \le Q\,,
\end{aligned}
$$

for some positive function γ_ϵ. Then the following lower bound holds:

$$r_\epsilon \geq \sum_{k=1}^{\infty} \frac{\epsilon^2 \sigma_k^2 m_k^2}{m_k^2 + \epsilon^2 \sigma_k^2 (1+\delta_k'')} - 4\gamma_\epsilon \sum_{k=1}^{\infty} R^2(\delta_k', \delta_k'') m_k^2 .$$

Proof of Theorem 2.3. By Theorem 2.1, we have the following upper bound for the minimax risk:

$$r_\epsilon \leq r_\epsilon^l = d_\epsilon = \epsilon^2 \sum_{k=1}^{N} \sigma_k^2 - \epsilon^2 c_\epsilon \sum_{k=1}^{N} \sigma_k^2 a_k . \tag{2.37}$$

Introduce

$$\alpha_\epsilon = \epsilon^2 c_\epsilon^{-1} \left(\log \epsilon^{-1} \sum_{k=1}^{N} a_k^2 \sigma_k^4 (1 - c_\epsilon a_k)^2 \right)^{1/2} ,$$

where c_ϵ and $N = N_\epsilon$ are defined by (2.12) and (2.21). Note that $\alpha_\epsilon > 0$ and $\lim_{\epsilon\to 0} \alpha_\epsilon = 0$ because of the condition of the theorem. Take a positive function β_ϵ such that $\lim_{\epsilon\to 0} \beta_\epsilon = \infty$ and $\alpha_\epsilon \beta_\epsilon^{1/2} \to 0$ as $\epsilon \to 0$; for example, we can choose $\beta_\epsilon = \alpha_\epsilon^{-1}$.

Take now the sequence

$$m_k^2 = \frac{\tilde{\theta}_k^2}{1+\eta_\epsilon}, \qquad k = 1, 2, \dots ,$$

with

$$\eta_\epsilon = Q^{-1} \alpha_\epsilon \beta_\epsilon^{1/2} ,$$

and $\tilde{\theta}_k^2$ defined by (2.16). We now show that the equation (2.25) is satisfied for these m_k's with $\gamma_\epsilon = \epsilon$. First, by (2.12), note that

$$\sum_{k=1}^{\infty} a_k^2 \tilde{\theta}_k^2 = \epsilon^2 c_\epsilon^{-1} \sum_{k=1}^{N} \sigma_k^2 a_k (1 - c_\epsilon a_k) = Q$$

and

$$\begin{aligned} &\left(\log \epsilon^{-1} \sum_{k=1}^{\infty} a_k^4 \tilde{\theta}_k^4 \right)^{1/2} \\ &= \epsilon^2 c_\epsilon^{-1} \left(\log \epsilon^{-1} \sum_{k=1}^{N} a_k^2 \sigma_k^4 (1 - c_\epsilon a_k)^2 \right)^{1/2} = \alpha_\epsilon . \end{aligned}$$

Therefore,

$$
\begin{aligned}
&\sum_{k=1}^{\infty} a_k^2 m_k^2 + \left(\beta_\epsilon \left(\log \epsilon^{-1}\right) \sum_{k=1}^{\infty} a_k^4 m_k^4\right)^{1/2} \\
&= \frac{1}{1+\eta_\epsilon} \sum_{k=1}^{\infty} a_k^2 \tilde{\theta}_k^2 \left(1 + \frac{\left(\beta_\epsilon \left(\log \epsilon^{-1}\right) \sum_{k=1}^{\infty} a_k^4 \tilde{\theta}_k^4\right)^{1/2}}{\sum_{k=1}^{\infty} a_k^2 \tilde{\theta}_k^2}\right) \\
&= \frac{1}{1+\eta_\epsilon} \sum_{k=1}^{\infty} a_k^2 \tilde{\theta}_k^2 (1+\eta_\epsilon) = Q\,.
\end{aligned}
$$

Apply now the same reasoning as in Theorem 2.2 to obtain, for sufficiently small ϵ and any fixed $\alpha > 0$,

$$
r_n \geq \sum_{k=1}^{\infty} \frac{\epsilon^2 \sigma_k^2 m_k^2 (1+\delta_2(\epsilon))^{-1}}{m_k^2 (1+\delta_2(\epsilon))^{-1} + \epsilon^2 \sigma_k^2} - 4R_\epsilon^2 \epsilon^{C_1+\alpha} \sum_{k=1}^{\infty} m_k^2\,, \tag{2.38}
$$

with $C_1 > 0$ such that $R_\epsilon^2 \epsilon^{C_1} \to 0$ as $\epsilon \to 0$. Denote

$$
\rho_\epsilon = (1+\eta_\epsilon)(1+\delta_2(\epsilon)) - 1\,.
$$

Substituting the chosen sequence $(m_k^2,\ k = 1, 2, \ldots)$, we calculate

$$
\begin{aligned}
r_\epsilon &\geq \frac{\epsilon^2}{(1+\rho_\epsilon)} \sum_{k=1}^{\infty} \frac{\sigma_k^2 \tilde{\theta}_k^2}{\tilde{\theta}_k^2 (1+\rho_\epsilon)^{-1} + \epsilon^2 \sigma_k^2} + O(\epsilon^\alpha) \\
&= \epsilon^2 \sum_{k=1}^{N} \sigma_k^2 - \epsilon^2 c_\epsilon \sum_{k=1}^{N} \sigma_k^2 a_k \\
&\quad - \epsilon^2 c_\epsilon \rho_\epsilon \sum_{k=1}^{N} \frac{\sigma_k^2 a_k (1 - c_\epsilon a_k)}{1 + \rho_\epsilon c_\epsilon a_k} + O(\epsilon^\alpha)\,.
\end{aligned}
$$

From (2.12) it follows that c_ϵ can not be of smaller order than ϵ^2. Picking now some $\alpha > 4$ and recalling that $\rho_\epsilon > 0$, $\rho_\epsilon \to 0$ as $\epsilon \to 0$, and $0 \leq c_\epsilon a_k \leq 1$ for $k = 1, 2, \ldots, N$ (see (2.22)), we conclude that the last lower bound, together with upper bound (2.37), proves the theorem. □

Remark 2.12. The statement of the theorem remains valid under the following weaker condition: there exist a positive constant α_0 and positive function γ_ϵ such that

$$
\gamma_\epsilon^{\alpha_0} = o(c_\epsilon)\,, \qquad \epsilon \to 0,
$$

and

$$\lim_{\epsilon\to 0} \epsilon^4 c_\epsilon^{-2} \log \gamma_\epsilon^{-1} \sum_{k=1}^{N} a_k^2 \sigma_k^4 (1 - c_\epsilon a_k)^2 = 0\,, \qquad \epsilon \to 0\,,$$

where c_ϵ and $N = N_\epsilon$ are defined by (2.12) and (2.21). Indeed, let $R_\epsilon = R(\delta_1(\epsilon), \delta_2(\epsilon))$ be chosen in such a way that $R_\epsilon^2 \epsilon^{C_1} \to 0$ as $\epsilon \to 0$ for some $C_1 > 0$. Then substituting the chosen sequence $(m_k^2,\ k = 1, 2, \dots)$ in (2.38), with $\gamma_\epsilon^{C_1+\alpha}$ instead of $\epsilon^{C_1+\alpha}$, leads to

$$\begin{aligned}
r_\epsilon &\geq \frac{\epsilon^2}{(1+\rho_\epsilon)} \sum_{k=1}^{\infty} \frac{\sigma_k^2 \tilde{\theta}_k^2}{\tilde{\theta}_k^2 (1+\rho_\epsilon)^{-1} + \epsilon^2 \sigma_k^2} - 4 R_\epsilon^2 \gamma_\epsilon^{C_1+\alpha} \sum_{k=1}^{\infty} \tilde{\theta}_k^2 \\
&= \epsilon^2 \sum_{k=1}^{N} \sigma_k^2 - \epsilon^2 c_\epsilon \sum_{k=1}^{N} \sigma_k^2 a_k - \epsilon^2 c_\epsilon \rho_\epsilon \sum_{k=1}^{N} \frac{\sigma_k^2 a_k (1 - c_\epsilon a_k)}{1 + \rho_\epsilon c_\epsilon a_k} \\
&\quad - 4 R_\epsilon^2 \gamma_\epsilon^{C_1+\alpha} \epsilon^2 \sum_{k=1}^{N} \frac{\sigma_k^2 (1 - c_\epsilon a_k)}{c_\epsilon a_k}\,. \\
&= \epsilon^2 \sum_{k=1}^{N} \sigma_k^2 - \epsilon^2 c_\epsilon \sum_{k=1}^{N} \sigma_k^2 a_k - \epsilon^2 c_\epsilon \rho_\epsilon \sum_{k=1}^{N} \frac{\sigma_k^2 a_k (1 - c_\epsilon a_k)}{1 + \rho_\epsilon c_\epsilon a_k} \\
&\quad - o(1) \epsilon^2 c_\epsilon \sum_{k=1}^{N} \sigma_k^2 a_k \\
&= \epsilon^2 \sum_{k=1}^{N} \sigma_k^2 - \epsilon^2 c_\epsilon \sum_{k=1}^{N} \sigma_k^2 a_k (1 + o(1))\,.
\end{aligned}$$

Proof of Theorem 2.4. Let m_k, $k = 1, 2, \dots$ be some sequence of positive numbers such that

$$\sum_{k=1}^{\infty} a_k^2 m_k^2 \leq Q\,, \tag{2.39}$$

i.e. $m = (m_k,\ k = 1, 2, \dots) \in \Theta$. Introduce

$$\nu_k(x) = m_k^{-1} \nu_0(m_k^{-1} x), \quad k = 1, 2, \dots\,,$$

where $\nu_0(x) = I\{|x| \leq 1\} \cos^2(\pi x/2)$. These are probability densities with supports $[-m_k, m_k]$ respectively. It is easy to calculate the Fisher information of the distribution defined by the density $\nu_k(x)$:

$$I(\nu_k) = m_k^{-2} I(\nu_0) = m_k^{-2} \pi^2\,, \tag{2.40}$$

where $\mathbf{E}_{\nu_k}$ denotes expectation with respect to the density ν_k.

We select a prior measure $d\mu(\theta)$ such that $\theta_k,\ k = 1, 2, \dots$, are distributed independently with densities $\nu_k(x),\ k = 1, 2, \dots$, respectively. Since (2.39) provides that $\operatorname{supp} \mu \subseteq \Theta$, we proceed estimating the minimax risk (2.23) from below as follows:

$$\begin{aligned} r_\epsilon &= \inf_{\hat\theta} \sup_{\Theta} \sum_{k=1}^{\infty} \mathbf{E}_\theta (\hat\theta_k - \theta_k)^2 \\ &\geq \inf_{\hat\theta} \int_\Theta \sum_{k=1}^{\infty} \mathbf{E}_\theta (\hat\theta_k - \theta_k)^2 d\mu(\theta) \\ &= \inf_{\hat\theta} \sum_{k=1}^{\infty} \mathbf{E} (\hat\theta_k - \theta_k)^2 . \end{aligned} \tag{2.41}$$

In view of (2.40), the inequality (2.24) yields

$$\mathbf{E}(\hat\theta_k - \theta_k)^2 \geq \frac{1}{\pi^2 m_k^{-2} + \epsilon^{-2}\sigma_k^{-2}} .$$

From this and (2.41), we have that for any m from the ellipsoid Θ the minimax risk r_ϵ satisfies

$$r_\epsilon \geq \epsilon^2 \sum_{k=1}^{\infty} \frac{\sigma_k^2 m_k^2/\pi^2}{m_k^2/\pi^2 + \epsilon^2 \sigma_k^2} . \tag{2.42}$$

Using Lemma A.1, one obtains the following lower bound:

$$\begin{aligned} r_\epsilon &\geq \sup_{m \in \Theta(Q)} \epsilon^2 \sum_{k=1}^{\infty} \frac{\sigma_k^2 m_k^2/\pi^2}{m_k^2/\pi^2 + \epsilon^2 \sigma_k^2} \\ &= \sup_{m \in \Theta(Q/\pi^2)} \epsilon^2 \sum_{k=1}^{\infty} \frac{\sigma_k^2 m_k^2}{m_k^2 + \epsilon^2 \sigma_k^2} = d_\epsilon(\Theta(Q/\pi^2)). \end{aligned}$$

Combining the last relation with Theorem 2.1 completes the proof. □

Proof of Corollary 2.1. The left hand side of the inequality (2.42) does not depend on m. Therefore, we can take any $m \in \Theta$. Now we make use of the vector $(\tilde\theta_k,\ k = 1, 2, \dots)$ defined by (2.16). Relation (2.20) provides that $\tilde\theta \in \Theta$. Substituting $\tilde\theta_k$ in (2.42), $k = 1, 2, \dots$, one calculates

$$r_\epsilon \geq \epsilon^2 \sum_{k=1}^{\infty} \frac{\sigma_k^2 \tilde\theta_k^2/\pi^2}{\tilde\theta_k^2/\pi^2 + \epsilon^2 \sigma_k^2}$$

$$= \epsilon^2 \sum_{k=1}^{N} \sigma_k^2 - \epsilon^2 \sum_{k=1}^{N} \frac{\sigma_k^2 c_\epsilon a_k}{(1-\pi^{-2}) c_\epsilon a_k + \pi^{-2}}.$$

Using now this and (2.22), we obtain that

$$r_\epsilon \geq \epsilon^2 \sum_{k=1}^{N} \sigma_k^2 - \epsilon^2 \pi^2 \sum_{k=1}^{N} \sigma_k^2 c_\epsilon a_k .$$

Combining the last relation with the condition of the corollary and the upper bound (2.37) completes the proof. □

2.6 Bibliographic remarks

Pinsker (1980) initiated the study of minimax estimation procedures for the filtration problem in Gaussian noise. In Pinsker (1980) it was shown that, for ellipsoids meeting certain regularity conditions, the quadratic minimax risk over the ellipsoids Θ coincides asymptotically with the minimax risk within the class of linear estimators. A procedure for obtaining the minimax linear estimators and evaluating their risks was given. This fact allowed the first description of exact asymptotics of the minimax risk in nonparametric curve estimation problems.

The observation model (2.5) arises as the limiting experiment in many other estimation problems. This model has been actively pursued recently, see Donoho and Johnstone (1994b), Donoho et al. (1990), Golubev and Levit (1996b), Golubev and Nussbaum (1990) and further references therein. These papers demonstrate amply the importance of asymptotic minimax estimators and their practical relevance.

The results of this chapter are also presented in Belitser and Levit (1995) and partly in Belitser and Levit (1994).

Chapter 3

Minimax nonparametric regression

In this chapter we are concerned with the problem of optimal estimation of a nonparametric regression function. Till recently, the notion of asymptotic optimality of an estimator was associated with the optimal convergence rate of the risk of this estimator. However, comparing estimators just on the basis of the convergence rates of their risks does not make it possible to distinguish among estimators optimal in that sense. Also from a more practical point of view, such an approach does not give a recipe for choosing parameters of the estimator involved: the bandwidth for the kernel method, the number of terms for the orthogonal series method, etc. Thus two estimators, optimal in the sense of the rate of convergence, can perform in actual applications quite differently. The minimax approach becomes more useful if the constants involved in the lower and upper bounds are found, especially when these constants happen to coincide. Presently, the problem of finding the exact constants is of increasing interest.

In this chapter we establish the exact asymptotics of the minimax risk and propose a kernel type estimator which, under some regularity conditions, attains this asymptotics, i.e. it is shown to be asymptotically minimax. We illustrate the main results by two examples and discuss the consistency questions.

3.1 The model

Consider the problem of estimating a nonparametric regression function $f(x)$, $x \in [0,1]$ on the basis of the observations

$$Y_i = f(t_{in}) + \epsilon_i \,, \quad i = 1, 2, \ldots, n \,, \tag{3.1}$$

where ϵ_i's are independent Gaussian random variables with zero mean and variance σ^2.

The design is assumed to be equidistant: $t_{in} = i/n, \quad i = 1, 2, \ldots, n.$ For simplicity, the dependence of some variables on subscript n will frequently be dropped from notation.

Let $L_2 = L_2([0,1])$ be the Hilbert space of square-integrable functions on $[0,1]$ and $\{\phi_k(x),\ k = 1, 2, \ldots\}$ be its orthonormal trigonometric basis, i.e.

$$\phi_j(x) = \begin{cases} 1, & j = 1 \\ \sqrt{2}\sin(2\pi k x), & j = 2k \\ \sqrt{2}\cos(2\pi k x), & j = 2k+1 \,. \end{cases}$$

We assume that $f(x) \in L_2[0,1]$. Hence it can be represented as follows:

$$f(x) = \sum_{k=1}^{\infty} \theta_k \phi_k(x) \,, \qquad \text{where} \quad \theta_k = \int_0^1 f(x)\phi_k(x)dx \,.$$

Here convergence is meant in L_2-sense.

Let $(a_k,\ k = 1, 2, \ldots)$ be a nonnegative numerical sequence. Now we define the nonparametric class:

$$\Theta = \Theta(Q) = \left\{ f(\cdot) \in L_2 : \ \sum_{k=1}^{\infty} a_k^2 \theta_k^2 \le Q \right\} \,. \tag{3.2}$$

We are interested in the asymptotic behaviour of the minimax risk

$$r_n = r_n(\Theta) = \inf_{\hat{f}_n} \sup_{\Theta} \mathbf{E}_f \|\hat{f}_n - f\|^2 \,,$$

where $\|\cdot\|$ denotes the usual norm in $L_2([0,1])$. Here infimum is taken over all estimators and supremum is taken over all regression curves from class Θ. We recall that an estimator $\hat{f}_n$ is called *asymptotically minimax* if

$$R_n(\hat{f}_n) \stackrel{\text{def}}{=} \sup_{\Theta} \mathbf{E}_f \|\hat{f}_n - f\|^2 = r_n(\Theta)(1 + o(1)) \quad \text{as} \quad n \to \infty \,.$$

All asymptotic equations in this chapter refer to, unless otherwise specified, $n \to \infty$.

Our approach is based essentially on "equivalence" of the initial nonparametric model to a sequence of linear models of increasing dimensions. Here by equivalence of two models we mean that the corresponding minimax risks coincide asymptotically. Namely, with the class of regression functions $f(x)$ under consideration, our problem of estimating $f(x)$ is equivalent to that of estimating an infinite-dimensional parameter $(\theta_i,\ i = 1, 2, \ldots)$ based on observations:

$$Z_i = \theta_i + \tilde{\theta}_i + n^{-1/2}\xi_i\,, \quad i = 1, 2, \ldots, n\,,$$

and resembles the estimation problem considered in the previous chapter. Here ξ_l's are Gaussian random variables, $\mathbf{E}\xi_i = 0$, $\mathbf{E}[\xi_l \xi_k] = \sigma^2 \delta_{lk}$ ($\delta_{kl} = 1$ if $k = l$ and $\delta_{kl} = 0$ if $k \neq l$), the $\tilde{\theta}_i$'s are "nuisance" parameters, which are negligibly small provided $f(x)$ belongs to appropriate classes of smooth functions.

3.2 Minimax consistency

In this section we employ a different notion of consistency than usually in the literature – uniform mean square consistency. We say that estimator $\hat{f}_n$ is *uniformly mean square consistent* (slightly abusing terminology, we will call it just consistent) if

$$\sup_{\Theta} \mathbf{E}_f \|\hat{f}_n - f\|^2 \to 0 \quad \text{as} \quad n \to \infty\,.$$

Suppose for the moment that instead of our original model we have a little more general one:

(i) ϵ_i's are not necessarily Gaussian, but independent random variables with zero mean, variance σ^2 and finite Fisher information I_ϵ;

(ii) fix some orthonormal basis $\{\phi_k(x),\ k = 1, 2, \ldots\}$ in $L_2([0,1])$ such that

$$\sup_x |\phi_k(x)| \le M \qquad k = 1, 2, \ldots;$$

(iii) the design is not necessarily equidistant, but "uniform" enough to satisfy

$$\lim_{n \to \infty} \max_{1 \le k \le n} |\Delta t_{kn}| = 0\,, \tag{3.3}$$

where $\Delta t_i = \Delta t_{in} = t_i - t_{i-1}$, $i = 1, \ldots, n$, $t_0 = 0$.

Then the following theorem gives necessary and sufficient conditions for the minimax risk to converge to zero.

Theorem 3.1. *Let (i), (ii), (iii) hold. Then if*

$$\liminf_{n\to\infty} r_n(\Theta) \to 0,$$

then $a_k \to \infty$ as $k \to \infty$. Conversely, if $a_k \to \infty$ as $k \to \infty$, then

$$\liminf_{n\to\infty} r_n(\Theta') \to 0,$$

where

$$\Theta' = \Theta'(Q, P) = \Theta(Q) \cap \{f(\cdot) : \|f\|^2 \le P\}.$$

Proof. The proof of the theorem includes two parts. First we prove that if the sequence $(a_k,\ k = 1, 2, \ldots)$, does not converge to infinity, then

$$\liminf_{n\to\infty} r_n(\Theta) > 0\,.$$

Introduce

$$\liminf_{k\to\infty} a_k = A\,,$$

which is finite since $a_k \not\to \infty$ as $k \to \infty$. Then there exists a subsequence $(a_{k_l},\ l = 1, 2, \ldots)$ such that

$$\lim_{l\to\infty} a_{k_l} = A\,.$$

For some $0 < \epsilon < A$ denote

$$\mathcal{A} = \{k_l : \ |a_{k_l} - A| \le \epsilon\}\,.$$

Let $\mathcal{N}$ be an N-element subset of $\mathcal{A}$.

Let $m = (m_1, m_2, \ldots)$ be a set of nonnegative numbers such that $m \in \Theta$, $m_k > 0$, $k \in \mathcal{N}$ and $m_k = 0$, $k \notin \mathcal{N}$. Introduce

$$\nu_k(x) = (1/m_k)\nu_0(x/m_k), \quad k \in \mathcal{N}\,,$$

where $\nu_0(x)$ is a probability density on the interval $[-1, 1]$ with a finite Fisher information

$$I_0 = \int_{-1}^{1} (\nu_0'(x))^2 \nu_0^{-1}(x)dx$$

such that

$$\nu_0(-1) = \nu_0(1) = 0$$

and $\nu_0(x)$ is continuously differentiable for $|x| < 1$. The functions $\nu_k(x)$ are probability densities with supports $[-m_k, m_k]$ respectively. It is easy to calculate the Fisher information of the distribution defined by the density $\nu_k(x)$:

$$I(\nu_k) = I_0 m_k^{-2}.$$

It is known that the minimum of $\int_{-1}^{1}(q'(t))^2 q^{-1}(t)dt$ over all differentiable densities $q(t)$ with support $[-1,1]$ is attained by function $q(t) = \cos^2(\pi t/2)$ (see Borovkov (1984)). Therefore, one always has $I_0 \geq \pi^2$.

Define the measure μ on l_2 such that $\theta_k = 0$, for $k \notin \mathcal{N}$ and θ_k, for $k \in \mathcal{N}$, are distributed independently with densities $\nu_k(x)$ respectively. Since by assumption $m \in \Theta$, the measure μ has $\operatorname{supp} \mu \subseteq \Theta$. We estimate the minimax risk by the Parseval identity from below as follows:

$$\begin{aligned} r_n &\geq \inf_{\hat{\theta}} \int_{\Theta} \sum_{k=1}^{\infty} \mathbf{E}_f(\hat{\theta}_k - \theta_k)^2 d\mu(\theta) \\ &= \inf_{\hat{\theta}} \sum_{k=1}^{\infty} \mathbf{E}(\hat{\theta}_k - \theta_k)^2 \geq \inf_{\hat{\theta}} \sum_{k \in \mathcal{N}} \mathbf{E}(\hat{\theta}_k - \theta_k)^2 . \end{aligned} \tag{3.4}$$

Here we write $\mathbf{E}$ for the expectation with respect to the joint distribution of $Y_1, \ldots, Y_n$ and $\theta_1, \theta_2, \ldots$. To estimate $\mathbf{E}(\hat{\theta}_k - \theta_k)^2$, $k \in \mathcal{N}$, we apply the van Trees inequality (see Theorem A.1):

$$\mathbf{E}(\hat{\theta}_k - \theta_k)^2 \geq \frac{1}{\mathbf{E}I(\theta_k) + I(\nu_k)}, \qquad k \in \mathcal{N},$$

where $I(\theta_k)$ is the Fisher information about parameter θ_k contained in observations $Y_1, \ldots, Y_n$. It is easy to evaluate

$$\begin{aligned} \mathbf{E}I(\theta_k) &= \int \mathbf{E}_f \left(\sum_{i=1}^{n} \frac{\partial \log p_\epsilon(Y_i - f(t_{in}))}{\partial \theta_k} \right)^2 d\nu_k(\theta_k) \\ &= I_\epsilon \sum_{i=1}^{n} \phi_k^2(t_{in}) \leq M^2 I_\epsilon n. \end{aligned}$$

Recalling that $I(\nu_k) = I_0 m_k^{-2}$, $k \in \mathcal{N}$, we obtain

$$\mathbf{E}(\hat{\theta}_k - \theta_k)^2 \geq \frac{1}{I_0 m_k^{-2} + M^2 I_\epsilon n}, \qquad k \in \mathcal{N}.$$

We make use of the last inequality and (3.4):

$$r_n \geq \sum_{k \in \mathcal{N}} \frac{1}{I_0 m_k^{-2} + M^2 I_\epsilon n}. \tag{3.5}$$

Now we choose $m = (m_1, m_2, \ldots) \in \Theta$ such that

$$m_l = \begin{cases} Q^{1/2} N^{-1/2} (A+\epsilon)^{-1}, & l \in \mathcal{N} \\ 0, & l \notin \mathcal{N}. \end{cases}$$

It is easy to verify that $m \in \Theta$. So, substituting this m in (3.5) results in

$$\begin{aligned} r_n &\geq \sum_{k \in \mathcal{N}} \frac{1}{I_0 m_k^{-2} + M^2 I_\epsilon n} \\ &\geq \frac{Q}{I_0 (A+\epsilon)^2 + M_2 I_\epsilon Q n / N}. \end{aligned}$$

The number N can be chosen arbitrarily large and ϵ arbitrarily small. Therefore we finally get

$$r_n \geq \frac{Q}{I_0 A^2},$$

which proves the first part of the theorem.

Now we prove the second part of the theorem. Let $I\{S\}$ denote the indicator of set S. Define the following estimator

$$\hat{f}_n(x) = \hat{f}_n(x, N) = \sum_{k=1}^{N} \hat{\theta}_k^t \phi_k(x),$$

where

$$\hat{\theta}_k^t = \hat{\theta}_k I\{|\hat{\theta}_k| \leq P\} + P \operatorname{sign}(\hat{\theta}_k) I\{|\hat{\theta}_k| > P\},$$

$$\hat{\theta}_k = \sum_{i=1}^{n} \phi_k(t_i) Y_i \Delta t_i,$$

i.e. $\hat{\theta}_k^t$ is the projection of $\hat{\theta}_k$ on $[-P, P]$.

Then, by the Parseval equality and (1.2),

$$\begin{aligned} \mathbf{E}_f \|\hat{f}_n - f\|^2 &= \sum_{k=1}^{N} \mathbf{E}_f (\hat{\theta}_k^t - \theta_k)^2 + \sum_{k=N+1}^{\infty} \theta_k^2 \\ &= \sum_{k=1}^{N} \mathbf{Var}\, \hat{\theta}_k^t + \sum_{k=1}^{N} (E_f \hat{\theta}_k^t - \theta_k)^2 + \sum_{k=N+1}^{\infty} \theta_k^2. \end{aligned}$$

Let us bound, uniformly over Θ', each term from above in the right hand side of the last inequality.

First, since $a_k \to \infty$ as $k \to \infty$, for any fixed $\epsilon > 0$ one can find N large enough to provide the following estimate for the third term:

$$\sup_{f \in \Theta'} \sum_{k=N+1}^{\infty} \theta_k^2 \le \epsilon .$$

Next, we obviously have that, as $n \to 0$,

$$\begin{aligned} \mathbf{Var}\, \hat{\theta}_k^t &\le \sigma^2 \sum_{i=1}^{n} \phi_k^2(t_i)(\Delta t_l)^2 \le C\sigma^2 \max_{1 \le l \le n} |\Delta t_l| \int \phi_k^2(x)dx \\ &\le C\sigma^2 \max_{1 \le l \le n} |\Delta t_l| \to 0 . \end{aligned}$$

So, we bound the first term: $\sum_{k=1}^{N} \mathbf{Var}\, \hat{\theta}_k^t \le \epsilon$ for sufficiently large n.

To evaluate the second term, we note first that $\sup_{f \in \Theta'} \sum_{k=1}^{N} (\mathbf{E}_f \hat{\theta}_k^t - \theta_k)^2$ is finite because both $|\hat{\theta}_k^t| \le P$ and $|\theta_k| \le P$. Therefore, for any $\epsilon > 0$ there exists $f_\epsilon \in \Theta'$ such that

$$\begin{aligned} \sup_{f \in \Theta'} \sum_{k=1}^{N} (\mathbf{E}_f \hat{\theta}_k^t - \theta_k(f))^2 &\le \sum_{k=1}^{N} (\mathbf{E}_{f_\epsilon} \hat{\theta}_k^t - \theta_k(f_\epsilon))^2 + \epsilon \\ &\le \sum_{k=1}^{N} \left\{ \sum_{i=1}^{n} \phi_k(t_i) f_\epsilon(t_i) \Delta t_i - \theta_k(f_\epsilon) \right\}^2 + \epsilon , \end{aligned}$$

where $\theta_k(f_\epsilon) = \int f_\epsilon(x)\phi_k(x)dx$. The condition that $\max_{1 \le l \le n} |\Delta t_l| \to 0$ as $n \to \infty$ implies that the integral sum $\sum_{i=1}^{n} \phi_k(t_i) f_\epsilon(t_i) \Delta t_i \to \theta_k(f_\epsilon)$ as $n \to \infty$. So, for sufficiently large n we have that

$$\sup_{f \in \Theta'} \sum_{k=1}^{N} (\mathbf{E}_f \hat{\theta}_k^t - \theta_k)^2 \le 2\epsilon .$$

Thus, for sufficiently large n, we obtain that

$$\sup_{f \in \Theta'} \mathbf{E}_f \|\hat{f}_n - f\|^2 \le 4\epsilon .$$

The last relation guarantees that the estimator $\hat{f}_n$ is consistent. □

Remark 3.1. As one can see from the proof of this theorem, the condition

$$\liminf_{k\to\infty} a_k > 0$$

is necessary for the minimax risk r_n to be finite.

Remark 3.2. Suppose now that we study the problem of the minimax estimation over some set $\mathcal{S} \subseteq L_2$. Then, as the proof of the theorem above implies, the condition that $\mathcal{S}$ has an empty interior is necessary for a consistent (over the set $\mathcal{S}$) estimator to exist. In fact, even stronger result is true. Let S be isomorphic representation of $\mathcal{S}$ in l_2, i.e.

$$S = \{\theta(f) : \ \theta(f) = (\theta_1(f), \theta_2(f), \ldots), \ f \in \mathcal{S}\}.$$

If $S = T \times B$, where B is infinite dimensional and has nonempty interior (in l_2-topology), then there are no uniformly mean square consistent estimators.

Remark 3.3. The first assertion of the theorem also holds with Θ' instead of Θ. Indeed, introduce

$$\Theta'' = \Theta''(Q, P) = \left\{ f(\cdot) \in L_2 : \ \sum_{k=1}^{\infty} b_k^2 \theta_k^2 \leq \min\{Q, P\} \right\},$$

where b_k's are all positive such that $b_k^2 = \max\{a_k^2, 1\}$. Then if $a_k \not\to \infty$ as $k \to \infty$, the sequence $(b_k, \ k = 1, 2 \ldots)$ does not converge to infinity either. Further, since $\Theta'' \subseteq \Theta'$, applying the theorem to $r_n(\Theta'')$, we obtain

$$r_n(\Theta') \geq r_n(\Theta'') > 0.$$

3.3 Main results

As Theorem 3.1 shows, the case $a_k \not\to \infty$ as $k \to \infty$ is not interesting. Therefore we suppose from this point that the sequence $(a_k, \ k = 1, 2, \ldots)$ converges to infinity as $k \to \infty$. Furthermore, since only finitely many zero a_k's are possible in this case, we suppose without loss of generality that this sequence is strictly positive. Indeed, all results below remain valid under the weaker condition $a_k \geq 0$, $k = 1, 2, \ldots$, with minor modifications of some proofs.

From now on we consider the model we started with. Before we formulate the results, let us introduce some notations. Since $a_k \to \infty$ as $k \to \infty$, the equation

$$\sum_{k=1}^{\infty} \sigma^2 a_k (1 - c_n a_k)_+ = c_n Q n \tag{3.6}$$

has a unique solution $c_n = c_n(\sigma^2, \Theta) > 0$. Here x_+ denotes the nonnegative part of x. Denote

$$\mathcal{I} = \mathcal{I}_n(\sigma^2, \Theta) = \{k : \ 0 \le c_n a_k < 1\}, \tag{3.7}$$

$$N = N_n(\sigma^2, \Theta) = \operatorname{card} \mathcal{I}, \tag{3.8}$$

$$d_n = d_n(\sigma^2, \Theta) = n^{-1} \sum_{k=1}^{\infty} \sigma^2 (1 - c_n a_k)_+. \tag{3.9}$$

Note that in fact $\mathcal{I}$ is the set of indices k for which corresponding terms in sum (3.6) are nonzero.

Remark 3.4. Suppose that the sequence $(a_k, \ k = 1, 2, \ldots)$ is nondecreasing. Then from (3.6) it is easy to see that $\mathcal{I} = \{1, 2, \ldots, N\}$ and

$$c_n = \frac{\sum_{k=1}^{N} a_k}{Q n \sigma^{-2} + \sum_{k=1}^{N} a_k^2},$$

where N is the number of nonzero terms in the sum (3.6). One can verify that

$$N = \max\{k : \ a_k \le c_n^{-1}\} = \max\left\{ l : \ \sum_{k=1}^{l} (a_k a_l - a_k^2) \le Q \sigma^{-2} n \right\}. \tag{3.10}$$

Denote next

$$\psi_n(\gamma) = \psi_n(\gamma, \sigma^2, \Theta) = \exp\left\{ \frac{-\gamma n^2 c_n^2}{\sigma^4 \sum_{k=1}^{\infty} a_k^2 (1 - c_n a_k)_+^2} \right\}. \tag{3.11}$$

We introduce also two conditions:

$$\mathcal{F}_1 = \mathcal{F}_1(\sigma^2) : \quad \text{for any } \gamma > 0 \quad \psi_n(\gamma) = o(c_n),$$

$$\mathcal{F}_2 = \mathcal{F}_2(\sigma^2) : \quad c_n \sum_{k \in \mathcal{I}} a_k = o(N).$$

Here c_n, $\mathcal{I}$ and N are defined by (3.6)–(3.8).

The next two theorems give the lower bounds for the minimax risk.

Theorem 3.2. *If condition $\mathcal{F}_1$ or $\mathcal{F}_2$ is fulfilled, then*

$$r_n(\Theta) \geq d_n(\sigma^2, \Theta)(1 + o(1)),$$

where d_n is defined by (3.6), (3.9).

Theorem 3.3. *For any ellipsoid $\Theta(Q)$,*

$$r_n \geq d_n(\sigma^2, \Theta(Q/\pi^2)),$$

where d_n is defined by (3.6), (3.9).

Remark 3.5. Consider the topology generated by the following norm:

$$\|g\| = \left(\sum_{k=1}^{\infty} a_k^2 g_k^2 \right)^{1/2},$$

where the g_k's are Fourier coefficients of $g(\cdot) \in L_2[0,1]$. If we substitute any ball S, $S \subseteq \Theta$, of radius Q in the definition of r_n instead of Θ, then Theorems 3.2 and 3.3 still hold. The proofs are in essence the same. Note that the lower bounds do not depend on the center of the ball S.

Remark 3.6. Although the lower bound in Theorem 3.3 is worse than that in Theorem 3.2, it has the advantage of being nonasymptotic. One can apply this bound to the cases when d_n does not depend on Q at least in the first order term.

Now we construct the estimator which is going to be asymptotically minimax for ellipsoids satisfying certain regularity conditions. Define

$$\hat{f}_n^M(x) = \sum_{k=1}^{n} \lambda_k \hat{\theta}_k \phi_k(x), \tag{3.12}$$

where

$$\hat{\theta}_k = n^{-1} \sum_{i=1}^{n} \phi_k(i/n) Y_i, \tag{3.13}$$

$$\lambda_k = (1 - c_n a_k)_+ \tag{3.14}$$

and c_n is defined by (3.6). We see that the estimator $\hat{f}_n^M(x)$ is a generalized kernel estimator

$$\hat{f}_n^M(x) = \sum_{i=1}^{n} K_n(x, i/n) Y_i,$$

where the kernel $K_n(x, i/n)$ is given by

$$K_n(x, i/n) = n^{-1} \sum_{k=1}^{n} (1 - c_n a_k)_+ \, \phi_k(i/n) \phi_k(x) \, . \tag{3.15}$$

We introduce conditions under either of which we derive an upper bound for the minimax risk:

$$\mathcal{F}_3 : \quad \max_{1 \le k \le n} \sum_{l=1}^{\infty} a_{k+ln}^{-2} = o(n^{-1}) \, ,$$

$$\mathcal{F}_4 = \mathcal{F}_4(\sigma^2) : \quad \sum_{k=n}^{\infty} a_k^{-2} = o(d_n) \, ,$$

where $d_n = d_n(\sigma^2, \Theta)$ is defined by (3.6), (3.9).

Theorem 3.4. *If the condition $\mathcal{F}_3$ or $\mathcal{F}_4$ is fulfilled, then*

$$\sup_{\Theta} \mathbf{E}_f \| \hat{f}_n^M - f \|^2 \le d_n(\sigma^2, \Theta)(1 + o(1)) \, ,$$

where the estimator $\hat{f}_n^M$ is defined by (3.12)–(3.14) and d_n is defined by (3.6), (3.9).

Remark 3.7. Suppose the ϵ_k's are independent random variables (not necessarily Gaussian), all with zero mean and variance σ^2 (the ϵ_k's are, for example, identically distributed), then Theorem 3.4 remains unchanged. If the ϵ_k's are independent random variables, all with zero mean, distribution densities $p_{\epsilon_k}(x)$ and finite Fisher information

$$I_\epsilon = \int \left(\frac{\partial \log p_{\epsilon_k}(x)}{\partial x} \right)^2 p_{\epsilon_k}(x) dx,$$

then Theorems 3.2 and 3.3 still hold with I_ϵ^{-1} in place of σ^2. Thus, only for Gaussian errors do the lower and upper bounds coincide asymptotically. In the general case the lower bound is apparently asymptotically exact and the minimax estimator is likely to be no longer linear (cf. Efromovich (1996) for a related model).

Remark 3.8. Denote $Y = (Y_1, \ldots, Y_n)^T$, $\epsilon = (\epsilon_1, \ldots, \epsilon_n)^T$ and

$$\phi_k = (\phi_k(1/n), \ldots, \phi_k(n/n))^T .$$

We rewrite the model (3.1) in vector form:

$$Y = \sum_{k=1}^{\infty} \theta_k \phi_k + \epsilon .$$

Now we multiply this equality by ϕ_l^T/n, $l = 1, 2, \ldots, n$. Then, using Proposition 3.1 in section 3.5, we get:

$$Z_l = \theta_l + \tilde{\theta}_l + n^{-1/2}\xi_l , \quad l = 1, 2, \ldots, n ,$$

where (see proof of Theorem 3.4)

$$\tilde{\theta}_k = \sum_{l=1}^{\infty}(\theta_{k+2ln} + (-1)^{k+1}\theta_{2ln-k}) , \qquad 1 \le k \le n-1 ,$$

$$\tilde{\theta}_n = \sum_{l=1}^{\infty} \theta_{(2l+1)n} ,$$

$$Z_l = \phi_l^T Y/n \quad \text{and} \quad \xi_l = n^{-1/2}\phi_l^T \epsilon ,$$

i.e. ξ_l, $l = 1, \ldots, n$, are Gaussian random variables with zero means and covariances $\mathbf{E}[\xi_l \xi_k] = \sigma^2 \delta_{lk}$. The regularity conditions $(\mathcal{F}_1 - \mathcal{F}_4)$ imply that, as the proofs of Theorems 3.2 and 3.4 show, the original model and the model (cf. Chapter 2)

$$Z_l' = \theta_l + n^{-1/2}\xi_l , \quad l = 1, 2, \ldots, n ,$$

are asymptotically equivalent in the sense that the best linear estimators and the minimax risks for both models coincide asymptotically. Note that

$$\sum_{k=1}^{\infty}\left(\frac{\sigma^2\lambda_k^2}{n} + (1-\lambda_k)^2\theta_k^2\right)$$

appearing in (3.36) is nothing else than the risk of the linear estimator

$$\hat{\theta}_k' = \lambda_k Z_k' , \qquad k = 1, 2, \ldots, n .$$

We immediately conclude from Theorems 3.2 and 3.4 the following result.

Corollary 3.1. *Let either of conditions $\mathcal{F}_1$, $\mathcal{F}_2$ and either of conditions $\mathcal{F}_3$, $\mathcal{F}_4$ be fulfilled. Then*

$$r_n(\Theta) = d_n(\sigma^2, \Theta)(1 + o(1))$$

and the estimator $\hat{f}_n^M$ is asymptotically minimax.

Consider the problem of robust estimation of the unknown regression function $f(x) \in \Theta$.

Corollary 3.2. *Let either of conditions $\mathcal{F}_1$, $\mathcal{F}_2$ and either of conditions $\mathcal{F}_3$, $\mathcal{F}_4$ be fulfilled. Then*

$$\inf_{\hat{f}_n} \sup_{\Theta} \sup_{p_\epsilon \in \Pi} \mathbf{E}_{f,p_\epsilon} \|\hat{f}_n - f\|^2 = d_n(\sigma^2, \Theta)(1 + o(1)),$$

where Π is the set of all distributions of noises with zero mean and variance σ^2.

Proof. On the one hand,

$$\begin{aligned}\inf_{\hat{f}_n} \sup_{\Theta} \sup_{p_\epsilon \in \Pi} \mathbf{E}_{f,p_\epsilon} \|\hat{f}_n - f\|^2 &\geq \inf_{\hat{f}_n} \sup_{\Theta} \mathbf{E}_f \|\hat{f}_n - f\|^2 \\ &\geq d_n(\sigma^2, \Theta)(1 + o(1)),\end{aligned}$$

where p_ϵ in the right-hand side is taken to be Gaussian. On the other hand, according to Remark 3.7, we have

$$\begin{aligned}\inf_{\hat{f}_n} \sup_{\Theta} \sup_{p_\epsilon \in \Pi} \mathbf{E}_{f,p_\epsilon} \|\hat{f}_n - f\|^2 &\leq \sup_{\Theta} \sup_{p_\epsilon \in \Pi} \mathbf{E}_{f,p_\epsilon} \|\hat{f}_n^M - f\|^2 \\ &\leq d_n(\sigma^2, \Theta)(1 + o(1)),\end{aligned}$$

where $\hat{f}_n^M$ is the estimator defined by (3.12)–(3.14). □

3.4 Examples

If an ellipsoid Θ is such that for some positive constant $C = C(\Theta)$ and positive decreasing to zero sequence ϕ_n the asymptotic relation

$$r_n(\Theta) = C(\Theta)\phi_n^2(1 + o(1))$$

holds, then, clearly, ϕ_n is the rate of convergence and the constant $C(\Theta)$ is optimal. We describe below examples where this is the case.

Example 3.1. Let, for a given α, $\alpha > 1/2$,

$$\Theta = \Theta(Q) = \left\{ f(\cdot) \in L_2 : \sum_{k=1}^{\infty} k^{2\alpha}(\theta_{2k}^2 + \theta_{2k+1}^2) \leq Q \right\}. \tag{3.16}$$

We have to impose the condition $\alpha > 1/2$ in order to ensure Θ satisfies conditions $\mathcal{F}_1$ and $\mathcal{F}_3$.

Corollary 3.3. *Let the ellipsoid* Θ *be defined by (3.16). Then*

$$r_n = Q^{\frac{1}{2\alpha+1}} \sigma^{\frac{4\alpha}{2\alpha+1}} \left(\frac{2\alpha}{\alpha+1}\right)^{\frac{2\alpha}{2\alpha+1}} (2\alpha+1)^{\frac{1}{2\alpha+1}} n^{-\frac{2\alpha}{2\alpha+1}} (1+o(1))$$

and the estimator $\hat{f}_n^M$ *is asymptotically minimax.*

Proof. Condition $\mathcal{F}_3$ can be verified straightforwardly:

$$\max_{1\le k\le n} \sum_{k=1}^{n} a_{k+ln}^{-2} \le 2^{2\alpha} \sum_{l=1}^{\infty} (ln)^{-2\alpha} \le Cn^{-2\alpha} = o(n^{-1}).$$

We calculate now the asymptotic value of d_n. Since in this case $N \to \infty$, it is easy to see from the first equality of (3.10) that $c_n N^\alpha \to 1$ as $n \to \infty$. Therefore, $N = c_n^{-1/\alpha}(1+o(1))$. The equation (3.6) to define c_n is as follows:

$$2\sum_{k=1}^{N} (k^\alpha - c_n k^{2\alpha}) = Q\sigma^{-2} n c_n .$$

Note that here N corresponds in fact to $N/2$ for N defined by general formula (3.8) (or (3.10)). This is more convenient for computations.

Making use of the asymptotic equality

$$\sum_{k=1}^{M} k^\alpha = \frac{M^{\alpha+1}}{\alpha+1}(1+o(1)) \quad \text{as} \quad M \to \infty, \quad \alpha > -1, \tag{3.17}$$

we obtain the asymptotic relations:

$$c_n = \left(\frac{2\alpha\sigma^2}{(\alpha+1)(2\alpha+1)Qn}\right)^{\frac{\alpha}{2\alpha+1}} (1+o(1)),$$

$$N = \left(\frac{(2\alpha+1)(\alpha+1)Qn}{2\alpha\sigma^2}\right)^{\frac{1}{2\alpha+1}} (1+o(1)). \tag{3.18}$$

Using this, (3.9) and again (3.17), we find that

$$d_n = n^{-\frac{2\alpha}{2\alpha+1}} Q^{\frac{1}{2\alpha+1}} (2\alpha+1)^{\frac{1}{2\alpha+1}} \left(\frac{2\alpha\sigma^2}{\alpha+1}\right)^{\frac{2\alpha}{2\alpha+1}} (1+o(1)).$$

Next, one makes sure easily that $\Theta \in \mathcal{F}_1$:

$$\psi_n(\gamma) = O\left(\exp\left\{-\gamma_1 n^{1/(2\alpha+1)}\right\}\right) = o(c_n)$$

for some $\gamma_1 > 0$. Finally, applying Corollary 3.1, we get the statement of this corollary. □

Remark 3.9. Let, for a natural number α and $f(\cdot) \in L_2([0,1])$, $D^\alpha f$ denote the derivative of order α in distributional sense, and let

$$\widetilde{W}_2^\alpha(Q) = \left\{ f \in L_2 : \ \|D^\alpha f\|^2 \le Q, \ D^l f(0) = D^l f(1), \ l = 0, \dots, \alpha - 1 \right\}$$

be the αth order periodic Sobolev space on the unit interval. Then the following asymptotic equation holds (cf. Nussbaum (1985) and Golubev and Nussbaum (1990)):

$$\begin{aligned} r_n(\widetilde{W}_2^\alpha(Q)) &= r_n(\Theta(Q/(2\pi)^{2\alpha}))(1+o(1)) \\ &= \gamma(\alpha, Q)\sigma^{\frac{4\alpha}{2\alpha+1}} n^{-\frac{2\alpha}{2\alpha+1}}(1+o(1)), \end{aligned}$$

where $\Theta(Q)$ is defined by (3.16) and

$$\gamma(\alpha, Q) = (Q(2\alpha+1))^{\frac{1}{2\alpha+1}} \left(\frac{\alpha}{\pi(\alpha+1)} \right)^{\frac{2\alpha}{2\alpha+1}}$$

is Pinsker's constant. Indeed, the upper bound follows from the relation:

$$\widetilde{W}_2^\alpha(Q) \subseteq \Theta(Q/(2\pi)^{2\alpha}).$$

The proof of the lower bound carries through literally since

$$\sum_{k=1}^N \theta_k \phi_k(x) \in \widetilde{W}_2^\alpha(Q) \qquad \text{if} \qquad \theta \in \Theta(Q/(2\pi)^{2\alpha}).$$

Thus, the nonparametric class (3.16) can be viewed as an extension of the class $\widetilde{W}_2^\alpha$ for nonperiodic functions and nonnatural α.

For the ellipsoid defined by (3.16), it is not difficult to get the following expression for the kernel (3.15):

$$K_n(x, i/n) = n^{-1} \left(1 + 2 \sum_{k=1}^N (1 - c_n k^\alpha) \cos(2\pi k(x - i/n)) \right).$$

Consider estimator $\tilde{f}_n^M$ defined by (3.12)-(3.14) with $c_n = N_n^{-\alpha}$, where N_n is an arbitrary sequence satisfying (3.18). We claim that this estimator is also asymptotically minimax over the class (3.16). Indeed, following the proof of Theorem 3.4, one obtains

$$R_n(\tilde{f}_n^M) = \sup_\Theta \left\{ \sum_{k=1}^n \left(n^{-1}\sigma^2\lambda_k^2 + (1-\lambda_k)^2\theta_k^2 \right) \right\} + o(n^{-1}).$$

Further, (3.2) and (3.14) imply that

$$\begin{aligned}
&\sup_{\Theta}\left\{\sum_{k=1}^{n}\left(n^{-1}\sigma^2\lambda_k^2+(1-\lambda_k)^2\theta_k^2\right)\right\}\\
&= n^{-1}\sum_{k=1}^{n}\sigma^2\lambda_k^2+\sup_{\Theta}\left\{\sum_{k=1}^{n}\frac{(1-\lambda_k)^2}{a_k^2}\theta_k^2 a_k^2\right\}\\
&\le n^{-1}\sum_{k=1}^{\infty}\sigma^2\lambda_k^2+Q\sup_{k\ge 1}(1-\lambda_k)^2/a_k^2\\
&\le n^{-1}\sum_{k=1}^{\infty}\sigma^2(1-c_n a_k)_+^2+Qc_n^2\,.
\end{aligned}$$

Combining the last relations and using (3.17), one computes

$$\begin{aligned}
R_n(\tilde f_n^M) &\le QN_n^{-2\alpha}+2n^{-1}\sum_{k=1}^{N_n}\sigma^2(1-k^\alpha N_n^{-\alpha})^2+O(n^{-1})\\
&= n^{-\frac{2\alpha}{2\alpha+1}}(Q(2\alpha+1))^{\frac{1}{2\alpha+1}}\left(\frac{2\alpha\sigma^2}{\alpha+1}\right)^{\frac{2\alpha}{2\alpha+1}}(1+o(1))\\
&= d_n(1+o(1))\,.
\end{aligned}$$

So, we have shown that the estimator $\tilde f_n^M$ is asymptotically minimax. The necessity of considering this estimator stems from the fact that, for the ellipsoid defined by (3.16), in principle one can deduce the formula for the kernel corresponding to the estimator $\tilde f_n^M$.

For the ellipsoid (3.16) with $\alpha = 1$, by routine calculations we obtain the expression for the kernel corresponding to the estimator $\tilde f_n^M$:

$$K_n(x,i/n)=\frac{\sin^2(N_n\pi(x-i/n))}{nN_n\sin^2(\pi(x-i/n))},$$

which is the well known Feier kernel. For $\alpha = 2$, the kernel of the estimator $\tilde f_n^M$ is as follows:

$$\begin{aligned}
K_n(x,i/n) &= \frac{\sin(2N_n\pi(x-i/n))\cos(\pi(x-i/n))}{2nN_n^2\sin^3(\pi(x-i/n))}\\
&\quad-\frac{\cos(2N_n\pi(x-i/n))}{nN_n\sin^2(\pi(x-i/n))}\,.
\end{aligned}$$

As α increases, the calculations become more involved.

Example 3.2. For some $\beta > 0$, let

$$\Theta = \left\{ f(\cdot) \in L_2 : \sum_{k=1}^{\infty} e^{2\beta k} (\theta_{2k}^2 + \theta_{2k+1}^2) \le Q \right\}. \tag{3.19}$$

In this case it has been possible to describe the minimax risk up to the rate of the second order term.

Corollary 3.4. *Let the ellipsoid Θ be defined by (3.19). Then*

$$r_n = \frac{\sigma^2 \log n}{\beta n} + O(n^{-1}) \tag{3.20}$$

and the estimator $\hat{f}_n^M$ is asymptotically minimax and also second order minimax with respect to the rate of convergence.

Proof. From (3.7) and (3.19) it follows that

$$e^{\beta} \le c_n e^{N\beta} \le 1\,. \tag{3.21}$$

Write equation (3.6) for this case:

$$2 \sum_{k=1}^{N} (e^{\beta k} - c_n e^{2\beta k}) = Q \sigma^{-2} n c_n\,.$$

The last two relations yield the following asymptotics for N:

$$N = \frac{\log n}{2\beta} + O(1).$$

According to Theorem 3.3 and the proof of Theorem 3.4, we have

$$d_n(\sigma^2, \Theta(Q/\pi^2)) \le r_n \le d_n(\sigma^2, \Theta(Q)) + \delta_n\,, \tag{3.22}$$

where

$$|\delta_n| \le Q a_n^{-2} + 2Q \sum_{k=n+1}^{\infty} a_k^{-2} = O\left(e^{-\beta n}\right).$$

Further, the asymptotics for N and (3.21) imply that

$$\begin{aligned} d_n &= \frac{2\sigma^2}{n} \sum_{k=1}^{N} (1 - c_n e^{\beta k}) \\ &= \frac{\sigma^2}{\beta} \frac{\log n}{n} + O(n^{-1})\,. \end{aligned}$$

From this and (3.22) we finally obtain

$$r_n = \frac{\sigma^2 \log n}{\beta n} + O(n^{-1}).$$

□

The first order term of the minimax risk does not depend on the "size" Q of ellipsoid Θ. That is why a stronger result is available. Namely,

Corollary 3.5. *Let the ellipsoid Θ be defined by (3.19). Then for any neighbourhood $V \subseteq \Theta$*

$$\inf_{\hat{f}_n} \sup_{V} \mathbf{E}_f \|\hat{f}_n - f\|^2 = \frac{\sigma^2 \log n}{\beta n} + O(n^{-1}),$$

where the meant topology is generated by the norm defined in Remark 3.5.

Proof. Indeed, let S be a ball such that $S \subseteq V$. Then, on the one hand, according to Remark 3.5, we have

$$\begin{aligned} \inf_{\hat{f}_n} \sup_{V} \mathbf{E}_f \|\hat{f}_n - f\|^2 &\geq \inf_{\hat{f}_n} \sup_{S} \mathbf{E}_f \|\hat{f}_n - f\|^2 \\ &= \frac{\sigma^2 \log n}{\beta n} + O(n^{-1}) \end{aligned}$$

and, on the other hand,

$$\begin{aligned} \inf_{\hat{f}_n} \sup_{V} \mathbf{E}_f \|\hat{f}_n - f\|^2 &\leq \inf_{\hat{f}_n} \sup_{\Theta} \mathbf{E}_f \|\hat{f}_n - f\|^2 \\ &= \frac{\sigma^2 \log n}{\beta n} + O(n^{-1}). \end{aligned}$$

□

Note that the second order term of the local minimax risk

$$\inf_{\hat{f}_n} \sup_{V} \mathbf{E}_f \|\hat{f}_n - f\|^2$$

certainly depends on neighbourhood V.

Remark 3.10. Let the ellipsoid Θ be defined by (3.19). Consider the projection estimator $\hat{f}_n^P$ defined by (3.12) with

$$\lambda_k = \begin{cases} 1, & k \le N_n \\ 0, & k > N_n, \end{cases}$$

where N_n is any positive sequence satisfying the inequality:

$$\left| N_n - \beta^{-1} \log n \right| \le (1-\mu)\beta^{-1} \log\log n \quad \text{for some} \quad \mu > 0 .$$

The estimator $\hat{f}_n^P$, while being simpler than the estimator $\hat{f}_n^M$ above, is still asymptotically minimax, i.e.

$$R_n(\hat{f}_n^P) = \frac{\sigma^2 \log n}{\beta n}(1+o(1)) .$$

If $N_n = \beta^{-1} \log n$, then the estimator $\hat{f}_n^P$ is asymptotically second order minimax:

$$R_n(\hat{f}_n^P) = \frac{\sigma^2 \log n}{\beta n} + \frac{Q}{n}(1+o(1)) .$$

On the other hand, consider the estimator $\hat{f}_n^{VP}$ corresponding to the de la Vallee Poussin kernel (cf. Ibragimov and Hasminskii (1982)) which is estimator (3.12) with

$$\lambda_k = \begin{cases} 1, & k \le N_n/2, \\ \dfrac{N_n - k}{N_n/2}, & 1 + N_n/2 \le k \le N_n, \\ 0, & k > N_n . \end{cases}$$

One can choose the sequence N_n optimally as

$$N_n = \frac{\log n}{2\beta} .$$

It is well known (see Ibragimov and Hasminskii (1982)) that such an estimator allows one to obtain the optimal rates, with properly chosen N_n, for all nonparametric classes considered above. However, this estimator is not asymptotically minimax as one can see by comparing (3.20) to the maximal risk of the estimator $\hat{f}_n^{VP}$:

$$R_n(\hat{f}_n^{VP}) = \frac{4}{3} \frac{\sigma^2 \log n}{\beta n} (1+o(1)) .$$

Corollary 3.6. *Let the ellipsoid Θ be defined by (3.19). Then the estimator $\hat{f}_n^P$ defined in Remark 3.10 is locally asymptotically minimax and adaptive with respect to σ^2.*

Proof. We take for example $N_n = \beta^{-1} \log n$ and while constructing the estimator $\hat{f}_n^P$ we need not know σ^2 and neighbourhood. Now the statement of this corollary follows from the previous corollary and the expression for the maximal risk $R_n(\hat{f}_n^P)$. □

The kernel corresponding to the estimator $\hat{f}_n^P$ has the following form:

$$K_n(x, i/n) = \frac{\sin((2N_n + 1)\pi(x - i/n))}{n \sin(\pi(x - i/n))}.$$

3.5 Proofs

Denote

$$\phi_k = (\phi_k(1/n), \ldots, \phi_k(n/n))^T.$$

Proposition 3.1.

$$\begin{aligned} \phi_{k+mn} &= \begin{cases} \phi_k, & m = 2l, \quad l = 1, 2, \ldots \\ (-1)^{n-k+1}\phi_{n-k}, & m = 2l - 1, \quad l = 1, 2, \ldots ; \end{cases} \\ \phi_k^T \phi_l &= n\,\delta_{kl}, \quad 1 \le k \le n, \qquad 1 \le l \le n, \end{aligned}$$

where $\phi_0 = 0$ and δ_{kl} is defined as follows

$$\delta_{kl} = \begin{cases} 1, & k = l \\ 0, & k \neq l. \end{cases}$$

We skip the elementary proof of this proposition.

Proof of Theorem 3.2. The proof of this theorem is closely similar to that of corresponding assertions in Chapter 2. We assume that $\hat{f}$ is an estimator with realizations in L_2 because otherwise the statement of the theorem becomes trivial. Then we have by the Parseval identity:

$$r_n = \inf_{\hat{f}} \sup_{\Theta} \mathbf{E}_f \|\hat{f} - f\|^2 = \inf_{\hat{\theta}} \sup_{\Theta} \mathbf{E}_f \sum_{k=1}^{\infty} (\hat{\theta}_k - \theta_k)^2. \qquad (3.23)$$

First we consider a somewhat simpler case: condition $\mathcal{F}_2$ is fulfilled. Now we repeat the reasoning as in the proof of Theorem 3.1. Let m_k, $k = 1, 2, \ldots$, be a set of positive numbers such that

$$\sum_{k=1}^{\infty} a_k^2 m_k^2 \leq Q,$$

i.e. $m = (m_1, m_2, \ldots) \in \Theta$. Introduce

$$\nu_k(x) = (1/m_k)\nu_0(x/m_k), \quad k = 1, 2, \ldots ,$$

where $\nu_0(x)$ is a probability density on the interval $[-1, 1]$ with a finite Fisher information I_0,

$$\nu_0(-1) = \nu_0(1) = 0$$

and $\nu_0(x)$ is continuously differentiable for $|x| < 1$. Recall that, under these conditions, the minimal Fisher information is π^2. So, $I_0 \geq \pi^2$. The functions $\nu_k(x)$ are probability densities with supports $[-m_k, m_k]$ and the Fisher informations

$$I(\nu_k) = I_0 m_k^{-2},$$

respectively.

Let the measure μ be such that θ_k, $k = 1, 2, \ldots$, are distributed independently with densities $\nu_k(x)$, $k = 1, 2, \ldots$, respectively. The assumption $m \in \Theta$ ensures that $\operatorname{supp} \mu \subseteq \Theta$.

Write $\mathbf{E}$ for the expectation with respect to the joint distribution of $Y_1, \ldots, Y_n$ and $\theta_1, \theta_2, \ldots$. From (3.23) it follows that

$$r_n \geq \inf_{\hat{\theta}} \int_{\Theta} \sum_{k=1}^{\infty} \mathbf{E}_f(\hat{\theta}_k - \theta_k)^2 d\mu(\theta) = \inf_{\hat{\theta}} \sum_{k=1}^{\infty} \mathbf{E}(\hat{\theta}_k - \theta_k)^2 . \tag{3.24}$$

To estimate $\mathbf{E}(\hat{\theta}_k - \theta_k)^2$, we apply the van Trees inequality (Theorem A.1):

$$\mathbf{E}(\hat{\theta}_k - \theta_k)^2 \geq \frac{1}{\mathbf{E}I(\theta_k) + I(\nu_k)} , \tag{3.25}$$

where $I(\theta_k)$ is the Fisher information about parameter θ_k contained in observations $Y_1, \ldots, Y_n$ and

$$\begin{aligned} \mathbf{E}I(\theta_k) &= \int \mathbf{E}_f \left[\sum_{i=1}^{n} \frac{\partial \log p_\epsilon(Y_i - f(i/n))}{\partial \theta_k} \right]^2 d\nu_k(\theta_k) \\ &= \sigma^{-2} \sum_{i=1}^{n} \phi_k^2(i/n) = \sigma^{-2} n , \end{aligned}$$

which follows from Proposition 3.1. Recalling that $I(\nu_k) = I_0 m_k^{-2}$, we obtain

$$\mathbf{E}(\hat{\theta}_k - \theta_k)^2 \geq \frac{1}{I_0 m_k^{-2} + \sigma^{-2} n} = \frac{\sigma^2 n^{-1} m_k^2 I_0^{-1}}{m_k^2 I_0^{-1} + \sigma^2 n^{-1}} .$$

We make use of the last inequality and (3.24):

$$r_n \geq \frac{\sigma^2}{n} \sum_{k=1}^{\infty} \frac{m_k^2 I_0^{-1}}{m_k^2 I_0^{-1} + \sigma^2 n^{-1}} . \tag{3.26}$$

The inequality (3.26) holds for any $m \in \Theta$. At this point take

$$m_k^2 = \frac{\sigma^2 (1 - c_n a_k)_+}{c_n a_k n}, \tag{3.27}$$

where c_n is defined by (3.6). Now note that equation (3.6) can be also rewritten as

$$\sum_{k=1}^{\infty} a_k^2 m_k^2 = Q. \tag{3.28}$$

So, $m \in \Theta$. Substituting this particularly chosen m in (3.26) results in

$$\begin{aligned} r_n &\geq \frac{\sigma^2}{n} \sum_{k=1}^{\infty} \frac{m_k^2 I_0^{-1}}{m_k^2 I_0^{-1} + \sigma^2 n^{-1}} \\ &= \frac{N\sigma^2}{n} - \frac{\sigma^2}{n} \sum_{k \in \mathcal{I}} \frac{c_n a_k}{(1 - I_0^{-1}) c_n a_k + I_0^{-1}} , \end{aligned}$$

where set $\mathcal{I}$ is defined by (3.7). Combining this with (3.9) and condition $\Theta \in \mathcal{F}_2$, we finally get

$$\begin{aligned} r_n &\geq \frac{N\sigma^2}{n} - \frac{c_n I_0 \sigma^2}{n} \sum_{k \in \mathcal{I}} a_k \\ &= \frac{N\sigma^2}{n} (1 + o(1)) \\ &= d_n(\sigma^2, \Theta)(1 + o(1)), \end{aligned}$$

which proves the first part of the theorem.

Suppose now that condition $\mathcal{F}_1$ is fulfilled. For arbitrary $0 < \delta < 1$ we can find $R_\delta > 0$ and a probability density $\nu_\delta(x)$ such that $\nu_\delta(x)$ is positive and continuously differentiable inside the interval $(-R_\delta, R_\delta)$, equals to zero outside this interval, has finite Fisher information $I(\nu_\delta)$ and satisfies the following properties:

$$\mathbf{E}X^2 = 1 - \delta/2$$

and

$$I(\nu_\delta) = \int_{-R_\delta}^{R_\delta} \frac{(\nu_\delta'(x))^2}{\nu_\delta(x)} dx \le 1 + \delta\,,$$

where X is a random variable with probability density $\nu_\delta(x)$. For a complete explanation why it is possible to choose a density with such properties, see the proof of Theorem 2.2.

Introduce for arbitrary $m_k > 0$, $k = 1, 2, \ldots$,

$$\nu_k(x) = m_k^{-1}\nu_0(m_k^{-1}x), \quad k = 1, 2, \ldots\,.$$

These are probability densities with supports $(-R_\delta m_k, m_k R_\delta)$ respectively and if $X_k = m_k X$ then X_k is a random variable with density $\nu_k(x)$. We have

$$\begin{aligned} \mathbf{E}X_k^2 &= m_k^2(1 - \delta/2) &\quad (3.29)\\ I(\nu_k) &= m_k^{-2} I(\nu_\delta) \le m_k^{-2}(1 + \delta). &\quad (3.30) \end{aligned}$$

Let θ be distributed according to a prior measure μ such that θ_k, $k = 1, 2, \ldots$, are distributed independently with the densities $\nu_k(x)$, $k = 1, 2, \ldots$, respectively. In view of (3.23), we evaluate the minimax risk

$$\begin{aligned} r_n &\ge \inf_{\hat\theta} \int_\Theta \sum_{k=1}^\infty \mathbf{E}_\theta(\hat\theta_k - \theta_k)^2 d\mu(\theta)\\ &= \inf_{\hat\theta \in supp\,\mu} \int_\Theta \sum_{k=1}^\infty \mathbf{E}_\theta(\hat\theta_k - \theta_k)^2 d\mu(\theta)\\ &\ge \inf_{\hat\theta} \sum_{k=1}^\infty \mathbf{E}(\hat\theta_k - \theta_k)^2 - \sup_{\hat\theta \in supp\,\mu} \int_{\Theta^C} \sum_{k=1}^\infty \mathbf{E}_\theta(\hat\theta_k - \theta_k)^2 d\mu(\theta)\\ &\ge \inf_{\hat\theta} \sum_{k=1}^\infty \mathbf{E}(\hat\theta_k - \theta_k)^2 - 4R_\delta^2 \mu(\Theta^C) \sum_{k=1}^\infty m_k^2\,. &\quad (3.31) \end{aligned}$$

Due to the assumptions on probability density $\nu_k(x)$, we can apply the van Trees inequality (3.25) to the Bayes risk $\mathbf{E}(\hat\theta_k - \theta_k)^2$. Thus, by (3.25) and (3.30), we obtain

$$\inf_{\hat\theta}\sum_{k=1}^{\infty}\mathbf{E}(\hat\theta_k-\theta_k)^2 \geq \frac{\sigma^2}{n(1+\delta)}\sum_{k=1}^{\infty}\frac{m_k^2}{m_k^2(1+\delta)^{-1}+\sigma^2 n^{-1}}\,. \tag{3.32}$$

Suppose now that $\sum_{k=1}^{\infty} a_k^2 m_k^2 \leq Q$, i.e. $m=(m_1,m_2,\ldots)\in\Theta$. Then, by (3.29), we have

$$\begin{aligned} |a_k^2(\theta_k^2-\mathbf{E}\theta_k^2)| &\leq a_k^2 m_k^2 |R_\delta^2 - 1 + \delta/2|,\\ Q-\sum_{k=1}^{\infty} a_k^2\mathbf{E}\theta_k^2 &= Q-(1-\delta/2)\sum_{k=1}^{\infty} a_k^2 m_k^2 \geq Q\delta/2\,. \end{aligned}$$

Using these relations and the Hoeffding's inequality (see Pollard (1984)), we evaluate $\mu(\Theta^C)$:

$$\begin{aligned} \mu(\Theta^C) &= \mu\left\{\sum_{k=1}^{\infty} a_k^2(\theta_k^2-\mathbf{E}\theta_k^2) > Q-\sum_{k=1}^{\infty} a_k^2\mathbf{E}\theta_k^2\right\}\\ &\leq \exp\left\{\frac{-\gamma}{\sum_{k=1}^{\infty} a_k^4 m_k^4}\right\}, \end{aligned} \tag{3.33}$$

where

$$\gamma=\frac{(Q\delta)^2}{8(R_\delta^2-1+\delta/2)^2}.$$

Take now m according to (3.27). Recall that, by (3.28), $m\in\Theta$. By the definition of $\psi_n(\gamma)$ (3.11), we see that the right hand side of the inequality (3.33) becomes $\psi_n(\gamma)$. Therefore, combining (3.31), (3.32) and (3.33) gives

$$r_n \geq \frac{\sigma^2}{n(1+\delta)}\sum_{k=1}^{\infty}\frac{m_k^2}{m_k^2(1+\delta)^{-1}+\sigma^2 n^{-1}} - 4R_\delta^2\psi_n(\gamma)\sum_{k=1}^{\infty} m_k^2\,. \tag{3.34}$$

According to the condition $\Theta\in\mathcal{F}_1$,

$$\psi_n(\gamma)=o(c_n)$$

as $n\to\infty$. From this, (3.9) and (3.27) it follows that, as $n\to\infty$,

$$\begin{aligned} \psi_n(\gamma)\sum_{k=1}^{\infty} m_k^2 &= \frac{\psi_n(\gamma)}{n}\sum_{k=1}^{\infty}\frac{\sigma^2(1-c_n a_k)_+}{c_n a_k}\\ &= o(1)n^{-1}\sum_{k=1}^{\infty}\sigma^2(1-c_n a_k)_+ = o(d_n)\,. \end{aligned}$$

Therefore, substituting now m defined by (3.27) in (3.34) leads to

$$\begin{aligned} r_n &\geq \frac{\sigma^2}{n(1+\delta)} \sum_{k=1}^{\infty} \frac{m_k^2}{m_k^2(1+\delta)^{-1} + \sigma^2 n^{-1}} + R_\delta^2 o(d_n) \\ &\geq \frac{\sigma^2}{n(1+\delta)} \sum_{k=1}^{\infty} \frac{m_k^2}{m_k^2 + \sigma^2 n^{-1}} + R_\delta^2 o(d_n) \\ &= (1+\delta)^{-1} d_n + R_\delta^2 o(d_n) \end{aligned}$$

as $n \to \infty$. Since this inequality holds for any $\delta \in (0,1)$, the theorem follows. □

Proof of Theorem 3.3. Because we do not use any condition of Theorem 3.2 up to (3.26), we invoke the inequality (3.26) with $I_0 = \pi^2$. Recall that (3.26) holds for any $m \in \Theta(Q)$ and therefore

$$\begin{aligned} r_n &\geq \sup_{m \in \Theta(Q)} \frac{\sigma^2}{n} \sum_{k=1}^{\infty} \frac{m_k^2/\pi^2}{m_k^2/\pi^2 + \sigma^2 n^{-1}} \\ &= \sup_{m \in \Theta(Q/\pi^2)} \frac{\sigma^2}{n} \sum_{k=1}^{\infty} \frac{m_k^2}{m_k^2 + \sigma^2 n^{-1}} . \end{aligned}$$

Finally, applying Lemma A.1 with $\epsilon^2 = n^{-1}$ and $\sigma_k^2 = \sigma^2$, completes the proof of the theorem. □

Proof of Theorem 3.4. Since, by (3.12), (3.13), (3.14) and the second property of Proposition 3.1,

$$\begin{aligned} &\mathbf{E}_f(\lambda_k \hat{\theta}_k - \theta_k)^2 \\ &= \frac{\sigma^2 \lambda_k^2}{n} + (1-\lambda_k)^2 \theta_k^2 + 2(\lambda_k - 1)\lambda_k \theta_k \tilde{\theta}_k + \lambda_k^2 \tilde{\theta}_k^2 , \end{aligned} \tag{3.35}$$

where

$$\tilde{\theta}_k = \mathbf{E}_f \hat{\theta}_k - \theta_k = n^{-1} \sum_{m=1}^{n} \phi_k(m/n) f(m/n) - \theta_k ,$$

we bound the risk of the estimator $\hat{f}_n^M$ as follows:

$$\sup_\Theta \mathbf{E}_f \|\hat{f}_n^M - f\|^2 = \sup_\Theta \left\{ \sum_{k=1}^{n} \mathbf{E}_f(\lambda_k \hat{\theta}_k - \theta_k)^2 + \sum_{k=n+1}^{\infty} \theta_k^2 \right\}$$

$$\begin{aligned} \leq\ & \sup_{\Theta}\left\{\sum_{k=1}^{\infty}\left(\frac{\sigma^2\lambda_k^2}{n}+(1-\lambda_k)^2\theta_k^2\right)\right\} \\ & + 2\sup_{\Theta}\left\{\sum_{k=1}^{n}\left((\lambda_k-1)\lambda_k\theta_k\tilde{\theta}_k+\lambda_k^2\tilde{\theta}_k^2\right)\right\} \\ & + \sup_{\Theta}\left\{\sum_{k=n+1}^{\infty}\theta_k^2\right\}. \end{aligned} \tag{3.36}$$

According to Lemma A.1, the first term of the right hand side of the last inequality is exactly $d_n(\sigma^2, \Theta)$. For completeness, we give a direct evaluation of this term. Rewrite (3.6) in the form

$$n^{-1}\sum_{k=1}^{\infty}\sigma^2 c_n a_k(1-c_n a_k)_+ = c_n^2 Q$$

and notice

$$\begin{aligned} & Qc_n^2 + n^{-1}\sum_{k=1}^{\infty}\sigma^2(1-c_n a_k)_+^2 \\ = \ & n^{-1}\sum_{k=1}^{\infty}\sigma^2((c_n a_k(1-c_n a_k)_+ + (1-c_n a_k)_+^2)) \\ = \ & n^{-1}\sum_{k=1}^{\infty}\sigma^2(1-c_n a_k)_+ = d_n\,. \end{aligned}$$

Therefore, by (3.9) and (3.14),

$$\begin{aligned} \sup_{\Theta}\left\{\sum_{k=1}^{\infty}\left(\frac{\sigma^2\lambda_k^2}{n}+(1-\lambda_k)^2\theta_k^2\right)\right\} & \leq Q\sup_{k\geq 1}(1-\lambda_k)^2/a_k^2 + \frac{\sigma^2}{n}\sum_{k=1}^{\infty}\lambda_k^2 \\ & \leq Qc_n^2 + \frac{\sigma^2}{n}\sum_{k=1}^{\infty}(1-c_n a_k)_+^2 \\ & = d_n\,. \end{aligned} \tag{3.37}$$

Consequently, it is sufficient to show

$$\sup_{\Theta}\left\{\sum_{k=1}^{n}\left((\lambda_k-1)\lambda_k\theta_k\tilde{\theta}_k+\lambda_k^2\tilde{\theta}_k^2\right)\right\} = o(d_n)\,, \tag{3.38}$$

$$\sup_{\Theta}\left\{\sum_{k=n+1}^{\infty}\theta_k^2\right\} = o(d_n)\,.$$

The last relation follows immediately from the conditions $\mathcal{F}_3$ and $\mathcal{F}_4$:

$$\sup_{\Theta}\left\{\sum_{k=n+1}^{\infty}\theta_k^2\right\} \le \sup_{\Theta}\left\{\sum_{k=n+1}^{\infty}\theta_k^2 a_k^2\right\}\max_{k>n} a_k^{-2} = o(d_n)\,.$$

Suppose we have the following relations:

$$\sup_{\Theta}\sum_{k=1}^{n}\lambda_k^2\tilde{\theta}_k^2 = o(d_n), \tag{3.39}$$

$$Qc_n^2 \le d_n. \tag{3.40}$$

Then, taking into account that $\lambda_k = (1-c_n a_k)_+$, we prove (3.38) by the Cauchy-Schwarz inequality:

$$\begin{aligned}
&\sup_{\Theta}\left\{\sum_{k=1}^{n}\left((\lambda_k-1)\lambda_k\theta_k\tilde{\theta}_k + \lambda_k^2\tilde{\theta}_k^2\right)\right\} \\
&\le \sup_{\Theta}\left\{\left(\sum_{k=1}^{\infty}a_k^2\theta_k^2\right)^{1/2} c_n\left(\sum_{k=1}^{n}\lambda_k^2\tilde{\theta}_k^2\right)^{1/2}\right\} + \sup_{\Theta}\sum_{k=1}^{n}\lambda_k^2\tilde{\theta}_k^2 \\
&\le \sqrt{d_n}\sup_{\Theta}\left(\sum_{k=1}^{n}\lambda_k^2\tilde{\theta}_k^2\right)^{1/2} + \sup_{\Theta}\sum_{k=1}^{n}\lambda_k^2\tilde{\theta}_k^2 = o(d_n)\,.
\end{aligned}$$

It remains to show (3.39) and (3.40). The relation (3.40) follows immediately from (3.37). Let us prove (3.39). For any k such that $1 \le k \le n-1$, by Proposition 3.1 and the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
\tilde{\theta}_k &= n^{-1}\sum_{m=1}^{n}\phi_k(m/n)f(m/n) - \theta_k \\
&= n^{-1}\sum_{l=1}^{\infty}\theta_l\sum_{m=1}^{n}\phi_k(m/n)\phi_l(m/n) - \theta_k \\
&= \sum_{l=n+1}^{\infty}\theta_l\, n^{-1}\sum_{m=1}^{n}\phi_k(m/n)\phi_l(m/n) \\
&= \sum_{l=1}^{\infty}(\theta_{k+2ln} + (-1)^{k+1}\theta_{2ln-k})
\end{aligned}$$

and

$$\tilde{\theta}_n = \sum_{l=1}^{\infty} \theta_{(2l+1)n} \,.$$

Therefore,

$$\begin{aligned}\tilde{\theta}_k &\le \left(\sup_{\Theta} \sum_{l=1}^{\infty} \theta_{k+2ln}^2 a_{k+2ln}^2 + \theta_{2ln-k}^2 a_{2ln-k}^2\right)^{1/2} \\ &\quad \times \left(\sum_{l=1}^{\infty} a_{k+2ln}^{-2} + a_{2ln-k}^{-2}\right)^{1/2} \\ &\le Q^{1/2}\left(\sum_{l=1}^{\infty} a_{k+2ln}^{-2} + a_{2ln-k}^{-2}\right)^{1/2}\end{aligned}$$

and hence, we arrive to (3.39) by condition $\mathcal{F}_3$

$$\begin{aligned}\sup_{\Theta} \sum_{k=1}^{n} \lambda_k^2 \tilde{\theta}_k^2 &\le Q\left(2 \max_{1\le k\le n} \sum_{l=1}^{\infty} a_{k+ln}^{-2}\right) \sum_{k=1}^{n} \lambda_k \\ &= o(n^{-1}) \sum_{k=1}^{n} \lambda_k = o(d_n)\end{aligned}$$

or by condition $\mathcal{F}_4$

$$\begin{aligned}\sup_{\Theta} \sum_{k=1}^{n} \lambda_k^2 \tilde{\theta}_k^2 &\le \sup_{\Theta} \sum_{k=1}^{n} \tilde{\theta}_k^2 \\ &\le 2Q \sum_{k=n+1}^{\infty} a_k^{-2} = o(d_n)\,.\end{aligned}$$

This completes the proof of the theorem. □

3.6 Bibliographic remarks

The nonparametric regression estimation problem was studied by, among others, Speckman (1985), Nussbaum (1985), Golubev and Nussbaum (1990), Korostelev (1994), Efromovich (1996).

In the paper of Speckman (1985) the minimax linear estimator is a spline. The first result about precise asymptotics of the minimax risk within the class of all estimators in a regression context is due to Nussbaum (1985), where normality of the errors was assumed, the nonparametric class is a Sobolev class and a smoothing spline proved to be asymptotically minimax among all estimators. Exact lower bounds for the minimax risk were obtained in the paper of Golubev and Nussbaum (1990) for nonequidistant designs of observations without assumption of normality of the errors. In a recent paper Efromovich (1996) studied exact asymptotic behaviour of the minimax risk for random design nonparametric regression models also without assumption of normality. The result of Korostelev (1994) is described in Example 1.2.

Our treatment of the lower bound is based on the elementary but rather powerful van Trees inequality which is due to van Trees (1968). For further references and applications of the van Trees inequality see Borovkov (1984), Gill and Levit (1995). Another approach for obtaining lower bounds based on asymptotic equivalence of the original model and the white noise model has been also actively pursued recently, see Brown and Low (1996), Nussbaum (1996), Korostelev and Nussbaum (1995).

The main results of this chapter can be found in Belitser and Levit (1996).

Chapter 4

Efficient density estimation with censored data

Suppose one observes a sample from some unknown probability distribution on some measurable space. The general statistical problem in this context is to gain information about some features of the underlying distribution, using the observed data. In this respect the problem of probability density estimation can be viewed as one of the basic problems in statistics. By far the most frequently used type of observation model is based on a sample of growing size of independent identically distributed random variables. It is not however appropriate when the observations are incomplete. This is the case, for example, when the observations have been censored. In real life applications in demography, actuarial science, epidemiology, survival analysis and other fields, the observations are typically at risk of being censored from the right. In such situations, the observation model we study, the so called *random censorship model*, is often a realistic one.

In this chapter we consider the problem of nonparametric minimax density estimation when the observation model is complicated by the presence of censoring and the density is assumed to belong to the class of "infinitely smooth" functions.

4.1 Introduction

Let $(X_1, Y_1), \dots (X_n, Y_n)$ be independent identically distributed pairs of random variables where X_1 and Y_1 are independent and have the distribution functions F and G respectively. We assume also that the distribution of X_1 is absolutely continuous with density f. The following model is known as *random censorship model.* We observe only the pairs (Z_i, Δ_i), $i = 1, 2, \ldots, n$, with $Z_i = \min(X_i, Y_i)$ and $\Delta_i = I\{X_i \leq Y_i\}$. In survival analysis the X_i's are called survival times and the Y_i's censoring times. We suppose F and G are unknown and our goal is, using the observed data, to estimate the density $f(x)$ at a given point x.

To elucidate the ideas of the results, we give here some heuristic arguments. The unknown underlying density f is assumed to belong to the class of densities with exponentially decreasing Fourier transformations – "infinitely smooth" densities. This nonparametric class has the advantage that one can treat the problem of estimating $f(x)$ as if a smooth functional was to be estimated. In particular, it turns out that any density from this nonparametric class can be represented in the following asymptotic form (see the Approximation Lemma 4.9 below):

$$f(x) = \int \phi_n(x-y)f(y)dy + O(n^{-1/2}), \qquad \text{as} \quad n \to \infty,$$

locally uniformly in f in a proper topology, where $\phi_n(y)$ is some sequence of functions (the exact definitions are given in the next section) which we will call *kernels*, treating this notion in a broader sense than is usual in the literature. The local minimax quadratic risk proves to be of order $(\log n)^{1/r}/n$ ($0 < r \leq 1$ and appears later in the definition of the nonparametric class) and therefore the remainder term can be neglected, while the first term resembles a smooth functional (in fact it is a sequence of functionals because of its dependence on n) to which one can apply well developed methods for deriving an optimal estimator and its asymptotic behavior.

So, in case there is no censoring one can expect the estimator

$$\tilde{f}_n(x) = \int \phi_n(x-y)dF_n(y),$$

with the empirical distribution function F_n, to be optimal in some sense. If for independent identically distributed observations the estimator of the density is some functional of the empirical distribution function $T(F_n)$, then in the case of censoring one tries usually to use the estimator $T(\widetilde{F}_n)$, with the well known Kaplan-Meier estimator $\widetilde{F}_n(y)$ instead of the empirical distribution function F_n. Thus, it is natural to propose the estimator

$$\tilde{f}_n = \tilde{f}_n(x) = \int \phi_n(x-y)d\tilde{F}_n(y). \tag{4.1}$$

In this chapter, we establish, under conditions that the censoring is not too severe and the density belongs to the class of "infinitely smooth" functions, the exact limiting behavior of the local minimax risk up to a constant. We show also that a kernel type estimator, with a properly chosen kernel, is locally asymptotically efficient. We emphasize here that the choice of nonparametric class has made this possible. We propose a wide class of kernels on which the estimator can be based, which turns out to be important in the estimation problem with censored observations.

In the definition of our class we have a smoothness parameter r, $0 < r \leq 1$. Varying the parameter r, we obtain two essentially different cases: $0 < r < 1$ and $r = 1$ which we will call the infinitely smooth and the analytic cases respectively. The analytic case has a distinctive feature. Namely, it turns out that there are no efficient kernel estimators with finitely supported kernels – to estimate efficiently one has to use, roughly speaking, all observations within a distance of $\log n$ from x, due to the phenomenon of the long-range reciprocal memory contained in two separated sets of values of an analytic function. However, we show that, under a condition that censoring is not too severe, one can at least choose a kernel with exponentially decreasing tails. The proof of efficiency of the estimator in this case is based on the martingale technique. In the infinitely smooth case, there are efficient estimators with compactly supported kernels. This facilitates also the use of strong approximation results for the Kaplan-Meier estimator in the proof of the exact upper bound for the minimax risk. The lower bound for the local minimax risk is based on the elementary van Trees inequality in either case.

4.2 Definitions and main results

In this section we summarize the main results. First we recall in brief the notion of efficiency.

Prior information about an unknown density f is usually formalized by assuming $f \in \mathcal{F}$, for some class of densities $\mathcal{F}$. Suppose now that we have some topology on $\mathcal{F}$. For each neighbourhood V define the *local minimax risk*:

$$r_n(V) = r_n(V, x) = \inf_{\tilde{f}_n} \sup_{f \in V} \mathbf{E}_f(\tilde{f}_n(x) - f(x))^2, \tag{4.2}$$

where the infimum is taken over all estimators $\tilde{f}_n$. The estimator $\tilde{f}_n$ is called *asymptotically efficient*, or just *efficient*, if for any sufficiently small neighbourhood V, for some positive sequence ψ_n,

$$\limsup_{n\to\infty} \psi_n^{-2} \left(\sup_{f\in V} \mathbf{E}_f(\tilde{f}_n(x) - f(x))^2 - r_n(V) \right) = 0$$

while

$$\liminf_{n\to\infty} \psi_n^{-2} r_n(V) > 0.$$

Recall that the sequence ψ_n is the minimax rate of convergence. Note also that one can write lim instead of limsup.

Denote from now on the Fourier transform of an absolutely integrable function f by $\hat{f}$:

$$\hat{f}(t) = \int e^{ity} f(y)dy.$$

Define now the nonparametric class $\mathcal{F}_\delta$ of underlying densities.

Definition 4.1. For given $P, \delta > 0$, $0 < r \le 1$ denote

$$\tilde{\mathcal{F}}_\delta = \tilde{\mathcal{F}}_\delta(P,r) = \left\{ f(\cdot) : \ (2\pi)^{-1} \int \exp(2\delta|t|^r)|\hat{f}(t)|^2 dt < P \right\}.$$

Remark 4.1. For $0 < r < 1$, the functions in $\tilde{\mathcal{F}}_\delta(P,r)$ are infinitely differentiable, while $\tilde{\mathcal{F}}_\delta(P,1)$ is a class of analytic functions. Below we describe it more precisely. Let the class $\mathcal{A}_\delta = \mathcal{A}_\delta(Q)$ consist of functions admitting bounded analytic continuation into the strip $\{y + iu, \ |u| \le \delta\}$ and $\int |f(y+i\delta)|^2 dy \le Q < \infty$. In case there is no censoring the nonparametric classes of the type $\mathcal{A}_\delta$ were considered first in Ibragimov and Hasminskii (1983), where the minimax rates of convergence in L_p were derived. There is a close relationship between the 1classes $\tilde{\mathcal{F}}_\delta(P,1)$ and $\mathcal{A}_\delta$: for any β, $0 < \beta < \delta$, there exist positive constants Q_1 and Q_2 such that

$$\mathcal{A}_\delta(Q_1) \subseteq \tilde{\mathcal{F}}_\delta(P,1) \subseteq \mathcal{A}_{\delta-\beta}(Q_2).$$

Indeed, if a density $f \in \mathcal{A}_\delta$, then, according to Timan (1963), p. 137, the limit

$$\lim_{u\to\delta} \mathrm{Re} f(y+iu) = g_f(y)$$

exists for almost all y and $f(y)$ can be represented as a convolution:

$$f(y) = \frac{1}{2\delta} \int \cosh^{-1}\left(\frac{\pi(y-u)}{2\delta}\right) g_f(u)du.$$

Furthermore, because of the relation (see Gradshtein and Ryzhik (1980), equation 3.983.1)

$$\frac{1}{2\delta}\int e^{itu}\cosh^{-1}\left(\frac{\pi u}{2\delta}\right)du = \frac{1}{\cosh(\delta t)},$$

$$\hat{f}(t) = \frac{\hat{g}(t)}{\cosh(\delta t)}.$$

By the Parseval formula,

$$\frac{1}{2\pi}\int \cosh^2(\delta t)|\hat{f}(t)|^2 dt = \int g_f^2(y)dy \leq Q_1$$

and hence the first inclusion holds. The second inclusion follows immediately from the Paley-Wiener theorem (see, for example, Katznelson (1976), p. 174).

Note also that the class $\tilde{\mathcal{F}}_\delta$ is quite broad: the Gauss, Student and Cauchy distributions are, among many others, for appropriate δ, in this class, as well as their finite mixtures.

Definition 4.2. Let $\mathcal{S}_\delta$ and $\mathcal{U}_\delta = \mathcal{U}_\delta(r)$ be the topologies on $\tilde{\mathcal{F}}_\delta(P,r)$ induced by the distances

$$\begin{aligned}
\rho_s(f,g) &= \sup_y |f(y)-g(y)| + \sup_y |f'(y)-g'(y)| + \int |f(y)-g(y)|\,dy,\\
\rho_u(f,g) &= \left(\int \exp(2\delta|t|^r)|\hat{f}(t)-\hat{g}(t)|^2 dt\right)^{1/2} + \int |f(y)-g(y)|dy,
\end{aligned}$$

respectively. Let $\tilde{\mathcal{T}}_\delta = \tilde{\mathcal{T}}_\delta(r)$ be any topology on $\tilde{\mathcal{F}}_\delta(P,r)$ such that $\mathcal{U}_\delta \subseteq \tilde{\mathcal{T}}_\delta \subseteq \mathcal{S}_\delta$.

Remark 4.2. $\mathcal{U}_\delta$ is a strong topology – closeness with respect to ρ_u implies, by the formula for the inverse Fourier transform, closeness of all derivatives in the uniform topology: for $g, h \in \mathcal{F}_\delta$,

$$\begin{aligned}
\sup_y |g^{(m)}(y) - h^{(m)}(y)| &\leq C_1\left(\int t^{2m}|\hat{g}(t)-\hat{h}(t)|^2 dt\right)^{1/2}\\
&\leq C_2\,\rho_u(g,h).
\end{aligned}$$

Remark 4.3. $\mathcal{S}_\delta$ and $\mathcal{U}_\delta$ are possible choices of weak and strong topologies respectively, for which the properties stated in the assertions below hold locally uniformly, i.e. for each $f \in \mathcal{F}_\delta$ there exists a neighbourhood $V(f)$ such that these properties hold uniformly over this vicinity. In fact, in assertions concerning the upper bound for the local minimax risk, one need prove the local uniformity only for the topology $\mathcal{S}_\delta$, and in assertions concerning the lower bound only for the topology $\mathcal{U}_\delta$.

In almost every estimation problem with censored data one faces the well known unstable behavior of the Kaplan-Meier process $\sqrt{n}(\tilde{F}_n(y) - F(y))$ (here $\tilde{F}_n$ is the Kaplan-Meier estimator, see definition below) in the right tails of F and G. Recall that we are going to use a kernel type estimator of the form (4.1). Therefore, the lighter are the tails of the kernel, the less restrictive conditions on the censoring mechanism are needed. On the other hand, it turns out that, in the analytic case $r = 1$, one has to use observations distant from x as well as those close to x when constructing an efficient estimator for $f(x)$. Loosely, this corresponds to the fact that even for y's distant from x the values $f(y)$ still carry some information about $f(x)$ – analytic functions have "long memory". This is formalized by imposing the following restriction on the nonparametric class $\tilde{\mathcal{F}}_\delta(P, 1)$.

Definition 4.3. For given $\alpha, \tau_0 > 0$, $k \geq 1$, denote

$$\mathcal{F}_\delta = \mathcal{F}_\delta(P, r) = \begin{cases} \tilde{\mathcal{F}}_\delta(P, r), & 0 < r < 1 \\ \tilde{\mathcal{F}}_\delta(P, r) \cap \mathcal{R}(\tau_0, k, \alpha, G), & r = 1, \end{cases}$$

where

$$\mathcal{R}(\tau_0, k, \alpha, G) = \left\{ f : \inf_y \left\{ e^{\tau_0 y^{2k}} (1 - F(y))(1 - G(y)) \right\} > \alpha \right\}. \quad (4.3)$$

Here $\tilde{\mathcal{F}}_\delta(P, r)$ is defined by Definition 4.1 and F is the distribution function corresponding to the density f.

Remark 4.4. The restriction on the original class $\tilde{\mathcal{F}}_\delta(P, 1)$ expresses the requirement for the censoring mechanism to allow sufficiently large observations with positive probability as the number of observations tends to infinity. Indeed, for some $0 < p < 1$ let

$$y_n = \left(\frac{1}{\tau_0} \log \left(\frac{-\alpha n}{\log(1 - p)} \right) \right)^{1/(2k)},$$

then

$$\begin{aligned}
\mathbf{P}_f\left\{Z_{(n)} > y_n\right\} &= 1-(1-(1-F(y_n))(1-G(y_n)))^n \\
&\geq 1-(1-\alpha e^{-\tau_0 y_n^{2k}})^n \\
&= 1-\left(1+\frac{\log(1-p)}{n}\right)^n \geq p,
\end{aligned}$$

where $Z_{(n)} = \max_{1\leq i\leq n} Z_i$.

Remark 4.5. Without loss of generality we suppose that k is integer. Indeed, we will see later that both the upper bound and the lower bound for the local minimax risk do not depend on k, nor on α, P and τ_0.

Definition 4.4. Let $\mathcal{T}_\delta$ be the topology induced by $\tilde{\mathcal{T}}_\delta$ on $\mathcal{F}_\delta$.

Let us establish several conventions throughout this chapter:

- sometimes we will write $F \in \mathcal{F}_\delta$ meaning actually that the corresponding density $f \in \mathcal{F}_\delta$;
- in the proofs we will denote generic positive constants by $C_1, C_2, \ldots$ and they are assumed to be different in the proofs of different assertions;
- all symbols O and o correspond to the asymptotics $n \to \infty$ unless otherwise specified;
- if we say that a particular property holds locally uniformly, this means that for each $f \in \mathcal{F}_\delta$ there exists a neighbourhood V of f such that this property holds uniformly over V.

Now we describe a class of kernels to be used in the construction of the estimator. Denote, for some positive b, β, τ, A, $m \geq 0$,

$$v(y) = v(y,b,\beta) = \begin{cases} A\exp\left(-\frac{1}{(b^2-y^2)^\beta}\right), & -b<y<b \\ 0, & y \notin (-b,b), \end{cases} \tag{4.4}$$

$$q_r(y) = \begin{cases} v(y), & 0<r<1 \\ e^{-\tau y^{2k}}, & r=1, \end{cases} \tag{4.5}$$

$$a_n = a_n(m,\delta,r) = \left(\frac{\log n + m\log\log n}{2\delta}\right)^{1/r} \tag{4.6}$$

$$s_n(y) = s_n(y,m,\delta,r) = \frac{\sin(a_n y)}{\pi y}, \tag{4.7}$$

where the constants k, δ, r appear in the definition of the class $\mathcal{F}_\delta$ and the constant A is defined by the requirement:

$$q_r(0) = 1\,. \tag{4.8}$$

The other constants are chosen according to the following conditions:

(i) the constant b is any fixed number such that $b + x < \tau_G$, where $\tau_G = \inf\{y :\ G(y) = 1\}$ and x is the point at which we want to estimate the density f;

(ii) the constant β is any fixed number such that $\beta/(\beta+1) > r$, where r is the parameter in the definition of the class $\mathcal{F}_\delta$;

(iii) the constant m is any fixed number such that

$$\frac{m}{2} > \frac{1}{r} - 1;$$

(iv) the constant τ is any fixed number such that $\tau > 3\tau_0/2$, where the constant τ_0 appears in the definition of the class $\mathcal{F}_\delta(P, 1)$.

Note that, τ_G is necessarily infinite in case $r = 1$ (see Remark 4.9 below); if $x > \tau_G$, then even consistent estimation of $f(x)$ is not possible. Next introduce the kernel

$$\phi_n(y) = \phi_n(y, \tau, \delta, m) = q_r(y) s_n(y). \tag{4.9}$$

and define the following estimator

$$\tilde{f}_n = \tilde{f}_n(x) = \int \phi_n(x - y) d\tilde{F}_n(y), \tag{4.10}$$

where $\tilde{F}_n(y)$ is the Kaplan-Meier estimator, a well known nonparametric efficient estimator of the distribution function $F(y)$:

$$\tilde{F}_n(y) = 1 - \prod_{i:\, Z_{(i)} < y} \left(\frac{n - i}{n - i + 1} \right)^{\Delta_{(i)}}, \tag{4.11}$$

with the convention $0^0 = 1$. Here the $Z_{(i)}$ denote the ordered sequence of Z_i's and the $\Delta_{(i)}$'s are correspondent indicators. A rich literature is devoted to this estimator and its properties (see Andersen et al. (1993) and further references therein).

Remark 4.6. As is shown in Weits (1993), in case of a Hölder type class, the Kaplan-Meier estimator is not optimal with respect to the convergence rate of the second order minimax risk. The problem of the second order efficiency of a smoothed version of the Kaplan-Meier estimator for the infinitely smooth class is studied in Belitser (1997).

Remark 4.7. Since, by the standard formula for the Fourier transform of the product of two functions,

$$\hat{\phi}_n(t) = \frac{1}{2\pi}(\hat{q}_r * I_{[-a_n,a_n]})(t), \tag{4.12}$$

$\hat{\phi}_n(t)$ is nothing else but a smoothed indicator of $[-a_n, a_n]$. Here $*$ is the convolution operation and I_S denotes the indicator function of set S. In words, convolution of a function with the kernel ϕ_n in the time domain corresponds to thresholding the Fourier transform of the function in the frequency domain.

Note also that the function $\hat{q}_r(t)$ is even. The asymptotic behavior of $\hat{q}_r(t)$ for $0 < r < 1$ and $r = 1$, as $|t| \to \infty$, is described in Fedoruk (1977), pp. 213–214, 220, 229. We adapt these results in a simplified form, suitable for our purposes: for some $A_1, A_2, A_3, A_4 > 0$,

$$|\hat{q}_r(t)| \leq A_1 \exp\left\{-A_2|t|^{\frac{\beta}{\beta+1}}\right\}, \quad 0 < r < 1, \tag{4.13}$$

$$|\hat{q}_r(t)| \leq A_3 \exp\left\{-A_4|t|^{\frac{2k}{2k-1}}\right\}, \quad r = 1. \tag{4.14}$$

The constants A_1, A_2 depend in general on b, β and A_3, A_4 on k and τ.

Denote $a \wedge b = \min\{a, b\}$. In the next theorem the local asymptotic performance of the estimator $\tilde{f}_n$ with respect to the topology $\mathcal{T}_\delta$ is established. The proofs of the theorems are given in the last section of this chapter.

Theorem 4.1. *Let $f_0 \in \mathcal{F}_\delta$ be such that $x < \tau_G \wedge \tau_{F_0}$ and distribution function G is continuous at point x. Then, for any sufficiently small neighbourhood $V(f_0)$, the relation*

$$\limsup_{n\to\infty} \frac{n}{(\log n)^{1/r}} \mathbf{E}_f(\tilde{f}_n(x) - f(x))^2 \leq \sigma^2(f)$$

holds uniformly over $f \in V(f_0)$, where

$$\sigma^2(f) = \sigma^2(f, x) = \frac{f(x)}{(2\delta)^{1/r}\pi(1 - G(x))} \tag{4.15}$$

and the estimator $\tilde{f}_n(x)$ is defined by (4.10).

Remark 4.8. In the proof of the theorem we have to assume that the constant b from (4.4) is chosen in such a way that $x+b \leq \tau_{F_0}$. Although this seems to be rather restrictive at first sight because we do not know the density f_0, we can assume this without loss of generality. The point is that we can let the constant b depend on n so that $b_n \to 0$ as $n \to \infty$. Provided b_n converges to zero slowly enough, one needs to modify only slightly the proof of the Approximation Lemma 4.9. All the other proofs remain unchanged.

Theorem 4.1 gives an upper bound for the local minimax risk (4.2): for a sufficiently small neighbourhood $V(f_0)$

$$\limsup_{n\to\infty} \frac{n}{(\log n)^{1/r}} r_n(V) \leq \sup_{f\in V} \sigma(f).$$

If we can provide a lower bound for the local minimax risk, coinciding asymptotically with the upper one, then we clearly determine the asymptotic behavior of the local minimax risk. The next theorem describes the lower bound for the local minimax risk.

Theorem 4.2. *Let $f_0 \in \mathcal{F}_\delta$ be such that $x < \tau_G \wedge \tau_{F_0}$ and distribution function G is continuous at point x. Then, for any sufficiently small neighbourhood $V(f_0)$,*

$$\liminf_{n\to\infty} \frac{n}{(\log n)^{1/r}} r_n(V) \geq \sup_{f\in V} \sigma^2(f),$$

where the local minimax risk $r_n(V)$ and $\sigma^2(f)$ are defined by (4.2) and (4.15) respectively.

Remark 4.9. Note that the condition $x < \tau_G \wedge \tau_F$ is always fulfilled in the analytic case $r = 1$ due to the restriction (4.3). Therefore, as is apparent from the proof, the statement of the last theorem is valid for any neighbourhood $V \subseteq \mathcal{F}_\delta$ in the analytic case.

In view of Theorems 4.1 and 4.2, the estimator $\tilde{f}_n$ is efficient. Indeed, for each $f \in \mathcal{F}_\delta$ and for any sufficiently small neighbourhood $V(f)$,

$$\lim_{n\to\infty} \frac{n}{(\log n)^{1/r}} \left(\sup_{f\in V} \mathbf{E}_f(\tilde{f}_n(x) - f(x))^2 - r_n(V) \right) = 0\,.$$

Moreover, as an immediate consequence of Theorems 4.1 and 4.2, we obtain the asymptotic behavior of the local minimax risk.

Corollary 4.1. *Let $f_0 \in \mathcal{F}_\delta$ be such that $x < \tau_G \wedge \tau_{F_0}$ and suppose the distribution function G is continuous at the point x. Then for any sufficiently small neighbourhood $V(f_0)$*

$$\lim_{n\to\infty} \frac{n}{(\log n)^{1/r}} r_n(V) = \sup_{f\in V} \sigma^2(f).$$

Remark 4.10. Since $\sigma^2(\cdot)$ is a continuous functional, this implies also that

$$\lim_{V\downarrow f_0} \lim_{n\to\infty} \frac{n}{(\log n)^{1/r}} r_n(V) = \lim_{V\downarrow f_0} \sup_{f\in V} \sigma^2(f) = \sigma^2(f_0).$$

Remark 4.11. Note that the smaller α and the bigger are P, k, τ_0 in the definition of the class $\mathcal{F}_\delta$, the less restrictive is this class, while the asymptotic behavior of the local minimax risk in no way depends on α, P, k and τ_0.

Remark 4.12. Compared to the result of Golubev and Levit (1996a) in the analytic case ($r = 1$), we see that the fact of censorship does not influence the convergence rate, but it does influence the optimal constant.

Remark 4.13. Since the Kaplan-Meier estimator is asymptotically normal, it seems plausible that a central limit theorem for the estimator $\tilde{f}_n(x)$ can be given:

$$\sqrt{\frac{n}{\log n}}(\tilde{f}_n(x) - f(x)) \xrightarrow{D} \mathcal{N}(0, \sigma^2(f,x)) \quad \text{as} \quad n \to \infty,$$

where $\sigma^2(f,x)$ is defined by (4.15). For a related result, see Yang (1994) where a central limit theorem for the functional $\int \psi dF$ is established. However we can not apply the methods of Yang (1994) to our functional $\tilde{f}_n(x)$ directly because the kernel ϕ_n depends on n. In the case $0 < r < 1$, one can prove a central limit theorem by using strong approximation results for the Kaplan-Meier estimator. This problem will not be treated here.

4.3 Auxiliary results

In this section we provide technical results which we will need below.

Lemma 4.1. *Let $q_r(y)$ and $\phi_n(y)$ be defined by (4.5) and (4.9) respectively, and let the function $h(y)$ be continuous at x and satisfy*

$$\int q_r^2(x-y)|h(y)|dy < \infty.$$

Then, as $n \to \infty$, the relation

$$\begin{aligned}
&\int \phi_n^2(x-y)h(y)dF(y) \\
&= h(x)f(x)\int \phi_n^2(y)dy + \left(h(x)f(x)\int \phi_n^2(y)dy\right)o(1) \\
&= \frac{h(x)f(x)}{\pi}\left(\frac{\log n}{2\delta}\right)^{1/r} + o\left((\log n)^{1/r}\right)
\end{aligned}$$

holds locally uniformly in $f \in \mathcal{F}_\delta$.

Proof. Let us prove the first relation. Let $O_\epsilon(x) = \{y : |x-y| < \epsilon\}$ be the open interval around x of radius $\epsilon = \epsilon_n$, $\epsilon_n \to 0$,

$$\frac{1}{\epsilon_n^2(\log n)^{1/r}} = o(1)$$

as $n \to \infty$. We have obviously

$$\begin{aligned}
&\int \phi_n^2(x-y)h(y)dF(y) - h(x)f(x)\int \phi_n^2(y)dy \\
&= \int_{O_\epsilon(x)} \phi_n^2(x-y)\left(h(y)f(y) - h(x)f(x)\right)dy \\
&\quad + \int_{(O_\epsilon(x))^C} \phi_n^2(x-y)\left(h(y)f(y) - h(x)f(x)\right)dy\,.
\end{aligned}$$

So it is enough to prove that the right hand side of the last identity is of order $o\left((\log n)^{1/r}\right)$ locally uniformly.

According to (4.7) and (4.9), one can bound the function $\phi_n^2(x-y)$ outside the interval $O_\epsilon(x)$ as follows:

$$\phi_n^2(x-y) \le \frac{q_r^2(x-y)}{\pi^2\epsilon_n^2}.$$

Therefore, the inequality

$$\int_{(O_\epsilon(x))^C} \phi_n^2(x-y)\left|h(y)f(y) - h(x)f(x)\right|dy$$

$$\begin{aligned} &\le \epsilon_n^{-2} C_1 \int_{(O_\epsilon(x))^C} q_r^2(x-y)(|h(y)|+1)dy \\ &= o\left((\log n)^{1/r}\right) \end{aligned}$$

holds locally uniformly due to the fact that our topology is stronger than the topology induced by the sup-norm (see Definition 4.2). Next, owing to this fact again, it is easy to see that

$$\int_{O_\epsilon(x)} \phi_n^2(x-y)\left(h(y)f(y)-h(x)f(x)\right)dy = o(1)\int \phi_n^2(y)dy$$

locally uniformly and the first equality is proved.

To prove the second equality, by (4.7) and (4.8), write

$$\begin{aligned} &\int \phi_n^2(y)dy \\ &= \int_{|y|\le a_n^{-1/2}} \phi_n^2(y)dy + \int_{|y|>a_n^{-1/2}} \phi_n^2(y)dy \\ &= (1+o(1))\int_{|y|\le a_n^{-1/2}} s_n^2(y)dy + O(1)\int_{|y|>a_n^{-1/2}} s_n^2(y)dy \\ &= (1+o(1))a_n\pi^{-2}\int_{|y|\le a_n^{1/2}} \frac{\sin^2(y)}{y^2}dy \\ &\quad + O(1)a_n\int_{|y|>a_n^{1/2}} \frac{\sin^2(y)}{y^2}dy \\ &= \frac{a_n}{\pi}(1+o(1)) = \frac{1}{\pi}\left(\frac{\log n}{2\delta}\right)^{1/r}(1+o(1)), \end{aligned}$$

where a_n is defined by (4.6). □

Lemma 4.2. *The functional*

$$\psi_n(y) = \psi_n(y,x,F) = \int_y^\infty \phi_n(x-u)dF(u) \tag{4.16}$$

is bounded locally uniformly in $F \in \mathcal{F}_\delta$ and uniformly in y.

Proof. Denote $D_1(y) = O_\epsilon(x)\cap[y,+\infty) = (b_1,b_2)$, $D_2(y) = (O_\epsilon(x))^C \cap [y,+\infty)$, where $O_\epsilon(x)$ is the open interval around x of radius ϵ. Then

$$\psi_n(y) = \int_{D_1(y)} \phi_n(x-u)dF(u) + \int_{D_2(y)} \phi_n(x-u)dF(u).$$

The second term, the integral over $D_2(y)$, is clearly bounded. For the first term, we have that

$$\begin{aligned} \left| \int_{D_1(y)} \phi_n(x-u) dF(u) \right| &\le \left| f(x) \int_{D_1(y)} \phi_n(x-u) du \right| \\ &+ \left| \int_{D_1(y)} f'(u^*)(x-u)\phi_n(x-u) du \right| \\ &\le C_1 + C_2 \int_{D_1(y)} q_r(x-u) du \le C_3 \end{aligned}$$

locally uniformly because $\sup_{u \in O_\epsilon(x)} |f'(u)|$ is bounded locally uniformly (see Definition 4.2) and

$$\left| \int_{D_1(y)} \phi_n(x-u) du \right| = \left| \int_{a_n(b_1-x)}^{a_n(b_2-x)} \frac{q_r(a_n^{-1}u) \sin u}{\pi u} du \right| \le C_4.$$

□

Lemma 4.3. *As $n \to \infty$, the relations*

$$\begin{aligned} \mathbf{E}_f \left(\frac{\psi_n(Z_{(n)})}{1 - F(Z_{(n)})} \right)^2 &= O(n^{-1}), \\ \mathbf{E}_f \left(\psi_n(Z_{(n)}) \right)^2 &= O(n^{-1}) \end{aligned}$$

hold locally uniformly in $F \in \mathcal{F}_\delta(P,1)$, where $\psi_n(y)$ is defined by (4.16).

Proof. Denote

$$H(y) \stackrel{\text{def}}{=} \mathbf{P}_f\{Z_1 \le y\} = 1 - (1 - F(y))(1 - G(y)).$$

From restriction (4.3) on the nonparametric class $\mathcal{F}_\delta(P,1)$ it follows that

$$H(y) \le 1 - \alpha e^{-\tau_0 y^{2k}} \tag{4.17}$$

for each F from this class. Fix some $\epsilon > 0$ and notice that for all $y \le x + \epsilon$

$$H(y) \le H(x+\epsilon) \le 1 - \alpha e^{-\tau_0 (x+\epsilon)^{2k}} = q < 1\,.$$

Further, by condition (iv) and the definition of the kernel function (4.9), it is easy to see that, with some constant $C_1 > 0$,

$$\phi_n^2(x-y) \le (\pi\epsilon)^{-2} e^{-2\tau(x-y)^{2k}} \le C_1 e^{-3\tau_0 y^{2k}} \tag{4.18}$$

for all $y \geq x + \epsilon$. Besides, we have obviously that

$$\phi_n^2(y) \leq a_n^2 \pi^{-2}.$$

Now, using the Hölder inequality and all the inequalities above, we obtain the second assertion of the lemma:

$$\begin{aligned}
&\mathbf{E}_f \left(\int_{Z_{(n)}}^{\infty} \phi_n(x-y)dF(y) \right)^2 \\
&\leq \mathbf{E}_f \int_{-\infty}^{\infty} I\{Z_{(n)} \leq y\} \phi_n^2(x-y)dF(y) \\
&= \int_{-\infty}^{\infty} H^n(y)\phi_n^2(x-y)dF(y) \\
&= \int_{-\infty}^{x+\epsilon} H^n(y)\phi_n^2(x-y)dF(y) + \int_{x+\epsilon}^{\infty} H^n(y)\phi_n^2(x-y)dF(y) \\
&\leq q^n \pi^{-2} a_n^2 F(x+\epsilon) + C_2 \int_{x+\epsilon}^{\infty} \left(1 - \alpha e^{-\tau_0 y^{2k}}\right)^n \phi_n^2(x-y)dy \\
&\leq q^n \pi^{-2} a_n^2 + C_3 \int_{x+\epsilon}^{\infty} e^{-3\tau_0 y^{2k}} \left(1 - \alpha e^{-\tau_0 y^{2k}}\right)^n dy \\
&\leq C_4 e^{-C_5 n} (\log n)^2 + C_6 \int_0^{C_7} (1-u)^n du = O(n^{-1})
\end{aligned}$$

locally uniformly because $f(y)$ is bounded locally uniformly and $0 < C_7 \leq 1$.

To prove the first relation, note that from (4.3)

$$\begin{aligned}
&\mathbf{E}_f \left(\frac{\psi_n(Z_{(n)})}{1 - F(Z_{(n)})} \right)^2 \\
&= \mathbf{E}_f \left(\frac{\psi_n(Z_{(n)})}{1 - F(Z_{(n)})} \left(I\left\{Z_{(n)} \leq x+\epsilon\right\} + I\left\{Z_{(n)} > x+\epsilon\right\} \right) \right)^2 \\
&\leq C_8\, \mathbf{E}_f \left(\psi_n(Z_{(n)})\right)^2 \\
&\quad + 2\alpha^{-2} \mathbf{E}_f \left(\exp\left\{2\tau_0 (Z_{(n)})^{2k}\right\} I\left\{Z_{(n)} > x+\epsilon\right\} \psi_n^2(Z_{(n)}) \right).
\end{aligned}$$

It remains only to show that the second term in the right hand side of the last inequality is of order $O(n^{-1})$ locally uniformly. By the Hölder inequality and (4.18), we obtain

$$\mathbf{E}_f \left(\exp\left\{2\tau_0 (Z_{(n)})^{2k}\right\} I\left\{Z_{(n)} > x+\epsilon\right\} \psi_n^2(Z_{(n)}) \right)$$

$$
\begin{aligned}
&\le \mathbf{E}_f\left(\exp\left\{2\tau_0(Z_{(n)})^{2k}\right\} I\left\{Z_{(n)} > x+\epsilon\right\} \int_{Z_{(n)}}^{\infty} \phi_n^2(x-y)dF(y)\right) \\
&\le C_9\mathbf{E}_f\left(\exp\left\{2\tau_0(Z_{(n)})^{2k}\right\} I\left\{Z_{(n)} > x+\epsilon\right\} \int_{Z_{(n)}}^{\infty} e^{-3\tau_0 y^{2k}}dy\right) \\
&\le C_9\mathbf{E}_f \int_{Z_{(n)}}^{\infty} e^{-\tau_0 y^{2k}}dy = C_9\int H^n(y)e^{-\tau_0 y^{2k}}dy \\
&\le C_9\int\left(1-\alpha e^{-\tau_0 y^{2k}}\right)^n e^{-\tau_0 y^{2k}}dy \\
&\le C_{10}\int_0^1 (1-u)^n du = O(n^{-1})
\end{aligned}
$$

locally uniformly. □

Lemma 4.4. *Let the function $h_1(u)$ be an integrable function, let the function $h_2(u)$ be of bounded variation such that $h_2(-\infty)=0$. Then*

$$\int\int h_1(u)h_1(v)h_2(u\wedge v)dudv = \int\left(\int_v^{\infty} h_1(u)du\right)^2 dh_2(v),$$

provided at least one of these two integrals exists.

Proof. Denote

$$H_1(u) = \int_u^{\infty} h_1(v)dv.$$

Integrating by parts twice, we obtain

$$
\begin{aligned}
&\int\int h_1(u)h_1(v)h_2(u\wedge v)dudv \\
&= \int h_1(u)\left(\int_{-\infty}^{u} h_1(v)h_2(v)dv\right)du + \int h_1(u)\left(\int_u^{\infty} h_1(v)h_2(u)dv\right)du \\
&= \int h_1(u)\left(-\int_{-\infty}^{u} h_2(v)dH_1(v)\right)du + \int h_1(u)h_2(u)H_1(u)du \\
&= \int h_1(u)\left(-h_2(u)H_1(u)\right)du + \int h_1(u)\left(\int_{-\infty}^{u} H_1(v)dh_2(v)\right)du \\
&\quad + \int h_1(u)h_2(u)H_1(u)du \\
&= \int h_1(u)\left(\int_{-\infty}^{u} H_1(v)dh_2(v)\right)du
\end{aligned}
$$

$$= -\int \left(\int_{-\infty}^{u} H_1(v) dh_2(v) \right) dH_1(u)$$
$$= \int (H_1(u))^2 \, dh_2(u).$$

□

4.4 Preliminaries: the Kaplan-Meier estimator

Our treatment of the upper bound for the minimax risk in the analytic case relies heavily on the martingale approach to the Kaplan-Meier estimator (Gill (1980)). Below we present the necessary preliminaries, beginning with a suitable adaptation from Gill (1980).

Let N_n be the process counting observed X_i's and Y be the process giving the number at risk:

$$\begin{aligned} N_n(u) &= \#\{i : \, Z_i \le u, \, \Delta_i = 1\}, \\ Y_n(u) &= \#\{i : \, Z_i \ge u\}, \\ J_n(u) &= I\{Y_n(u) > 0\}, \end{aligned}$$

where symbol # denotes the number of elements in a set. Let $X(u-)$ denote left hand limit of X at point u and $\bar{F}(y) = 1 - F(y)$. It is known (see, for example, Gill (1980)) that for y such that $F(y) < 1$ and $Y_n(y) > 0$, i.e. $y < \tau_F$ and $y \le Z_{(n)}$,

$$\begin{aligned} \tilde{F}_n(y) - F(y) &= (1 - F(y)) \int_{-\infty}^{y} \frac{(1 - \tilde{F}_n(u-))}{(1 - F(u))} \frac{J_n(u)}{Y_n(u)} \\ &\quad \times \left(dN_n(u) - \frac{Y_n(u) \, dF(u)}{1 - F(u-)} \right) \\ &= \bar{F}(y) \int_{-\infty}^{y} \frac{(1 - \tilde{F}_n(u-))}{(1 - F(u))} \frac{J_n(u)}{Y_n(u)} dM_n(u), \quad (4.19) \end{aligned}$$

where $M_n(u)$ is a square integrable martingale with the predictable variation process

$$\langle M, M \rangle (y) = \int_{-\infty}^{y} \frac{Y_n(u)(1 - F(u))}{(1 - F(u-))^2} dF(u), \qquad (4.20)$$

while $J_n(u)$, $\tilde{F}_n(u-)$, $Y_n(u)$ are left continuous adapted processes.

Now we give several technical results which will be needed in the proof of the theorems.

For the following result we refer to Weits (1993).

Lemma 4.5. *Let*

$$\mathcal{B} = \mathcal{B}(C_1, C_2) = \{F : \ (1 - F(C_1))(1 - G(C_1)) \geq C_2 > 0\}.$$

Then, as $n \to \infty$, the relation

$$\mathbf{E}\frac{(1 - \tilde{F}_n(y-))^2}{(1 - F(y-))^2}\frac{J_n(y)}{Y_n(y)} = \frac{1}{n\bar{F}(y-)\bar{G}(y-)} + O(n^{-2})$$

holds uniformly over $\mathcal{B}$ and y, $y \leq C_1$.

The proof of this lemma is essentially contained in the proof of Lemma 4 in Weits (1993). In the paper of Weits Y_n/n corresponds to our Y_n.

Lemma 4.6. *For all $n \geq 2$,*

$$\mathbf{E}\left(\frac{J_n(y)}{Y_n(y)}\right) \leq p(y)\,(1 - p(y))^{n-1} + \frac{(1 - p(y))^{n/2}}{p(y)} + \frac{2}{np(y)},$$

where $p = p(y) = (1 - F(y-))(1 - G(y-))$.

Proof. Denote

$$\mu(n) = \mu(n, y) = n\mathbf{E}\left(\frac{J_n(y)}{Y_n(y)}\right) = \sum_{l=1}^{n} \frac{1}{l/n}\begin{pmatrix} n \\ l \end{pmatrix} p^l(1 - p)^{n-l}.$$

Reasoning as in Weits (1993), we obtain the following recursive equation for $\mu(n)$:

$$\begin{aligned}
\mu(n) &= n\sum_{l=1}^{n-1}\frac{1}{l}\left(\begin{pmatrix} n-1 \\ l-1 \end{pmatrix} + \begin{pmatrix} n-1 \\ l \end{pmatrix}\right)p^l(1-p)^{n-l} + p^n \\
&= n\left(\frac{1}{n}\sum_{l=1}^{n}\frac{n!}{(n-l)!l!}p^l(1-p)^{n-l}\right) \\
&\quad + n\left((1-p)\sum_{l=1}^{n-1}\frac{1}{l}\begin{pmatrix} n-1 \\ l \end{pmatrix}p^l(1-p)^{n-l-1}\right) \\
&= n\left(\frac{1}{n}(1-(1-p)^n) + (1-p)\frac{1}{n-1}\mu(n-1)\right) \\
&= 1 + (1-p)\mu(n-1)\frac{n}{n-1} - (1-p)^n.
\end{aligned}$$

Certainly $\mu(n) \le \lambda(n)$, where $\lambda(n)$ satisfies the following recursive equation:

$$\lambda(n) = 1 + (1-p)\frac{n}{n-1}\lambda(n-1), \qquad n \ge 2,$$

with initial condition $\lambda(1) = \mu(1) = p$. Let $C(n)$ be a solution of the corresponding homogeneous equation:

$$C(n) = (1-p)\frac{n}{n-1}C(n-1), \qquad n \ge 2,$$

with initial condition $C(1) = 1$. Obviously,

$$C(n) = (1-p)^{n-1}n.$$

Let $B(n)$ be such that $\lambda(n) = C(n)B(n)$. Then $B(n)$ satisfies

$$B(n) = B(n-1) + \frac{n-1}{n(1-p)}\frac{1}{C(n-1)}, \qquad n \ge 2,$$

with $B(1) = p$. It is easy to see that

$$B(n) = p + \sum_{k=2}^{n}(1-p)^{1-k}k^{-1}$$

and consequently

$$\lambda(n) = p(1-p)^{n-1}n + n\sum_{k=2}^{n}(1-p)^{n-k}k^{-1}.$$

Denote by $\lfloor c \rfloor$ the whole part of c. Now we bound the second term in the last relation

$$\begin{aligned}
\sum_{k=2}^{n}(1-p)^{n-k}k^{-1} &\le \sum_{k=1}^{\lfloor n/2 \rfloor}(1-p)^{n-k} \\
&\quad + \frac{1}{\lfloor n/2 \rfloor + 1}\sum_{k=\lfloor n/2 \rfloor + 1}^{n}(1-p)^{n-k} \\
&\le \frac{(1-p)^{n/2}}{p} + \frac{2}{np},
\end{aligned}$$

and hence the claim follows

$$\begin{aligned}
\mathbf{E}\left(\frac{J_n(y)}{Y_n(y)}\right) &= \frac{\mu(n)}{n} \le \frac{\lambda(n)}{n} \\
&\le p\,(1-p)^{n-1} + \frac{(1-p)^{n/2}}{p} + \frac{2}{np}.
\end{aligned}$$

□

Lemma 4.7. *Let the distribution function G be continuous at the point x. Then, as $n \to \infty$, the relation*

$$\begin{aligned}\frac{n}{\log n}\mathbf{E}_f \int_{-\infty}^{Z_{(n)}} & \left(\phi_n(x-y)\bar{F}(y) - \psi_n(y)\right)^2 \\ & \times \frac{(1-\tilde{F}_n(y-))^2}{(1-F(y-))^4}\frac{J_n(y)}{Y_n(y)}\bar{F}(y)dF(y) \\ \leq \; & \sigma^2(f)(1+o(1)) \end{aligned} \qquad (4.21)$$

holds locally uniformly in $F \in \mathcal{F}_\delta(P,1)$, where $\sigma^2(f)$ and ψ_n are defined by (4.15) and (4.16) respectively.

Proof. By continuity of $F(y)$, we write the the left-hand side of (4.21), for some positive ϵ, as a sum of two terms

$$\begin{aligned} I_1 \;=\; & \mathbf{E}_f \int_{-\infty}^{Z_{(n)}} \left(\phi_n(x-y)\bar{F}(y) - \psi_n(y)\right)^2 \\ & \times \frac{(1-\tilde{F}_n(y-))^2}{(1-F(y-))^3}\frac{J_n(y)}{Y_n(y)} I\{Z_{(n)} \leq x+\epsilon\} dF(y) \end{aligned}$$

and

$$\begin{aligned} I_2 \;=\; & \mathbf{E}_f \int_{-\infty}^{Z_{(n)}} \left(\phi_n(x-y)\bar{F}(y) - \psi_n(y)\right)^2 \\ & \times \frac{(1-\tilde{F}_n(y-))^2}{(1-F(y-))^3}\frac{J_n(y)}{Y_n(y)} I\{Z_{(n)} > x+\epsilon\} dF(y)\,. \end{aligned}$$

To evaluate the first term, observe first that by (4.17)

$$\begin{aligned} \mathbf{P}_f\{Z_1 \leq x+\epsilon\} = H(x+\epsilon) \;&\leq\; 1 - \alpha e^{-\tau_0 (x+\epsilon)^{2k}} = q < 1\,, \\ \left(\phi_n(x-y)\bar{F}(y) - \psi_n(y)\right)^2 \;&\leq\; 2\phi_n^2(x-y)\bar{F}^2(y) + 2\psi_n^2(y) \\ &\leq\; 4\pi^{-2} a_n^2 e^{-2\tau(x-y)^{2k}} \bar{F}^2(y) \\ &=\; C_1 (\log n)^2 e^{-2\tau(x-y)^{2k}} \bar{F}^2(y) \\ &\leq\; C_2 (\log n)^2 e^{-3\tau_0 y^{2k}} \bar{F}^2(y) \end{aligned}$$

and $J_n(y)/Y_n(y) \leq 1$. Thus, recalling (4.3), we bound the first term as follows:

$$I_1 \;\leq\; C_2 (\log n)^2 H^n(x+\epsilon) \int_{-\infty}^{\infty} \frac{e^{-3\tau_0 y^{2k}} dF(y)}{(1-F(y))}.$$

$$
\begin{aligned}
&\le C_2(\log n)^2 q^n \alpha^{-1} \int_{-\infty}^{\infty} e^{-3\tau_0 y^{2k}+\tau_0 y^{2k}} dF(y) \\
&\le C_3 e^{-C_4 n}
\end{aligned}
$$

uniformly.

For the second term, we split the integral I_2 into two parts: the integral over $(-\infty, x+\epsilon]$ and the integral over $(x+\epsilon, Z_{(n)}]$. Since for $y \ge x+\epsilon$

$$
\begin{aligned}
\left(\phi_n(x-y)\bar{F}(y) - \psi_n(y)\right)^2 &\le 2\phi_n^2(x-y)\bar{F}^2(y) + 2\psi_n^2(y) \\
&\le 4(\pi\epsilon)^{-2} e^{-2\tau(x-y)^{2k}} \bar{F}^2(y) \\
&\le C_5 e^{-3\tau_0 y^{2k}} \bar{F}^2(y),
\end{aligned}
$$

we bound the expectation of the integral over $(x+\epsilon, Z_{(n)}]$ merely by

$$
C_5 \int_{x+\epsilon}^{\infty} \mathbf{E}_f \left(\frac{J_n(y)}{Y_n(y)}\right) \frac{e^{-3\tau_0 y^{2k}} dF(y)}{(1-F(y))}.
$$

Thus,

$$
\begin{aligned}
I_2 \le\ & \mathbf{E}_f \int_{-\infty}^{x+\epsilon} \left(\phi_n(x-y)\bar{F}(y) - \psi_n(y)\right)^2 \\
& \times \frac{(1-\tilde{F}_n(y-))^2}{(1-F(y-))^3} \frac{J_n(y)}{Y_n(y)} dF(y) \\
& + C_5 \int_{x+\epsilon}^{\infty} \mathbf{E}_f \left(\frac{J_n(y)}{Y_n(y)}\right) \frac{e^{-3\tau_0 y^{2k}} dF(y)}{(1-F(y))} = S_1 + S_2,
\end{aligned}
$$

say.

To evaluate S_2, we make use of (4.3), Lemma 4.6 and (4.17):

$$
\begin{aligned}
S_2 &= C_5 \int_{x+\epsilon}^{\infty} e^{-3\tau_0 y^{2k}} \mathbf{E}_f \left(\frac{J_n(y)}{Y_n(y)}\right) \frac{dF(y)}{(1-F(y))} \\
&\le 2C_5 \int_{x+\epsilon}^{\infty} \frac{e^{-3\tau_0 y^{2k}} (H^{n/2}(y) + n^{-1}) dF(y)}{(1-F(y))^2 (1-G(y))} \\
&\le C_6 \int_{x+\epsilon}^{\infty} e^{-3\tau_0 y^{2k}+2\tau_0 y^{2k}} \left(\left(1-\alpha e^{-\tau_0 y^{2k}}\right)^{n/2} + n^{-1}\right) f(y) dy \\
&\le C_7 \int_0^{C_8} (1-u)^{n/2} du + C_6 n^{-1} \\
&= O(n^{-1})
\end{aligned}
$$

locally uniformly. Therefore, to complete the proof, it remains only to prove that the relation

$$\frac{n}{\log n} S_1 \leq \sigma^2(f)(1+o(1)) \tag{4.22}$$

holds locally uniformly.

Since

$$(1-F(x+\epsilon))(1-G(x+\epsilon)) \geq \alpha e^{-\tau_0 (x+\epsilon)^{2m}} = \gamma > 0$$

uniformly over $\mathcal{F}_\delta(P,1)$, we have, by (4.9) and Lemma 4.5, that

$$\begin{aligned} S_1 &= \mathbf{E}_f \int_{-\infty}^{x+\epsilon} \left(\phi_n(x-y)\bar{F}(y) - \psi_n(y)\right)^2 \\ &\quad \times \frac{(1-\tilde{F}_n(y-))^2}{(1-F(y-))^3} \frac{J_n(y)}{Y_n(y)} dF(y) \\ &\leq \int_{-\infty}^{x+\epsilon} \frac{\left(\phi_n(x-y)\bar{F}(y) - \psi_n(y)\right)^2 dF(y)}{n(1-F(y))^2(1-G(y-))} + O\left((\log n)^2 n^{-2}\right) \end{aligned}$$

uniformly. Finally, the last inequality, Lemma 4.1 and 4.2 imply that

$$S_1 \leq \int_{-\infty}^{x+\epsilon} \frac{\phi_n^2(x-y)dF(y)}{n(1-G(y-))} + O(n^{-1}) = \frac{\log n}{n}\sigma^2(f)(1+o(1))$$

locally uniformly. Thus (4.22) is fulfilled and the proof of the lemma is complete. □

Remark 4.14. In fact, in view of the lower bound, equality holds true instead of inequality in the statement of the last lemma.

Remark 4.15. A similar to the last lemma's assertion can be proved in case $0 < r < 1$. The proof is somewhat simpler and relies on the fact that in this case kernel ϕ_n is finitely supported.

The following result which is due to Lo et al. (1989) gives a representation of the Kaplan-Meier estimator as an average of independent random variables. First introduce some notations:

$$\begin{aligned} g(y) &= \int_{-\infty}^{y} \frac{dF(u)}{(\bar{F}(u))^2 \bar{G}(u-)}, \qquad (4.23) \\ \xi_i(t) &= \xi(Z_i,\Delta_i,t) = -\bar{F}(t)g(Z_i \wedge t) + \frac{\bar{F}(t)}{\bar{H}(t)} I\{Z_i \leq t, \Delta_i = 1\}. \end{aligned}$$

Lemma 4.8 (Lo, Mack and Wang (1989)). *Let the distribution func tion F be continuous. Then*

$$\tilde{F}_n(y) = F(y) + \frac{1}{n}\sum_{i=1}^{n} \xi_i(y) + R_n(y),$$

where for any $T < \tau_F \wedge \tau_G$ and any $\alpha \geq 1$

$$\sup_{y \leq T} \mathbf{E}|R_n(y)|^{\alpha} = O\left((\log n/n)^{\alpha}\right) \qquad as \qquad n \to \infty.$$

Remark 4.16. Actually, the result of Lo et al. (1989) concerns the case of nonnegative "lifetimes" $X_1, \ldots, X_n$. It is however a straightforward matter to extend this to any continuous distribution function F.

Remark 4.17. Tracing the proof of this lemma, one can show that this representation holds locally uniformly over a sufficiently small neighbourhood of any F such that $T < \tau_G \wedge \tau_F$, in the topology generated by the distance in variation.

Remark 4.18. Note that uniformly in $y \leq T$ the random variables $\xi_i(y)$'s are bounded, independent and by routine calculations,

$$\mathbf{E}\xi_i(y) = 0, \qquad \mathbf{E}\left(\xi_i(y)\xi_i(u)\right) = \bar{F}(y)\bar{F}(u)g(y \wedge u). \tag{4.24}$$

4.5 Approximation Lemma

The following lemma is of particular importance. It reflects the fact that each function from the class $\mathcal{F}_\delta$ can be approximated with a negligible error by a sequence of "smooth functionals", which exhibits a close resemblance of our density estimation problem with the problem of estimating a smooth functional.

Lemma 4.9 (Approximation Lemma). *As $n \to \infty$, the relation*

$$\left(\int \phi_n(x-y)dF(y) - f(x)\right)^2 = O\left(n^{-1}\right)$$

holds uniformly over $\mathcal{F}_\delta$.

Proof. Recalling Definition 4.3, we obtain the following uniform bound:

$$\begin{aligned}
&\left(\int \phi_n(x-y)dF(y) - f(x)\right)^2 \\
&= \left(\frac{1}{2\pi}\int e^{-itx}(\hat{\phi}_n(t)-1)\hat{f}(t)dt\right)^2 \\
&\leq \frac{1}{2\pi}\int \exp\{2\delta|t|^r\}|\hat{f}(t)|^2 dt \cdot \frac{1}{2\pi}\int \exp\{-2\delta|t|^r\}|\hat{\phi}_n(t)-1|^2 dt \\
&\leq C_1 \int |\hat{\phi}_n(t)-1|^2 \exp\{-2\delta|t|^r\}dt \\
&\leq C_1 \int_{-a_n}^{a_n} |\hat{\phi}_n(t)-1|^2 \exp\{-2\delta|t|^r\}dt + C_2\int_{|t|\geq a_n} \exp\{-2\delta|t|^r\}dt \\
&= 2C_1 \int_0^{a_n} |\hat{\phi}_n(t)-1|^2 \exp\{-2\delta|t|^r\}dt + O\left(n^{-1}\right)
\end{aligned}$$

since, due to condition (iii),

$$\begin{aligned}
\int_{|t|\geq a_n} \exp(-2\delta|t|^r)dt &= 2\int_{a_n}^{\infty} \exp(-2\delta t^r)dt \\
&= \frac{1}{\delta r}\int_{\exp\{2\delta a_n^r\}}^{\infty} \left(\frac{\log u}{2\delta}\right)^{\frac{1-r}{r}} \frac{du}{u^2} \\
&= \frac{\exp\{-2\delta a_n^r\}}{r\delta a_n^{r-1}}(1+o(1)) \\
&= \frac{2(\log n)^{\frac{1-r}{r}}}{r\,(2\delta)^{1/r} n(\log n)^m}(1+o(1)) \\
&= O\left(n^{-1}\right).
\end{aligned}$$

Hence it suffices to prove that

$$\int_0^{a_n} |\hat{\phi}_n(t)-1|^2 \exp\{-2\delta|t|^r\} = O(n^{-1}). \tag{4.25}$$

Since function $\hat{q}_r(u)$ is even,

$$\int_{u>t+a_n} |\hat{q}_r(u)|du \leq \int_{u<t-a_n} |\hat{q}_r(u)|du = \int_{u>a_n-t} |\hat{q}_r(u)|du$$

for $t \in [0, a_n]$. By (4.8), we have also that

$$\int \hat{q}_r(u)du = q_r(0)2\pi = 2\pi.$$

Consider first the case $0 < r < 1$. By using the two last inequalities, (4.12) and (4.13), we obtain that, for $t \in [0, a_n]$,

$$\begin{aligned}
|\hat{\phi}_n(t) - 1| &= \left|(2\pi)^{-1}(\hat{q}_r * I_{(-a_n, a_n)})(t) - 1\right| \\
&= \left|(2\pi)^{-1}\int (I_{(-a_n,a_n)})(t-u) - 1)\hat{q}_r(u)du\right| \\
&= (2\pi)^{-1}\left|\int_{|t-u|>a_n} \hat{q}_r(u)du\right| \\
&\le 2(2\pi)^{-1}\int_{u>a_n-t} |\hat{q}_r(u)|du \\
&\le C_3\int_{a_n-t}^{\infty} \exp\left\{-A_2 u^{\frac{\beta}{\beta+1}}\right\} du \\
&\le C_4 \exp\left\{-C_5(a_n - t)^{\frac{\beta}{\beta+1}}\right\}. \qquad (4.26)
\end{aligned}$$

Similar reasoning applies in the case $r = 1$ by (4.14): for $t \in [0, a_n]$,

$$\begin{aligned}
|\hat{\phi}_n(t) - 1| &= \left|(2\pi)^{-1}(\hat{q}_r * I_{(-a_n, a_n)})(t) - 1\right| \\
&\le C_6\int_{u>a_n-t} \exp\left\{-A_4 u^{\frac{2k}{2k-1}}\right\} du \\
&\le C_7 \exp\left\{-C_8(a_n - t)^{\frac{2k}{2k-1}}\right\}. \qquad (4.27)
\end{aligned}$$

Recall now the c_r-inequality (see Loève (1963), p. 155):

$$|h_1 + h_2|^r \le c_r|h_1|^r + c_r|h_2|^r, \qquad r > 0, \qquad (4.28)$$

where $c_r = 1$ or $c_r = 2^{r-1}$ according as $r \le 1$ or $r > 1$. So, in case $0 < r < 1$, we prove (4.25) by combining (4.26) with the c_r-inequality and the fact that $\beta/(\beta + 1) > r$ (see condition (ii)):

$$\begin{aligned}
&\int_0^{a_n} |\hat{\phi}_n(t) - 1|^2 \exp\{-2\delta|t|^r\}\, dt \\
&\le C_4^2 \int_0^{a_n} \exp\left\{-2C_5(a_n - t)^{\frac{\beta}{\beta+1}} - 2\delta t^r\right\} dt \\
&= C_9 \int_0^{a_n} \exp\left\{-2C_5 t^{\frac{\beta}{\beta+1}} - 2\delta(a_n - t)^r\right\} dt \\
&\le C_9 \int_0^{a_n} \exp\left\{-2C_5 t^{\frac{\beta}{\beta+1}} - 2\delta(a_n^r - t^r)\right\} dt \\
&\le C_9 e^{-2\delta a_n^r} \int_0^{\infty} \exp\left\{-2C_5 t^{\frac{\beta}{\beta+1}} + 2\delta t^r\right\} dt = \frac{C_{10}}{n(\log n)^m}
\end{aligned}$$

uniformly. Finally, to prove (4.25) in case $r = 1$, we use (4.27):

$$\begin{aligned}
&\int_0^{a_n} |\hat{\phi}_n(t) - 1|^2 \exp(-2\delta|t|)dt \\
\leq\ & C_7^2 \int_0^{a_n} \exp\left\{-2C_8(a_n - t)^{\frac{2k}{2k-1}} - 2\delta t\right\} dt \\
=\ & C_{11} \int_0^{a_n} \exp\left\{-2C_8 t^{\frac{2k}{2k-1}} - 2\delta(a_n - t)\right\} dt \\
\leq\ & C_{11} e^{-2\delta a_n} \int_0^{\infty} \exp\left\{-2C_8 t^{\frac{2k}{2k-1}} + 2\delta t\right\} dt \\
=\ & \frac{C_{12}}{n(\log n)^m}
\end{aligned}$$

uniformly. □

Remark 4.19. As one can see from the proof of this lemma, a stronger bound on the approximation error is in fact valid. Namely, the relation

$$\left(\int \phi_n(x-y)dF(y) - f(x)\right)^2 = O\left(\frac{1}{n(\log n)^{m+1-1/r}}\right)$$

holds uniformly over $\mathcal{F}_\delta$. Thus, we can make the error of approximation smaller by choosing a larger m in (4.6).

Remark 4.20. Certainly, the proof of this approximation property is almost trivial if $\phi_n(y) = s_n(y)$, where the function s_n is defined by (4.7). Let us clarify why this is a bad choice of the kernel function for the estimator (4.10). The risk of the estimator is bounded from above by a sum of two terms (see the proof of Theorem 4.1) which we call the approximation term and the stochastic term. The first term is analogous to the bias term in the noncensored case and comes from the approximation error. The second term has a stochastic origin and is analogous to the variance of an estimator in the noncensored case. So, choosing $\phi_n(y) = s_n(y)$ provides a small approximation error, while leading to a bigger stochastic term since this function is badly localized in the time domain. The idea is to find a proper localizing factor, the function $q_r(y)$, such that both the stochastic term becomes smaller and the approximation property remains valid.

4.6 Proofs of Theorems

Upper bound: proof of Theorem 4.1. The proof consists of two parts. First we consider the case $0 < r < 1$.

Without loss of generality we suppose that the constant b in (4.4) is chosen in such a way that $x + b < \tau_{F_0} \wedge \tau_G$; see Remark 4.8. Now using integration by parts, Lemma 4.8 (see also Remark 4.17 and (4.24)) and the elementary inequality

$$(a+b)^2 \leq (1+\gamma)\, a^2 + (1+\gamma^{-1})\, b^2, \qquad 0 < \gamma \leq 1, \tag{4.29}$$

we have that, uniformly over a sufficiently small neighbourhood of f_0,

$$\begin{aligned} &\mathbf{E}_f \left(\int \phi_n(x-y) d(\tilde{F}_n(y) - F(y)) \right)^2 \\ &= \mathbf{E}_f \left(\int (\tilde{F}_n(y) - F(y)) d\phi_n(x-y) \right)^2 \\ &\leq \frac{(1+\gamma_n)}{n} \iint \bar{F}(t)\bar{F}(u) g(u \wedge t) d\phi_n(x-t) d\phi_n(x-u) \\ &\quad + (1+\gamma_n^{-1})\, \mathbf{E}_f \int \left(R_n(y)\phi_n'(x-y)\right)^2 dy \\ &\leq \frac{(1+\gamma_n)}{n} \iint \bar{F}(t)\bar{F}(u) g(u \wedge t) d\phi_n(x-t) d\phi_n(x-u) \\ &\quad + (1+\gamma_n^{-1}) C_1 n^{-2} (\log n)^{2+2/r}, \end{aligned} \tag{4.30}$$

where g is defined by (4.23) and γ_n is to be chosen later. We can apply Lemma 4.8 because the kernel $\phi_n(x-y)$ has finite support $[x-b, x+b]$ such that $x + b < \tau_G \wedge \tau_F$ uniformly in a neighbourhood of f_0.

Tedious but straightforward calculations lead to

$$\begin{aligned} &\iint \bar{F}(t)\bar{F}(u) g(u \wedge t) d\phi_n(x-t) d\phi_n(x-u) \\ &= \int \frac{\phi_n^2(x-t) dF(t)}{1 - G(t-)} + \iint \phi_n(x-t)\phi_n(x-u) h(t \wedge u) dF(u) dF(t), \end{aligned}$$

where

$$h(y) = \int_{-\infty}^{y} \frac{dF(u)}{(\bar{F}(u))^2 \bar{G}(u-)} - \frac{1}{\bar{F}(y)\bar{G}(y-)}.$$

By Lemma 4.4, we have

$$\iint \phi_n(x-t)\phi_n(x-u) h(t \wedge u) dF(u) dF(t)$$

$$
\begin{aligned}
&= \int \left(\int_t^\infty \phi_n(x-u)dF(u) \right)^2 dh(t) \\
&= -\int \left(\int_t^\infty \phi_n(x-u)dF(u) \right)^2 \frac{dG(t-)}{\bar{F}(t)(\bar{G}(t-))^2}.
\end{aligned}
$$

Therefore,

$$
\iint \bar{F}(t)\bar{F}(u)g(u \wedge t)d\phi_n(x-t)d\phi_n(x-u) \le \int \frac{\phi_n^2(x-t)dF(t)}{1-G(t-)}.
$$

Now we evaluate the risk of the estimator (4.10). From the last relation, (4.30) and again the elementary inequality (4.29) it follows that, uniformly in a neighbourhood of f_0,

$$
\begin{aligned}
&\mathbf{E}_f \left(\tilde{f}_n(x) - f(x) \right)^2 \\
&= \mathbf{E}_f \left(\int \phi_n(x-y)d(\tilde{F}_n(y) - F(y)) + \int \phi_n(x-y)dF(y) - f(x) \right)^2 \\
&\le \frac{(1+\gamma_n)^2}{n} \int \frac{\phi_n^2(x-t)dF(t)}{1-G(t-)} + \frac{(\gamma_n^{-1} + 2 + \gamma_n)C_1(\log n)^{2+2/r}}{n^2} \\
&\quad + (1+\gamma_n^{-1}) \left(\int \phi_n(x-y)dF(y) - f(x) \right)^2.
\end{aligned}
$$

We choose now γ_n such that $\gamma_n \to 0$ and $(\gamma_n(\log n)^{1/r})^{-1} = o(1)$ as $n \to \infty$. Using the last relation, Lemma 4.1 and the Approximation Lemma 4.9, we obtain that

$$
\limsup_{n\to\infty} \frac{n}{(\log n)^{1/r}} \mathbf{E}_f(\tilde{f}_n(x) - f(x))^2 \le \frac{f(x)}{(2\delta)^{1/r}\pi(1-G(x))}
$$

uniformly over a sufficiently small neighbourhood of f_0.

Now consider the case $r = 1$. First we provide necessary preliminaries. In view of (4.19), we have

$$
\begin{aligned}
&\int_{-\infty}^{Z_{(n)}} \phi_n(x-y)d(\tilde{F}_n(y) - F(y)) \\
&= \int_{-\infty}^{Z_{(n)}} \phi_n(x-y)\bar{F}(y)\frac{(1-\tilde{F}_n(y-))}{(1-F(y-))}\frac{J_n(y)}{Y_n(y)}dM_n(y) \\
&\quad - \int_{-\infty}^{Z_{(n)}} \phi_n(x-y) \left(\int_{-\infty}^{y} \frac{(1-\tilde{F}_n(u-))}{(1-F(u-))}\frac{J_n(u)}{Y_n(u)}dM_n(u) \right) dF(y)
\end{aligned}
$$

$$= \int_{-\infty}^{Z_{(n)}} \left(\phi_n(x-y)\bar{F}(y) - \psi_n(y)\right) \frac{(1-\tilde{F}_n(y-))}{(1-F(y-))} \frac{J_n(y)}{Y_n(y)} dM_n(y)$$
$$+ \frac{\tilde{F}_n(Z_{(n)}) - F(Z_{(n)})}{1-F(Z_{(n)})} \psi_n(Z_{(n)}) \tag{4.31}$$

because, by integration by parts and again (4.19),

$$\int_{-\infty}^{Z_{(n)}} \phi_n(x-y) \left(\int_{-\infty}^{y} \frac{(1-\tilde{F}_n(u-))}{(1-F(u-))} \frac{J_n(u)}{Y_n(u)} dM_n(u) \right) dF(y)$$
$$= -\int_{-\infty}^{Z_{(n)}} \frac{(1-\tilde{F}_n(u-))}{(1-F(u-))} \frac{J_n(u)}{Y_n(u)} dM_n(u) \int_{Z_{(n)}}^{\infty} \phi_n(x-u)dF(u)$$
$$+ \int_{-\infty}^{Z_{(n)}} \left(\int_{y}^{\infty} \phi_n(x-u)dF(u) \right) \frac{(1-\tilde{F}_n(y-))}{(1-F(y-))} \frac{J_n(y)}{Y_n(y)} dM_n(y)$$
$$= -\frac{\tilde{F}_n(Z_{(n)}) - F(Z_{(n)})}{1-F(Z_{(n)})} \psi_n(Z_{(n)})$$
$$+ \int_{-\infty}^{Z_{(n)}} \psi_n(y) \frac{(1-\tilde{F}_n(y-))}{(1-F(y-))} \frac{J_n(y)}{Y_n(y)} dM_n(y).$$

Since the first term of the right hand side of the relation (4.31) is the integral of a predictable locally bounded process (almost all its sample paths are locally bounded) with respect to a square integrable martingale with the predictable variation process (4.20), one can represent its second moment as follows (see, for example, Gill (1980)):

$$\mathbf{E}_f \left(\int_{-\infty}^{Z_{(n)}} \left(\phi_n(x-y)\bar{F}(y) - \psi_n(y)\right) \frac{(1-\tilde{F}_n(y-))}{(1-F(y-))} \frac{J_n(y)}{Y_n(y)} dM_n(y) \right)^2$$
$$= \mathbf{E}_f \int_{-\infty}^{Z_{(n)}} \left(\phi_n(x-y)\bar{F}(y) - \psi_n(y)\right)^2$$
$$\times \frac{(1-\tilde{F}_n(y-))^2}{(1-F(y-))^4} \frac{J_n(y)}{Y_n(y)} \bar{F}(y) dF(y). \tag{4.32}$$

Note that $\tilde{F}_n(y)$ is constant on $[Z_{(n)}, \infty)$. So, using (4.31), (4.32) and (twice) the elementary inequality (4.29), we obtain

$$\mathbf{E}_f \left(\int_{-\infty}^{\infty} \phi_n(x-y) d(\tilde{F}_n(y) - F(y)) \right)^2$$
$$\leq (1+\gamma_n) \mathbf{E}_f \left(\int_{-\infty}^{Z_{(n)}} \phi_n(x-y) d(\tilde{F}_n(y) - F(y)) \right)^2$$

$$
\begin{aligned}
&+ (1+\gamma_n^{-1})\,\mathbf{E}_f\left(\psi_n(Z_{(n)})\right)^2 \\
\le\;& (1+\gamma_n)^2\,\mathbf{E}_f\int_{-\infty}^{Z_{(n)}} \left(\phi_n(x-y)\bar F(y) - \psi_n(y)\right)^2 \\
&\times \frac{(1-\tilde F_n(y-))^2}{(1-F(y-))^4}\frac{J_n(y)}{Y_n(y)}\bar F(y)dF(y) \\
&+ (\gamma_n^{-1}+2+\gamma_n)\,\mathbf{E}\left(\frac{\psi_n(Z_{(n)})}{1-F(Z_{(n)})}\right)^2 \\
&+ (1+\gamma_n^{-1})\,\mathbf{E}_f\left(\psi_n(Z_{(n)})\right)^2,
\end{aligned}
$$

where γ_n is to be chosen later.

Let us bound now the risk of the estimator (4.10) from above. The last inequality and the elementary inequality (4.29) yield

$$
\begin{aligned}
&\mathbf{E}_f(\tilde f_n(x) - f(x))^2 \\
=\;& \mathbf{E}_f\left(\tilde f_n(x) - \int \phi_n(x-y)dF(y) + \int \phi_n(x-y)dF(y) - f(x)\right)^2 \\
\le\;& (1+\gamma_n)\,\mathbf{E}_f\left(\int \phi_n(x-y)d(\tilde F_n(y) - F(y))\right)^2 \\
&+ (1+\gamma_n^{-1})\left(\int \phi_n(x-y)dF(y) - f(x)\right)^2 \\
\le\;& (1+\gamma_n)^3\,\mathbf{E}_f\int_{-\infty}^{Z_{(n)}} \left(\phi_n(x-y)\bar F(y) - \psi_n(y)\right)^2 \\
&\times \frac{(1-\tilde F_n(y-))^2}{(1-F(y-))^4}\frac{J_n(y)}{Y_n(y)}\bar F(y)dF(y) \\
&+ (\gamma_n^{-1}+3+3\gamma_n+\gamma_n^2)\,\mathbf{E}_f\left(\frac{\psi_n(Z_{(n)})}{1-F(Z_{(n)})}\right)^2 \\
&+ (\gamma_n^{-1}+2+\gamma_n)\mathbf{E}_f\left(\psi_n(Z_{(n)})\right)^2 \\
&+ (1+\gamma_n^{-1})\left(\int \phi_n(x-y)dF(y) - f(x)\right)^2.
\end{aligned}
$$

We choose now γ_n such that $\gamma_n \to 0$ and $(\gamma_n \log n)^{-1} = o(1)$ as $n \to \infty$. Combining the last relation with Lemmas 4.3, 4.7 and the Approximation Lemma 4.9 proves the theorem. □

Remark 4.21. In the case $0 < r < 1$ one can also apply Theorem A.2 instead of Lemma 4.8. In our case we take $T_n = x + b$, $\epsilon = 2$. Then we derive relation (4.30) as follows:

$$\begin{aligned}
&\mathbf{E}_f\left(\int \phi_n(x-y)d(\tilde{F}_n(y)-F(y))\right)^2 \\
&= \mathbf{E}_f\left(n^{-1/2}\int \bar{F}(y)W_n(g(y))d\phi_n(x-y)+\int R_n(y)d\phi_n(x-y)\right)^2 \\
&\le \frac{(1+\gamma_n)}{n}\int\!\!\int \bar{F}(t)\bar{F}(u)g(u\wedge t)d\phi_n(x-t)d\phi_n(x-u) \\
&\quad + (1+\gamma_n^{-1})\,\mathbf{E}_f\left(\int R_n(y)d\phi_n(x-y)\right)^2
\end{aligned}$$

and

$$\begin{aligned}
&\mathbf{E}_f\left(\int R_n(y)d\phi_n(x-y)\right)^2 \\
&\le 2\mathbf{E}_f\int \left(\phi_n'(x-y)R_n(y)\right)^2 I\{|R_n(y)|>C_1n^{-1}\log n\}dy \\
&\quad +2\mathbf{E}_f\int \left(\phi_n'(x-y)R_n(y)\right)^2 I\{|R_n(y)|\le C_1n^{-1}\log n\}dy \\
&\le \frac{C_2(\log n)^{2/r}}{n^2}+\frac{C_3(\log n)^{2+2/r}}{n^2}.
\end{aligned}$$

Lower bound: proof of Theorem 4.2. Let f_0 be an arbitrary density from the neighbourhood V and let F_0 be the corresponding distribution function. Consider the following family of densities:

$$f_\theta(y)=f_\theta(y,x,\phi_n,f_0)=f_0(y)(1+\theta(\phi_n(x-y)-\bar{\phi}_n(x)),$$

where $|\theta|\le\theta_n$,

$$\bar{\phi}_n(x)=\int \phi_n(x-y)f_0(y)dy$$

and $\phi_n(y)$ is defined by (4.9) with

$$a_n=a_n(m,\delta,r)=\left(\frac{\log n - m\log\log n}{2\delta}\right)^{1/r} \tag{4.33}$$

instead of a_n defined by (4.6). Let θ_n be such that $\epsilon_n\le\theta_n\le\rho_n$, where the positive sequences ϵ_n and ρ_n satisfy

$$\frac{1}{\epsilon_n^2 n\,(\log n)^{1/r}}=o(1),\qquad \rho_n^2 n=o(1).$$

One can choose for example $\theta_n=n^{-1/2}(\log n)^{-1/(4r)}$.

The proof of the theorem will proceed via the following two claims.

Proposition 4.1. *For sufficiently large n, $f_\theta \in V$.*

Proof. Take $\epsilon > 0$ such that $O_\epsilon(f_0) \subseteq V$, where

$$O_\epsilon(f_0) = \{f \in \mathcal{F}_\delta : \ \rho(f, f_0) < \epsilon\}.$$

We prove now that $f_\theta \in O_\epsilon(f_0)$ for sufficiently large n where $\rho = \rho_u$ generates the strong topology $\mathcal{U}_\delta$.

It is easy to check condition (4.3) on the nonparametric class in case $r = 1$: for sufficiently large n,

$$\begin{aligned} &\inf_y \left\{e^{\tau_0 y^{2k}}(1 - F_\theta(y))(1 - G(y))\right\} \\ &= \inf_y \left\{e^{\tau_0 y^{2k}}(1 - F_0(y))(1 - G(y))\right\}(1 + o(1)) > \alpha \end{aligned}$$

since $F_0 \in \mathcal{F}_\delta$.

Denote

$$\psi(y) = \psi(y, x) = f_0(y)\phi_n(x - y).$$

First, by the Minkowski inequality, we have

$$\begin{aligned} \rho_u(f_\theta, f_0) &= |\theta| \left(\int \exp\{2\delta|t|^r\} |\hat{\psi}(t, x) - \bar{\phi}_n(x)\hat{f}_0(t)|^2 dt\right)^{1/2} \\ &\quad + |\theta| \int |f_0(y)(\phi_n(x - y) - \bar{\phi}_n(x))| dy \\ &\leq 2\theta_n \left(\int \exp\{2\delta|t|^r\} |\hat{\psi}(t, x)|^2 dt\right)^{1/2} \\ &\quad + 2\theta_n |\bar{\phi}_n(x)| \left(\int \exp\{2\delta|t|^r\} |\hat{f}_0(t)|^2 dt\right)^{1/2} \\ &\quad + \theta_n \int |f_0(y)(\phi_n(x - y) - \bar{\phi}_n(x))| dy \\ &\leq 2\theta_n \left(\int \exp\{2\delta|t|^r\} |\hat{\psi}(t, x)|^2 dt\right)^{1/2} + C_1\theta_n(\log n)^{1/r}. \end{aligned}$$

Since $\theta_n = o(n^{-1/2})$, it suffices to show that the the first term on the right hand side of the last inequality converges to zero as $n \to \infty$.

Note that

$$\hat{\psi}(t, x) = (2\pi)^{-1} \int e^{ixu} \hat{f}_0(t + u)\hat{\phi}_n(u) du.$$

By the generalized Minkowski inequality (Nikol'skii (1975), p. 20)), the c_r-inequality (4.28), Definition 4.3 and (4.12), it follows that

$$\begin{aligned}
&\left(\int \exp\{2\delta|t|^r\}\,|\hat{\psi}(t,x)|^2 dt\right)^{1/2}\\
&\le C_2\left(\int\left|\int \exp\{\delta|t|^r\}\,e^{ixu}\hat{f}_0(t+u)\hat{\phi}_n(u)du\right|^2 dt\right)^{1/2}\\
&\le C_2\int\left(\int\left|\exp\{\delta|t|^r\}\,e^{ixu}\hat{f}_0(t+u)\hat{\phi}_n(u)\right|^2 dt\right)^{1/2} du\\
&\le C_2\int\left(\int\left|\exp\{\delta|t+u|^r\}\,\hat{f}_0(t+u)\exp\{\delta|u|^r\}\,\hat{\phi}_n(u)\right|^2 dt\right)^{1/2} du\\
&\le C_3\left(\int \exp\{2\delta|t|^r\}\,|\hat{f}_0(t)|^2 dt\right)^{1/2}\int \exp\{\delta|u|^r\}\,|\hat{\phi}_n(u)|du\\
&\le C_4\int \exp\{\delta|u|^r\}\,|\hat{\phi}_n(u)|du\\
&\le C_4\left(\int \exp\{2\delta|u|^r\}\left|\left(\hat{q}_r * I_{(-a_n,a_n)}\right)(u)\right|^2 du\right)^{1/2}\\
&= C_4\left(\int\left|\int \exp\{\delta|u|^r\}\,\hat{q}_r(u-t)I_{(-a_n,a_n)}(t)dt\right|^2 du\right)^{1/2}\\
&\le C_4\int\left(\int\left|\exp\{\delta|u|^r\}\,\hat{q}_r(u-t)I_{(-a_n,a_n)}(t)\right|^2 du\right)^{1/2} dt\\
&\le C_5\left(\int \exp\{2\delta|t|^r\}\,|\hat{q}_r(t)|^2 dt\right)^{1/2}\int e^{\delta|u|^r} I_{(-a_n,a_n)}(u)du\\
&= C_6 n^{1/2}\left(\int \exp\{2\delta|t|^r\}\,|\hat{q}_r(t)|^2 dt\right)^{1/2}
\end{aligned}$$

since, by (4.33) and (iii),

$$\begin{aligned}
\int e^{\delta|u|^r} I_{(-a_n,a_n)}(u)du &= \frac{2}{\delta r}\int_1^{\exp\{\delta a_n^r\}}\left(\frac{\log u}{\delta}\right)^{\frac{1-r}{r}} du\\
&= \frac{\exp\{\delta a_n^r\}\,a_n^{1-r}}{\delta r}(1+o(1))\\
&= \frac{2\,n^{1/2}(\log n)^{1/r-1}}{r\,(2\delta)^{1/r}(\log n)^{m/2}}(1+o(1))\\
&\le C_7\sqrt{n}.
\end{aligned}$$

Now evaluate, by (4.13),

$$\int e^{2\delta|t|^r}|\hat{q}_r(t)|^2 dt \;\le\; A_1 \int \exp\left\{2\delta|t|^r - 2A_2|t|^{\frac{\beta}{\beta+1}}\right\} dt \le C_8$$

in case $0 < r < 1$ and, by (4.14),

$$\int e^{2\delta|t|}|\hat{q}_r(t)|^2 dt \;\le\; A_3 \int \exp\left\{2\delta|t| - 2A_4|t|^{\frac{2k}{2k-1}}\right\} dt \le C_9$$

in case $r = 1$.

Recalling the condition on the θ_n, we obtain finally that

$$\begin{aligned}\rho_u(f_\theta, f_0) &\le C_{10}\theta_n n^{1/2} + C_1\theta_n(\log n)^{1/r} \\ &\le C_{10}\rho_n n^{1/2} + C_1\rho_n(\log n)^{1/r} = o(1)\end{aligned}$$

as $n \to \infty$. □

If X_i is distributed with density $f_\theta(y)$, then the corresponding observation (Z_i, Δ_i) has the density

$$f_\theta(y,\tau) = (f_\theta(y)(1-G(y)))^\tau\,(g(y)(1-F_\theta(y)))^{1-\tau}, \qquad \tau \in \{0,1\}.$$

The following proposition describes the Fisher information $I(\theta)$ about θ contained in the observation (Z, Δ).

Proposition 4.2. *As $n \to \infty$, the relation*

$$I(\theta) \stackrel{\text{def}}{=} \mathbf{E}_f\left[\frac{\partial \log f_\theta(Z,\Delta)}{\partial\theta}\right]^2 = \frac{f_0(x)\bar{G}(x)}{\pi}\left(\frac{\log n}{2\delta}\right)^{1/r}(1+o(1)). \tag{4.34}$$

holds uniformly in θ, $|\theta| < \theta_n$.

Proof. By straightforward calculations,

$$\begin{aligned}I(\theta) &= \mathbf{E}_f\left[\frac{\partial \log f_\theta(Z,\Delta)}{\partial\theta}\right]^2 \\ &= \int \frac{(\phi_n(x-y) - \bar{\phi}_n(x))^2 f_0(y)(1-G(y))dy}{1+\theta(\phi_n(x-y)-\bar{\phi}_n(x))} \\ &\quad + \int \frac{\left(\int_{-\infty}^{y} f_0(u)(\phi_n(x-u)-\bar{\phi}_n(x))du\right)^2 dG(y)}{1-F_\theta(y)}.\end{aligned}$$

Split the second term in the right hand side of the last inequality into two parts: the integral over $(-\infty, x+\epsilon]$ and the integral over $(x+\epsilon, \infty)$:

$$\begin{aligned}\int \frac{\left(\int_{-\infty}^{y} f_0(u)(\phi_n(x-u) - \bar{\phi}_n(x))du\right)^2 dG(y)}{1 - F_\theta(y)} &= \int_{-\infty}^{x+\epsilon} + \int_{x+\epsilon}^{\infty} \\ &= I_1 + I_2.\end{aligned}$$

The integral $\int_{-\infty}^{y} f_0(u)(\phi_n(x-u) - \bar{\phi}_n(x))du$ is bounded for $y \in (-\infty, x+\epsilon]$ by Lemma 4.2 and the Approximation Lemma 4.9. Obviously,

$$\begin{aligned}1 - F_\theta(y) &= (1 - F_0(y))(1 + o(1)) \\ &\geq (1 - F_0(x+\epsilon))(1 + o(1))\end{aligned}$$

for $y \in (-\infty, x+\epsilon]$. Therefore, the integral I_1 is bounded.

Further note that

$$\int_{-\infty}^{y} f_0(u)(\phi_n(x-u) - \bar{\phi}_n(x))du = -\int_{y}^{\infty} f_0(u)(\phi_n(x-u) - \bar{\phi}_n(x))du$$

and the function $\phi_n(x-y)$ is bounded for $y \in (x + \quad \infty)$. Therefore, for $y \in (x+\epsilon, \infty)$ and sufficiently large n,

$$\begin{aligned}&\frac{\left(\int_{-\infty}^{y} f_0(u)(\phi_n(x-u) - \bar{\phi}_n(x))du\right)^2}{1 - F_\theta(y)} \\ &= \frac{\left(\int_{y}^{\infty} f_0(u)(\phi_n(x-u) - \bar{\phi}_n(x))du\right)^2}{1 - F_\theta(y)} \\ &\leq \frac{C_1\left(\int_{y}^{\infty} f_0(u)du\right)^2}{1 - F_\theta(y)} \\ &= \frac{C_1(1 - F_0(y))^2}{(1 - F_0(y))(1 + o(1))} \\ &\leq C_2(1 - F_0(y)) \leq C_2.\end{aligned}$$

Thus, we obtained that

$$\int \frac{\left(\int_{-\infty}^{y} f_0(z)(\phi_n(x-z) - \bar{\phi}_n(x))dz\right)^2 dG(y)}{1 - F_\theta(y)} = O(1)$$

uniformly in θ, $|\theta| < \theta_n$.

According to Lemmas 4.1 and 4.2, it is not difficult to see that

$$\begin{aligned}
&\int \frac{(\phi_n(x-y) - \bar{\phi}_n(x))^2 f_0(y)(1-G(y))dy}{1+\theta(\phi_n(x-y) - \bar{\phi}_n(x))} \\
&= \int \phi_n^2(x-y) f_0(y)(1-G(y))dy\,(1+o(1)) \\
&= \frac{f_0(x)\bar{G}(x)}{\pi}\left(\frac{\log n}{2\delta}\right)^{1/r}(1+o(1)).
\end{aligned}$$

uniformly in θ, $|\theta| < \theta_n$. Relation (4.34) is proved. □

Now we proceed to prove the theorem. Introduce

$$\nu(x) = \nu_n(x) = \theta_n^{-1}\nu_0(\theta_n^{-1}x),$$

where $\nu_0(x)$ is a probability density on the interval $[-1,1]$ with finite Fisher information

$$I_0 = \int_{-1}^{1} (\nu_0'(x))^2 \nu_0^{-1}(x)dx,$$

such that $\nu_0(-1) = \nu_0(1) = 0$ and $\nu_0(x)$ is continuously differentiable for $|x| < 1$. The function $\nu(x)$ is a probability density with support $[-\theta_n, \theta_n]$. It is easy to calculate the Fisher information of the distribution defined by density $\nu(x)$:

$$I(\nu) = I_n(\nu) = I_0\theta_n^{-2}.$$

Applying now the van Trees inequality for the Bayes risk below (Theorem A.1) and Propositions 4.1 and 4.2, we obtain that, for sufficiently large n,

$$\begin{aligned}
r_n(V) &= \inf_{\tilde{f}_n} \sup_{f\in V} \mathbf{E}_f(\tilde{f}_n - f(x))^2 \\
&\geq \inf_{\tilde{f}_n} \sup_{|\theta|\leq\theta_n} \mathbf{E}_{f_\theta}(\tilde{f}_n - f_\theta(x))^2 \\
&\geq \inf_{\tilde{f}_n} \int \mathbf{E}_{f_\theta}(\tilde{f}_n - f_\theta(x))^2 \nu(\theta)d\theta \\
&\geq \frac{\left(\int (\partial f_\theta(x)/\partial\theta)\,\nu(\theta)d\theta\right)^2}{n\int I(\theta)\nu(\theta)d\theta + I(\nu)}
\end{aligned}$$

$$= \frac{(f_0(x)\phi_n(0))^2 (1+o(1))}{n\int I(\theta)\nu(\theta)d\theta + I_0\theta_n^{-2}}$$

$$\geq \frac{\left(\frac{f_0(x)}{\pi}\right)^2 \left(\frac{\log n}{2\delta}\right)^{2/r} (1+o(1))}{\frac{nf_0(x)\bar{G}(x)}{\pi}\left(\frac{\log n}{2\delta}\right)^{1/r}(1+o(1)) + I_0\epsilon_n^{-2}}$$

$$\geq \frac{f_0(x)}{n\pi\bar{G}(x)}\left(\frac{\log n}{2\delta}\right)^{1/r}(1+o(1))$$

or

$$\liminf_{n\to\infty} \frac{n}{(\log n)^{1/r}} r_n(V) \geq \frac{f_0(x)}{(2\delta)^{1/r}\pi\bar{G}(x)} = \sigma^2(f_0).$$

The function f_0 was chosen arbitrarily from the neighbourhood V and hence, by the same reasoning, this relation is valid for any function $f \in V$:

$$\liminf_{n\to\infty} \frac{n}{(\log n)^{1/r}} r_n(V) \geq \sigma^2(f).$$

Therefore

$$\liminf_{n\to\infty} \frac{n}{(\log n)^{1/r}} r_n(V) \geq \sup_{f\in V} \sigma^2(f),$$

which proves the theorem. □

4.7 Bibliographic remarks

The problem of density estimation under random censorship has been treated by a number of authors (see for example Diehl and Stute (1988), Mielniczuk (1986), Lo et al. (1989), Hentzschel (1992), Kulasekera (1995), Huang and Wellner (1995), Liu (1996)). In Hentzschel (1992) an estimator based on the orthonormal system of the Laguerre series on the positive line is investigated and under some assumptions the rates of the mean integrated square error and the mean square error are obtained. In Kulasekera (1995) upper L_1-bounds for the kernel-type estimator are given for two classes of densities: monotonically decreasing densities on $[0,\infty)$ and densities which are of bounded variation on $[0,1]$. For a decreasing density function, Huang and Wellner (1995) showed that the nonparametric maximum likelihood estimator of the density is asymptotically equivalent to the estimator obtained by differentiating the least concave majorant of the Kaplan-Meier estimator and established the asymptotic distributions of the different estimators at a fixed point.

However, in all the above mentioned papers the question of optimality has not been touched on. In a recent paper Liu (1996) considered the general problem of estimating functionals of a distribution F for some nonparametric classes defined in terms of the Hellinger modulus of continuity. With regard to density estimation, minimax Kaplan-Meier based kernel procedures were shown, under some conditions, to be of optimal rate and, moreover, within certain lower and upper bounds.

In the nonparametric minimax estimation context, the notion of asymptotic optimality is usually associated with the optimal rate of convergence of the minimax risk. In order to derive the exact asymptotics of the minimax risk and to be able to compare the estimators with the optimal rate of convergence, one may strengthen the optimal rate results by finding optimal constants when they exist. Results about the optimal constants in minimax density estimation have only been obtained in a limited number of works for models with independent identically distributed observations. The majority of authors has considered the global minimax risk. However, studying the so called local minimax risk yields more exact results. We mention the work of Golubev and Levit (1996a) whose results motivated the present study. In the problem of estimation of an analytic density at a given point, with independent identically distributed observations, they derived the exact limiting behavior of the local minimax risk and proposed an efficient estimator.

The main results of this chapter can also be found in Belitser (1998a) and Belitser and Levit (1997) (cf. also Belitser (1998b)).

Appendix A

A.1 A technical lemma

In this section we present a technical lemma which is just a constructive version of Theorem 2.1. We use this lemma also in Chapter 3. A related assertion is given in Pinsker (1980), see Lemma 1 in this paper.

Let $(a_k,\ k = 1, 2, \dots)$ and $(\sigma_k,\ k = 1, 2, \dots)$ be two positive sequence and let $(a_k,\ k = 1, 2, \dots)$ converge to infinity. The following ellipsoid is a compact set in l_2:

$$\Theta = \Theta(Q) = \{\theta : \sum_{k=1}^{\infty} a_k^2 \theta_k^2 \le Q\}. \tag{A.1}$$

Define the functional on the ellipsoid:

$$R_\epsilon(x, \theta) = \sum_{k=1}^{\infty} \left(\epsilon^2 \sigma_k^2 x_k^2 + (1 - x_k)^2 \theta_k^2\right).$$

Let c_ϵ be a solution of the equation

$$\epsilon^2 \sum_{k=1}^{\infty} \sigma_k^2 a_k (1 - c a_k)_+ = cQ \tag{A.2}$$

and

$$d_\epsilon = d_\epsilon(\Theta) = \epsilon^2 \sum_{k=1}^{\infty} \sigma_k^2 (1 - c_\epsilon a_k)_+ , \tag{A.3}$$

where b_+ denotes the nonnegative part of b.

Lemma A.1. *Let c_ϵ and d_ϵ be defined by* (A.2) *and* (A.3). *Then*

$$\inf_x \sup_{\theta \in \Theta} R_\epsilon(x, \theta) = \sup_{\theta \in \Theta} \inf_x R_\epsilon(x, \theta), \tag{A.4}$$

with the saddle point $(\tilde{x}, \tilde{\theta})$ *given by*

$$\tilde{x}_k = (1 - c_\epsilon a_k)_+ , \tag{A.5}$$

$$\tilde{\theta}_k^2 = \frac{\epsilon^2 \sigma_k^2 (1 - c_\epsilon a_k)_+}{c_\epsilon a_k} \tag{A.6}$$

and

$$\inf_x \sup_{\theta \in \Theta} R_\epsilon(x, \theta) = d_\epsilon = \sup_{\theta \in \Theta} \sum_{k=1}^{\infty} \frac{\epsilon^2 \sigma_k^2 \theta_k^2}{\theta_k^2 + \epsilon^2 \sigma_k^2}. \tag{A.7}$$

Proof. Let us show first that

$$\sup_{\theta \in \Theta} R_\epsilon(x, \theta) = \sum_{k=1}^{\infty} \epsilon^2 \sigma_k^2 x_k^2 + Q \sup_{k \geq 1} (1 - x_k)^2 a_k^{-2}. \tag{A.8}$$

Denote the right hand side of (A.8) by $R_\epsilon(x)$. Clearly, $\sup_{\theta \in \Theta} R_\epsilon(x, \theta) \leq R_\epsilon(x)$. For an arbitrary $\kappa > 0$, let l be such that

$$(1 - x_l)^2 a_l^{-2} \geq \sup_{k \geq 1} (1 - x_k)^2 a_k^{-2} - \kappa .$$

Then by choosing $\theta \in \Theta$ such that $\theta_l^2 = Q a_l^{-2}$ and $\theta_k^2 = 0$, $k \neq l$, we get $R_\epsilon(x, \theta) \geq R_\epsilon(x) - \kappa$, which proves (A.8).

Define the set

$$X_t = \Big\{ x : \sup_{k \geq 1} (1 - x_k)^2 a_k^{-2} = t^2 \Big\}, \qquad t \geq 0 .$$

Obviously, for any $t \geq 0$,

$$\inf_{X_t} R_\epsilon(x) = R_\epsilon(\tilde{x}(t)) = Qt^2 + \epsilon^2 \sum_{k=1}^{\infty} \sigma_k^2 (1 - t a_k)_+^2 \stackrel{\text{def}}{=} S_\epsilon(t) , \tag{A.9}$$

where $\tilde{x}_k(t) = (1 - t a_k)_+$. Thus, according to (A.9), we have

$$\inf_x R_\epsilon(x) = \inf_{t \geq 0} \inf_{x \in X_t} R_\epsilon(x) = \inf_{t \geq 0} R_\epsilon(\tilde{x}(t)) = \inf_{t \geq 0} S_\epsilon(t) ,$$

Since the function $S_\epsilon(t)$, $t \geq 0$, is continuous, strictly convex, differentiable and $S_\epsilon(\infty) = \infty$, there exists a unique point in which $\inf_{t \geq 0} S_\epsilon(t)$ is attained. Therefore, from (A.8) and the last relations we have

$$\inf_x \sup_{\theta \in \Theta} R_\epsilon(x, \theta) = \sup_{\theta \in \Theta} R_\epsilon(\tilde{x}, \theta) = \inf_{t \geq 0} S_\epsilon(t) = S_\epsilon(c_\epsilon) \tag{A.10}$$

because $S'_\epsilon(c_\epsilon) = 0$ is exactly the equation (A.2) and $\tilde{x}(c_\epsilon) = \tilde{x}$.

From (A.2) it follows that

$$Qc_\epsilon^2 = \epsilon^2 \sum_{k=1}^{\infty} \sigma_k^2 c_\epsilon a_k (1 - c_\epsilon a_k)_+ .$$

Therefore,

$$\begin{aligned} S_\epsilon(c_\epsilon) &= Qc_\epsilon^2 + \epsilon^2 \sum_{k=1}^{\infty} \sigma_k^2 (1 - c_\epsilon a_k)_+^2 \\ &= \epsilon^2 \sum_{k=1}^{\infty} \sigma_k^2 ((c_\epsilon a_k (1 - c_\epsilon a_k)_+ + (1 - c_\epsilon a_k)_+^2)) \\ &= \epsilon^2 \sum_{k=1}^{\infty} \sigma_k^2 (1 - c_\epsilon a_k)_+ = d_\epsilon . \end{aligned}$$

Consider now $\inf_x R_\epsilon(x, \theta)$. It is clear that

$$\inf_x R_\epsilon(x, \theta) = \sum_{k=1}^{\infty} \frac{\epsilon^2 \sigma_k^2 \theta_k^2}{\theta_k^2 + \epsilon^2 \sigma_k^2} ,$$

where the infimum is attained for

$$x_k = x_k(\theta_k) = \frac{\theta_k^2}{\epsilon^2 \sigma_k^2 + \theta_k^2} .$$

Note that $\tilde{x}_k = x_k(\tilde{\theta}_k)$, where the $\tilde{x}_k$'s and the $\tilde{\theta}_k$'s are given by (A.5) and (A.6). Therefore, the only feasible saddle point is $(\tilde{x}, \tilde{\theta})$. It follows from (A.2) that $\tilde{\theta} \in \Theta$. Therefore, by direct calculations,

$$\sup_{\theta \in \Theta} \inf_x R_\epsilon(x, \theta) \geq \inf_x R_\epsilon(x, \tilde{\theta}) = d_\epsilon .$$

□

A.2 The van Trees inequality

In this section we give a version of van Trees inequality we use in chapters 2, 3 and 4. This inequality is due to van Trees (1968), for a more recent reference see Gill and Levit (1995), where some interesting applications of this inequality are given. Further applications of the van

Trees inequality can be found in Bobrovsky et al. (1987), Klaassen (1989), Brown and Gajek (1990), Golubev and Levit (1996a), (1996b) and Schipper (1996).

The van Trees inequality applies to the quadratic Bayes risk in the problem of estimating a finite dimensional parameter, while we are interested in nonparametric situations; for example, in estimating a curve ranging over a nonparametric class. Therefore, to obtain a reasonable lower bound, one has to choose carefully the finite dimensional subfamilies in the nonparametric class under consideration. The idea is to select the subfamilies of growing dimensions such that the difficulty of the problem of estimating a parameter of growing dimension is asymptotically equal to the difficulty of the original nonparametric estimation problem. This requires an understanding how the underlying nonparametric class can be best approximated by a finite dimensional subset.

We follow Gill and Levit (1995).

Let Θ be a closed interval on the real line and let $f(x, \theta)$ be the density of a measure P_θ with respect to the dominating measure μ. Further, let π be some probability distribution on Θ with a density λ with respect to Lebesgue measure. Suppose that λ and $f(x, \cdot)$ are both absolutely continuous, and that λ converges to zero at the endpoints of the interval Θ.

Suppose Y is drawn from the distribution π, and, conditional on $Y = \theta$, let $X_1, \ldots, X_n$ be a sample from the probability distribution P_θ. We write $\mathbf{E}_\theta$ for expectation with respect to the conditional distribution P_θ, and $\mathbf{E}$ for expectation with respect to the joint distribution of $X_1, \ldots, X_n$ and Y. Let $\hat{\theta}_n = \hat{\theta}_n(X_1, \ldots, X_n)$ be any estimator of θ based on the observations. Define the Fisher information for θ

$$I(\theta) = \mathbf{E}_\theta \left(\frac{\partial \log f(x, \theta)}{\partial \theta} \right)^2$$

and the Fisher information for a location parameter in the distribution π

$$I(\lambda) = \mathbf{E} \left(\frac{\partial \log \lambda(x)}{\partial x} \right)^2 .$$

Theorem A.1. *Suppose that $f(x, \theta)$ is absolutely continuous in θ. Suppose further that $I(\theta)$ is continuous. Then, for any absolutely continuous function ψ and any estimator $\hat{\theta}_n$, the following relation holds:*

$$\mathbf{E} \left(\hat{\theta}_n - \psi(Y) \right)^2 \geq \frac{(\mathbf{E}\, \psi'(Y))^2}{n \mathbf{E} I(Y) + I(\lambda)}.$$

The proof of this theorem can be found in Gill and Levit (1995).

A.3 An approximation of the Kaplan-Meier estimator

Throughout this section we use the notations from Chapter 4, i.e. F and $\tilde{F}_n$ denote the distribution function of "survival times" and the Kaplan-Meier estimator respectively.

In the following theorem of Burke et al. (1988) (see also Burke et al. (1981)) an approximation of the Kaplan-Meier estimator is given in terms of Wiener processes.

Theorem A.2 (Burke, Csörgő and Horváth (1988)). *Let the distribution function F be continuous and, for each $n \geq 1$, let $T_n = T_n(\epsilon)$ be a number such that*

$$T_n < \tau_F \qquad \text{and} \qquad 1 - F(T_n) \geq (2\epsilon n^{-1} \log n)^{1/2},$$

where ϵ is any given positive number. Then on a suitable probability space there exists a sequence $\{W_n(t), 0 \leq t < \infty\}_{n=1}^{\infty}$ of standard Wiener processes such that

$$P\left\{ \sup_{t \leq T_n} |R_n(y)| > B_1(1 - F(T_n))^{-4} n^{-1} \log n \right\} \leq B_2 n^{-\epsilon},$$

where

$$R_n(y) = R_n(y, F, G) = (\tilde{F}_n(y) - F(y)) - n^{-1/2} \bar{F}(y) W_n(g(y)),$$

$B_1 = B_1(\epsilon)$ and B_2 are some positive constants.

The proof of this theorem can be found in Burke et al. (1988).

Bibliography

Andersen, P. K., O. Borgan, R. D. Gill, and N. Keiding (1993). *Statistical models based on counting processes.* Springer Verlag, New York.

Belitser, E. (1993). An adaptive minimax filtering algorithm for a functional of the nonparametric median. Unpublished manuscript.

Belitser, E. (1997). On second order efficiency of a smoothed Kaplan-Meier estimator. Under preparation.

Belitser, E. (1998a). Efficient estimation of analytic density under random censorship. *Bernoulli 4*, 519–543.

Belitser, E. (1998b). Local minimax pointwise estimation of a multivariate density. Accepted by Statistica Neerlandica.

Belitser, E. N. and A. P. Korostelev (1992). Pseudovalues and minimax filtering algorithms for the nonparametric median. In R. Z. Hasminskii (Ed.), *Advances in Soviet Mathematics*, Volume 12, pp. 115–124. Providence: American Mathematical Society.

Belitser, E. N. and B. Y. Levit (1994). On minimax estimation over ellipsoids. In P. Lachout and J. A. Víšek (Eds.), *Information theory, statistical decision functions, random processes*, pp. 28–31. Academy of sciences of the Czech republic, Institute of information theory and automation, Prague.

Belitser, E. N. and B. Y. Levit (1995). On minimax filtering over ellipsoids. *Math. Methods Statist. 4*(3), 259–273.

Belitser, E. N. and B. Y. Levit (1996). Asymptotically minimax nonparametric regression in L_2. *Statistics 28*, 105–122.

Belitser, E. N. and B. Y. Levit (1997). Asymptotically local minimax estimation of infinitely smooth density with censored data. Accepted by AISM.

Bickel, P. J., C. A. J. Klaasen, Y. Ritov, and J. A. Wellner (1993). *Efficient and Adaptive Estimation for Semiparametric Models.* John Hopkins University Press, Baltimore.

Birgé, L. (1983). Approximation dans les espaces métriques et théorie de l'estimation. *Z. Wahrsch. Verw. Gebiete 65*, 181–237.

Birgé, L. (1985). Nonasymptotic minimax risk for Hellinger balls. *Probab. Math. Statist. 5*, 21–29.

Birgé, L. (1986). On estimating a density using Hellinger distance and some other strange facts. *Probab. Theory Related Fields 71*, 271–291.

Birgé, L. (1987). Estimating a density under order restrictions: nonasymptotic minimax risk. *Ann. Statist. 15*, 995–1012.

Bobrovsky, B. Z., E. Mayer-Wolf, and M. Zakai (1987). Some classes of global Cramér-Rao bounds. *Ann. Statist. 15*, 1421–1438.

Boente, G. and R. Fraiman (1989). Robust nonparametric regression estimation for dependent observation. *Ann. Statist. 17*, 1242–1256.

Borovkov, A. A. (1984). *Mathematical Statistics. Parameter Estimation.* Nauka, Moscow. In Russian.

Bretagnolle, J. and C. Huber (1979). Estimation des densitiés: risque minimax. *Z. Wahrsch. Verw. Gebiete 47*, 119–137.

Brown, L. D. and L. Gajek (1990). Information inequalities for the Bayes risk. *Ann. Statist. 18*, 1578–1594.

Brown, L. D. and M. G. Low (1996). Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist. 24*, 2384–2398.

Burke, M. D., S. Csörgő, and L. Horváth (1981). Strong approximations of some biometric estimates under random censorship. *Z. Wahrsch. Verw. Gebiete 56*, 87–112.

Burke, M. D., S. Csörgő, and L. Horváth (1988). A correction to and improvement of 'Strong approximations of some biometric estimates under random censorship'. *Probab. Theory Related Fields 79*, 51–57.

Diehl, S. and W. Stute (1988). Kernel density and hazard function estimation in the presence of censoring. *J. Mult. Anal. 25*, 299–310.

Donoho, D. L. (1994). Asymptotic minimax risk for sup-norm loss: Solution via optimal recovery. *Probab. Theory Related Fields 99*, 145–170.

Donoho, D. L. and I. M. Johnstone (1994a). Ideal spatial adaptation by wavelet shrinkage. *Biometrika 81*, 425–455.

Donoho, D. L. and I. M. Johnstone (1994b). Minimax risk over l_p-balls for l_q-error. *Probab. Theory Related Fields 99*, 277–303.

Donoho, D. L. and I. M. Johnstone (1995). Adapting to unknown smoothness via wavelet shrinkage. *J. Amer. Statist. Assoc. 90*, 1200–1224.

Donoho, D. L., I. M. Johnstone, G. G. Kerkyacharian, and D. Picard (1995). Wavelet shrinkage: Asymptopia? *J. Roy. Statist. Soc. Ser. B 57*, 301–369. With discussion.

Donoho, D. L. and R. C. Liu (1991). Geometrizing rates of convergence, III. *Ann. Statist. 19*, 668–701.

Donoho, D. L., R. C. Liu, and B. MacGibbon (1990). Minimax risk over hyperrectangles, and implications. *Ann. Statist. 18*, 1416–1437.

Efromovich, S. Y. (1985). Nonparametric estimation of a density of unknown smoothness. *Theory Probab. Appl. 30*(3), 557–568.

Efromovich, S. Y. (1996). On nonparametric regression for iid observations in a general setting. *Ann. Statist. 24*, 1126–1144.

Efromovich, S. Y. and M. S. Pinsker (1982). Estimation of square-integrable density of a random variable. *Problems Inform. Transmission 18*, 175–189.

Efromovich, S. Y. and M. S. Pinsker (1984). Learning algorithm for nonparametric filtering. *Autom. Remote Contr. 11*, 1434–1440.

Fedoruk, M. V. (1977). *Metod perevala.* Nauka, Moscow. In Russian.

Gill, R. D. (1980). *Censoring and Stochastic Integrals*, Volume 124. Mathematisch Centrum, Amsterdam.

Gill, R. D. and B. Y. Levit (1995). Applications of the van Trees inequality: a Bayesian Cramér-Rao bound. *Bernoulli 1*(1), 59–79.

Golubev, G. K. (1987). Adaptive asymptotically minimax estimators of smooth signals. *Problems Inform. Transmission 23*, 47–55.

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

Golubev, G. K. (1992). Asymptotically minimax regression estimation in additive model. *Problems Inform. Transmission 28*, 3–15. In Russian.

Golubev, G. K. and B. Y. Levit (1996a). Asymptotically efficient estimation for analytic distributions. *Math. Methods Statist. 5*(3), 357–368.

Golubev, G. K. and B. Y. Levit (1996b). On the second order minimax esimation of distribution functions. *Math. Methods Statist. 5*(1), 1–31.

Golubev, G. K., B. Y. Levit, and A. B. Tsybakov (1996). Asymptotically efficient estimation of analytic functions in Gaussian noise. *Bernoulli 2*(2), 167–182.

Golubev, G. K. and M. Nussbaum (1990). A risk bound in Solobev class regression. *Ann. Statist. 18*(2), 758–778.

Golubev, G. K. and M. Nussbaum (1992). Adaptive spline estimation for nonparametric regression models. *Theory Probab. Appl. 37*, 512–529.

Gradshtein, I. S. and I. M. Ryzhik (1980). *Table of Integrals, Series, and Products.* Academic press, New York.

Groeneboom, P. (1996). Lectures on inverse problems. Technical report, Delft University of Technology.

Groeneboom, P. and J. A. Wellner (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation.* Birkhäuser, New York.

Hall, P. (1989). On convergence rates in nonparametric problems. *Internat. Statist. Rev. 57*(1), 45–58.

Härdle, W. and J. S. Marron (1985). Optimal bandwidth selection in nonparametric regression function estimation. *Ann. Statist. 13*, 1465–1481.

Hentzschel, J. (1992). Density estimation with Laguerre series and censored samples. *Statistics 23*, 49–61.

Huang, J. and J. A. Wellner (1995). Estimation of a monotone density and monotone hazard under random censoring. *Scand. J. Statist. 22*, 3–33.

Ibragimov, I. A. and R. Z. Hasminskii (1980). On nonparametric estimation of regression. *Soviet Math. Dokl. 21*(3), 810–814.

Ibragimov, I. A. and R. Z. Hasminskii (1981). *Statistical Estimation: Asymptotic Theory.* Springer Verlag, New York–Heidelberg–Berlin.

Ibragimov, I. A. and R. Z. Hasminskii (1982). Bounds for the risks of non-parametric regression estimates. *Theory Probab. Appl. 27*(1), 84–99.

Ibragimov, I. A. and R. Z. Hasminskii (1983). Estimation of distribution density. *Journ. Sov. Math. 21*(2), 40–57. Originally publushed in Russian in 1980.

Ibragimov, I. A. and R. Z. Hasminskii (1984a). Asymptotic bounds on the quality of the nonparametric regression estimation in L_p. *Journ. Sov. Math. 24*(5), 540–550. Originally publushed in Russian in 1980.

Ibragimov, I. A. and R. Z. Hasminskii (1984b). More on the estimation of distribution densities. *Journ. Sov. Math. 25*(3), 1155–1165. Originally publushed in Russian in 1981.

Ibragimov, I. A. and R. Z. Hasminskii (1984c). On nonparametric estimation of the value of a linear functional in Gaussian white noise. *Theory Probab. Appl. 29*, 18–32.

Ibragimov, I. A. and R. Z. Hasminskii (1990). On density estimation in the view of Kolmogorov's ideas in approximation theory. *Ann. Statist. 18*(4), 999–1010.

Ibragimov, I. A. and R. Z. Hasminskii (1991). Asymptotically normal families of distributions and efficient estimation. *Ann. Statist. 19*(4), 1681–1724.

Katznelson, Y. (1976). *An introduction to harmonic analysis.* Dover publications, New York. Second corrected edition.

Klaassen, C. A. J. (1989). The asymptotic spread of estimators. *J. Statist. Plannning and Inference 23*, 267–285.

Korostelev, A. P. (1994). An asymptotically minimax regression estimator in the uniform norm up to exact constant. *Theory Probab. Appl. 38*, 737–743.

Korostelev, A. P. and M. Nussbaum (1995). Density estimation in the uniform norm and white noise approximation. Technical Report 153, Weierstrass Institute, Berlin.

Koshevnik, Y. A. and B. Y. Levit (1976). On a non-parametric analogue of the information matrix. *Theory Probab. Appl. 21*, 738–753.

Kulasekera, K. B. (1995). A bound on the L_1-error of a nonparametric density estimator with censored data. *Statist. Probab. Lett. 23*, 233–238.

Lepski, O. V. (1990). One problem of adaptive estimation in Gaussian white noise. *Theory Probab. Appl. 35*, 459–470.

Lepski, O. V. (1991). Asymptotic minimax adaptive estimation. 1. Upper bounds. *Theory Probab. Appl. 36*, 645–659.

Lepski, O. V. (1992). Asymptotic minimax adaptive estimation. 2. Statistical model without optimal adaptation. Adaptive estimators. *Theory Probab. Appl. 37*, 468–481.

Lepski, O. V. and V. G. Spokoiny (1996). Optimal pointwise adaptive methods in noparametric estimation. Technical Report 229, Weierstrass Institute, Berlin.

Levit, B. Y. (1978). Infinite-dimensional informational lower bounds. *Theory Probab. Appl. 23*, 388–394.

Liu, R. C. (1996). Optimal rates, minimax estimations, and K-M based procedures for estimating functionals of a distribution under censoring. Technical report, Department of Mathematics, Cornell University.

Lo, S. H., Y. P. Mack, and J. L. Wang (1989). Density and hazard rate estimation for censored data via strong representation of the Kaplan-Meier estimator. *Prob. Theory and Rel. Fields 80*, 461–473.

Marron, J. S. (1988). Automatic smoothing parameter selection: a survey. *Empirical Economics 13*, 187–208.

Marron, J. S. (1989). Automatic smoothing parameter selection: a survey. In A. Ullah (Ed.), *Semiparametric and Nonparametric Econometrics*, pp. 65–86. Physika, Heidelberg.

Mielniczuk, J. (1986). Some asymptotic properties of kernel estimators of a density function in case of censored data. *Ann. Statist. 14*, 766–773.

Nikol'skii, S. M. (1975). *Approximation of Functions of Several Variables and Imbedding Theorems.* Springer Verlag, Berlin–Heidelberg–New York.

Nussbaum, M. (1985). Spline smoothing in regression models and asymptotic efficiency in L_2. *Ann. Statist. 13*, 984–997.

Nussbaum, M. (1996). Asymptotic equivalence of density estimation and Gaussian white noise. *Ann. Statist. 24*, 2399–2430.

Oudshoorn, C. G. M. (1996). *Optimality and Adaptivity in Nonparametric Regression.* Ph. D. thesis, University of Utrecht, Department of Mathematics.

Pfanzagl, J. (1982). *Contribution to a General Asymptotic Statistical Theory*, Volume 13. Lecture Notes in Statist., Springer, New York.

Pinsker, M. S. (1980). Optimal filtration of square-integrable signals in Gaussian noise. *Problems Inform. Transmission 16*(2), 120–133.

Pollard, D. (1984). *Convergence of Stochastic Processes.* Springer Verlag, New York.

Rosenblatt, M. (1956). On some nonparametric estimates of a density function. *Ann. Math. Statist. 27*, 832–837.

Schipper, M. (1996). Optimal rates and constants in L_2-minimax esimation of probability density functions. *Math. Methods Statist. 5*(3), 253–274.

Schipper, M. (1997). *Sharp Asymptotics in Nonparametric Estimation.* Ph. D. thesis, University of Utrecht, Department of Mathematics.

Speckman, P. (1985). Spline smoothing and optimal rates of convergence in nonparametric regression models. *Ann. Statist. 13*, 970–983.

Stone, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *Ann. Statist. 8*, 1348–1360.

Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist. 10*(4), 1040–1053.

Stone, C. J. (1984). An asymptotically optimal window selection rule for kernel density estimates. *Ann. Statist. 12*, 1285–1297.

Stone, C. J. (1985). Additive regression and other nonparametric models. *Ann. Statist. 13*, 689–705.

Timan, A. F. (1963). *Theory of Approximation of Functions of a Real Variable.* Pergamon Press, Oxford–London–New York–Paris.

Truong, Y. K. and C. J. Stone (1992). Nonparametric function estimation involving time series. *Ann. Statist. 20*, 77–97.

Tsybakov, A. B. (1997). Asymptotically efficient signal estimation in L_2 under general loss functions. To appear in Problems Inform. Transmission.

van der Vaart, A. W. (1988). *Statistical Estimation in Large Parameter Spaces*, Volume 44. Centrum Wisk. Inform., Amsterdam.

van der Vaart, A. W. (1991). On differentiable functionals. *Ann. Statist. 19*, 178–204.

van Trees, H. L. (1968). *Detection, Estimation and Modulation Theory, Part 1.* Wiley, New York.

Wald, A. (1950). *Statistical Decision Functions.* John Wiley & Sons, New York.

Weits, E. (1993). The second order optimality of a smoothed Kaplan-Meier estimator. *Scand. J. Statist. 20*, 111–132.

Yang, S. (1994). A central limit theorem for functionals of the Kaplan-Meier estimator. *Statist. Probab. Lett. 21*, 337–345.

CWI TRACTS

1 D.H.J. Epema. *Surfaces with canonical hyperplane sections.* 1984.
2 J.J. Dijkstra. *Fake topological Hilbert spaces and characterizations of dimension in terms of negligibility.* 1984.
3 A.J. van der Schaft. *System theoretic descriptions of physical systems.* 1984.
4 J. Koene. *Minimal cost flow in processing networks, a primal approach.* 1984.
5 B. Hoogenboom. *Intertwining functions on compact Lie groups.* 1984.
6 A.P.W. Böhm. *Dataflow computation.* 1984.
7 A. Blokhuis. *Few-distance sets.* 1984.
8 M.H. van Hoorn. *Algorithms and approximations for queueing systems.* 1984.
9 C.P.J. Koymans. *Models of the lambda calculus.* 1984.
10 C.G. van der Laan, N.M. Temme. *Calculation of special functions: the gamma function, the exponential integrals and error-like functions.* 1984.
11 N.M. van Dijk. *Controlled Markov processes; time-discretization.* 1984.
12 W.H. Hundsdorfer. *The numerical solution of nonlinear stiff initial value problems: an analysis of one step methods.* 1985.
13 D. Grune. *On the design of ALEPH.* 1985.
14 J.G.F. Thiemann. *Analytic spaces and dynamic programming: a measure theoretic approach.* 1985.
15 F.J. van der Linden. *Euclidean rings with two infinite primes.* 1985.
16 R.J.P. Groothuizen. *Mixed elliptic-hyperbolic partial differential operators: a case-study in Fourier integral operators.* 1985.
17 H.M.M. ten Eikelder. *Symmetries for dynamical and Hamiltonian systems.* 1985.
18 A.D.M. Kester. *Some large deviation results in statistics.* 1985.
19 T.M.V. Janssen. *Foundations and applications of Montague grammar, part 1: Philosophy, framework, computer science.* 1986.
20 B.F. Schriever. *Order dependence.* 1986.
21 D.P. van der Vecht. *Inequalities for stopped Brownian motion.* 1986.
22 J.C.S.P. van der Woude. *Topological dynamix.* 1986.
23 A.F. Monna. *Methods, concepts and ideas in mathematics: aspects of an evolution.* 1986.
24 J.C.M. Baeten. *Filters and ultrafilters over definable subsets of admissible ordinals.* 1986.
25 A.W.J. Kolen. *Tree network and planar rectilinear location theory.* 1986.
26 A.H. Veen. *The misconstrued semicolon: Reconciling imperative languages and dataflow machines.* 1986.
27 A.J.M. van Engelen. *Homogeneous zero-dimensional absolute Borel sets.* 1986.
28 T.M.V. Janssen. *Foundations and applications of Montague grammar, part 2: Applications to natural language.* 1986.
29 H.L. Trentelman. *Almost invariant subspaces and high gain feedback.* 1986.
30 A.G. de Kok. *Production-inventory control models: approximations and algorithms.* 1987.
31 E.E.M. van Berkum. *Optimal paired comparison designs for factorial experiments.* 1987.
32 J.H.J. Einmahl. *Multivariate empirical processes.* 1987.
33 O.J. Vrieze. *Stochastic games with finite state and action spaces.* 1987.
34 P.H.M. Kersten. *Infinitesimal symmetries: a computational approach.* 1987.
35 M.L. Eaton. *Lectures on topics in probability inequalities.* 1987.
36 A.H.P. van der Burgh, R.M.M. Mattheij (editors). *Proceedings of the first international conference on industrial and applied mathematics (ICIAM 87).* 1987.
37 L. Stougie. *Design and analysis of algorithms for stochastic integer programming.* 1987.
38 J.B.G. Frenk. *On Banach algebras, renewal measures and regenerative processes.* 1987.
39 H.J.M. Peters, O.J. Vrieze (eds.). *Surveys in game theory and related topics.* 1987.
40 J.L. Geluk, L. de Haan. *Regular variation, extensions and Tauberian theorems.* 1987.
41 Sape J. Mullender (ed.). *The Amoeba distributed operating system: Selected papers 1984-1987.* 1987.
42 P.R.J. Asveld, A. Nijholt (eds.). *Essays on concepts, formalisms, and tools.* 1987.
43 H.L. Bodlaender. *Distributed computing: structure and complexity.* 1987.
44 A.W. van der Vaart. *Statistical estimation in large parameter spaces.* 1988.
45 S.A. van de Geer. *Regression analysis and empirical processes.* 1988.
46 S.P. Spekreijse. *Multigrid solution of the steady Euler equations.* 1988.
47 J.B. Dijkstra. *Analysis of means in some non-standard situations.* 1988.
48 F.C. Drost. *Asymptotics for generalized chi-square goodness-of-fit tests.* 1988.
49 F.W. Wubs. *Numerical solution of the shallow-water equations.* 1988.
50 F. de Kerf. *Asymptotic analysis of a class of perturbed Korteweg-de Vries initial value problems.* 1988.
51 P.J.M. van Laarhoven. *Theoretical and computational aspects of simulated annealing.* 1988.
52 P.M. van Loon. *Continuous decoupling transformations for linear boundary value problems.* 1988.
53 K.C.P. Machielsen. *Numerical solution of optimal control problems with state constraints by sequential quadratic programming in function space.* 1988.
54 L.C.R.J. Willenborg. *Computational aspects of survey data processing.* 1988.
55 G.J. van der Steen. *A program generator for recognition, parsing and transduction with syntactic patterns.* 1988.
56 J.C. Ebergen. *Translating programs into delay-insensitive circuits.* 1989.
57 S.M. Verduyn Lunel. *Exponential type calculus for linear delay equations.* 1989.
58 M.C.M. de Gunst. *A random model for plant cell population growth.* 1989.
59 D. van Dulst. *Characterizations of Banach spaces not containing l^1.* 1989.
60 H.E. de Swart. *Vacillation and predictability properties of low-order atmospheric spectral models.* 1989.

61 P. de Jong. *Central limit theorems for generalized multilinear forms.* 1989.

62 V.J. de Jong. *A specification system for statistical software.* 1989.

63 B. Hanzon. *Identifiability, recursive identification and spaces of linear dynamical systems, part I.* 1989.

64 B. Hanzon. *Identifiability, recursive identification and spaces of linear dynamical systems, part II.* 1989.

65 B.M.M. de Weger. *Algorithms for diophantine equations.* 1989.

66 A. Jung. *Cartesian closed categories of domains.* 1989.

67 J.W. Polderman. *Adaptive control & identification: Conflict or conflux?.* 1989.

68 H.J. Woerdeman. *Matrix and operator extensions.* 1989.

69 B.G. Hansen. *Monotonicity properties of infinitely divisible distributions.* 1989.

70 J.K. Lenstra, H.C. Tijms, A. Volgenant (eds.). *Twenty-five years of operations research in the Netherlands: Papers dedicated to Gijs de Leve.* 1990.

71 P.J.C. Spreij. *Counting process systems. Identification and stochastic realization.* 1990.

72 J.F. Kaashoek. *Modeling one dimensional pattern formation by anti-diffusion.* 1990.

73 A.M.H. Gerards. *Graphs and polyhedra. Binary spaces and cutting planes.* 1990.

74 B. Koren. *Multigrid and defect correction for the steady Navier-Stokes equations. Application to aerodynamics.* 1991.

75 M.W.P. Savelsbergh. *Computer aided routing.* 1992.

76 O.E. Flippo. *Stability, duality and decomposition in general mathematical programming.* 1991.

77 A.J. van Es. *Aspects of nonparametric density estimation.* 1991.

78 G.A.P. Kindervater. *Exercises in parallel combinatorial computing.* 1992.

79 J.J. Lodder. *Towards a symmetrical theory of generalized functions.* 1991.

80 S.A. Smulders. *Control of freeway traffic flow.* 1996.

81 P.H.M. America, J.J.M.M. Rutten. *A parallel object-oriented language: design and semantic foundations.* 1992.

82 F. Thuijsman. *Optimality and equilibria in stochastic games.* 1992.

83 R.J. Kooman. *Convergence properties of recurrence sequences.* 1992.

84 A.M. Cohen (ed.). *Computational aspects of Lie group representations and related topics. Proceedings of the 1990 Computational Algebra Seminar at CWI, Amsterdam.* 1991.

85 V. de Valk. *One-dependent processes.* 1994.

86 J.A. Baars, J.A.M. de Groot. *On topological and linear equivalence of certain function spaces.* 1992.

87 A.F. Monna. *The way of mathematics and mathematicians.* 1992.

88 E.D. de Goede. *Numerical methods for the three-dimensional shallow water equations.* 1993.

89 M. Zwaan. *Moment problems in Hilbert space with applications to magnetic resonance imaging.* 1993.

90 C. Vuik. *The solution of a one-dimensional Stefan problem.* 1993.

91 E.R. Verheul. *Multimedians in metric and normed spaces.* 1993.

92 J.L.M. Maubach. *Iterative methods for non-linear partial differential equations.* 1994.

93 A.W. Ambergen. *Statistical uncertainties in posterior probabilities.* 1993.

94 P.A. Zegeling. *Moving-grid methods for time-dependent partial differential equations.* 1993.

95 M.J.C. van Pul. *Statistical analysis of software reliability models.* 1993.

96 J.K. Scholma. *A Lie algebraic study of some integrable systems associated with root systems.* 1993.

97 J.L. van den Berg. *Sojourn times in feedback and processor sharing queues.* 1993.

98 A.J. Koning. *Stochastic integrals and goodness-of-fit tests.* 1993.

99 B.P. Sommeijer. *Parallelism in the numerical integration of initial value problems.* 1993.

100 J. Molenaar. *Multigrid methods for semiconductor device simulation.* 1993.

101 H.J.C. Huijberts. *Dynamic feedback in nonlinear synthesis problems.* 1994.

102 J.A.M. van der Weide. *Stochastic processes and point processes of excursions.* 1994.

103 P.W. Hemker, P. Wesseling (eds.). *Contributions to multigrid.* 1994.

104 I.J.B.F. Adan. *A compensation approach for queueing problems.* 1994.

105 O.J. Boxma, G.M. Koole (eds.). *Performance evaluation of parallel and distributed systems - solution methods. Part 1.* 1994.

106 O.J. Boxma, G.M. Koole (eds.). *Performance evaluation of parallel and distributed systems - solution methods. Part 2.* 1994.

107 R.A. Trompert. *Local uniform grid refinement for time-dependent partial differential equations.* 1995.

108 M.N.M. van Lieshout. *Stochastic geometry models in image analysis and spatial statistics.* 1995.

109 R.J. van Glabbeek. *Comparative concurrency semantics and refinement of actions.* 1996.

110 W. Vervaat, H. Holwerda (ed.). *Probability and lattices.* 1997.

111 I. Helsloot. *Covariant formal group theory and some applications.* 1995.

112 R.N. Bol. *Loop checking in logic programming.* 1995.

113 G.J.M. Koole. *Stochastic scheduling and dynamic programming.* 1995.

114 M.J. van der Laan. *Efficient and inefficient estimation in semiparametric models.* 1995.

115 S.C. Borst. *Polling models.* 1996.

116 G.D. Otten. *Statistical test limits in quality control.* 1996.

117 K.G. Langendoen. *Graph reduction on shared-memory multiprocessors.* 1996.

118 W.C.A. Maas. *Nonlinear $\mathcal{H}_\infty$ control: the singular case.* 1996.

119 A. Di Bucchianico. *Probabilistic and analytical aspects of the umbral calculus.* 1997.

120 M. van Loon. *Numerical methods in smog prediction.* 1997.

121 B.J. Wijers. *Nonparametric estimation for a windowed line-segment process.* 1997.

122 W.K. Klein Haneveld, O.J. Vrieze, L.C.M. Kallenberg (editors). *Ten years LNMB – Ph.D. research and graduate courses of the Dutch Network of Operations Research.* 1997.

123 R.W. van der Hofstad. *One-dimensional random polymers.* 1998.

124 W.J.H. Stortelder. *Parameter estimation in nonlinear dynamical systems.* 1998.

125 M.H. Wegkamp. *Entropy methods in statistical estimation.* 1998.

126 K. Aardal, J.K. Lenstra, F. Maffioli, D.B. Shmoys (eds.) *Selected publications of Eugene L. Lawler.* 1999.

127 E. Belitser. *Minimax Estimation in Regression and Random Censorship Models.* 2000.

MATHEMATICAL CENTRE TRACTS

1 T. van der Walt. *Fixed and almost fixed points.* 1963.

2 A.R. Bloemena. *Sampling from a graph.* 1964.

3 G. de Leve. *Generalized Markovian decision processes, part I: model and method.* 1964.

4 G. de Leve. *Generalized Markovian decision processes, part II: probabilistic background.* 1964.

5 G. de Leve, H.C. Tijms, P.J. Weeda. *Generalized Markovian decision processes, applications.* 1970.

6 M.A. Maurice. *Compact ordered spaces.* 1964.

7 W.R. van Zwet. *Convex transformations of random variables.* 1964.

8 J.A. Zonneveld. *Automatic numerical integration.* 1964.

9 P.C. Baayen. *Universal morphisms.* 1964.

10 E.M. de Jager. *Applications of distributions in mathematical physics.* 1964.

11 A.B. Paalman-de Miranda. *Topological semigroups.* 1964.

12 J.A.Th.M. van Berckel, H. Brandt Corstius, R.J. Mokken, A. van Wijngaarden. *Formal properties of newspaper Dutch.* 1965.

13 H.A. Lauwerier. *Asymptotic expansions.* 1966, out of print: replaced by MCT 54.

14 H.A. Lauwerier. *Calculus of variations in mathematical physics.* 1966.

15 R. Doornbos. *Slippage tests.* 1966.

16 J.W. de Bakker. *Formal definition of programming languages with an application to the definition of ALGOL 60.* 1967.

17 R.P. van de Riet. *Formula manipulation in ALGOL 60, part 1.* 1968.

18 R.P. van de Riet. *Formula manipulation in ALGOL 60, part 2.* 1968.

19 J. van der Slot. *Some properties related to compactness.* 1968.

20 P.J. van der Houwen. *Finite difference methods for solving partial differential equations.* 1968.

21 E. Wattel. *The compactness operator in set theory and topology.* 1968.

22 T.J. Dekker. *ALGOL 60 procedures in numerical algebra, part 1.* 1968.

23 T.J. Dekker, W. Hoffmann. *ALGOL 60 procedures in numerical algebra, part 2.* 1968.

24 J.W. de Bakker. *Recursive procedures.* 1971.

25 E.R. Paërl. *Representations of the Lorentz group and projective geometry.* 1969.

26 European Meeting 1968. *Selected statistical papers, part I.* 1968.

27 European Meeting 1968. *Selected statistical papers, part II.* 1968.

28 J. Oosterhoff. *Combination of one-sided statistical tests.* 1969.

29 J. Verhoeff. *Error detecting decimal codes.* 1969.

30 H. Brandt Corstius. *Exercises in computational linguistics.* 1970.

31 W. Molenaar. *Approximations to the Poisson, binomial and hypergeometric distribution functions.* 1970.

32 L. de Haan. *On regular variation and its application to the weak convergence of sample extremes.* 1970.

33 F.W. Steutel. *Preservations of infinite divisibility under mixing and related topics.* 1970.

34 I. Juhász, A. Verbeek, N.S. Kroonenberg. *Cardinal functions in topology.* 1971.

35 M.H. van Emden. *An analysis of complexity.* 1971.

36 J. Grasman. *On the birth of boundary layers.* 1971.

37 J.W. de Bakker, G.A. Blaauw, A.J.W. Duijvestijn, E.W. Dijkstra, P.J. van der Houwen, G.A.M. Kamsteeg-Kemper, F.E.J. Kruseman Aretz, W.L. van der Poel, J.P. Schaap-Kruseman, M.V. Wilkes, G. Zoutendijk. *MC-25 Informatica Symposium.* 1971.

38 W.A. Verloren van Themaat. *Automatic analysis of Dutch compound words.* 1972.

39 H. Bavinck. *Jacobi series and approximation.* 1972.

40 H.C. Tijms. *Analysis of (s,S) inventory models.* 1972.

41 A. Verbeek. *Superextensions of topological spaces.* 1972.

42 W. Vervaat. *Success epochs in Bernoulli trials (with applications in number theory).* 1972.

43 F.H. Ruymgaart. *Asymptotic theory of rank tests for independence.* 1973.

44 H. Bart. *Meromorphic operator valued functions.* 1973.

45 A.A. Balkema. *Monotone transformations and limit laws.* 1973.

46 R.P. van de Riet. *ABC ALGOL, a portable language for formula manipulation systems, part 1: the language.* 1973.

47 R.P. van de Riet. *ABC ALGOL, a portable language for formula manipulation systems, part 2: the compiler.* 1973.

48 F.E.J. Kruseman Aretz, P.J.W. ten Hagen, H.L. Oudshoorn. *An ALGOL 60 compiler in ALGOL 60, text of the MC-compiler for the EL-X8.* 1973.

49 H. Kok. *Connected orderable spaces.* 1974.

50 A. van Wijngaarden, B.J. Mailloux, J.E.L. Peck, C.H.A. Koster, M. Sintzoff, C.H. Lindsey, L.G.L.T. Meertens, R.G. Fisker (eds.). *Revised report on the algorithmic language ALGOL 68.* 1976.

51 A. Hordijk. *Dynamic programming and Markov potential theory.* 1974.

52 P.C. Baayen (ed.). *Topological structures.* 1974.

53 M.J. Faber. *Metrizability in generalized ordered spaces.* 1974.

54 H.A. Lauwerier. *Asymptotic analysis, part 1.* 1974.

55 M. Hall, Jr., J.H. van Lint (eds.). *Combinatorics, part 1: theory of designs, finite geometry and coding theory.* 1974.

56 M. Hall, Jr., J.H. van Lint (eds.). *Combinatorics, part 2: graph theory, foundations, partitions and combinatorial geometry.* 1974.

57 M. Hall, Jr., J.H. van Lint (eds.). *Combinatorics, part 3: combinatorial group theory.* 1974.

58 W. Albers. *Asymptotic expansions and the deficiency concept in statistics.* 1975.

59 J.L. Mijnheer. *Sample path properties of stable processes.* 1975.

60 F. Göbel. *Queueing models involving buffers.* 1975.

63 J.W. de Bakker (ed.). *Foundations of computer science.* 1975.

64 W.J. de Schipper. *Symmetric closed categories.* 1975.

65 J. de Vries. *Topological transformation groups, 1: a categorical approach.* 1975.

66 H.G.J. Pijls. *Logically convex algebras in spectral theory and eigenfunction expansions.* 1976.

68 P.P.N. de Groen. *Singularly perturbed differential operators of second order.* 1976.

69 J.K. Lenstra. *Sequencing by enumerative methods.* 1977.

70 W.P. de Roever, Jr. *Recursive program schemes: semantics and proof theory.* 1976.

71 J.A.E.E. van Nunen. *Contracting Markov decision processes.* 1976.

72 J.K.M. Jansen. *Simple periodic and non-periodic Lamé functions and their applications in the theory of conical waveguides.* 1977.

73 D.M.R. Leivant. *Absoluteness of intuitionistic logic.* 1979.

74 H.J.J. te Riele. *A theoretical and computational study of generalized aliquot sequences.* 1976.

75 A.E. Brouwer. *Treelike spaces and related connected topological spaces.* 1977.

76 M. Rem. *Associons and the closure statements.* 1976.

77 W.C.M. Kallenberg. *Asymptotic optimality of likelihood ratio tests in exponential families.* 1978.

78 E. de Jonge, A.C.M. van Rooij. *Introduction to Riesz spaces.* 1977.

79 M.C.A. van Zuijlen. *Empirical distributions and rank statistics.* 1977.

80 P.W. Hemker. *A numerical study of stiff two-point boundary problems.* 1977.

81 K.R. Apt, J.W. de Bakker (eds.). *Foundations of computer science II, part 1.* 1976.

82 K.R. Apt, J.W. de Bakker (eds.). *Foundations of computer science II, part 2.* 1976.

83 L.S. van Benthem Jutting. *Checking Landau's "Grundlagen" in the AUTOMATH system.* 1979.

84 H.L.L. Busard. *The translation of the elements of Euclid from the Arabic into Latin by Hermann of Carinthia (?), books vii-xii.* 1977.

85 J. van Mill. *Supercompactness and Wallmann spaces.* 1977.

86 S.G. van der Meulen, M. Veldhorst. *Torrix I, a programming system for operations on vectors and matrices over arbitrary fields and of variable size.* 1978.

88 A. Schrijver. *Matroids and linking systems.* 1977.

89 J.W. de Roever. *Complex Fourier transformation and analytic functionals with unbounded carriers.* 1978.

90 L.P.J. Groenewegen. *Characterization of optimal strategies in dynamic games.* 1981.

91 J.M. Geysel. *Transcendence in fields of positive characteristic.* 1979.

92 P.J. Weeda. *Finite generalized Markov programming.* 1979.

93 H.C. Tijms, J. Wessels (eds.). *Markov decision theory.* 1977.

94 A. Bijlsma. *Simultaneous approximations in transcendental number theory.* 1978.

95 K.M. van Hee. *Bayesian control of Markov chains.* 1978.

96 P.M.B. Vitányi. *Lindenmayer systems: structure, languages, and growth functions.* 1980.

97 A. Federgruen. *Markovian control problems; functional equations and algorithms.* 1984.

98 R. Geel. *Singular perturbations of hyperbolic type.* 1978.

99 J.K. Lenstra, A.H.G. Rinnooy Kan, P. van Emde Boas (eds.). *Interfaces between computer science and operations research.* 1978.

100 P.C. Baayen, D. van Dulst, J. Oosterhoff (eds.). *Proceedings bicentennial congress of the Wiskundig Genootschap, part 1.* 1979.

101 P.C. Baayen, D. van Dulst, J. Oosterhoff (eds.). *Proceedings bicentennial congress of the Wiskundig Genootschap, part 2.* 1979.

102 D. van Dulst. *Reflexive and superreflexive Banach spaces.* 1978.

103 K. van Harn. *Classifying infinitely divisible distributions by functional equations.* 1978.

104 J.M. van Wouwe. *GO-spaces and generalizations of metrizability.* 1979.

105 R. Helmers. *Edgeworth expansions for linear combinations of order statistics.* 1982.

106 A. Schrijver (ed.). *Packing and covering in combinatorics.* 1979.

107 C. den Heijer. *The numerical solution of nonlinear operator equations by imbedding methods.* 1979.

108 J.W. de Bakker, J. van Leeuwen (eds.). *Foundations of computer science III, part 1.* 1979.

109 J.W. de Bakker, J. van Leeuwen (eds.). *Foundations of computer science III, part 2.* 1979.

110 J.C. van Vliet. *ALGOL 68 transput, part I: historical review and discussion of the implementation model.* 1979.

111 J.C. van Vliet. *ALGOL 68 transput, part II: an implementation model.* 1979.

112 H.C.P. Berbee. *Random walks with stationary increments and renewal theory.* 1979.

113 T.A.B. Snijders. *Asymptotic optimality theory for testing problems with restricted alternatives.* 1979.

114 A.J.E.M. Janssen. *Application of the Wigner distribution to harmonic analysis of generalized stochastic processes.* 1979.

115 P.C. Baayen, J. van Mill (eds.). *Topological structures II, part 1.* 1979.

116 P.C. Baayen, J. van Mill (eds.). *Topological structures II, part 2.* 1979.

117 P.J.M. Kallenberg. *Branching processes with continuous state space.* 1979.

118 P. Groeneboom. *Large deviations and asymptotic efficiencies.* 1980.

119 F.J. Peters. *Sparse matrices and substructures, with a novel implementation of finite element algorithms.* 1980.

120 W.P.M. de Ruyter. *On the asymptotic analysis of large-scale ocean circulation.* 1980.

121 W.H. Haemers. *Eigenvalue techniques in design and graph theory.* 1980.

122 J.C.P. Bus. *Numerical solution of systems of nonlinear equations.* 1980.

123 I. Yuhász. *Cardinal functions in topology - ten years later.* 1980.

124 R.D. Gill. *Censoring and stochastic integrals.* 1980.

125 R. Eising. *2-D systems, an algebraic approach.* 1980.

126 G. van der Hoek. *Reduction methods in nonlinear programming.* 1980.

127 J.W. Klop. *Combinatory reduction systems.* 1980.

128 A.J.J. Talman. *Variable dimension fixed point algorithms and triangulations.* 1980.

129 G. van der Laan. *Simplicial fixed point algorithms.* 1980.

130 P.J.W. ten Hagen, T. Hagen, P. Klint, H. Noot, H.J. Sint, A.H. Veen. *ILP: intermediate language for pictures.* 1980.

131 R.J.R. Back. *Correctness preserving program refinements: proof theory and applications.* 1980.

132 H.M. Mulder. *The interval function of a graph.* 1980.

133 C.A.J. Klaassen. *Statistical performance of location estimators.* 1981.

134 J.C. van Vliet, H. Wupper (eds.). *Proceedings international conference on ALGOL 68.* 1981.

135 J.A.G. Groenendijk, T.M.V. Janssen, M.J.B. Stokhof (eds.). *Formal methods in the study of language, part I.* 1981.

136 J.A.G. Groenendijk, T.M.V. Janssen, M.J.B. Stokhof (eds.). *Formal methods in the study of language, part II.* 1981.

137 J. Telgen. *Redundancy and linear programs.* 1981.

138 H.A. Lauwerier. *Mathematical models of epidemics.* 1981.

139 J. van der Wal. *Stochastic dynamic programming, successive approximations and nearly optimal strategies for Markov decision processes and Markov games.* 1981.

140 J.H. van Geldrop. *A mathematical theory of pure exchange economies without the no-critical-point hypothesis.* 1981.

141 G.E. Welters. *Abel-Jacobi isogenies for certain types of Fano threefolds.* 1981.

142 H.R. Bennett, D.J. Lutzer (eds.). *Topology and order structures, part 1.* 1981.

143 J.M. Schumacher. *Dynamic feedback in finite- and infinite-dimensional linear systems.* 1981.

144 P. Eijgenraam. *The solution of initial value problems using interval arithmetic; formulation and analysis of an algorithm.* 1981.

145 A.J. Brentjes. *Multi-dimensional continued fraction algorithms.* 1981.

146 C.V.M. van der Mee. *Semigroup and factorization methods in transport theory.* 1981.

147 H.H. Tigelaar. *Identification and informative sample size.* 1982.

148 L.C.M. Kallenberg. *Linear programming and finite Markovian control problems.* 1983.

149 C.B. Huijsmans, M.A. Kaashoek, W.A.J. Luxemburg, W.K. Vietsch (eds.). *From A to Z, proceedings of a symposium in honour of A.C. Zaanen.* 1982.

150 M. Veldhorst. *An analysis of sparse matrix storage schemes.* 1982.

151 R.J.M.M. Does. *Higher order asymptotics for simple linear rank statistics.* 1982.

152 G.F. van der Hoeven. *Projections of lawless sequencies.* 1982.

153 J.P.C. Blanc. *Application of the theory of boundary value problems in the analysis of a queueing model with paired services.* 1982.

154 H.W. Lenstra, Jr., R. Tijdeman (eds.). *Computational methods in number theory, part I.* 1982.

155 H.W. Lenstra, Jr., R. Tijdeman (eds.). *Computational methods in number theory, part II.* 1982.

156 P.M.G. Apers. *Query processing and data allocation in distributed database systems.* 1983.

157 H.A.W.M. Kneppers. *The covariant classification of two-dimensional smooth commutative formal groups over an algebraically closed field of positive characteristic.* 1983.

158 J.W. de Bakker, J. van Leeuwen (eds.). *Foundations of computer science IV, distributed systems, part 1.* 1983.

159 J.W. de Bakker, J. van Leeuwen (eds.). *Foundations of computer science IV, distributed systems, part 2.* 1983.

160 A. Rezus. *Abstract AUTOMATH.* 1983.

161 G.F. Helminck. *Eisenstein series on the metaplectic group, an algebraic approach.* 1983.

162 J.J. Dik. *Tests for preference.* 1983.

163 H. Schippers. *Multiple grid methods for equations of the second kind with applications in fluid mechanics.* 1983.

164 F.A. van der Duyn Schouten. *Markov decision processes with continuous time parameter.* 1983.

165 P.C.T. van der Hoeven. *On point processes.* 1983.

166 H.B.M. Jonkers. *Abstraction, specification and implementation techniques, with an application to garbage collection.* 1983.

167 W.H.M. Zijm. *Nonnegative matrices in dynamic programming.* 1983.

168 J.H. Evertse. *Upper bounds for the numbers of solutions of diophantine equations.* 1983.

169 H.R. Bennett, D.J. Lutzer (eds.). *Topology and order structures, part 2.* 1983.