12-2007

Issues in Model Selection, Minimax Estimation, and Censored Data Analysis

Meng Zhao
Clemson University, mzhao@clemson.edu

Issues in Model Selection, Minimax Estimation, and Censored Data Analysis

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Mathematics

by
Meng Zhao
December 2007

Accepted by:
Dr. Karunarathna B Kulasekera, Committee Chair
Dr. Robert L Taylor
Dr. Herman F Senter
Dr. Chanseok Park

Abstract

In this dissertation, we address several research problems in statistical inference. We obtain results in the following four directions: linear model selection, minimax estimation of linear functionals, Bayes type estimators for the survival functions based on right censored data, and estimation of survival functions based on doubly censored data.

Table of Contents

Title Page i

Abstract ii

1 Introduction 1

1.1 Linear Model Selection 1

1.2 Pointwise Bayes Type Estimators of the Survival Probability 2

1.3 Minimax Estimation of Linear Functionals 2

1.4 Estimation of survival functions based on doubly censored data 3

2 Linear Model Selection 5

2.1 Linear Model Selection 5

2.2 Main Results 7

2.3 Proofs 13

3 Pointwise Bayes Type Estimator of the Survival Probability 20

3.1 Introduction 20

3.2 Estimation 22

3.3 Empirical Studies 27

3.4 Discussion 29

3.5 Proofs 33

4 Minimax Estimation of Linear Functionals 38

4.1 Introduction 38

4.2 The Hardest one dimensional sub-problem and the Full Problem 42

4.3 Applications 49

4.4 Proofs 55

5 Estimation of survival functions based on doubly censored data 102

5.1 Introduction 102

5.2 The Main Results 105

5.3 Conclusion 111

5.4 Proofs 111

Chapter 1

Introduction

In this dissertation, we address several research problems in statistical inference. We obtain results in the following four directions: linear model selection, minimax estimation of linear functionals, Bayes type estimators for the survival functions based on right censored data, and estimation of survival functions based on doubly censored data.

1.1 Linear Model Selection

In statistical inference, the underlying statistical model behind the data is generally unknown. It is sometimes assumed that the true model comes from a collection of candidate models. The problem of model selection is choosing the model that, according to certain criterion, best explains the data. In this work, we discuss a special case of a model selection: linear model selection. In a linear model, a set of predictors is used to predict the value of a response variable. Without knowing what predictors actually contribute to the response, generally, in the initial stage of linear modelling, a large number of predictors are included in the model. Using too many predictors is not only computationally expensive, but also results in over fitting, which means the model fits the data very well but performs poorly in predicting new outcomes.

Linear model selection procedures use various critera to judge which model is appropriate. Many of these critera are in the form of the sum of two terms, one of which is determined by residual sum of squares and the other is related to the "size" of the model. The second term can be considered a penalty on the size of the model. Such critera include Akaike information criterion (Akaike, 1970), Bayesian information criterion (Schwarz, 1978), C_p (Mallows, 1973), and GIC (Rao and Wu, 1989). One major difference among these critera is the different penalty terms that they use. It is generally required that a model selection procedure be consistent; i.e. as the sample size goes to infinity, the probability that the right model will be chosen converges to one. For a model selection criterion that relies on a penalty term, the penalty term has to be properly chosen in order for the procedure to be consistent; penalizing too much will result in under estimation while not penalizing enough will result in over estimation.

In this work, under a fairly general setting, we discuss linear model selection procedures that use a penalty term. We focus on the choice of the penalty term, providing sufficient conditions on the penalty term that ensures consistency of the model selection procedure.

1.2 Pointwise Bayes Type Estimator of the Survival Probability with Censored Data

Censored data frequently arise in practical situations such as reliability and medical follow-up studies. When analyzing right censored data, the Kaplan-Meier estimator (Kaplan and Meier, 1958) is generally used for estimating the survival function. The K-M estimator is consistent and asymptotically normal. One limitation of the K-M estimator is that it tends to be unstable in the tails.

We propose Bayes type estimators for the survival functions evaluated at a fixed point. This approach allows accommodation of a prior, which can be considered as weights assigned to different time points. We show that these Bayes type estimators, under mild conditions on the priors, are consistent and asymptotically normal. Simulation results show that these Bayes type estimators are superior to the K-M estimator over some parts of the support of the survival function.

1.3 Minimax Estimation of Linear Functionals

Estimation of linear functionals deals with the situation in which scientists observe an unknown signal corrupted with noise. It is generally assumed that the signal is from a predefined set of candidates. Also, one assumes certain properties of the noise structure. The objective is to estimate a linear functional of the signal. In most instances, there is an additional complication that the signal undergoes certain operations before being mixed with the noise. In other words, what is observed is the combination of the image of the signal under an operator and the noise. One example of estimation of linear functionals is the estimation of the value of a density function at a fixed point.

There has been numerous discussions on minimax estimation of such linear functionals in the literature (Ibragimov and Khasminskii, 1984; Donoho and Liu, 1991; Donoho, 1994; Zhao, 1997). In these articles, minimax risk of linear estimators and the rate of convergence of their minimax risk for the white noise model has been established. In a white noise model, it is assumed that the noise is a Gaussian white noise. Gaussion white noise model can be considered as a limit of a sequence of nonparametric regression models with independent errors or weakly associated errors (the errors have summable covariances). The minimax risk gives a guideline as to what is the best that an estimator can achieve, while the minimax rate of convergence provides an insight into the asymptotic behavior of the minimax estimators.

In our research, we generalize the results in the literature. We considered a general model which not only include the white noise model as a special case, but also the fractional Brownian motion model. The fractional Brownian motion model can be considered as a limit of a sequence of nonparametric regression models with strongly associated errors. Thus, our results can be applied to situations in which the assumption of short-range dependency is not appropriate.

1.4 Estimation of survival functions based on doubly censored data

Doubly censored data consist of observations that can be censored from two sides, both above and below. Estimation of the underlying survival function is one of the essential items in lifetime data analysis. In the literature, the self-consistent estimators – solutions of the self-consistent equation, are generally used as estimators for the survival function when the data are censored. For right censored data, the self-consistent estimator is the same as the K-M estimator. For doubly censored data, under certain conditions, the existence of self-consistent estimators has been proved (Turnbull, 1974a). These self-consistent estimators are solutions to the self-consistent equation based on empirical data. It has been shown that the nonparametric maximum likelihood estimator (NPMLE) is also self-consistent. Consistency and asymptotic normality of the self-consistent estimators above have been established in the literature by numerous researchers including Chang and Yang (1987), Samuelsen (1989), Gu and Zhang (1993), and Yu and Li (2001).

Although the available results in the literature address certain problems concerning estimation of the survival function based on doubly censored data, many scenarios are left unaddressed. For example, the above mentioned self-consistent estimators are solutions to the self-consistency equations , give discrete estimators. Such estimators may not be the optimal choice for estimating a smooth survival function. When estimating smooth survival functions, it may be desirable to consider smooth self-consistent estimators. Although it has been shown that solutions for the discrete self-consistency equations always exists, it has not been verified for more general situations.

To address these problems, we propose modified and more general self-consistent equations. We show that the solutions for such equations always exist. Under certain smoothness conditions, we also show that the self-consistent estimators are smooth. In addition, we define a generalized nonparametric maximum likelihood estimator (GNPMLE), and show that a GNPMLE is also self-consistent. Finally, we prove the consistency and asymptotic normality of the proposed self-consistent estimators.

Chapter 2

Linear Model Selection

2.1 Linear Model Selection

Consider the regression model

$$\mathbf{y}_n = \boldsymbol{\mu}_n + \mathbf{e}, \tag{2.1}$$

where $\mathbf{y}_n = (y_1, y_2, \ldots, y_n)'$ is a vector of n independent responses, with unknown mean vector $\boldsymbol{\mu}_n = (\mu_1, \ldots, \mu_n)'$, and $\mathbf{e}_n = (e_1, \ldots, e_n)'$ is a vector of n independent, identically distributed errors with common mean 0 and variance σ^2. Suppose that associated with y_i there is a p_n vector of covariates $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip_n})'$, and let $\mathbf{X}_n = (\mathbf{x}_1, \ldots, \mathbf{x}_n)'$ be the $n \times p_n$ design matrix, which for simplicity, is assumed to have full rank. Generally, p_n is assumed to be less than n. To estimate $\boldsymbol{\mu}_n$ we propose the linear model $\boldsymbol{\mu}_n = \mathbf{X}_n\boldsymbol{\beta}_n$ where $\boldsymbol{\beta}_n = (\beta_1, \ldots, \beta_{p_n})$ is a p_n vector of real valued parameters. It is possible that some of the components of $\boldsymbol{\beta}_n$ are 0. Thus, we may consider sub of the form

$$\boldsymbol{\mu}_n = \mathbf{X}_n(\alpha)\boldsymbol{\beta}(\alpha), \tag{2.2}$$

where α is a subset of $\{1, \ldots, p_n\}$ and $\boldsymbol{\beta}(\alpha)$ $(\mathbf{X}_n(\alpha))$ is the subvector (submatrix) containing the components of $\boldsymbol{\beta}_n$ (columns of $\mathbf{X}_n$) that are indexed by the integers in α.

There are $2^{p_n} - 1$ possible different models of the form $\mathbf{X}_n(\alpha)\boldsymbol{\beta}(\alpha)$. Instead of taking all of the $2^{p_n} - 1$ models into account, we assume that there is a class $\mathcal{A}_n$ of subsets of $\{1, \ldots, p_n\}$ and only models that correspond to members of $\mathcal{A}_n$ are considered to be candidate models. Let α_0 be the true model, that is, $\beta_i \neq 0$ for $i \in \alpha_0$ and $\beta_i = 0$ for $i \notin \alpha_0$. A model $\alpha \in \mathcal{A}_n$ is said to be correct if $\boldsymbol{\mu}_n = \mathbf{X}_n(\alpha)\boldsymbol{\beta}_n(\alpha)$ holds. This is equivalent to $\alpha_0 \subset \alpha$. Denote the set of all correct models in $\mathcal{A}_n$ by $\mathcal{AC}_n$, and define $\mathcal{AI}_n = \mathcal{A}_n \setminus \mathcal{AC}_n$. For each model in $\mathcal{A}_n$, we use least squares method to estimate $\boldsymbol{\mu}_n$. Let $\mathbf{M}_n(\alpha) = \mathbf{X}_n(\alpha)\left(\mathbf{X}_n(\alpha)'\mathbf{X}_n(\alpha)\right)^{-1}\mathbf{X}_n(\alpha)'$ be the hat matrix. Then $\hat{\boldsymbol{\mu}}_n(\alpha) = \mathbf{M}_n(\alpha)\mathbf{y}_n$ is the LSE of $\boldsymbol{\mu}_n$ under model α.

The dimension of a model α is defined to be $|\alpha|$, the number of elements in α. We are interested in procedures for finding correct model(s) that have the smallest dimension (we assume that such "smallest" correct model(s) exist). Many of the widely used procedures including AIC (Akaike, 1970) BIC (Schwarz, 1978) C_p (Mallows, 1973) and GIC (Rao and Wu, 1989) choose the minimizer $\hat{\alpha}_n$ of a criterion that has the form

$$T_{n,\lambda_n}(\alpha) = \frac{1}{n}RSS_n(\alpha) + \frac{\lambda_n\,|\mathbf{M}_n(\alpha)|\,\hat{\sigma}^2}{n}, \tag{2.3}$$

with respect to all α in $\mathcal{A}_n$, where $RSS_n(\alpha) = \left(\mathbf{y}_n - \hat{\boldsymbol{\mu}}_n(\alpha)\right)'\left(\mathbf{y}_n - \hat{\boldsymbol{\mu}}_n(\alpha)\right)$ is the residual sum of squares, $|\mathbf{M}_n(\alpha)|$ is the rank of the matrix $\mathbf{M}_n(\alpha)$, $\hat{\sigma}^2$ is a consistent estimator of the error variance using the full model and λ_n controls the degree of penalization. The term λ_n has a major role in the consistency of the selection where we say that a model selection procedure is to be consistent if

$$\hat{\alpha}_n \xrightarrow{P} \alpha_n^c, \tag{2.4}$$

where $\hat{\alpha}_n$ is the index set of the model chosen by the procedure and $\alpha_n^c \in \mathcal{A}_n$ is a smallest correct model.

In this chapter we discuss the role of λ_n in model selection. The literature on model selection is very extensive. Shao (1997), Eubank (1998), Fan and Li (2001), Shi and Tsai (2002) and the citations therein give a good account on the available methods and properties of these methods. Some of these authors discuss the conditions on λ_n for consistent model selection. In particular, Shao (1997) shows that the selection is consistent if $\lambda_n \to \infty$ and if the error distribution has at least eight finite moments. Zheng and Loh (1997) discuss model selection for errors with finite variance under the assumptions that $p_n/n \to 0$ and the penalty term diverges faster than p_n where p_n is the size of the full model. Shi and Tsai (2002) establish that λ_n should grow at a rate of $\log(n)$ under the condition that the error distribution has at least four finite moments and other restrictions on the model (see discussion after Theorem 4 below).

In this chapter, we establish conditions on the penalty term λ_n in (2.3) in order to satisfy (2.4) with less restrictions. In particular, for cases with finite error variance, we determine the rate of divergence of the penalty term for consistent model selection with less restrictions (compared with those in Zheng and Loh (1997)) on the model size p_n and for a larger class of candidate models. We also show that λ_n can diverge to ∞ at any rate if the error distribution has a finite fourth moment, a more flexible condition than that of Shao (1997) and Shi and Tsai (2002). Fan and Li (2001) deal with the model selection using a penalized likelihood where the distribution of the responses is assumed to have regularity properties (Lehmann and Caeslla, 1998). The penalty term in their approach is in the form of a sum of suitable functions. We show that their results agree with those in this chapter for normal log-likelihood and an appropriate class of penalty functions. In the process we show that for the results in their paper to hold, they actually need a slightly stronger differentiability condition than what was assumed.

In the remainder of this chapter, we shall use $\hat{\alpha}_n$ to denote the model selected by minimizing (2.3). We will discuss the consistency of $\hat{\alpha}_n$ in the next section.

2.2 Main Results

We first list several assumptions that are used in the sequel. Most of these have been used in Shao (1997) among others. The error distribution is assumed to be unknown with zero mean and finite variance.

Assumption 2.1. *The smallest correct model denoted by α_n^c exists and for every model α in $\mathcal{AC}_n$, we have $\alpha_n^c \subset \alpha$. Let $r_n = |\alpha_n^c|$*

Assumption 2.2. $\sup_n r_n < \infty$.

For a set of models $\mathcal{A}$, let $\mathcal{A}^{(k)} = \{\alpha | \alpha \in \mathcal{A}, |\alpha| \leq k\}$, and $\mathcal{A}_n^{[k]} = \mathcal{A}^{(k)} \setminus \mathcal{A}^{(k-1)}$ and $\alpha_{\mathcal{A}} = \cup_{\alpha \in \mathcal{A}} \alpha$.

Assumption 2.3. *There exists a sequence of positive numbers $\{d_1, d_2, \ldots\}$ such that $\left|\alpha_{\mathcal{A}_n^{(k)}}\right| \leq d_n k$ for all k and n.*

Note that this is a weaker condition than the hierarchical model assumption in Zheng and Loh (1997) allowing $\mathcal{A}_n$ to be created without any preordering of covariates. Now we give a series of Lemmas and Theorems establishing the consistency of model selection under various conditions on λ_n and the error distribution. The proofs of these statements are deferred to an Appendix.

Lemma 2.1. *Suppose Assumptions 2.1 and 2.2 hold and*

$$\lim_{n\to\infty} \frac{\sum_{1\leq k} \left|\alpha_{\mathcal{A}_n^{(k)}} \setminus \alpha_{\mathcal{A}_n^{(k-1)}}\right| /k}{\lambda_n} = 0. \tag{2.5}$$

Let $\mathcal{D}_n = \mathcal{A}_n \setminus \mathcal{A}_n^{(r_n)}$. *Then* $P(\hat{\alpha}_n \in \mathcal{D}_n) \to 0$.

The term $\sum_{1\leq k} \left|\alpha_{\mathcal{A}_n^{(k)}} \setminus \alpha_{\mathcal{A}_n^{(k-1)}}\right| /k$ in condition (2.5) determines how fast the number of variables included in the models grows with k and how much the models overlap. If Assumption 2.3 is satisfied and the penalty goes to infinity faster than $d_n \log(p_n)$, then, it can be shown that (2.5) is satisfied. Then, we have the following lemma.

Lemma 2.2. *Under Assumptions 2.1, 2.2, and 2.3, and under the condition that*

$$d_n \log p_n/\lambda_n \to 0$$

we have

$$P\left(\hat{\alpha}_n \in \mathcal{D}_n\right) \to 0.$$

For a collection of models $\mathcal{A}$, let $\overline{\mathcal{A}}$ denote the set of all maximum elements in $\mathcal{A}$ (an element α in $\mathcal{A}$ is said to be maximum if for any $\beta \in \mathcal{A}$ such that $\alpha \subset \beta$, we have $\beta = \alpha$). Let $\Delta_n(\alpha) = \boldsymbol{\mu}_n' \left(\mathbf{I} - \mathbf{M}_n(\alpha)\right) \boldsymbol{\mu}_n/n$. Then we can prove

Lemma 2.3. *Suppose that Assumption 2.1 holds. Let* $\mathcal{B}_n \subset \mathcal{AI}_n, n = 1, \ldots$. *If*

$$\lim_n \min_{\alpha\in\mathcal{B}_n} n\Delta(\alpha)/\lambda_n = \infty, \tag{2.6}$$

and

$$\sum_{\alpha\in\overline{\mathcal{B}_n}} |\alpha|/n\Delta_n(\alpha) \to 0, \tag{2.7}$$

then, $P\left(\hat{\alpha}_n \in \mathcal{B}_n\right) \to 0$.

If Assumptions 2.2 and 2.3 hold, and the $\mathcal{B}_n$ in Lemma 2.3 is taken to be $\mathcal{AI}_n^{(r_n)}$, then, $\left|\alpha_{\mathcal{AI}_n^{(r_n)}}\right| \le d_n r_n$. Thus, $\max_{\alpha \in \overline{\mathcal{AI}_n^{(r_n)}}} |\alpha| \le r_n$ and $\left|\overline{\mathcal{AI}_n^{(r_n)}}\right| = O\left(d_n^{r_n}\right)$. Since r_n is bounded we see that (2.7) with $\mathcal{B}_n = \mathcal{AI}_n^{(r_n)}$ is satisfied if

$$\min_{\alpha \in \mathcal{AI}_n^{(r_n)}} n\Delta_n(\alpha)/d_n^{r_n} \to \infty. \tag{2.8}$$

If d_n is of order $\log n$ or n^α for some $\alpha < \inf_n 1/r_n$, then (2.8) is a weaker constraint than (2.5) in Shao (1997) With this observation and Lemma 2.2 we have

Theorem 2.4. *Suppose Assumptions 2.1, 2.2, and 2.3 hold, and* (2.8) *is satisfied. Also suppose that*

$$\min_{\alpha \in \mathcal{AI}_n^{(r_n)}} n\Delta_n(\alpha)/\lambda_n \to \infty, \tag{2.9}$$

and

$$d_n \log p_n/\lambda_n \to 0 \tag{2.10}$$

Then, $P(\hat{\alpha}_n = \alpha_n^c) \to 1.$

Remark 2.5. Zheng and Loh (1997) use a penalty term in the form $h_n(k)$, where k is the size of the model in a hierarchical model selection with random covariates. Their $h_n(k)$ plays the same role as $\lambda_n k$ in this chapter. If we allow a general form $\lambda_n(k)$, where $\lambda_n(k)$ is increasing for every n, by using the same techniques in the proofs of Lemmas 2.1, 2.2, 2.3 and Theorem 2.4, we can show the following.

Corollary 2.6. *Suppose Assumptions 2.1, 2.2 are met and Assumption 2.3 is satisfied with a bounded* d_n. *Furthermore, suppose*

$$\liminf_n \min_{\alpha \in \mathcal{AI}_n^{(r_n)}} n\Delta_n(\alpha)/\lambda_n(r_n) = \infty, \tag{2.11}$$

and

$$\sum_{r_n+1 \le k \le p_n} 1/(\lambda_n(k) - \lambda_n(r_n)) \to 0. \tag{2.12}$$

Then $P\left(\hat{\alpha}_n = \alpha_n^c\right) \to 1.$

Note that (2.12) is a weaker condition on the penalty terms than the combination of conditions B2 and B3 in Zheng and Loh (1997). Also, $p_n/n \to 0$ is not required.

Remark 2.7. Comparing Theorem 2.4 with Theorem 1 of Shi and Tsai (2002) (ST hereafter), we find that the assumptions here are weaker. The Assumption 1 of ST requires the errors to have at least four finite moments, while we only require finite variance. We do not need a constraint like the Assumption 2 of ST . Also, we only need $\min_{\alpha \in \mathcal{AI}_n^{(r_n)}} n\Delta_n(\alpha)/\lambda_n \to \infty$. Since λ_n is allowed to go to infinity faster than $\log n$ for a reasonably chosen λ_n, this is a weaker condition on $\Delta_n(\alpha)$ than that of Assumption 3 in ST, which required $\min_{\alpha \in \mathcal{AI}_n^{(r_n)}} n\Delta_n(\alpha)/n \to \infty$. If the size of the models, p_n, is bounded, then it is easy to check that (2.5) is satisfied, and then, by Lemma 1 above, consistency is achieved as long as $\lambda_n \to \infty$, a more general result than Theorem 1 of ST. Even with stronger assumptions they only showed that the selection avoids over estimation in the weak sense. However, these stronger assumptions were used to show that the procedure will not under-estimate in the strong sense.

Example 2.8. Suppose that for each n, the p_n explanatory variables are arranged in decreasing order of importance, and the true model is $\alpha_0 = \{1, \dots, p\}$ for some positive integer p. We are interested in models that have the form $\alpha_i = \{1, \dots, i\}$ for $i \le p_n$. Let $\mathcal{A}_n = \{\alpha_1, \dots, \alpha_{p_n}\}, n = 1, 2, \dots$. Then $\mathcal{AC}_n = \{\alpha_1, \dots, \alpha_{p_n}\}$ for $p_n \ge p$, $r_n = p$, and $\mathcal{AI}_n = \{\alpha_1, \dots, \alpha_{p-1}\}$. It is easy to see that Assumptions 2.1, 2.2 are satisfied and 2.3 is satisfied with $d_n = 1, n = 1, 2, \dots$. So, if conditions (2.9) and (2.10) are satisfied, then $\hat{\alpha}_n$ is consistent in the sense of (2.4). It should be noted that Saho's technique gives a similar result if we assume that e_1 has finite eighth moment.

Example 2.9. Suppose that the explanatory variables are indexed by two integers, which means that the set of variables can be represented as $\{x_1, \dots, x_{p_n}\} = \{z_{1,1}, \dots, z_{1,s_n}, \dots, z_{t_n,1}, \dots, z_{t_n,s_n}\}$. This is the case when we need not only to select the right number of variables, but also the right "order" for each variable. In this setting, again we assume that the variables are arranged in decreasing order of importance, and if we include a certain order of a variable in a model, we also include all the lower orders of that variable. Thus, a model in this case has the form

$$\alpha_{j:i_1,\dots,i_j} = \{z_{1,1}, \dots, z_{1,i_1}, \dots, z_{j,1}, \dots, z_{j,i_j}\},$$

and

$$\mathcal{A}_n = \{a_{j:i_1,\dots,i_j} : j \le t_n \text{ and } i_k \le s_n \text{ for } k = 1, ..., t_n\}$$

It can be shown that

$$\left|\alpha_{\mathcal{A}_n^{(k)}}\right| \leq \begin{cases} k(k+1), \text{ for } k \leq t_n; \\ \\ t_n(t_n+1) + (k-t_n)t_n \text{ for } k > t_n. \end{cases}$$

and

$$\left|\alpha_{\mathcal{A}_n^{(k)}} \setminus \alpha_{\mathcal{A}_n^{(k-1)}}\right| \leq \begin{cases} k, \text{ for } k \leq t_n; \\ \\ t_n, \text{ for } t_n < k \leq t_n + s_n. \end{cases}$$

Hence,

$$\begin{aligned} \sum_{k=1}^{p_n} \left|\alpha_{\mathcal{A}_n^{(k)}} \setminus \alpha_{\mathcal{A}_n^{(k-1)}}\right| / k &\leq t_n + t_n \sum_{k=t_n+1}^{t_n+s_n} 1/k \\ &\leq t_n + t_n \log(t_n + s_n). \end{aligned}$$

In particular, if we take s_n to be of order $\log n$ and t_n to be of order $\log n / \log\log n$, then (2.5) is satisfied if we have $\lambda_n / \log n \to \infty$. It the true model α_0 is finite, that is $|\alpha_0| < \infty$, it is clear that Assumptions 2.1 and 2.2 are satisfied. If in addition (2.6) is true, then it can be shown that (2.7) with $\mathcal{B}_n = \mathcal{AI}_n^{(r_n)}$ is also true. And thus, by Lemma 2.1, Lemma 2.3 and the subsequent discussion, consistency in the sense of (2.4) is achieved.

Note that for the above results it is only necessary for the errors to have finite variance. Now suppose that we know that the errors also have finite fourth moment. We will show that this information will enable us to choose a λ_n diverging to infinity at any rate for consistent model selection.

Assumption 2.4. *The iid errors* $e_1, \ldots, e_n$ *in* (2.1) *satisfy* $E(e_1^4) = \tau < \infty$.

Lemma 2.10. *Suppose Assumptions 2.1, 2.2, 2.4 are satisfied. Also suppose that there exist constants* a, b, c *and a strictly increasing sequence of nonnegative numbers* $u_{n,1} = 0, u_{n,2}, \ldots$ *for every* n *such that*

$$\sum_{u_{n,k} < i < u_{n,k+1}} \frac{\left|\alpha_{\mathcal{A}_n^{(i)}} \setminus \alpha_{\mathcal{A}_n^{(i-1)}}\right|}{i} \leq a, \text{ for } k = 1, \ldots, n = 1, \ldots, \tag{2.13}$$

$$\left|\alpha_{\mathcal{A}_n^{(u_{n,k})}}\right|/u_{n,k} \le b, \; for \; k = 2, \ldots, \tag{2.14}$$

and

$$\sum_{1<k} 1/u_{n,k} \le c \; for \; n = 1, \ldots, . \tag{2.15}$$

Then, $P(\hat{\alpha}_n \in \mathcal{D}_n) \to 0$ *if* $\lambda_n \to \infty$.

Now suppose that we are given Assumptions 2.2 and 2.3. Also suppose that the sequence $\{d_n\}$ in Assumption 2.3 is bounded by d. Let $u_{n,k} = 2^{k-1} - 1$ for $k = 1, 2, \ldots$. It is not hard to check that conditions (2.13), (2.14), and (2.15) of Lemma 2.10 are all satisfied if we set $a = 2d, b = d,$ and $c = 2$. Notice also that the discussion after Lemma 2.3 is valid here. Then, by Lemma 2.10 we have the following theorem.

Theorem 2.11. *Suppose that Assumptions 2.1, 2.2, 2.3, 2.4, and* (2.9) *hold. Moreover, let the sequence* $\{d_n\}$ *in Assumption 2.3 be bounded and* $\lambda_n \to \infty$. *Then,* $P(\hat{\alpha}_n = \alpha_n^c) \to 1$.

Now reconsider the model selection problem described in Example 2.8. If $E(e_1^4) < \infty$, then, by Theorem 2.11, it is only necessary for the penalty to go to infinity to get consistency.

Remark 2.12. In Fan and Li (2001) model selection using the penalized likelihood function $L(\boldsymbol{\beta}) - n\sum_{j=1}^{d} p_{\eta_n}(|\beta_j|)$, where p_{η_n} are suitable penalty functions, is discussed. We can relate their penalty term $\sum p_{\eta_n}(|\beta_j|)$ and the penalty term in this chapter if we assume that the errors are normal. Then the log likelihood function will be proportional to the $-RSS_n$, and it will be maximized by the least squares estimator. If in addition, the value of the function p_{η_n} is constant outside an interval $(-\delta_n, \delta_n)$, where $\delta_n \to 0$, then the $n\sum p_{\eta_n}(|\beta_j|)$ term is comparable to the λ_n in this chapter.

We note that for the argument in the proof of Lemma 1 in Fan and Li (2001) to hold, one needs a slightly stronger condition on $p_{\eta_n}(\theta)$ than was assumed. The following example shows that the argument about the sign of the derivative of Q in the proof of Lemma 1 in Fan and Li (2001) fails under the assumption

$$\liminf_{n\to\infty} \liminf_{\theta\to 0^+} \eta_n^{-1} p'_{\eta_n}(\theta) > 0.$$

Let $\eta_n = u_n n^{-1/2}$, where $u_n \to \infty$ and

$$p_{\eta_n}(\theta) = \begin{cases} \sin(n^{2/3}\theta), & \text{for } |\theta| \le \pi/2n^{2/3}; \\ 1, & \text{otherwise .} \end{cases}$$

We see that $\liminf_{n\to\infty} \liminf_{\theta\to 0^+} \eta_n^{-1} p'_{\eta_n}(\theta) > 0, n^{-1/2}/\eta_n \to 0$, and $\max\{|p''_{\eta_n}(|\beta_{j0}|) : \beta_{j0} \neq 0\} \to 0$. However, for $\beta_j = cn^{-1/2}$, we have $\eta_n^{-1} p'_{\eta_n}(|\beta_j|) \to 0$ as $n \to \infty$. Thus a condition like

$$\liminf_{n\to\infty} \inf_{0<\theta<cn^{-1/2}} \eta_n^{-1} p'_{\eta_n}(\theta) > 0$$

for some $c > 0$ has to be used. This requires $p'_{\eta_n}(\theta)$ to be at least of the same order as η_n in the interval $(0, cn^{-1/2})$, and this will force the value of $p_{\eta_n}(cn^{-1/2})$ to be of order $\eta_n n^{-1/2} = u_n/n$. This means that $np_{\eta_n}(cn^{-1/2}) \to \infty$, which coincides with the result in this chapter, that $\lambda_n \to \infty$.

2.3 Proofs

In these proofs, for two matrices A and B, $A \ge B$ means that $A - B$ is positive semi-definite.

Proof of Lemma 2.1

Note that $\hat{\alpha}_n \in \mathcal{D}_n$ implies that $\min_{\alpha\in\mathcal{D}_n} T_{n,\lambda_n}(\alpha) - T_{n,\lambda_n}(\alpha_n^c) < 0$. Let $\mathbf{A}_n(\alpha) = \mathbf{I} - \mathbf{M}_n(\alpha)$. For $\alpha \in \mathcal{D}_n^{[k]}, r_n < k \le p_n$, we have

$$\begin{aligned}
& T_{n,\lambda_n}(\alpha) - T_{n,\lambda_n}(\alpha_n^c) \\
&= \frac{1}{n}\left(\mathbf{y}_n'\left(\mathbf{M}_n\left(\alpha_n^c\right) - \mathbf{M}_n\left(\alpha\right)\right)\mathbf{y}_n - \lambda_n\left(|\alpha_n^c| - |\alpha|\right)\hat{\sigma}^2\right) \\
&\ge \frac{1}{n}\left(\mathbf{y}_n'\left(\mathbf{M}_n\left(\alpha_n^c\right) - \mathbf{M}_n\left(\alpha_{\mathcal{A}_n^{(k)}}\right)\right)\mathbf{y}_n + \lambda_n\left(k - r_n\right)\hat{\sigma}^2\right) \\
&= \Delta_n\left(\alpha_{\mathcal{A}_n^{(k)}}\right) + \frac{2}{n}\boldsymbol{\mu}_n'\mathbf{A}_n\left(\alpha_{\mathcal{A}_n^{(k)}}\right)\mathbf{e} \\
&+ \frac{1}{n}\mathbf{e}'\left(\mathbf{M}_n\left(\alpha_n^c\right) - \mathbf{M}_n(\alpha_{\mathcal{A}_n^{(k)}})\right)\mathbf{e} + \lambda_n\left(k - r_n\right)\hat{\sigma}^2/n \\
&= \frac{1}{n}\mathbf{e}'\left(\mathbf{M}_n\left(\alpha_n^c\right) - \mathbf{M}_n(\alpha_{\mathcal{A}_n^{(k)}})\right)\mathbf{e} + \lambda_n\left(k - r_n\right)\hat{\sigma}^2/n \\
&\ge -\frac{1}{n}\mathbf{e}'\mathbf{M}_n\left(\alpha_{\mathcal{A}_n^{(k)}}\right)\mathbf{e} + \lambda_n\left(k - r_n\right)\hat{\sigma}^2/n
\end{aligned} \tag{2.16}$$

Thus

$$\min_{\alpha\in\mathcal{D}_n} T_{n,\lambda_n}(\alpha) < T_{n,\lambda_n}(\alpha_n^c) \Rightarrow$$

$$\max_{r_n<k\le p_n} \frac{\mathbf{M}_n\left(\alpha_{\mathcal{A}_n^{(k)}}\right)}{k-r_n} > \lambda_n\hat{\sigma}^2 \tag{2.17}$$

Define $\alpha_{\mathcal{A}_n^{(0)}}$ to be the empty set. Set $\mathbf{M}\left(\alpha_{\mathcal{A}_n^{(0)}}\right) = \mathbf{0}$ and

$$\mathscr{M}_n = \sum_{1\le k\le p_n} \left(\mathbf{M}_n\left(\alpha_{\mathcal{A}_n^{(k)}}\right) - \mathbf{M}_n\left(\alpha_{\mathcal{A}_n^{(k-1)}}\right)\right)/k.$$

For any integer $r_n < j \le p_n$ we have

$$\begin{aligned}(r_n+1)\mathscr{M}_n &\ge (r_n+1)\sum_{r_n<k\le j}\left(\mathbf{M}_n\left(\alpha_{\mathcal{A}_n^{(k)}}\right) - \mathbf{M}_n\left(\alpha_{\mathcal{A}_n^{(k-1)}}\right)\right)/k\\ &\ge (r_n+1)\sum_{r_n<k\le j}\left(\mathbf{M}_n\left(\alpha_{\mathcal{A}_n^{(k)}}\right) - \mathbf{M}_n\left(\alpha_{\mathcal{A}_n^{(k-1)}}\right)\right)/j\end{aligned}$$

$$\begin{aligned}&= (r_n+1)\mathbf{M}_n\left(\alpha_{\mathcal{A}_n^{(j)}}\right)/j\\ &\ge \mathbf{M}_n\left(\alpha_{\mathcal{A}_n^{(j)}}\right)/(j-r_n).\end{aligned}$$

Thus,

$$(r_n+1)\mathbf{e}'\mathscr{M}_n\mathbf{e} \ge \max_{r_n<k\le p_n} \mathbf{e}'\mathbf{M}_n\left(\alpha_{\mathcal{A}_n^{(k)}}\right)\mathbf{e}/(k-r_n). \tag{2.18}$$

Since

$$\begin{aligned}E\left[\mathbf{e}'\mathscr{M}_n\mathbf{e}\right] &= \sigma^2\sum_{1\le k\le p_n}\frac{tr\left(\mathbf{M}_n\left(\alpha_{\mathcal{A}_n^{(k)}}\right) - \mathbf{M}_n\left(\alpha_{\mathcal{A}_n^{(k-1)}}\right)\right)}{k}\\ &= \sigma^2\sum_{1\le k\le p_n}\frac{\left|\alpha_{\mathcal{A}_n^{(k)}}\setminus\alpha_{\mathcal{A}_n^{(k-1)}}\right|}{k},\end{aligned}$$

we have

$$P\left((r_n+1)\mathbf{e}'\mathscr{M}_n\mathbf{e} > \lambda_n\sigma^2/2\right) \le \frac{2(r_n+1)\sum_{1\le k\le p_n}\left|\alpha_n^{(k)} \setminus \alpha_n^{(k-1)}\right|/k}{\lambda_n}$$
$$\to 0,$$

and

$$P\left(\hat{\sigma}^2 < \sigma^2/2\right) \to 0.$$

Thus,

$$\begin{aligned} P\left((r_n+1)\mathbf{e}'\mathscr{M}_n\mathbf{e} > \lambda_n\hat{\sigma}^2\right) &\le P\left((r_n+1)\mathbf{e}'\mathscr{M}_n\mathbf{e} > \lambda_n\sigma^2/2\right) \\ &\quad + P\left(\hat{\sigma}^2 < \sigma^2/2\right) \\ &\to 0. \end{aligned}$$

This, combined with (2.18) and gives

$$\begin{aligned} P\left(\hat{\alpha}_n \in \mathcal{AC}_n \setminus \mathcal{AC}_n^{(r_n)}\right) &\le P\left(\min_{\alpha\in\mathcal{D}_n} T_{n,\lambda_n}(\alpha) \le T_{n,\lambda_n}(\alpha_n^c)\right) \\ &\le P\left((r_n+1)\mathbf{e}'\mathscr{M}_n\mathbf{e} \ge \lambda_n\hat{\sigma}^2\right) \\ &\to 0, \end{aligned}$$

giving the required result. □

Proof of Lemma 2.2

For simplicity we write $\left|\alpha_{\mathcal{A}_n^{(i)}} \setminus \alpha_{\mathcal{A}_n^{(i-1)}}\right|$ as a_i for $i = 1, 2, \ldots$. It is easy to see that $a_i = 0$ for $i > p_n$. First we will prove that

$$\sum_{i=1}^{p_n} a_i/i \le d_n \sum_{i=1}^{p_n} 1/i$$

In fact, by Assumption 2.3 we have

$$
\begin{aligned}
\sum_{i=1}^{p_n} a_i/i = a_1\left(1-\frac{1}{2}\right) + (a_1+a_2)\left(\frac{1}{2}-\frac{1}{3}\right) + \ldots \\
&+ (a_1+\ldots+a_{p_n-1})\left(\frac{1}{p_n-1}-\frac{1}{p_n}\right) + (a_1+\ldots+a_{p_n})\left(\frac{1}{p_n}\right) \\
&= \sum_{j=1}^{p_n-1}\left|\alpha_{\mathcal{A}_n^{(j)}}\right|\left(\frac{1}{j}-\frac{1}{j+1}\right) + \left|\alpha_{\mathcal{A}_n^{(p_n)}}\right|\frac{1}{p_n} \\
&\le \sum_{j=1}^{p_n-1} d_n j\left(\frac{1}{j}-\frac{1}{j+1}\right) + d_n p_n \frac{1}{p_n} \\
&= d_n \sum_{i=1}^{p_n} 1/i.
\end{aligned}
$$

Since $\sum_{i=1}^{p_n} 1/i = O(\log p_n)$. Thus (2.5) in Lemma 2.1 is satisfied. □

Proof of Lemma 2.3

We know that $\hat{\alpha}_n \in \mathcal{B}_n$ implies that

$$\min_{\alpha\in\mathcal{B}_n} T_{n,\lambda_n}(\alpha) \le T_{n,\lambda_n}(\alpha_n^c) \tag{2.19}$$

For $\alpha \in \mathcal{B}_n$, we can find a $\beta \in \overline{\mathcal{B}_n}$ such that $\alpha \subset \beta$. We have

$$
\begin{aligned}
& T_{n,\lambda_n}(\alpha) - T_{n,\lambda_n}(\alpha_n^c) \\
&= \frac{1}{n}\left(\mathbf{y}_n'\left(\mathbf{M}_n(\alpha_n^c) - \mathbf{M}_n(\alpha)\right)\mathbf{y}_n - \lambda_n\left(|\alpha_n^c| - |\alpha|\right)\hat{\sigma}^2\right) \\
&\ge \frac{1}{n}\left(\mathbf{y}_n'\left(\mathbf{M}_n(\alpha_n^c) - \mathbf{M}_n(\beta)\right)\mathbf{y}_n - \lambda_n r_n \hat{\sigma}^2\right) \\
&= \frac{1}{n}\boldsymbol{\mu}_n'\left(\mathbf{M}_n(\alpha_n^c) - \mathbf{M}_n(\beta)\right)\boldsymbol{\mu}_n + \frac{2}{n}\boldsymbol{\mu}_n'\left(\mathbf{M}_n(\alpha_n^c) - \mathbf{M}_n(\beta)\right)\mathbf{e} \\
&+ \frac{1}{n}\mathbf{e}'\left(\mathbf{M}_n(\alpha_n^c) - \mathbf{M}_n(\beta)\right)\mathbf{e} - \frac{1}{n}\lambda_n r_n \hat{\sigma}^2 \\
&\ge \Delta_n(\beta) + \frac{2}{n}\boldsymbol{\mu}_n'\mathbf{A}_n(\beta)\mathbf{e} - \lambda_n r_n \hat{\sigma}^2/n - \frac{1}{n}\mathbf{e}'\left(\mathbf{M}_n(\beta)\right)\mathbf{e}
\end{aligned}
$$

(2.19) implies that at least one of the three inequalities:

$$\max_{\beta\in\overline{\mathcal{B}_n}}\left|\frac{\boldsymbol{\mu}_n'\mathbf{A}_n(\beta)\mathbf{e}}{n\Delta_n(\beta)}\right| \ge 1/3$$

,

$$\max_{\beta\in\overline{\mathcal{B}_n}} \frac{\mathbf{e}'\left(\mathbf{M}_n(\beta)\right)\mathbf{e}}{n\Delta_n(\beta)} \geq 1/3$$

$$\max_{\beta\in\overline{\mathcal{B}_n}} \lambda_n r_n \hat{\sigma}^2/n\Delta_n(\beta) \geq 1/3$$

holds. Since $Var\left(\boldsymbol{\mu}_n'\mathbf{A}_n(\beta)\mathbf{e}\right) = n\sigma^2\Delta_n(\beta)$, we have for any positive δ

$$P\left(\max_{\beta\in\overline{\mathcal{B}_n}} \left|\frac{\boldsymbol{\mu}_n'\mathbf{A}_n(\beta)\mathbf{e}}{n\Delta_n(\beta)}\right| \geq \delta\right) \leq \sum_{\beta\in\overline{\mathcal{B}_n}} \frac{\sigma^2}{\delta^2 n\Delta_n(\beta)} \to 0. \tag{2.20}$$

Also, $E[\mathbf{e}'(\mathbf{M}_n(\beta))\mathbf{e}] = \sigma^2|\beta|$, so that

$$P\left(\max_{\beta\in\overline{\mathcal{B}_n}} \frac{\mathbf{e}'(\mathbf{M}_n(\beta))\mathbf{e}}{n\Delta_n(\beta)} \geq \delta\right) \leq \sum_{\beta\in\overline{\mathcal{B}_n}} \frac{\sigma^2|\beta|}{\delta n\Delta_n(\beta)} \to 0. \tag{2.21}$$

By condition (2.6), we also have $\lambda_n r_n\hat{\sigma}^2/n\Delta_n(\beta) \to 0$. Thus $P(\min_{\alpha\in\overline{\mathcal{B}_n}} T_{n,\lambda_n}(\alpha) > T_{n,\lambda_n}(\alpha_n^c)) \to 1$. □

Proof of Lemma 2.10

For $k = 1, 2, \ldots$, define

$$\mathbf{N}_k = \frac{\mathbf{M}_n\left(\alpha_{\mathcal{A}_n^{(u_{n,k})}}\right)}{u_{n,k}} + \sum_{i=u_{n,k}+1}^{u_{n,k+1}-1} \frac{\mathbf{M}_n\left(\alpha_{\mathcal{A}_n^{(i)}}\right) - \mathbf{M}_n\left(\alpha_{\mathcal{A}_n^{(i-1)}}\right)}{i}$$

Define 0/0=0. We have

$$\begin{aligned} E(\mathbf{e}'\mathbf{N}_k\mathbf{e}) &= \sigma^2\left(\frac{\left|\alpha_{\mathcal{A}_n^{(u_{n,k})}}\right|}{u_{n,k}} + \sum_{u_{n,k}<i<u_{n,k+1}} \frac{\left|\alpha_{\mathcal{A}_n^{(i)}} \setminus \alpha_{\mathcal{A}_n^{(i-1)}}\right|}{i}\right) \\ &\leq \sigma^2(a+b). \end{aligned}$$

For $\delta > 2\sigma^2(a+b)$, we have

$$\begin{aligned} P(\mathbf{e}'\mathbf{N}_k\mathbf{e} > \delta) &\leq P\left(\left|\mathbf{e}'\mathbf{N}_k\mathbf{e} - E\left(\mathbf{e}'\mathbf{N}_k\mathbf{e}\right)\right| > \delta/2\right) \\ &= P\left(\left|\mathbf{e}'\mathbf{N}_k\mathbf{e} - E\left(\mathbf{e}'\mathbf{N}_k\mathbf{e}\right)\right|^2 > \delta^2/4\right) \\ &\leq \frac{E\left(\left|\mathbf{e}'\mathbf{N}_k\mathbf{e} - E\left(\mathbf{e}'\mathbf{N}_k\mathbf{e}\right)\right|^2\right)}{\delta^2/4}. \end{aligned}$$

By Theorem 2 of Whittle (1960), we have

$$E\left(\left|\mathbf{e}'\mathbf{N}_k\mathbf{e} - E\left(\mathbf{e}'\mathbf{N}_k\mathbf{e}\right)\right|^2\right) \leq C\left(tr\left(\mathbf{N}_k'\mathbf{N}_k\right)\tau\right).$$

It is easy to see that

$$\mathbf{N}_k'\mathbf{N}_k = \frac{\mathbf{M}_n\left(\alpha_{\mathcal{A}_n^{(u_{n,k})}}\right)}{u_{n,k}^2} + \sum_{i=u_{n,k}+1}^{u_{n,k+1}-1} \frac{\mathbf{M}_n\left(\alpha_{\mathcal{A}_n^{(i)}}\right) - \mathbf{M}_n\left(\alpha_{\mathcal{A}_n^{(i-1)}}\right)}{i^2}.$$

Thus,

$$\begin{aligned} tr\left(\mathbf{N}_k'\mathbf{N}_k\right) &= \frac{\left|\alpha_{\mathcal{A}_n^{(u_{n,k})}}\right|}{u_{n,k}^2} + \sum_{i=u_{n,k}+1}^{u_{n,k+1}-1} \frac{\left|\alpha_{\mathcal{A}_n^{(i)}} \setminus \alpha_{\mathcal{A}_n^{(i-1)}}\right|}{i^2} \\ &\leq \begin{cases} a, & \text{for } k=1; \\ \frac{a+b}{u_{n,k}}, & \text{for } 2 \leq k. \end{cases} \end{aligned}$$

Hence

$$P\left(\mathbf{e}'\mathbf{N}_k\mathbf{e} > \delta\right) \leq \begin{cases} \frac{aC\tau}{\delta^2}, & \text{for } k=1; \\ \frac{(a+b)C\tau}{\delta^2 u_{n,k}}, & \text{for } k>1, \end{cases}$$

for some constant C. Notice that for any $u_{n,k} \leq i < u_{n,k+1}$ we have

$$\begin{aligned} \mathbf{N}_k &\geq \frac{\mathbf{M}_n\left(\alpha_{\mathcal{A}_n^{(u_{n,k})}}\right)}{u_k} + \ldots + \frac{\mathbf{M}_n\left(\alpha_{\mathcal{A}_n^{(i)}}\right) - \mathbf{M}_n\left(\alpha_{\mathcal{A}_n^{(i-1)}}\right)}{i} \\ &\geq \frac{\mathbf{M}_n\left(\alpha_{\mathcal{A}_n^{(u_{n,k})}}\right)}{i} + \ldots + \frac{\mathbf{M}_n\left(\alpha_{\mathcal{A}_n^{(i)}}\right) - \mathbf{M}_n\left(\alpha_{\mathcal{A}_n^{(i-1)}}\right)}{i} \\ &= \frac{\mathbf{M}_n\left(\alpha_{\mathcal{A}_n^{(i)}}\right)}{i} \geq \frac{\mathbf{M}_n(\alpha)}{i} \text{ for any } \alpha \in \mathcal{A}_n^{[i]}. \end{aligned}$$

For $i > r_n$, suppose that $u_{n,k} \leq i < u_{n,k+1}$. Then,

$$(r_n+1)\mathbf{e}'\mathbf{N}_k\mathbf{e} \geq \max_{\alpha \in \mathcal{A}_n^{[i]}} \frac{(r_n+1)\mathbf{e}'\mathbf{M}_n(\alpha)\mathbf{e}}{i} \geq \max_{\alpha \in \mathcal{A}_n^{[i]}} \frac{\mathbf{e}'\mathbf{M}_n(\alpha)\mathbf{e}}{i - r_n}.$$

¿From this we get

$$\max_{r_n < i \leq p_n} \max_{\alpha \in \mathcal{A}_n^{[i]}} \frac{\mathbf{e}'\mathbf{M}_n(\alpha)\mathbf{e}}{i - r_n} \leq (r_n+1) \max_{1 \leq k} \mathbf{e}'\mathbf{N}_k\mathbf{e}.$$

Noting $\mathcal{D}_n^{[k]} = \mathcal{A}_n^{[k]}$ for $r_n < k \le p_n$ we have

$$\max_{r_n<k\le p_n} \max_{\alpha\in\mathcal{D}_n^{[k]}} \frac{\mathbf{e}'\mathbf{M}_n(\alpha)\mathbf{e}}{k-r_n} \le (r_n+1)\max_{1\le k}\mathbf{e}'\mathbf{N}_k\mathbf{e}.$$

Thus, when λ_n is large enough, (2.17) gives

$$\begin{aligned}
P(\hat{\alpha}_n \in \mathcal{D}_n) &\le P\left((r_n+1)\max_{1\le k}\mathbf{e}'\mathbf{N}_k\mathbf{e} > \lambda_n\hat{\sigma}^2\right) \\
&\le P\left((r_n+1)\max_{1\le k}\mathbf{e}'\mathbf{N}_k\mathbf{e} > \lambda_n\sigma^2/2\right) + P\left(\hat{\sigma}^2 < \sigma^2/2\right) \\
&\le \sum_{1\le k} P\left((r_n+1)\mathbf{e}'\mathbf{N}_k\mathbf{e} > \lambda_n\sigma^2/2\right) + P\left(\hat{\sigma}^2 < \sigma^2/2\right) \\
&\le \frac{4(r_n+1)^2C\tau}{\lambda_n^2\sigma^4}\left(a + \sum_{2\le k}\frac{a+b}{u_{n,k}}\right) + P\left(\hat{\sigma}^2 < \sigma^2/2\right) \\
&\le \frac{(r_n+1)^2C(a+(a+b)c)\tau}{\lambda_n^2\sigma^4} + P\left(\hat{\sigma}^2 < \sigma^2/2\right) \\
&\to 0. \quad \square
\end{aligned}$$

Chapter 3

Pointwise Bayes Type Estimator of the Survival Probability with Censored Data

3.1 Introduction

The analysis of incomplete or censored data in medical or reliability studies has been investigated in great detail over the past several years. This fact is brought out by the vast number of articles in the recent literature. Of major importance in many of the studies is the estimation of the lifetime survival function, $S(.)$.

In this chapter we consider the random censorship model for estimating the survival function $S(t)$ at a given t. Let T be a random variable corresponding to the lifetime on study. Then, let $T_1, \ldots, T_n$ be *i.i.d.* positive lifetimes (realizations of T) from the distribution function F(t) and, independent of the T_i's, let $Y_1, \ldots, Y_n$ be *i.i.d.* positive censoring random variables from a distribution *G(t)*. Under the model of random censorship from the right, T_i is censored on the right by Y_i, so that we only observe the pairs $(X_i, \delta_i), i = 1, \ldots, n$, where X_i is the minimum of T_i and

Y_i, and δ_i is the indicator of the event $T_i \leq Y_i$, i.e. for i=1,...,n

$$X_i = min(T_i, Y_i), \;\; \delta_i = \begin{cases} 1 & \text{if } T_i \leq Y_i \\ 0 & \text{otherwise.} \end{cases} \tag{3.1}$$

Thus X_i (i=1,...,n) are i.i.d. random variables with distribution function H given by $1 - H(u) = (1 - F(u))(1 - G(u)), -\infty < u < \infty$. In estimating $S = 1 - F$, perhaps, the most widely used estimator is the Kaplan-Meier (K-M) estimator (Kaplan and Meier, 1958) defined as

$$\hat{S}_n(t) = \begin{cases} \prod_{j:Z_j \leq t} (\frac{n-j}{n-j+1})^{\delta_j^*} & \text{if } t < Z_n \\ 0 & \text{if } t \geq Z_n \end{cases} \tag{3.2}$$

where $Z_1 \leq \cdots \leq Z_n$ are the ordered values of the X_i's defined in (3.1) and $\delta_i^*, i = 1, \ldots, n$ are the corresponding δ's for each Z value.

The K-M estimator can sometimes be inefficient in estimating the tails of the life distribution and some improvements can be made by smoothing it. Properties of such estimators are discussed extensively in the literature, a partial list being, Blum and Susarla (1980), Foldes and Retjo (1981), Padgett and McNichols (1984), Burke and L. (1984), Uzunogullari and Wang (1992), Padgett (1986), Csorgo et al. (1991), Falk (1984) and Kulasekera et al. (2001).

Nonparametric Bayesian estimation of S was first discussed by Susarla and Van Ryzin (1976) where they estimate the function using an integrated squared error loss of type $\int [D(t) - F(t)]^2 dw(t)$ where $D(t)$ is an estimator for F. In this approach, they examine the estimation of the object F rather than the value of F at a given time point. However, their estimator is evaluated by minimizing the Bayes risk in the integrand at each t, and it is therefore free of the weight function w above. Regardless of the underlying cdf, their approach places a prior on the whole object (F) rather than a prior on $F(t)$ at a given time point of interest. In a follow-up article, Susarla and Van Ryzin (1978) discussed the asymptotic properties of their estimator. Other related work can be found in Phadia (1980), Tiwari (1988) and Kvam et al. (2000) and the references therein.

In this chapter we discuss the use of a Bayes-type argument in estimating the survival function S at a given time t. Our approach can accommodate different weights at different time points by allowing for a user to choose a prior depending on the importance of the particular t at which the estimation is done. It can be shown that the K-M estimator is a member of this family. Furthermore, we prove the asymptotic normality of the proposed estimator for a given t for all priors satisfying mild conditions. In the special case where there is no censoring, we analytically compare our estimator (which will be a Bayes estimator of a binomial parameter) to the nonparametric Bayes estimator of the cdf F for iid samples (Ferguson, 1973) showing that judicious choices of the prior in our method can outperform (in a Mean Squared Error sense) the classical nonparametric Bayes estimator over a large portion of the support of F. An empirical study shows that the performance of the new estimator can be superior to that of K-M estimator and the Bayes estimator by Susarla and Van Ryzin (1978) (SVR hereafter) over most of the support of the lifetime distribution. Not surprisingly, the superiority tends to decrease as the sample size increases.

Section 3.2 describes the poinwise Bayesian estimator and discusses the asymptotic properties. A small simulation study highlights the superior performance of the proposed approach over the K-M and SV estimators. The simulation results and the application of the proposed method to a real data set are presented in Section 3.3.

3.2 Estimation

We describe the development of the Bayesian estimator in this section and discuss large sample properties of such estimators. For $t > 0$, our goal is to estimate $\theta = S(t)$ based on right censored data described above in Section 3.1. Let $\beta = 1 - G(t)$ and let $\Delta_i = I[X_i > t]$, $i = 1, ..., n$. We assume that the functions F and G are right continuous and define $T_F = \inf\{t : F(t) = 1\}$ for the distribution function F. Also, let $p_1 = \int_t^\infty S(u)dG(u)$ and $p_2 = \int_0^t S(u)dG(u)$. Now, consider the data $(\Delta_i, \delta_i), i = 1, ..., n$. It is easy to show that the joint mass function of (Δ, δ) is

$$f(\delta, \Delta|\theta, \beta) = (\theta\beta - p_1)^{\delta\Delta}(1 - p_2 - \theta\beta)^{\delta(1-\Delta)}C_1^{(1-\delta)\Delta}C_2^{(1-\delta)(1-\Delta)}$$

for some constants $C_i, i = 1, 2$ that do not depend on θ and β. The constraints $\theta\beta > p_1$ and $1 - p_2 > \theta\beta$ will be satisfied whenever the probability of an observation being censored is less than 1, an obvious assumption in any lifetime study. Now suppose we impose a prior density g_n for θ while keeping β and $p_i, i = 1, 2$ fixed (i.e. degenerate priors for these parameters). Then, using iid data $(\Delta_i, \delta_i), i = 1, ..., n$, for a prior pdf of type

$$g_n(t) = \frac{\beta}{1 - p_1 - p_2} g\left(\frac{\beta t - p_1}{1 - p_1 - p_2}\right)$$

for some density g on $[0, 1]$, we can obtain the estimator $\hat{\theta}$ of θ as the posterior expectation $E[\theta|(\Delta_i, \delta_i), i = 1, ..., n]$. We have

$$\begin{aligned}
&E\left[\theta|\left(\Delta_i, \delta_i\right), i = 1, \ldots, n\right] \\
&= \int_0^1 \theta f\left(\theta|\Delta_1, \delta_1, \ldots, \Delta_n, \delta_n\right) d\theta \\
&= \int_0^1 \frac{\theta f\left(\theta, \Delta_1, \delta_1, \ldots, \Delta_n, \delta_n\right)}{f\left(\Delta_1, \delta_1, \ldots, \Delta_n, \delta_n\right)} d\theta \\
&= \frac{\int_{\frac{p_1}{\beta}}^{\frac{1-p_2}{\beta}} \theta\left(\theta\beta - p_1\right)^{\sum \delta_i \Delta_i}\left(1 - p_2 - \theta\beta\right)^{\sum \delta_i (1-\Delta_i)} g_n(\theta)\, d\theta}{\int_{\frac{p_1}{\beta}}^{\frac{1-p_2}{\beta}} \left(\theta\beta - p_1\right)^{\sum \delta_i \Delta_i}\left(1 - p_2 - \theta\beta\right)^{\sum \delta_i (1-\Delta_i)} g_n(\theta)\, d\theta}
\end{aligned}$$

This can be written as

$$E[\theta|(\Delta_i, \delta_i), i = 1, ..., n] = \frac{\int_{\frac{p_1}{\beta}}^{\frac{1-p_2}{\beta}} \theta\left(\theta\beta - p_1\right)^{\sum \delta_i \Delta_i}\left(1 - p_2 - \theta\beta\right)^{\sum \delta_i (1-\Delta_i)} g_n(\theta)\, d\theta}{\int_{\frac{p_1}{\beta}}^{\frac{1-p_2}{\beta}} \left(\theta\beta - p_1\right)^{\sum \delta_i \Delta_i}\left(1 - p_2 - \theta\beta\right)^{\sum \delta_i (1-\Delta_i)} g_n(\theta)\, d\theta}.$$

Since in practice, we do not know $p_i, i = 1, 2$ and β, we can use some suitable estimators of these parameters, say $\hat{p}_1, \hat{p}_2$ and $\hat{\beta}$ in the above expression to calculate the estimator $\hat{\theta}$. A possible estimator for β is $\hat{\beta} = 1 - \hat{G}(t)$ where $\hat{G}$ is the Kaplan-Meier estimator of G obtained by reversing the role of the censored and uncensored observations. Likewise, suitable estimators for p_1 and p_2 are

$$\hat{p}_1 = \int_t^\infty \hat{S}_n(u) d\hat{G}(u)$$

and

$$\hat{p}_2 = \int_0^t \hat{S}_n(u) d\hat{G}(u)$$

respectively, where $\hat{S}_n$ is the K-M estimator for S. Using these estimators and after simplification, we can write $\hat{\theta}$ as

$$\hat{\theta} = \frac{\displaystyle\int_0^1 \frac{u\left(1-\hat{p}_1-\hat{p}_2\right)+\hat{p}_1}{\hat{\beta}} u^{\sum \delta_i \Delta_i} \left(1-u\right)^{\sum \delta_i (1-\Delta_i)} g\left(u\right) du}{\displaystyle\int_0^1 u^{\sum \delta_i \Delta_i} \left(1-u\right)^{\sum \delta_i (1-\Delta_i)} g\left(u\right) du}. \tag{3.3}$$

The following theorem shows that the asymptotic properties of this estimator $\hat{\theta}$ are the same as those of Kaplan-Meier estimator. Here we assume that g is twice differentiable on $[0,1]$, g'' is bounded and that $g\left(\xi_0\right) > 0$ where

$$\xi_0 = P\left(T \leq Y, T > t\right) / P\left(T \leq Y\right).$$

Theorem 3.1. *Let $t < T_F$, and let F, G and g satisfy the above assumptions. Then $n^{1/2}\left(\hat{\theta} - S(t)\right)$ converges in distribution to a normal random variable with a zero mean and a variance*

$$\tau^2 = S^2(t) \int_0^{t-} \frac{[-dS(u)]}{S(u+)S^o(u)}$$

where $S^o(u)$ is the probability that an individual is alive and uncensored at time u.

The proof of this Theorem is given in the Appendix.

Remark 3.2. The implementation of the estimation procedure requires a specification of a prior distribution g with some parameters. In our numerical work we restricted g to a beta family with different indices. If g is the uniform distribution on $[0,1]$, then the resulting estimator is a slightly adjusted version of the K-M estimator. In particular, estimating $1-p_1-p_2$ with the proportion of uncensored observations $\sum \delta_i / n$, we get

$$\hat{\theta} = \frac{1}{\hat{\beta}} \left[\hat{p}_1 + \frac{\sum_{i=1}^n \delta_i \left(\sum_{i=1}^n \Delta_i \delta_i + 1\right)}{n\left(\sum_{i=1}^n \delta_i + 2\right)} \right].$$

Note that this estimator is asymptotically equivalent to

$$\hat{\theta}_1 = \frac{1}{\hat{\beta}} \left[\hat{p}_1 + \frac{\sum_{i=1}^n \Delta_i \delta_i}{n} \right].$$

Now if we use the same estimators above for β and p_1, we will can that $\hat{\theta}_1$ is actually the K-M estimator for $S(t)$. We have

$$\hat{S}_n(t) = \prod_{X_i \le t} \left(\frac{n - \sum_{j=1}^{n} I(X_j \le X_i)}{n - \sum_{j=1}^{n} I(X_j < X_i)} \right)^{\delta_i}$$

and

$$\hat{G}(t) = 1 - \prod_{X_i \le t} \left(\frac{n - \sum_{j=1}^{n} I(X_j \le X_i)}{n - \sum_{j=1}^{n} I(X_j < X_i)} \right)^{1-\delta_i}.$$

Thus,

$$\begin{aligned}
\hat{p}_1 &= \int_t^{\infty} \hat{S}_n(u) d\hat{G}(u) \\
&= \sum_{t < X_k < \infty} \prod_{X_i \le X_k} \left(\frac{n - \sum_{j=1}^{n} I(X_j \le X_i)}{n - \sum_{j=1}^{n} I(X_j < X_i)} \right)^{\delta_i} \left(\prod_{X_i < X_k} \left(\frac{n - \sum_{j=1}^{n} I(X_j \le X_i)}{n - \sum_{j=1}^{n} I(X_j < X_i)} \right)^{1-\delta_i} \right. \\
&\left. - \prod_{X_i \le X_k} \left(\frac{n - \sum_{j=1}^{n} I(X_j \le X_i)}{n - \sum_{j=1}^{n} I(X_j < X_i)} \right)^{1-\delta_i} \right)
\end{aligned}$$

$$\begin{aligned}
&= \sum_{t < X_k < \infty} \prod_{X_i \le X_k} \left(\frac{n - \sum_{j=1}^{n} I(X_j \le X_i)}{n - \sum_{j=1}^{n} I(X_j < X_i)} \right)^{\delta_i} \prod_{X_i < X_k} \left(\frac{n - \sum_{j=1}^{n} I(X_j \le X_i)}{n - \sum_{j=1}^{n} I(X_j < X_i)} \right)^{1-\delta_i} \\
&\left(1 - \left(\frac{n - \sum_{j=1}^{n} I(X_j \le X_k)}{n - \sum_{j=1}^{n} I(X_j < X_k)} \right)^{1-\delta_k} \right) \\
&= \sum_{t < X_k < \infty} \left(\frac{n - \sum_{j=1}^{n} I(X_j < X_k)}{n} \right) \left(1 - \left(\frac{n - \sum_{j=1}^{n} I(X_j \le X_k)}{n - \sum_{j=1}^{n} I(X_j < X_k)} \right)^{1-\delta_k} \right) \\
&= \frac{1}{n} \sum_{t < X_k < \infty} (1 - \delta_k) = \frac{1}{n} \sum (1 - \delta_k) \Delta_k.
\end{aligned}$$

Then

$$\begin{aligned}
\hat{\theta}_1 &= \frac{1}{\hat{\beta}} \left[\hat{p}_1 + \frac{\sum_{i=1}^{n} \Delta_i \delta_i}{n} \right] = \frac{1}{\hat{\beta}} \left[\frac{\sum (1 - \delta_k) \Delta_k}{n} + \frac{\sum_{i=1}^{n} \Delta_i \delta_i}{n} \right] \\
&= \frac{\sum \Delta_k}{n \hat{\beta}}
\end{aligned}$$

Let $\tilde{\theta}$ be the K-M estimator for θ. We have

$$\tilde{\theta} = \prod_{X_i \le t} \left(\frac{n - \sum_{j=1}^{n} I\left(X_j \le X_i\right)}{n - \sum_{j=1}^{n} I\left(X_j < X_i\right)} \right)^{\delta_i}.$$

Thus,

$$\begin{aligned} \tilde{\theta}\hat{\beta} &= \prod_{X_i \le t} \left(\frac{n - \sum_{j=1}^{n} I\left(X_j \le X_i\right)}{n - \sum_{j=1}^{n} I\left(X_j < X_i\right)} \right) \\ &= \frac{1}{n} \sum \Delta_i \end{aligned}$$

and $\tilde{\theta} = \frac{\sum \Delta_i}{n\hat{\beta}} = \hat{\theta}_1$.

Note that when there is no censoring, the estimator of the cdf $1 - \hat{\theta}$ with a uniform prior simplifies to

$$F_n(t) = \frac{1}{n+2} + \frac{n}{n+2} \hat{F}_n(t)$$

where $\hat{F}_n$ is the empirical cdf. This is a Bayes estimator of the binomial parameter $(1 - F(t))$ with a uniform prior. The Bayes estimator of Ferguson (1973) is

$$\tilde{F}_n(t) = \frac{\alpha((-\infty, t])}{n + \alpha((-\infty, \infty))} + (1 - u_n)\hat{F}_n(t)$$

where $u_n = \alpha((-\infty,\infty))/(\alpha((-\infty,\infty))+n)$ and α is a finite positive measure on the real line. Now, we can calculate the MSE for each of these estimators to compare their performance. For example, if $\alpha((-\infty,\infty)) = 2$, by making $\alpha((-\infty,t])$ to increase at a suitable rate we have $MSE(F_n(t)) < MSE(\tilde{F}_n(t))$ for all t satisfying the condition $|1 - 2F(t)| < |\alpha((-\infty,t]) - 2F(t)|$.

Remark 3.3. One possible drawback of the new estimator is that its monotonicity in t is not guaranteed. However, our simulations indicate that even for moderate sample sizes, the estimator is almost monotonic. The higher accuracy of the proposed method over the classical K-M estimator and the SVR estimator in a mean squared error sense for small to moderate sample sizes is very valuable in estimating tail probabilities. This benefit outweighs the possible slight non-monotonicity of the estimator at some places.

3.3 Empirical Studies

We illustrate the superiority of the proposed Bayes type estimator over the classical K-M estimator and the SVR estimator using a simulation study. The lifetime data are generated from a Weibull family with distribution function

$$F(t) = 1 - e^{-\frac{t^{\eta}}{\gamma}}, : \text{t} > 0,\ \gamma, \eta > 0\ . \tag{3.4}$$

For our simulations, we let $\gamma = 1$, and considered values of $\eta =$ *0.5*, *1.0*, and *1.5*. These values of η correspond to decreasing failure rate, constant failure rate, and increasing failure rate distributions which often arise in lifetime studies. These Weibull values are censored with an exponential variate with parameter λ. The values of λ were chosen so that the level of censoring was 10% and 20% for each Weibull distribution. We used several sample sizes with a beta prior

$$g(t) = B(a,b)t^{a}(1-t)^{b}$$

where $[B(a,b)]^{-1} = \int_0^1 t^a(1-t)^b dt$. We used multiple pairs of (a,b) values above and a few representative results corresponding to $(1,1), (0.5, 0.5)$ and $(0.1, 0.1)$ are provided in the sequel. The number of simulations was 200 in each case where all the calculations were carried out using R (Ihaka and Gentleman, 1996).

We compare the Bayes type survival probability estimator $\hat{\theta} = \hat{S}(t)$ at a time t with the K-M estimator $\hat{S}_n(t)$ and the SVR estimator $S_n(t)$ by looking at the estimated MSE (ESMSE) of the proposed estimator based on the simulations. Specifically, we consider the ratio

$$\frac{ESMSE(\hat{S}_n(t))}{ESMSE(\hat{\theta})}$$

and

$$\frac{ESMSE(S_n(t))}{ESMSE(\hat{\theta})}$$

over a set of values of t in a selected interval where

$$\text{ESMSE}(\phi_n(t)) = \sum_{i=1}^{N}(\phi_n(t) - \phi(t))^2/N$$

at any given t for a function ϕ, with N being the number of simulations. In computing the SVR estimator, we used the $\alpha(-\infty, t] = [1 - exp\{-(t-c)/d\}]I[t \geq c]$ for a variety of c and d values.

We give plots of t versus these ratios in Figures 3.1-3.6; only a few are presented for space considerations ($c = 2, d = 1$ for SVR estimator). The results with other sample sizes and parameter combinations were very similar.

In examining the Figures 3.1-3.4 we notice that the survival estimator using the Bayes argument outperforms the classical K-M and the SVR estimator for small to moderate t in all cases. For samples with $\eta = 1.5$, the K-M and SVR both beats the proposed method in the right tail in a few cases with high censoring. A careful examination revealed that when the data follow an IFR distribution censored by an exponential variable, the last few data points tend to be actual lifetime observations. Hence, the K-M estimator of β, $\hat{\beta}$, can be unstable in this situation leading to a slightly inferior Bayes estimator. In cases where the last few observations are censored, the proposed method seem to do better in the right tails.

The effect of the prior parameters on the performance is reflected in Figure 5. Except for a small range of t values, all examined combinations of the prior parameters seem to provide a more precise estimator of the survival probability than both the K-M estimator and the SVR estimator. When the sample size increases, as seen in Figure 6, the performance of the Bayes estimator becomes more and more comparable to the latter two as expected.

To illustrate the use of the proposed idea, we applied the Bayesian method in estimating the survival distribution of debond strength of carbon fibers without coating discussed by Kuhn and Padgett (1997). In an experiment by Harwell (1995), a droplet of epoxy resin was placed on a ribbon-fiber and cured by heat treatment. The fiber-in-droplet was then put in a "micro-vise", and the fiber was placed under tensile load to attempt to force it from the droplet. The stress at debonding of the fiber and droplet was recorded. Some specimens broke before debonding resulting in right censoring. This particular data consisted of 12 measurements of which 3 were censored. We used the beta prior with several (a, b) pairs and estimated the survival function of debonding strength. The Bayes estimated survival functions for several pairs of prior parameters for a beta distribution along with the corresponding K-M estimator are presented in Figure 3.7.

3.4 Discussion

The Bayesian estimation of the survival probabilities for right-censored lifetime data proposed here has several advantages as have been mentioned. The finite sample dominance of these estimators coupled with reasonable asymptotic properties make them desirable. However, there are several issues in the Bayesian approach that need further investigation. These include the use of other appropriate prior distributions as well as other estimators for the quantities like β. In addition, it is worthwhile to examine the impact of this idea when it is coupled with smoothing methods.

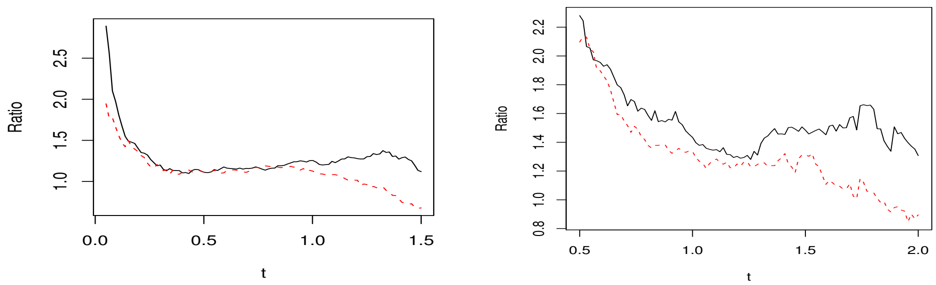

Figure 3.1: Ratio against t for Weibull with $\eta = 1.5, n = 25, a = .1, b = .1$; Solid :10%, Dashed :20%. Left:KM vs. Proposed; Right: SVR vs. Proposed

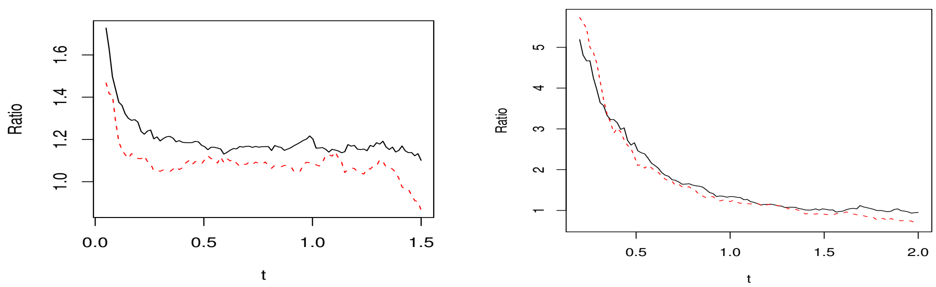

Figure 3.2: Ratio against t for Weibull with $\eta = 1.0, n = 25, a = .1, b = .1$; Solid :10%, Dashed :20%. Left: KM vs. Proposed; Right: SVR vs. Proposed

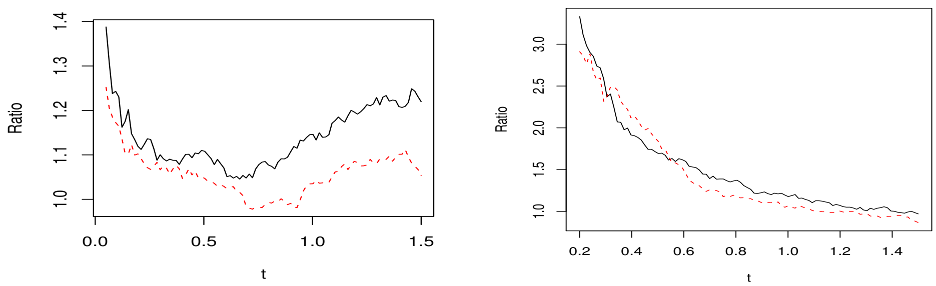

Figure 3.3: Ratio against t for Weibull with $\eta = 1.0, n = 50, a = .1, b = .1$; Solid :10%, Dashed :20%. Left: KM vs. Proposed; Right: SVR vs. Proposed

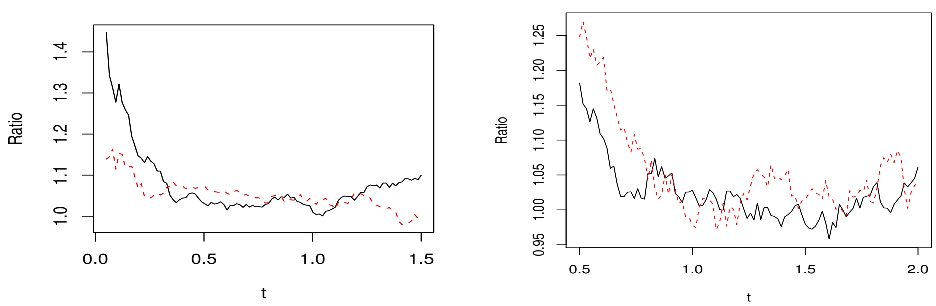

Figure 3.4: Ratio against t for Weibull with $\eta = 1.5, n = 100, a = .1, b = .1$; Solid :10%, Dashed :20%. Left: KM vs. Proposed; Right: SVR vs. Proposed

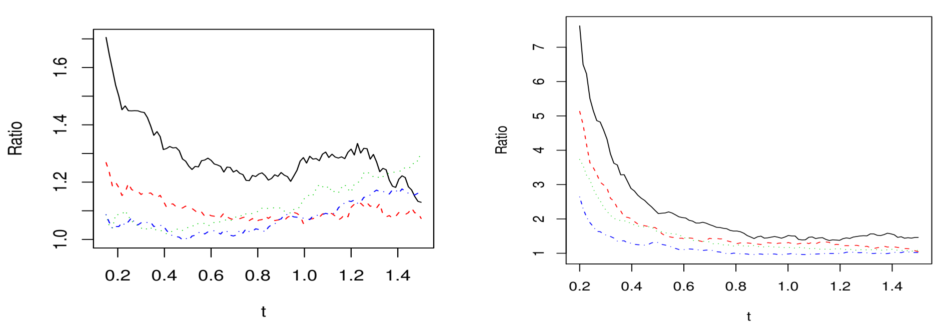

Figure 3.5: Ratio against t for Weibull with $\eta = 1.5, a = .1, b = .1$; 10% Censoring; Solid :$n = 25$, Dashed :$n = 50$ Dotted : $n = 75$,Dot-Dashed : $n = 100$. Left: KM vs. Proposed; Right: SVR vs. Proposed

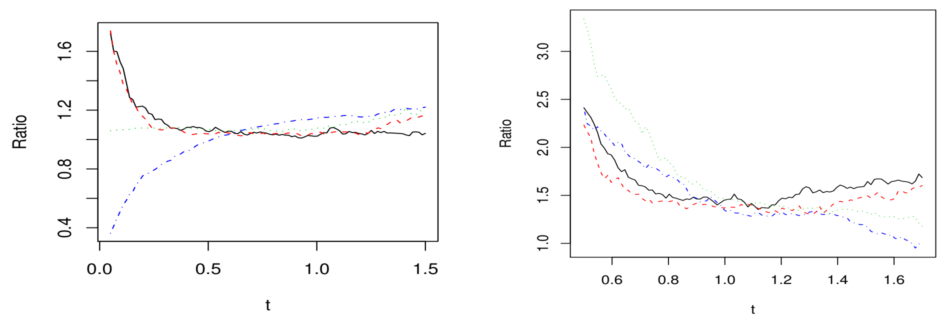

Figure 3.6: Ratio against t for Weibull with $\eta = 1.5, n = 25$, 10% Censoring; Solid :$a = .1, b = .1$; Dashed :$a = .2, b = .2$; Dotted :$a = .5, b = .5$; Dot-Dashed : $a = 1, b = 1$. Left: KM vs. Proposed; Right: SVR vs. Proposed

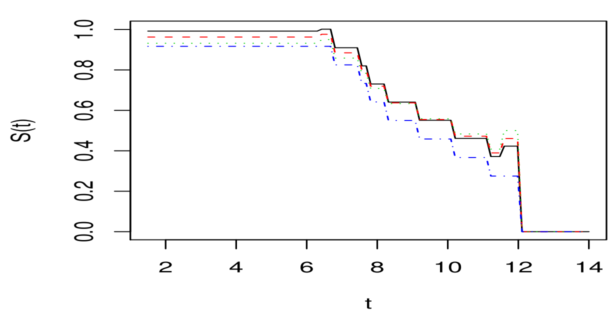

Figure 3.7: Estimated $S(t)$ for Fiber Debonding Data; Solid: Bayes $(a, b) = (0.1, 0.1)$; Dashed: Bayes $(a, b) = (0.5, 0.5)$;Dot: Bayes $(a, b) = (1, 1)$; Dot-Dashed: K-M.

3.5 Proofs

We prove the asymptotic normality of $\hat{\theta}$ here. We prove this using $\hat{\beta} = 1 - \hat{G}(t)$ where $\hat{G}$ is the Kaplan-Meier estimator of G obtained by reversing the role of the censored and uncensored observations and $\hat{p}_1 = \int_t^\infty \hat{S}_n(u)d\hat{G}(u)$ and $\hat{p}_2 = \int_0^t \hat{S}_n(u)d\hat{G}(u)$. Note that

$$\hat{\theta} = \frac{\frac{\sum \delta_i}{n\hat{\beta}} \int_0^1 u^{\sum \delta_i \Delta_i + 1} (1-u)^{\sum \delta_i (1-\Delta_i)} g(u)\, du}{\int_0^1 u^{\sum \delta_i \Delta_i} (1-u)^{\sum \delta_i (1-\Delta_i)} g(u)\, du} + \frac{\sum (1-\delta_i) \Delta_i}{n\hat{\beta}}$$

We will prove that

$$n^{1/2} \left(\hat{\theta} - \tilde{\theta} \right) \to 0$$

in probability, where $\tilde{\theta}$ is the KM estimator. By the argument in Remark 3.2, we have $\tilde{\theta} = \frac{\sum \Delta_i}{n\hat{\beta}}$. Now,

$$n^{1/2} \left(\hat{\theta} - \tilde{\theta} \right) = \frac{\sum \delta_i}{\sqrt{n}\hat{\beta}} \left(\frac{\int_0^1 u^{\sum \delta_i \Delta_i + 1} (1-u)^{\sum \delta_i (1-\Delta_i)} g(u)\, du}{\int_0^1 u^{\sum \delta_i \Delta_i} (1-u)^{\sum \delta_i (1-\Delta_i)} g(u)\, du} - \frac{\sum \delta_i \Delta_i}{\sum \delta_i} \right)$$

Let $a_n = \sum \delta_i \Delta_i + 1$ and $b_n = \sum \delta_i (1-\Delta_i) + 1$. Define

$$k_n(t) = t^{a_n - 1} (1-t)^{b_n - 1} g(t) \Big/ \int_0^1 u^{a_n - 1} (1-u)^{b_n - 1} g(u)\, du$$

We will prove that $\int_0^1 u k_n(u)\, du - t_n = o_p\left(n^{-1/2}\right)$, where $t_n = \frac{a_n - 1}{a_n + b_n - 2}$. Let

$$I_t(a_n, b_n) = t^{a_n - 1} (1-t)^{b_n - 1} \Big/ \int_0^1 u^{a_n - 1} (1-u)^{b_n - 1}\, du$$

be the beta density function. We first prove the following lemma.

Lemma 3.4. *Suppose that* $x(t)$ *is twice differentiable with* x'' *bounded on* $[0,1]$. *We have*

$$\int_0^1 I_t(a_n, b_n)\, x(t)\, dt - x(t_n) = o_p\left(n^{-1/2}\right).$$

Proof: It's easy to see that $a_n = O_p(n)$, $b_n = O_p(n)$. Now,

$$\sup_{t\in[0,1]} I_t(a_n, b_n) = I_{t_n}(a_n, b_n)$$
$$= (a_n + b_n - 1)\frac{(a_n + b_n - 2)!}{(a_n - 1)!\,(b_n - 1)!}(t_n)^{a_n-1}(1 - t_n)^{b_n-1}$$

$$\sim (a_n + b_n - 1)\frac{1}{\sqrt{2\pi}}\left(\frac{a_n + b_n - 2}{(a_n - 1)(b_n - 1)}\right)^{1/2}$$
$$= O_p\left(\sqrt{n}\right).$$

For any $\varepsilon \in (0, 1/6)$, we have

$$\frac{I_{t_n - n^{\varepsilon-1/2}}(a_n, b_n)}{I_{t_n}(a_n, b_n)}$$
$$= \exp\left(-\frac{1}{2}\left((a_n - 1)\left(\frac{n^{\varepsilon-1/2}}{t_n}\right)^2 + (b_n - 1)\left(\frac{n^{\varepsilon-1/2}}{1 - t_n}\right)^2\right) + o_p(1)\right)$$
$$= O_p\left(e^{-\left(n^{2\varepsilon}\right)}\right).$$

Similarly, it can be proved that

$$\frac{I_{t_n + n^{\varepsilon-1/2}}(a_n, b_n)}{I_{t_n}(a_n, b_n)} = O_p\left(e^{-\left(n^{2\varepsilon}\right)}\right)$$

and

$$\int_0^{t_n - 2n^{\varepsilon-1/2}} I_t(a_n, b_n)\, x(t)\, dt \le \|x\|_\infty \int_0^{t_n - 2n^{\varepsilon-1/2}} I_t(a_n, b_n)\, dt$$
$$\le \|x\|_\infty I_{t_n - n^{\varepsilon-1/2}}(a_n, b_n) = O_p\left(e^{-\left(n^{2\varepsilon}\right)}\sqrt{n}\right)$$

Also,

$$\int_{t_n + 2n^{\varepsilon-1/2}}^1 I_t(a_n, b_n)\, x(t)\, dt = O_p\left(e^{-\left(n^{2\varepsilon}\right)}\sqrt{n}\right).$$

Thus,

$$\begin{aligned}
&\int_0^1 I_t(a_n,b_n)\,x(t)\,dt - x(t_n)\\
&= \int_{t_n-2n^{\varepsilon-1/2}}^{t_n+2n^{\varepsilon-1/2}} I_t(a_n,b_n)\,(x(t)-x(t_n))\,dt + O_p\left(e^{-\left(n^{2\varepsilon}\right)}\sqrt{n}\right)\\
&= \int_{t_n-2n^{\varepsilon-1/2}}^{t_n+2n^{\varepsilon-1/2}} I_t(a_n,b_n)\,x'(t_n)\,(t-t_n)\,dt\\
&\quad + O\left(n^{2\varepsilon-1}\right) + O_p\left(e^{-\left(n^{2\varepsilon}\right)}\sqrt{n}\right).
\end{aligned}$$

Now,

$$\int_{t_n-2n^{\varepsilon-1/2}}^{t_n+2n^{\varepsilon-1/2}} I_t(a_n,b_n)\,x'(t_n)\,(t-t_n)\,dt = \int_{-2n^{\varepsilon-1/2}}^{2n^{\varepsilon-1/2}} I_{t+t_n}(a_n,b_n)\,tx'(t_n)\,dt$$

$$= x'(t_n)\left(\int_0^{2n^{\varepsilon-1/2}} t\left(I_{t_n+t}(a_n,b_n) - I_{t_n-t}(a_n,b_n)\right)dt\right),$$

and for $t \in \left[t_n - n^{\varepsilon-1/2}, t_n + n^{\varepsilon-1/2}\right]$,

$$\begin{aligned}
&\left(I_{t_n+t}(a_n,b_n) - I_{t_n-t}(a_n,b_n)\right)\\
&= I_{t_n}(a_n,b_n)\left(\frac{I_{t_n+t}(a_n,b_n)}{I_{t_n}(a_n,b_n)} - \frac{I_{t_n-t}(a_n,b_n)}{I_{t_n}(a_n,b_n)}\right).
\end{aligned}$$

We have

$$\begin{aligned}
\frac{I_{t_n+t}(a_n,b_n)}{I_{t_n}(a_n,b_n)} &= \frac{(t_n+t)^{\alpha_n-1}(1-t_n-t)^{b_n-1}}{t_n^{(\alpha_n-1)}(1-t_n)^{(b_n-1)}}\\
&= \left(1+\frac{t}{\frac{a_n-1}{a_n+b_n-2}}\right)^{\alpha_n-1}\left(1-\frac{t}{\frac{b_n-1}{a_n+b_n-2}}\right)^{b_n-1}
\end{aligned}$$

Thus,

$$\begin{aligned}
&\log\frac{I_{t_n+t}(a_n,b_n)}{I_{t_n}(a_n,b_n)}\\
&= (a_n-1)\log\left(1+\frac{t}{\frac{a_n-1}{a_n+b_n-2}}\right) + (b_n-1)\log\left(1-\frac{t}{\frac{b_n-1}{a_n+b_n-2}}\right)
\end{aligned}$$

$$= (a_n - 1)\left(\frac{t}{t_n} - \frac{1}{2}\left(\frac{t}{t_n}\right)^2 + \frac{1}{3}\left(\frac{1}{(1+u_1)^3}\right)\left(\frac{t}{t_n}\right)^3\right)$$
$$+ (b_n - 1)\left(-\frac{t}{1-t_n} - \frac{1}{2}\left(\frac{t}{1-t_n}\right)^2 - \frac{1}{3}\left(\frac{1}{(1+u_2)^3}\right)\left(\frac{t}{1-t_n}\right)^3\right)$$

for some $u_1 \in (-|t|/t_n, |t|/t_n)$, $u_2 \in (-|t|/(1-t_n), |t|/(1-t_n))$. Similarly,

$$\log \frac{I_{t_n - t}(a_n, b_n)}{I_{t_n}(a_n, b_n)} = (a_n - 1)\left(-\frac{t}{t_n} - \frac{1}{2}\left(\frac{t}{t_n}\right)^2 - \frac{1}{3}\left(\frac{1}{(1+v_1)^3}\right)\left(\frac{t}{t_n}\right)^3\right)$$
$$+ (b_n - 1)\left(\frac{t}{1-t_n} - \frac{1}{2}\left(\frac{t}{1-t_n}\right)^2 + \frac{1}{3}\left(\frac{1}{(1+v_2)^3}\right)\left(\frac{t}{1-t_n}\right)^3\right)$$

for some $v_1 \in (-|t|/t_n, |t|/t_n)$, $v_2 \in (-|t|/(1-t_n), |t|/(1-t_n))$. Thus

$$\begin{aligned}
&(I_{t_n+t}(a_n, b_n) - I_{t_n-t}(a_n, b_n)) \\
&= I_{t_n}(a_n, b_n)\left(\frac{I_{t_n+t}(a_n, b_n)}{I_{t_n}(a_n, b_n)} - \frac{I_{t_n-t}(a_n, b_n)}{I_{t_n}(a_n, b_n)}\right) \\
&= I_{t_n}(a_n, b_n)\left(\exp\left(\log \frac{I_{t_n+t}(a_n, b_n)}{I_{t_n}(a_n, b_n)}\right) - \exp\left(\log \frac{I_{t_n-t}(a_n, b_n)}{I_{t_n}(a_n, b_n)}\right)\right) \\
&= I_{t_n}(a_n, b_n)\, e^{\left(-\frac{1}{2}\left((a_n-1)\frac{t^2}{t_n^2} + (b_n-1)\left(\frac{t}{1-t_n}\right)^2\right)\right)}\left(e^{\alpha_n} - e^{\beta_n}\right)
\end{aligned}$$

where

$$\alpha_n = \frac{1}{3}\left((a_n - 1)\left(\frac{1}{(1+u_1)^3}\right)\left(\frac{t}{t_n}\right)^3 - (b_n - 1)\left(\frac{1}{(1+u_2)^3}\right)\left(\frac{t}{1-t_n}\right)^3\right)$$

and

$$\beta_n = \frac{1}{3}\left(-(a_n - 1)\left(\frac{1}{(1+v_1)^3}\right)\left(\frac{t}{t_n}\right)^3 + (b_n - 1)\left(\frac{1}{(1+v_2)^3}\right)\left(\frac{t}{1-t_n}\right)^3\right).$$

Note that a_n, b_n are of order $O(n)$, v_1, v_2, u_1, u_2, and t are of order $o(1)$, and t_n converges to $P(X_1 > t | T_1 < Y_1) \in (0, 1)$ in probability. Thus, α_n and β_n are of order $O_p\left(n^{3\varepsilon - 1/2}\right)$, which gives

$$(I_{t+t_n}(a_n, b_n) - I_{t_n-t}(a_n, b_n)) = I_{t_n+t}(a_n, b_n)\, O_p\left(n^{3\varepsilon - 1/2}\right),$$

Thus,

$$\int_0^1 I_t(a_n, b_n)\, x(t)\, dt - x(t_n) = O_p\left(n^{4\varepsilon - 1}\right) + O\left(n^{2\varepsilon - 1}\right) + O_p\left(e^{-\left(n^{2\varepsilon}\right)}\sqrt{n}\right)$$
$$= O_p\left(n^{4\varepsilon - 1}\right) = o_p\left(n^{-1/2}\right).$$

The following corollary follows directly from Lemma 3.4.

Corollary 3.5. $\int_0^1 uk_n(u)\,du - t_n = o_p\left(n^{-1/2}\right)$

Now, by Corollary 3.5

$$
\begin{aligned}
n^{1/2}\left(\hat{\theta}-\tilde{\theta}\right) &= \frac{\sum \delta_i}{\sqrt{n}\hat{\beta}}\left(\frac{\int_0^1 u^{\sum \delta_i \Delta_i + 1}(1-u)^{\sum \delta_i (1-\Delta_i)} g(u)\,du}{\int_0^1 u^{\sum \delta_i \Delta_i}(1-u)^{\sum \delta_i (1-\Delta_i)} g(u)\,du} - \frac{\sum \delta_i \Delta_i}{\sum \delta_i}\right) \\
&= \frac{\sum \delta_i}{\sqrt{n}\hat{\beta}} o_p\left(n^{-1/2}\right) = o_p(1),
\end{aligned}
$$

proving the Theorem.

Chapter 4

Minimax Estimation of Linear Functionals

4.1 Introduction

Suppose that we observe data of the form

$$\mathbf{y} = \mathbf{K}\mathbf{x} + \mathbf{Z} \tag{4.1}$$

where $\mathbf{x}$ belongs to a convex subset $\mathcal{F}$ of a separable real Hilbert space $\mathbf{X}$ and $\mathbf{K}$ is a linear operator from $\mathcal{F}$ to $\mathbf{U}$, another separable real Hilbert space. $\mathbf{Z}$ is a bounded linear operator from $\mathbf{U}$ to $\mathscr{L}_2(\Omega, \mathscr{F}, P)$, the space of all random variables define on the probability space $(\Omega, \mathscr{F}, P)$ that have finite variance. It is assumed that $\mathbf{Z}$ is invertible (the inverse might be unbounded) and $\mathbf{Z}\mathbf{w}$ has mean zero for $\mathbf{w} \in \mathbf{U}$. By defining $\mathbf{y} = \mathbf{K}\mathbf{x} + \mathbf{Z}$, we are treating $\mathbf{K}\mathbf{x}$ as an operator from $\mathbf{U}$ to $\mathscr{L}_2(\Omega, \mathscr{F}, P)$. This is justified since $\mathbf{K}\mathbf{x}$ defines a functional on $\mathbf{U}$, and real numbers can be treated as constant random variables in $\mathscr{L}_2(\Omega, \mathscr{F}, P)$. Suppose that L is a real affine functional on $\mathbf{X}$, that is $L = L_1 + l$ where L_1 is a linear functional on $\mathbf{X}$ and l is a constant. We consider the problem of estimating the value of L at some $\mathbf{x} \in \mathcal{F}$ by affine estimators of the form

$$\hat{L}(\mathbf{w}, d) = \mathbf{y}\mathbf{w} + d = \langle \mathbf{w}, \mathbf{K}\mathbf{x} \rangle + \mathbf{Z}\mathbf{w} + d, \tag{4.2}$$

where $\mathbf{w}$ is in $\mathbf{U}$. We will evaluate the performance of an affine estimator by the mean squared error. We want to obtain the minimax affine risk for estimating L

$$\inf_{\hat{L} \text{ affine}} R_{\mathcal{F}}\left(\hat{L}; L, K\right),$$

where

$$R_{\mathcal{A}}\left(\hat{L}; L, K\right) = \sup_{\mathbf{x}\in\mathcal{A}} E_{\mathbf{x}}\left(\hat{L} - L\left(\mathbf{x}\right)\right)^2.$$

Two special cases of model (4.1) are the white noise model

$$Y\left(t\right) = \int_{-1/2}^{t} f\left(s\right) ds + \sigma W\left(t\right), \qquad -1/2 \leq t \leq 1/2, \tag{4.3}$$

where $W\left(t\right)$ is a standard Brownian motion, and the regression model

$$y_i = f\left(t_i\right) + \sigma z_i, \qquad i = 1, \ldots, n; t_i \in [-1/2, 1/2] \tag{4.4}$$

where z_i's are i.i.d. noises and $f \in \mathcal{F}$, where $\mathcal{F}$ is a convex class of functions in each model. In Ibragimov and Khasminskii (1984) the minimax linear risk for the white noise model (4.3) was given together with the minimax linear estimator for a hypothesis set that is symmetric. A relationship between the minimax linear risk and the minimax risk was also established. Donoho and Liu (1991) removed the symmetry constraint on the set $\mathcal{F}$ and established the minimax affine risk and rate of convergence for the asymptotic minimax risk in the white noise model. In their work, the minimax affine risk was expressed in term of modulus of continuity. An interesting result of Donoho and Liu (1991) is that the minimax affine risk for the full problem is just the minimax affine risk of the hardest one dimensional subproblem. Using this result, it was readily shown that the ratio of the minimax affine risk to the minimax risk is bounded by 1.25. Donoho and Liu (1991) then applied their result on the white noise model to regression data with independent errors. Donoho (1994) showed the same results for a generalization of model (4.3) in which the data $\mathbf{y}$ have the form: $\mathbf{y} = \mathbf{K}\mathbf{x} + \mathbf{z}$ where $\mathbf{x}$ is from a convex subset $\mathbf{X}$ of l_2, the space of all square summable sequences, $\mathbf{K}$ is a linear operator and $\mathbf{z}$ is a noise vector. That generalization was in two senses: first, the data are not observed directly, but through the operator $\mathbf{K}$. Second, the Gaussian process $\mathbf{z}$ can be non-white noise. However, it was still required that the covariance has a bounded inverse. Note that this is not the case for fractional Brownian motion, which is a special cases of the model (1) above. Fan (1993) discussed the estimation of a regression function in a slightly different framework, and it was shown that under some restrictive conditions the local linear smoother, with a proper choice of kernel and bandwidth, is near minimax. By applying Ibragimov and Khasminskii (1984)'s and Donoho and Liu (1991)'s results, Zhao (1997) gave the exact linear minimax estimator of $f(0)$ under the white noise model with f known to be in the closure of the class $\{f : |f''(t)| \leq B\}$, and by comparing the kernel of this linear minimax estimator with the Epanechinikov kernel, it was shown that the Epanechinikov kernel is 99% efficient. This confirmed Ylvisaker's conjecture (see Sacks and Ylvisaker (1981)), which states that the Epanechinikov kernel is nearly minimax.

Several authors have considered the estimation of the function f itself, a different problem than estimating a linear functional of f, in the regression model. Among them are Wang (1996, 1997), Donoho and Johnstone (1998), and Johnstone (1999). In these papers, the performance of an estimator was evaluated by the expectation of the L_2 distance between the estimator and f. In Donoho and Johnstone (1998) it was shown that in the case of estimating f, the linear estimator no longer has near minimax risk. The authors proposed the use of a wavelet transformation to convert the function space $\mathcal{F}$ into a sequence space, and an estimator for f was then obtained in the wavelet domain by simple nonlinear shrinkage of the empirical wavelet coefficients. Besides showing the near minimaxity of the wavelet shrinkage estimator over a wide range of Triebel and Besov-type (Donoho and Johnstone, 1998) smoothness constraints and asymptotic minimaxity over certain Besov bodies, the authors also showed that more practical simple threshold nonlinear estimators are nearly minimax. This work of estimating f was extended by Wang (1996) and Wang (1997) for the case when long-range dependency appears in the data. In these works, Wang worked with the long-memory counterparts of models (4.3) and (4.4) – the fractional Gaussian noise model and the regression model with dependent errors. In the fractional Gaussian noise model, the white noise in model (4.3) was replaced by fractional Gaussian noise. In Wang (1996), it was shown that with proper scaling, the fractional Gaussian noise model is an approximation to the nonparametric regression model with correlated errors. Wavelet estimates with proper choices of thresholds were shown to achieve minimax rates over a wide range of function spaces. Wang (1997) extended the results of Wang (1996) to the case in which the data are indirect. Johnstone (1999) discussed the choice of the threshold in the wavelet estimators in Wang (1996) that adapts to a broad range of Besov classes. He also proposed an extension to the case of indirect data similar to (and independently from) Wang

(1997).

Deo (1997) discussed the estimation of a linear functional for data with long-memory errors. A kernel estimator was studied and, under some rather restrictive conditions, the asymptotic normality of the estimator was shown. We will show that their result coincides with our lower bound on the asymptotic rate for the minimax risk.

Other generalizations of the white noise model include Cai and Low (2003) and Cai and Low (2004). Based on the results of Donoho (1994), Cai and Low (2003) gave precise asymptotic descriptions of the minimax affine risks and bias variance trade-offs for estimating linear functionals for what the authors called regular modulus. Cai and Low (2004) extended the minimax theory for estimating linear functionals to the case of a finite union of convex parameter spaces. In this extension, an interesting contrast to the case of convex parameter spaces is that linear estimators no longer have optimal rates of convergence.

Here, we extend the results of Donoho and Liu (1991) and prove that, in our generalized setting (4.1),

$$\inf_{\hat{L} \text{ affine}} R_{\mathcal{F}}\left(\hat{L}; L, \mathbf{K}\right) = \sup_{\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{F}} \inf_{\hat{L} \text{ affine}} R_{[\mathbf{x}_1, \mathbf{x}_2]}\left(\hat{L}; L, \mathbf{K}\right) \tag{4.5}$$

where $[\mathbf{x}_1, \mathbf{x}_2] = \{\mathbf{x} : \mathbf{x} = \alpha\mathbf{x}_1 + \beta\mathbf{x}_2, 0 \le \alpha \le 1, \beta = 1 - \alpha\}$ is the convex hull of $\{\mathbf{x}_1, \mathbf{x}_2\}$. This is the same as saying that the minimax affine risk is the supremum of the minimax affine risks for one-dimenssional subproblems. To our knowledge, this is the most generalized version of the model for convex parameter spaces that has been considered and our results reduce to all existing results including those of Donoho and Liu (1991) and Donoho (1994). By defining the operator $\mathbf{Z}$ to be different stochastic processes, our results can be applied to a wide range of estimation problems. The operator $\mathbf{K}$ also give much flexibility to the model.

We apply the results to the fractional Gaussian noise model and the nonparametric regression model with correlated data, obtaining the rate of convergence for the asymptotic minimax affine risk of estimating the value of a function $f^{(k)}$ at a fixed point t_0. Under both these settings, we also show that the ratio of the minimax affine risk to the minimax risk is bounded above by 1.25. In both cases, there is a presence of long-range dependency in the responses. It is important to discuss responses with long-memory because we encounter them in many different fields of study; geology (Painter, 1996), hydrology (Turcotte, 1994), signal processing, (Wornell and Oppenheim, 1992) , Computer Science, (Park et al., 2005) , and finance (Andersen and Bollerslev, 1997) to name a few.

Notice that although Wang (1996), Wang (1997) and Johnstone (1999) discussed estimation with long-range dependent data, their discussion was focused on estimating f itself, and the performance was evaluated in L_2 distance, which does not tell us a lot about the error of estimation at a fixed point. Also, the relationship between the minimax affine risk and the minimax risk has not been derived in the literature for such settings.

We prove (4.5) in few steps. First, we notice that for an affine estimator $\hat{L}(\mathbf{w}, d)$, we can find $c \in \mathbb{R}$ and $\mathbf{w}_0 \in S(\mathbf{U})$, the unit sphere of $\mathbf{U}$, such that $\mathbf{w} = c\mathbf{w}_0$. Thus the affine estimator becomes $\hat{L}(c\mathbf{w}_0, d)$ for some $c, d \in \mathbb{R}$. For calculating the minimax affine risk, we can take the infimum over $\mathbf{w}_0 \in S(\mathbf{U})$ and $c, d \in \mathbb{R}$. Next, for technical reasons to be explained in the sequel, every vector $\mathbf{w} \in S(\mathbf{U})$ is approximated by some $c_1\mathbf{w}'$ with $\mathbf{w}' \in W_a$, and $c_1 \in \mathbb{R}$. Here

$$W_a = \{\mathbf{w} \in \mathbf{U} : \langle \mathbf{w}, \mathbf{v} \rangle = a, 0 < a < 1, \|\mathbf{w}\| \leq 1\} \tag{4.6}$$

for a fixed unit vector $\mathbf{v} \in S(\mathbf{U})$, thus solving the reduced problem of finding the minimax risk with respect to the subset $\left\{\hat{L}(c\mathbf{w}, d) : c, d \in R, \mathbf{w} \in W_a\right\}$ of all affine estimators. Finally, we extend the result to the full problem − finding the minimax affine risk (with the infimum taken over all $\mathbf{w} \in S(\mathbf{U})$).

This chapter is organized as follows. Section 4.2 gives some general results on the model described by (4.1). Section 4.3 applies the results to the fractional Brownian motion model and the nonparametric regression model with correlated errors respectively. Finally proofs for the results presented in these two sections are given in Section 4.4.

4.2 Minimax Risk for the Hardest one dimensional subproblem and the Full Problem

In this section we first consider the minimax risk for the hardest one dimensional problem. These results are then extended to the general $\mathcal{F}$. We start with a few definitions. Suppose that T is an operator from $\mathcal{F}$ to another Banach space H. The modulus of continuity of T is defined as

$$\omega(\epsilon; T, \mathcal{F}) = \sup\{\|T(\mathbf{x}_2) - T(\mathbf{x}_1)\| : \|\mathbf{x}_2 - \mathbf{x}_1\| \leq \epsilon \text{ and } \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{F}\}$$

where the $\| \|$'s are the norms of the respective normed spaces. As in Donoho (1994), we define the modulus of continuity of L with respect to the seminorm $\|v\|_{\mathbf{K}} \equiv \|\mathbf{K}v\|$ as

$$\omega(\epsilon; L, \mathbf{K}, \mathcal{F}) = \sup\left\{|L(\mathbf{x}_2) - L(\mathbf{x}_1)| : \|\mathbf{x}_2 - \mathbf{x}_1\|_{\mathbf{K}} \leq \epsilon \text{ and } \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{F}\right\}.$$

We assume that $\omega(\epsilon; L, \mathbf{K}, \mathcal{F})$ is finite for all $\epsilon > 0$. We also make the following assumptions

Assumption 4.5. *(a)* $\lim_{\epsilon\to 0} \omega(\epsilon; L, \mathcal{F}) \to 0$ *and (b)* $\lim_{\epsilon\to 0} \omega(\epsilon; \mathbf{K}, \mathcal{F}) \to 0$.

4.2.1 The hardest one dimensional sub-problem

To address the hardest one dimensional subproblem, we first look at a one dimensional sub-problem. Suppose that $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{F}$. Let $[\mathbf{x_1}, \mathbf{x_2}]$ denote the convex span of $\{\mathbf{x}_1, \mathbf{x}_2\}$. Since $\mathcal{F}$ is convex, this is a subfamily of $\mathcal{F}$. Now we consider the problem of estimating $L(\mathbf{x})$ with affine estimators when we know that $\mathbf{x}$ is in $[\mathbf{x}_1, \mathbf{x}_2]$. Let $\hat{L}$ be an affine estimator defined as in (4.2). The risk of $\hat{L}$ is

$$\begin{aligned} E_{\mathbf{x}}\left(\hat{L} - L(\mathbf{x})\right)^2 &= E\left(L(\mathbf{x}) - \langle \mathbf{w}, \mathbf{K}\mathbf{x}\rangle - \mathbf{Z}\mathbf{w} - d\right)^2 \\ &= \left(L(\mathbf{x}) - \langle \mathbf{w}, \mathbf{K}\mathbf{x}\rangle - d\right)^2 + \|\mathbf{Z}\mathbf{w}\|^2 \\ &= bias\left(\hat{L}, \mathbf{x}\right)^2 + \|\mathbf{Z}\mathbf{w}\|^2. \end{aligned}$$

Then the maximum risk for $\hat{L}$ over $[\mathbf{x}_1, \mathbf{x}_2]$ is

$$R_{[\mathbf{x}_1,\mathbf{x}_2]}(\hat{L}; L, \mathbf{K}) = \sup_{\mathbf{x}\in[\mathbf{x}_1,\mathbf{x}_2]} bias\left(\hat{L}, \mathbf{x}\right)^2 + \|\mathbf{Z}\mathbf{w}\|^2. \tag{4.7}$$

Later in the text, we may omit part or all of the secondary arguments of $R_{[\mathbf{x}_1,\mathbf{x}_2]}$ in (4.7) and simply write $R_{[\mathbf{x}_1,\mathbf{x}_2]}\left(\hat{L}\right)$ when they are clear from the context. We will also do the same in other notations with secondary arguments.

For any $\mathbf{x} \in [\mathbf{x}_1, \mathbf{x}_2]$, let $\mathbf{x} = \alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2$ with $\alpha \in [0, 1]$. We have

$$\begin{aligned} bias\left(\hat{L}, \mathbf{x}\right) &= E_{\mathbf{x}}\left(\hat{L} - L(\mathbf{x})\right) = E\left(\langle \mathbf{w}, \mathbf{K}\mathbf{x}\rangle + \mathbf{Z}\mathbf{w} + d - L(\mathbf{x})\right) \\ &= E\left(\langle \mathbf{w}, \mathbf{K}\left(\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2\right)\rangle + \mathbf{Z}\mathbf{w} + d - L\left(\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2\right)\right) \end{aligned}$$

$$= \alpha E\left(\langle \mathbf{w}, \mathbf{K}\mathbf{x}_1\rangle + \mathbf{Z}\mathbf{w} + d - L\left(\mathbf{x}_1\right)\right)$$
$$+ \left(1-\alpha\right) E\left(\langle \mathbf{w}, \mathbf{K}\mathbf{x}_2\rangle + \mathbf{Z}\mathbf{w} + d - L\left(\mathbf{x}_2\right)\right)$$
$$= \alpha bias\left(\hat{L}, \mathbf{x}_1\right) + \left(1-\alpha\right) bias\left(\hat{L}, \mathbf{x}_2\right).$$

Thus,

$$\begin{aligned} \sup_{\mathbf{x}\in[\mathbf{x}_1,\mathbf{x}_2]} \left|bias\left(\hat{L}, \mathbf{x}\right)\right| &= \sup_{\alpha\in[0,1]} \left|\left(\alpha bias\left(\hat{L}, \mathbf{x}_1\right) + \left(1-\alpha\right) bias\left(\hat{L}, \mathbf{x}_2\right)\right)\right| \\ &= \max\left\{\left|bias\left(\hat{L}, \mathbf{x}_1\right)\right|, \left|bias\left(\hat{L}, \mathbf{x}_2\right)\right|\right\}, \end{aligned}$$

and

$$R_{[\mathbf{x}_1,\mathbf{x}_2]}(\hat{L}) = \max\left\{bias\left(\hat{L}, \mathbf{x}_1\right)^2, bias\left(\hat{L}, \mathbf{x}_2\right)^2\right\} + \|\mathbf{Z}\mathbf{w}\|^2$$

For $\mathbf{w} \in \mathbf{U}, d \in \mathbb{R}$, let $\hat{L} = \hat{L}\left(\mathbf{w}, d\right)$. We have

$$\begin{aligned} &R_{[\mathbf{x}_1,\mathbf{x}_2]}(\hat{L}) \\ &= \max\left\{E\left(\langle \mathbf{w}, \mathbf{K}\mathbf{x}_1\rangle + \mathbf{Z}\mathbf{w} + d - L\left(\mathbf{x}_1\right)\right)^2 + E\left(\langle \mathbf{w}, \mathbf{K}\mathbf{x}_2\rangle + \mathbf{Z}\mathbf{w} + d - L\left(\mathbf{x}_2\right)\right)^2\right\} \\ &+ \|\mathbf{Z}\mathbf{w}\|^2 \end{aligned} \tag{4.8}$$

When $\mathbf{w}$ is fixed, from (4.8) we have

$$\inf_{d\in\mathbb{R}} R_{[\mathbf{x}_1,\mathbf{x}_2]}(\hat{L}(\mathbf{w}, d)) = \left(L_1\left(\mathbf{x}_2 - \mathbf{x}_1\right)/2 - \langle \mathbf{w}, \mathbf{K}\left(\mathbf{x}_2 - \mathbf{x}_1\right)/2\rangle\right)^2 + \|\mathbf{Z}\mathbf{w}\|^2,$$

and the above minimum is obtained if

$$d = L((\mathbf{x}_1 + \mathbf{x}_2)/2) - \langle \mathbf{w}, \mathbf{K}\left(\mathbf{x}_1 + \mathbf{x}_2\right)/2\rangle,$$

Notice that in the above equation, we extend the definition of L_1 and $\mathbf{K}$ to the vector space generated by $\mathcal{F}$, and we will do the same later on whenever necessary. Now, let $\mathbf{w} = c\mathbf{w}_0$, where $\mathbf{w}_0 \in S\left(\mathbf{U}\right) =$

$\{\mathbf{w} \in \mathbf{U} : \|\mathbf{w}\| = 1\}$. Then

$$\inf_{c,d} R_{[\mathbf{x}_1,\mathbf{x}_2]}\left(\hat{L}(c\mathbf{w}_0, d)\right) = \inf_{c}\Big\{c^2\left\|\mathbf{Z}\mathbf{w}_0\right\|^2 + \left[L_1\left(\left(\mathbf{x}_2 - \mathbf{x}_1\right)/2\right) - \langle c\mathbf{w_0}, \mathbf{K}\left(\mathbf{x}_2 - \mathbf{x}_1\right)/2\rangle\right]^2\Big\}.$$

A straightforward calculation shows that

$$\inf_{c,d} R_{[\mathbf{x}_1,\mathbf{x}_2]}\left(\hat{L}(c\mathbf{w}_0, d)\right) = \frac{\left[L_1(\left(\mathbf{x}_2 - \mathbf{x}_1\right)/2)\right]^2}{1 + \left[g_{\mathbf{w}_0}\left(\mathbf{x}_2 - \mathbf{x}_1\right)\right]^2}$$

with $g_{\mathbf{w}}\left(\mathbf{x}\right) = \langle \mathbf{w}, \mathbf{K}\mathbf{x}/2\rangle / \|\mathbf{Z}\mathbf{w}\|$, where the above minimum is achieved at

$$c = c_0 = \frac{L_1(\left(\mathbf{x}_2 - \mathbf{x}_1\right)/2)\langle \mathbf{w}_0, \mathbf{K}\left(\mathbf{x}_2 - \mathbf{x}_1\right)/2\rangle}{\langle \mathbf{w}_0, \mathbf{K}\left(\mathbf{x}_2 - \mathbf{x}_1\right)/2\rangle^2 + \left\|\mathbf{Z}\mathbf{w}_0\right\|^2}.$$

Hence, the minimax affine risk for the one-dimensional subfamily is

$$\begin{aligned}\inf_{\mathbf{w}\in\mathbf{U}, d\in\mathbb{R}} R_{[\mathbf{x}_1,\mathbf{x}_2]}\left(\hat{L}(\mathbf{w}, d)\right) &= \inf_{\mathbf{w}_0\in S(\mathbf{U}); c,d\in\mathbb{R}} R_{[\mathbf{x}_1,\mathbf{x}_2]}\left(\hat{L}(c\mathbf{w}_0, d)\right) \\ &= \inf_{\mathbf{w}_0\in S(\mathbf{U})} \frac{\left[L_1(\left(\mathbf{x}_2 - \mathbf{x}_1\right)/2)\right]^2}{1 + \left[g_{\mathbf{w}_0}\left(\mathbf{x}_2 - \mathbf{x}_1\right)\right]^2}. \end{aligned} \tag{4.9}$$

Thus finding the minimax risk for the one-dimensional subfamily is the same as finding the maximum of $\left|g_{\mathbf{w}}\left(\mathbf{x}_2 - \mathbf{x}_1\right)\right|$ over all $\mathbf{w} \in S(\mathbf{U})$. Also notice that in (4.9), $g_{\mathbf{w}_0}$ is determined by the "direction" of $\mathbf{w}_0$. This means that instead of taking the infimum in (4.9) over $S\left(\mathbf{U}\right)$, we can take it over a set W of vectors that "covers all the directions" in the sense that for any $\mathbf{w} \notin W$, we can find $\mathbf{w'} \in W$ such that $\mathbf{w} = c\mathbf{w'}$ for some $c \in \mathbb{R}$. Donoho (1994), under his setting, showed that the minimax affine risk is achieved by the estimator of the form $\hat{L}\left(c_0\mathbf{w}_0, d_0\right)$ with $\mathbf{w}_0 = \left(\mathbf{x}_2 - \mathbf{x}_1\right)/\|\mathbf{x}_2 - \mathbf{x}_1\|$ using a sufficiency argument. This approach is not generally possible in our setting. However, we can still find a sequence $\mathbf{w}_n \in S\left(\mathbf{U}\right)$ such that

$$\lim_{n} g_{\mathbf{w}_n}\left(\mathbf{x}_2 - \mathbf{x}_1\right) = \sup_{\mathbf{w}\in S(\mathbf{U})} g_{\mathbf{w}}\left(\mathbf{x}_2 - \mathbf{x}_1\right).$$

Now, by the weak sequential compactness of the unit ball of a separable Hilbert space, we can find a subsequence $\mathbf{w}_{n_k}$ that converge weakly to a $\mathbf{w'} \in \mathbf{U}$. A problem that can arise here is that this weak limit may be $\mathbf{0}$. As a way of getting around the zero limit issue, instead of taking the supremum over

$S(\mathbf{U})$, we take the supremum over an indexed collection of bounded closed convex subsets of $\mathbf{U}$ that do not contain $\mathbf{0}$. Of course such sets will not "cover all the directions" in the above sense. However, by taking a suitable limit on the indexing parameter, such condition can be "almost" satisfied. Our choice of these subsets are of the form $W_a = \{\mathbf{w} \in \mathbf{U} : \langle \mathbf{w}, \mathbf{v} \rangle = a, \|\mathbf{w}\| \le 1\}$, where $0 < a < 1$ and $\mathbf{v}$ is a fixed unit vector in $\mathbf{U}$. This collection of subsets has the property that every element in $S(\mathbf{U})$ can be approximated (in norm of $\mathbf{U}$) by some $c\mathbf{w}$ with $\mathbf{w} \in W_a$ and $c \in [-1, 1]$, and the error of this approximation goes to zero uniformly over $S(\mathbf{U})$ as $a \to 0$. Now we have the following result.

Lemma 4.1. *For every pair* $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{F}$, *and* $0 < a < 1$ *there exists a* $\mathbf{w}_a(\mathbf{x}_2 - \mathbf{x}_1) \in W_a$ *such that* $\left|g_{\mathbf{w}_a(\mathbf{x}_2-\mathbf{x}_1)}\left(\mathbf{x}_2 - \mathbf{x}_1\right)\right| = \sup_{\mathbf{w}\in W_a} \left|g_{\mathbf{w}}\left(\mathbf{x}_2 - \mathbf{x}_1\right)\right|.$

Next, for $0 < a < 1$ and $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{F}$, define

$$\begin{aligned} G_a\left(\mathbf{x}_2 - \mathbf{x}_1\right) &= G_a\left(\mathbf{x}_2 - \mathbf{x}_1; \mathbf{Z}, \mathbf{K}\right) \\ &= \sup_{\mathbf{w}\in W_a} \left|g_{\mathbf{w}}\left(\mathbf{x}_2 - \mathbf{x}_1\right)\right| = \left|g_{\mathbf{w}_a(\mathbf{x}_2-\mathbf{x}_1)}\left(\mathbf{x}_2 - \mathbf{x}_1\right)\right| \end{aligned}$$

and

$$\rho_a\left(\mathbf{x}_2 - \mathbf{x}_1\right) = \rho_a(\mathbf{x}_2 - \mathbf{x}_1; \mathbf{Z}, L, \mathbf{K}) = \inf_{\mathbf{w}\in W_a} \frac{\left(L_1((\mathbf{x}_2 - \mathbf{x}_1)/2)\right)^2}{1 + g_{\mathbf{w}}^2\left(\mathbf{x}_2 - \mathbf{x}_1\right)}.$$

Extending this notation we also define

$$G_0\left(\mathbf{x}\right) = \sup_{\mathbf{w}\in W_0} \left|g_{\mathbf{w}}\left(\mathbf{x}\right)\right|$$

and

$$\rho_0\left(\mathbf{x}\right) = \inf_{\mathbf{w}\in W_0} \frac{\left(L_1\left(\mathbf{x}/2\right)\right)^2}{1 + g_{\mathbf{w}}^2\left(\mathbf{x}\right)},$$

where $W_0 \equiv S\left(\mathbf{U}\right)$. Note that $\rho_a\left(\mathbf{x}\right)$ is non-increasing when $a \downarrow 0$. By Lemma 4.1,

$$\rho_a(\mathbf{x}) = \left(L_1(\mathbf{x}/2)\right)^2 \Big/ \left(1 + g_{\mathbf{w}_a(\mathbf{x})}^2\left(\mathbf{x}\right)\right) = \left(L_1\left(\mathbf{x}/2\right)\right)^2 / \left(1 + G_a^2\left(\mathbf{x}\right)\right), 0 < a < 1.$$

Thus

$$\inf_{\mathbf{w}\in W_a;c,d\in\mathbb{R}} R_{[\mathbf{x}_1,\mathbf{x}_2]}\left(\hat{L}(c\mathbf{w},d)\right) = \rho_a(\mathbf{x}_2-\mathbf{x}_1)$$
$$= \frac{\left[L_1\left(\left(\mathbf{x}_2-\mathbf{x}_1\right)/2\right)\right]^2}{1+\left[g_{\mathbf{w}_a(\mathbf{x}_2-\mathbf{x}_1)}\left(\mathbf{x}_2-\mathbf{x}_1\right)\right]^2}$$

for $a\in(0,1)$, where $R_{[\mathbf{x}_1,\mathbf{x}_2]}\left(\hat{L}\right)$ is defined in (4.7).

Having examined the minimax risk for a one dimensional sub-problem, we next find the minimax risk for the hardest 1-dimensional sub-problem. The lemma given below will be used in the sequel. It is a slightly modified version of Lemma 5 in Donoho (1994), and can be proven along the same lines.

Lemma 4.2. *Let* $\mathbf{V}$ *be a closed bounded convex set in a separable Hilbert space* $\mathbf{H}$*, and* $J(\mathbf{v})$ *a continuous convex functional on* $\mathbf{V}$*. Suppose that* $(\mathbf{v}_n)$ *is a sequence in* $\mathbf{V}$ *converging weakly to* $\mathbf{v}$*. Then* $J(\mathbf{v})\leq\liminf J(\mathbf{v}_n)$.

We define

$$\rho_a\left(\mathcal{F}\right)=\rho_a\left(\mathcal{F};\mathbf{Z},L,\mathbf{K}\right)=\sup_{\mathbf{x}_1,\mathbf{x}_2\in\mathcal{F}}\rho_a(\mathbf{x}_2-\mathbf{x}_1;\mathbf{Z},L,\mathbf{K})$$

for $0<a<1$ and

$$\rho_0\left(\mathcal{F}\right)=\rho_0\left(\mathcal{F};\mathbf{Z},L,\mathbf{K}\right)=\sup_{\mathbf{x}_1,\mathbf{x}_2\in\mathcal{F}}\rho_0\left(\mathbf{x}_2-\mathbf{x}_1;\mathbf{Z},L,\mathbf{K}\right).$$

It is easy to see that $\rho_a\left(\mathcal{F}\right)$ is non-increasing as $a\downarrow 0$. Now, we can prove the following lemma that finds the minimax risk for the hardest one dimensional sub-problem for the restricted case $\mathbf{w}\in W_a$.

Lemma 4.3. *If* $\mathcal{F}$ *is convex, closed, and bounded, then for each* $0<a<1$ *there exists a pair* $\mathbf{x}_1(a;\mathbf{Z},L,\mathbf{K},\mathcal{F})$*,* $\mathbf{x}_2(a;\mathbf{Z},L,\mathbf{K},\mathcal{F})$ *(which we simply write as* $\mathbf{x}_1\left(a\right),\mathbf{x}_2\left(a\right)$ *when there is no confusion) such that* $\rho_a(\mathbf{x}_2(a)-\mathbf{x}_1(a))=\rho_a\left(\mathcal{F}\right)$.

4.2.2 The Full Problem

We obtain the minimax risk for the full problem in this section. This is examined in several steps. First, we obtain the minimax risk for estimating L with affine estimators of the form $L\left(c\mathbf{w}, d\right)$, with $\mathbf{w} \in W_a$ over a bounded closed convex parameter space $\mathcal{F}$. Once the minimax risk over such sets are derived, our goal is to remove the boundedness and closedness for $\mathcal{F}$ and then extend the results to estimators with $\mathbf{w} \in S\left(\mathbf{U}\right)$. In particular, we now give the following result, which states that for a bounded closed convex parameter space, when we restrict our attention to affine estimators of the form $\hat{L}(c\mathbf{w} + d)$ with $\mathbf{w} \in W_a$, the minimax risk is the supremum of the minimax risks of one-dimensional subproblems.

Theorem 4.4. *If $\mathcal{F}$ is a bounded closed convex subset of $\mathbf{X}$, then*

$$\inf_{\mathbf{w}\in W_a; c,d\in\mathbb{R}} R_{\mathcal{F}}\left(\hat{L}\right) = \rho_a\left(\mathcal{F}\right) = \rho_a\left(\mathbf{x}_2\left(a\right) - \mathbf{x}_1\left(a\right)\right)$$

where $0 < a < 1$ and $\hat{L} = \hat{L}\left(c\mathbf{w}, d\right)$. The above infimum is achieved at $\mathbf{w}_0 = \mathbf{w}_a\left(\mathbf{x}_2\left(a\right) - \mathbf{x}_1\left(a\right)\right)$ with $c = c_0$ where

$$c_0 = \frac{L_1(\left(\mathbf{x}_2\left(a\right) - \mathbf{x}_1\left(a\right)\right)/2)\langle \mathbf{w}_0, \mathbf{K}\left(\mathbf{x}_2\left(a\right) - \mathbf{x}_1\left(a\right)\right)/2\rangle}{\langle \mathbf{w}_0, \mathbf{K}\left(\mathbf{x}_2\left(a\right) - \mathbf{x}_1\left(a\right)\right)/2\rangle^2 + \left\|\mathbf{Z}\mathbf{w}_0\right\|^2},$$

and

$$d = d_0 = L(\left(\mathbf{x}_1\left(a\right) + \mathbf{x}_2\left(a\right)\right)/2) - \langle \mathbf{w}_0, \mathbf{K}\left(\mathbf{x}_1\left(a\right) + \mathbf{x}_2\left(a\right)\right)/2\rangle,$$

and $\mathbf{x}_1\left(a\right)$ and $\mathbf{x}_2\left(a\right)$ are defined in Lemma 4.3.

Now we argue that the boundedness and closedness constraints for $\mathcal{F}$ in Theorem 4.4 can be removed. To this end, we have the following theorem which is proven in a fashion similar to the proof of Theorem 2 in Donoho (1994).

Theorem 4.5. *Let $\mathcal{F}$ be a convex subset of $\mathbf{X}$, and $0 < a < 1$. Then*

$$\inf_{\mathbf{w}\in W_a; c,d\in\mathbb{R}} R_{\mathcal{F}}\left(\hat{L}\right) = \rho_a\left(\mathcal{F}\right)$$

and there exists $\tilde{\mathbf{w}} \in W_a$ and $c, d \in \mathbb{R}$ such that the affine estimator $\hat{L}_0 = \hat{L}\left(c\tilde{\mathbf{w}}, d\right)$ achieves the above minimax risk.

Theorem 4.5 shows that when we restrict the affine estimator to be determined by some

$\mathbf{w} \in W_a$, the minimax affine risk of estimating $L(\mathbf{x})$ is just the supremum over the minimax affine risks (the infimum is taken over estimators of the form $\hat{L}(c\mathbf{w}, d)$ with $\mathbf{w} \in W_a$, and $c, d \in \mathbb{R}$) of all one-dimensional subproblems. We also notice that every vector in $S(\mathbf{U})$ can be approximated (in norm) by $c\mathbf{w}$ with a $c \in \mathbb{R}$ and $\mathbf{w} \in W_a$, and the error of this approximation goes to zero as a goes to zero. Thus it is natural to think that the minimax affine risk $\inf_{\mathbf{w}\in S(\mathbf{U});c,d\in\mathbb{R}} R_{\mathcal{F}}\left(\hat{L}(c\mathbf{w}, d)\right)$ can be obtained by taking the limit in a of $\inf_{\mathbf{w}\in W_a;c,d\in\mathbb{R}} R_{\mathcal{F}}\left(\hat{L}(c\mathbf{w}, d)\right)$. The following theorem, which is the main result of this chapter, will show that this is indeed the case.

Theorem 4.6. *Suppose that $\mathcal{F}$ is convex. Then the $\mathbf{v}$ in* (4.6) *can be chosen so that*

$$\inf_{\hat{L}\ affine} R_{\mathcal{F}}\left(\hat{L}\right) = \lim_{a\to 0} \inf_{\mathbf{w}\in W_a;c,d\in\mathbb{R}} R_{\mathcal{F}}\left(\hat{L}(c\mathbf{w}, d)\right) = \lim_{a\to 0} \rho_a(\mathcal{F}).$$

In Theorem 4.6, the minimax affine risk is expressed as the limit, $\lim_{a\to 0} \rho_a(\mathcal{F})$. It will be desirable if we can express the minimax affine risk in term of $\rho_0(\mathcal{F})$. This can be done for bounded $\mathcal{F}$ as given in the following lemma.

Lemma 4.7. *If $\mathcal{F}$ is convex, closed and bounded, then*

$$\lim_{a\to 0} \rho_a(\mathcal{F}) = \rho_0(\mathcal{F}). \tag{4.10}$$

The result proven in the next lemma shows that we can actually relax the boundedness restriction on $\mathcal{F}$. To proceed with this avenue, we make the following assumption.

Assumption 4.6. *For any positive M, we have*

$$\chi(M) = \sup_{\mathbf{x}_1,\mathbf{x}_2\in\mathcal{F},\|\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)\|\le M} \|\mathbf{x}_2 - \mathbf{x}_1\| < \infty$$

Lemma 4.8. *If $\mathcal{F}$ is closed and symmetric and Assumption 4.6 is satisfied, then* (4.10) *holds.*

With the above lemma and Theorem 4.6, and also the fact that $\rho_a(\mathcal{F}) = \rho_a(cl(\mathcal{F}))$, and $\rho_0(\mathcal{F}) = \rho_0(cl(\mathcal{F}))$, which are not hard to prove, the following corollary is immediate.

Corollary 4.9. *If Assumption 4.6 is satisfied and $\mathcal{F}$ is convex and symmetric, then* $\inf_{\hat{L}\ affine} R_{\mathcal{F}}\left(\hat{L}\right) = \rho_0(\mathcal{F})$.

4.3 Applications

Now we apply the above results to two examples. The first is the fractional Brownian motion model, and the second is the nonparametric regression model with correlated errors.

4.3.1 Fractional Brownian Motion Model

Let $\Omega = \mathscr{C}_0\left([0,T],\mathbb{R}\right)$ be the space of real-valued continuous functions on the interval $[0,T]$ with the topology of local uniform convergence and initial value zero. Let $\mathscr{F}$ be the Borel σ-algebra. There is a probability measure P on $(\Omega, \mathscr{F})$ under which the coordinate process $(Z_t, t \in [0,T])$ is a Gaussian process that has stationary increments and satisfies the following.

i. $Z_0 = 0$.

ii. $EZ_t = 0$ for all $t \geq 0$.

iii. $R(s,t) = EZ_tZ_s = \left(t^{2H} + s^{2H} - |s-t|^{2H}\right)/2$ for every $s, t \geq 0$.

The process $(Z_t, t \in [0,T])$ is called the fractional Brownian motion on the interval $[0,T]$ with the Hurst index H. Now consider the following model

$$y(t) = \int_0^t f(u)\,du + \sigma Z_t, t \in [0,T] \tag{4.11}$$

where $\sigma > 0$ and f belongs to $\mathbf{W}_{[0,T]}(m,p,C)$ (Donoho and Liu, 1991) with $m \geq 1$ and $1 \leq p \leq \infty$. That is, f is defined on $[0,T]$ and satisfies

a. $f, \ldots, f^{(m-1)}$ are absolutely continuous, and

b. $f \in \mathscr{L}_2[0,T]$ and $\left\|f^{(m)}\right\|_p \leq C$.

Cavalier (2004) discussed a model of similar to (4.11). But like Donoho and Johnstone (1998), Wang (1996,1997), and Johnstone (1999), the objective was to estimate f itself. The author assumed that the Hurst index H is in the interval $(0, 1/2)$. With this assumption, it is possible to connect this model with an inverse problem of the type

$$\tilde{Y} = Af + \sigma\xi. \tag{4.12}$$

To be specific, by acting a proper fractional integration operator A on $y(\cdot)$ in (4.11), it is possible transfer it to the inverse problem given by (4.12) with ξ being a standard Brownian motion. Thus solving the estimation problem defined by model (4.11) is equivalent to solving the corresponding inverse problem. However, for $H \in (1/2, 1)$, this approach will not work because under such situations, to convert the fractional Brownian motion Z_t to a standard Brownian motion, We need the integral operator $IZ_t = \int_0^t C(x-t)^{\frac{1}{2}-H-1} dZ_t$. But the integral operator $If(x) = \int C(x-t)^{\frac{1}{2}-H-1} f(t)\,dt$ is not well defined on $\mathbf{W}_{[0,T]}(m,p,C)$. From the results in the next subsection we can also see that the model (4.11) with $H \in (1/2, 1)$ is linked with nonparametric regression for long memory data through weak convergence of the probability measures involved. Thus we consider the more difficult case $H \in (1/2, 1)$ which is of more interest.

We are interested in the problem of estimating $f^{(k)}(t_0)$ for $0 \le k < m$ and $t_0 \in (0,T)$ from observing $y(t), t \in [0,T]$. It is assumed that either $k < m-1$ or $p > 1$. First we verify that this is a special case of the model and the estimation problem discussed in section 2. For a function $g \in \mathscr{L}_2[0,T]$, the integral $\int_0^T g(u)\,dZ_u$ is well defined, and it can be shown that

$$E\left(\left|\int_0^T g(u)\,dZ_u\right|^2\right) = \int_0^T \int_0^T g(u)\,g(v)\,\phi(u,v)\,dudv < \infty$$

where $\phi(u,v) = H(2H-1)|u-v|^{2H-2}$. Following Duncan et al. (2000), we define

$$|g|_\phi := \left(\int_0^T \int_0^T g(s)\,g(t)\,\phi(s,t)\,dsdt\right)^{1/2}$$

and let

$$\mathscr{L}_\phi^2 = \mathscr{L}_\phi^2([0,T]) = \left\{f | f : [0,T] \to \mathbb{R}, |f|_\phi^2 < \infty\right\}.$$

An inner product $\langle \cdot,\cdot \rangle_\phi$ can also be defined on $\mathscr{L}_\phi^2$:

$$\langle f,g \rangle_\phi = \int_0^T \int_0^T f(u)\,g(v)\,\phi(u,v)\,dudv.$$

Memin et al. (2001) showed that for $g \in \mathscr{L}_{1/H}[0,T]$,

$$E\left(\left|\int_0^T g(u)\,dZ_u\right|^2\right) \le c(H,2)\,\|g\|_{1/H}^2.$$

This enables us to treat the integration with respect to Z as a bounded operator from $\mathscr{L}_{1/H}[0,T]$ to $\mathscr{L}_2(\Omega, \mathscr{F}, P)$. For $f \in \mathscr{L}_2[0,T]$, $|f|^{1/H} \in \mathscr{L}_{2H}[0,T]$. Thus it can be shown that

$$\|f\|_{1/H} = \left(\int_0^T |f|^{1/H} \, dt \right)^H \leq T^{H-1/2} \|f\|_2 .$$

Thus the identity map I from $\mathscr{L}_2[0,T]$ to $\mathscr{L}_{1/H}[0,T]$ is bounded. Combining these two operators we get a bounded linear operator from $\mathscr{L}_2[0,T]$ to $\mathscr{L}_2(\Omega, \mathscr{F}, P)$. We write this operator as $\mathbf{Z}_1$ and let $\mathbf{Z}_\sigma = \sigma \mathbf{Z}_1$. To show that with $\mathcal{F} = \mathbf{W}_{[0,T]}(m,p,C)$, $\mathbf{Z} = \mathbf{Z}_\sigma$, and $\mathbf{K} = \mathbf{I}$, the model described by (4.11) is a special case of the model described by (4.1), now we only need to show that Assumption 4.5 is satisfied, which is immediate from the following lemma.

Lemma 4.10. $\sup_{f \in \mathbf{W}_{[0,T]}(m,p,C), \|f\|_r \leq \epsilon} \left\| f^{(k)} \right\|_\infty = O(\epsilon^{\alpha_k})$ *where*

$$\alpha_k = \left(m - k - p^{-1}\right) / \left(m - p^{-1} + r^{-1}\right) \; and 0 \leq k < m.$$

Now, we examine the rate of convergence of the minimax affine risk for estimating $f^{(k)}(t_0)$ for model (4.11). For $g \in \mathscr{L}_2[0,T]$ and $d \in \mathbb{R}$, we can define an affine estimator

$$\hat{L}(g,d) = \int_0^T g(u)\, y(u)\, du + d = \int_0^T g(u)\, f(u)\, du + \sigma \int_0^T g(u)\, dZ_t + d.$$

The main result in this section is to obtain the rate of convergence of the minimax affine risk when $\sigma \to 0$, and to show that the minimax affine risk and the minimax risk have the same rate of convergence. Define

$$\upsilon(\epsilon; \mathbf{Z}, L, \mathbf{K}, \mathcal{F}) = \inf \left\{ |G_0(\mathbf{x}_2 - \mathbf{x}_1)| : \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{F}, |L_1((\mathbf{x}_2 - \mathbf{x}_1)/2)| = \epsilon \right\}.$$

If $\{\mathbf{x}_2 - \mathbf{x}_1 : \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{F},\, L_1((\mathbf{x}_2 - \mathbf{x}_1)/2) = \epsilon\} = \emptyset$, then we let $\upsilon(\epsilon) = \infty$. It is easy to check that $\upsilon(\epsilon)$ is convex. Clearly,

$$\begin{aligned} \rho_0(\mathcal{F}) &= \sup_{\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{F}} \frac{\left[L_1((\mathbf{x}_2 - \mathbf{x}_1)/2)\right]^2}{1 + \left[G_0(\mathbf{x}_2 - \mathbf{x}_1)\right]^2} \\ &= \sup_{\epsilon > 0} \sup \left\{ \frac{\left[L_1(\mathbf{x}_2 - \mathbf{x}_1)/2\right]^2}{1 + \left[G_0(\mathbf{x}_2 - \mathbf{x}_1)\right]^2} : |L_1((\mathbf{x}_2 - \mathbf{x}_1)/2)| = \epsilon \right\} \end{aligned}$$

$$= \sup_{\epsilon>0} \frac{\epsilon^2}{1+\upsilon^2(\epsilon)}.$$

If $\mathcal{F}$ is symmetric and $\mathbf{K}$ satisfies Assumption 4.6, then by Corollary 4.9, we have

$$\inf_{\hat{L}\text{ affine}} R_{\mathcal{F}}\left(\hat{L}\right) = \sup_{\epsilon>0} \frac{\epsilon^2}{1+\upsilon^2(\epsilon)}. \tag{4.13}$$

In the fractional Brownian model described above, $\mathcal{F} = \mathbf{W}_{[0,T]}(m,p,C)$ and $\mathbf{K} = \mathbf{I}$. It is easy to check that $\mathbf{W}_{[0,T]}(m,p,C)$ is symmetric. For a function $h(\epsilon)$ of ϵ, we use the notation $h(\epsilon) \asymp \epsilon^\alpha$ to denote $A_1\epsilon^\alpha \le h(\epsilon) \le A_2\epsilon^\alpha$ for ϵ sufficiently small, and A_1, A_2 two constants free of ϵ. Now we have the following lemma that characterizes υ for this specific model.

Lemma 4.11. $\upsilon\left(\epsilon; \mathbf{Z}_\sigma, f^{(k)}(t_0), \mathbf{I}, \mathbf{W}_{[0,T]}(m,p,C)\right) \asymp \epsilon^{\gamma_k}/\sigma$, *with* $\gamma_k = \frac{m-p^{-1}+1-H}{m-k-p^{-1}}$.

With Lemma 4.11, we can now obtain the rate of the minimax affine risk given in (4.13) above. In particular, we have the following result.

Theorem 4.12. *The minimax affine risk for the model described in* (4.11) *satisfies*

$$\inf_{\hat{L}\ \textit{affine}} R_{\mathbf{W}_{[0,T]}(m,p,C)}\left(\hat{L}\right) \asymp \sigma^{2/\gamma_k}. \tag{4.14}$$

Next, we will get an upper bound for the ratio of the minimax affine risk and the minimax risk, and show that the rate of convergence for minimax risk for estimating $f^{(k)}(t_0)$ is also given by (4.14).

Theorem 4.13. *The minimax affine risk for the model described in* (4.11) *satisfies*

$$\frac{\inf_{\hat{L}\ \textit{affine}} R_{\mathbf{W}_{[0,T]}(m,p,C)}\left(\hat{L}\right)}{\inf_{\hat{T}\ \textit{measurable}} R_{\mathbf{W}_r(m,p,C)}\left(\hat{T}\right)} \le 1.25. \tag{4.15}$$

With Theorem 4.12 and Theorem 4.13, we have the following corollary.

Corollary 4.14. *The minimax risk for the model described in* (4.11) *satisfies*

$$\inf_{\hat{T}\ \textit{measurable}} R_{\mathbf{W}_{[0,T]}(m,p,C)}\left(\hat{T}\right) \asymp \sigma^{2/\gamma_k}.$$

A close look at the results of Theorem 4.12 and Corollary 4.14 reveals that if the index of the fractional Brownian motion H is taken to be $1/2$, in which case it reduces to Brownian motion, then our results agree with those presented in Donoho and Liu (1991).

4.3.2 Regression Model

Nonparametric regression with long-range dependent errors was studied in Wang (1996), who also established asymptotics for minimax risk with respect to the L_2 norm. In this chapter, we consider the pointwise minimax risk. The regression model is described as

$$y_i = f(t_i) + z_i, \quad i = 1, \dots, n, \tag{4.16}$$

where the t_i's are equispaced on $[0,T]$ and $z_1, \dots, z_n$ are observational errors with mean 0 and finite variance. We assume that $(z_i)_{1 \le i \le n}$ have long-range dependence (Wang, 1996),

$$R(j-i) = \mathrm{Cov}(z_i, z_j) \sim C_1 |j-i|^{-\alpha}, \quad j - i \to \infty$$

with $0 < \alpha < 1$. The regression function f is known to belong to $\mathbf{W}_{[0,T]}(m,p,C)$. The problem of interest is the estimation of $f^{(k)}(t_0)$ for some $t_0 \in (0,T)$, $0 \le k < m$, by affine estimators of the form (4.2) with $\mathbf{x} = f$, $\mathbf{y} = (y_1, \dots, y_n)' \in \mathbb{R}^n$, and $\mathbf{K}_n : \mathbf{W}_{[0,T]}(m,p,C) \to \mathbb{R}^n$ and $\tilde{\mathbf{Z}}_n : \mathbb{R}^n \to \mathscr{L}_2(\Omega, \mathscr{F}, P)$ being operators defined by $\mathbf{K}_n f = (f(t_1), \dots, f(t_n))'$ and $\tilde{\mathbf{Z}}_n((c_1, \dots, c_n)') = \frac{1}{n^{1/2}} \sum_{i=1}^n c_i z_i$ respectively. For two vectors $\mathbf{u} = (u_1, \dots, u_n)'$ and $\mathbf{v} = (v_1, \dots, v_n)'$ in $\mathbb{R}^n$, we define the inner product $\langle \mathbf{u}, \mathbf{v} \rangle = \frac{1}{n} \sum u_i v_i$ and the norm $\|\mathbf{u}\|_2 = \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle}$. By the smoothness of the functions in $\mathbf{W}_{[0,T]}(m,p,C)$, it is easy to see that $\mathbf{K}_n$ is well defined. We will also show that $\mathbf{K}_n$ also satisfies Assumption 4.5 and 4.6. Throughout this section we assume $L(f) = f^{(k)}(t_0)$ and $\mathcal{F} = \mathbf{W}_{[0,T]}(m,p,C)$.

By Lemma 4.10, the $\mathbf{K}_n$ and $\mathcal{F}$ defined above satisfies part (b) of Assumption 4.5. The fact that Assumption 4.6 is satisfied follows readily from the following lemma.

Lemma 4.15. *If $n \ge m$, then there exists a positive K independent of f such that*

$$\left\| f^{(k)} \right\|_\infty \le K \|\mathbf{K}_n f\|_2 + C T^{m-k-p^{-1}} \tag{4.17}$$

for $k = 0, \dots, m-1$.

The main results of this section is given in the following theorems. The proofs are very tedious, and hence we refer the reader to Zhao and Kulasekera (2005) for details.

Theorem 4.16. *Let* $\tilde{H} = 1 - \alpha/2$*.Then the minimax affine risk for the model described in* (4.16) *satisfies* $\inf_{\hat{L} \text{ affine}} R_{\mathbf{W}_{[0,T]}(m,p,C)}\left(\hat{L}; \mathbf{K}_n, \tilde{\mathbf{Z}}_n\right) \asymp n^{2\left(\tilde{H}-1\right)/\gamma_k}$ *where* $\gamma_k = \frac{m-p^{-1}+1-\tilde{H}}{m-k-p^{-1}}$.

By a sufficiency discussion similar to that in the proof of Theorem 4.13 (see Section 4.4), the following theorem can be proven.

Theorem 4.17. *For the model described in* (4.16)*, we have*

$$\frac{\inf_{\hat{L} \text{ affine}} R_{\mathbf{W}_{[0,T]}(m,p,C)}\left(\hat{L}; \mathbf{K}_n, \tilde{\mathbf{Z}}_n\right)}{\inf_{\hat{T} \text{ measurable}} R_{\mathbf{W}_{[0,T]}(m,p,C)}\left(\hat{T}; \mathbf{K}_n, \tilde{\mathbf{Z}}_n\right)} \leq 1.25.$$

By Theorem 4.16 and Theorem 4.17, we have

Corollary 4.18. *The minimax risk for the model described in* (4.16) *satisfies*

$$\inf_{\hat{T} \text{ measurable}} R_{\mathbf{W}_{[0,T]}(m,p,C)}\left(\hat{T}; \mathbf{K}_n, \tilde{\mathbf{Z}}_n\right) \asymp n^{2\left(\tilde{H}-1\right)/\gamma_k}.$$

Deo (1997) discussed two kernel estimators for estimating the value of a function $f : [0, 1] \to \mathbb{R}$ at $x \in (0, 1)$ from data with long-memory errors. According to Theorem 3 of Deo (1997), if f is assumed to be twice differentiable on $[0, 1]$ with both derivatives bounded, then the two kernel estimators studied have asymptotic risk of order $(nh_n)^{-\alpha}$ where h_n satisfies $nh_n^{1+4/\alpha} \to 0$, $nh_n^{1+\eta} \to \infty$ for some $\eta > 0$. It can be seen that in this case the parameter space is a subset of $W_{[0,1]}(2, \infty, C)$ for some $C > 0$, and according to our result, the rate of convergence for the minimax risk is of order $n^{-4\alpha/(4+\alpha)}$, which is a lower bound for the rate for the kernel estimators given by Deo (1997).

4.4 Proofs

In this section we provide proofs of some main results stated in the previous sections.

4.4.1 Proofs for Results in Section 4.2

Proof of Lemma 4.1

Suppose that $\mathbf{w}_n \in W_a$, $n = 1, 2, \ldots$, and

$$|g_{\mathbf{w}_n}(\mathbf{x})| \to \sup_{\mathbf{w} \in W_a} |g_{\mathbf{w}}(\mathbf{x})|.$$

By the weak sequential compactness of the unit ball $B(\mathbf{U})$, we can find a subsequence $\mathbf{w}_{n_k}$ which converge weakly to $\mathbf{w}_0 \in W_a$. Then, $\langle \mathbf{w}_{n_k}, \mathbf{K}\mathbf{x}/2 \rangle \to \langle \mathbf{w}_0, \mathbf{K}\mathbf{x}/2 \rangle$ and $\mathbf{Z}\mathbf{w}_{n_k} \xrightarrow{w} \mathbf{Z}\mathbf{w}_0$, which gives

$$\liminf \|\mathbf{Z}\mathbf{w}_{n_k}\| \geq \|\mathbf{Z}\mathbf{w}_0\|.$$

Thus,

$$\begin{aligned}
\sup_{\mathbf{w} \in W_a} |g_{\mathbf{w}}(\mathbf{x})| &= \lim_{k\to\infty} |\langle \mathbf{w}_{n_k}, \mathbf{K}\mathbf{x}/2 \rangle / \|\mathbf{Z}\mathbf{w}_{n_k}\| | \\
&= |\langle \mathbf{w}_0, \mathbf{K}\mathbf{x}/2 \rangle| / \liminf \|\mathbf{Z}\mathbf{w}_{n_k}\| \\
&\leq |\langle \mathbf{w}_0, \mathbf{K}\mathbf{x}/2 \rangle| / \|\mathbf{Z}\mathbf{w}_0\| \\
&= |g_{\mathbf{w}_0}(\mathbf{x})| \\
&\leq \sup_{\mathbf{w} \in W_a} |g_{\mathbf{w}}(\mathbf{x})|.
\end{aligned}$$

Letting $\mathbf{w}_a(\mathbf{x}) = \mathbf{w}_0$ finishes the proof.

Proof of Lemma 4.3

Suppose that $\mathbf{x}_1^n, \mathbf{x}_2^n, n = 1, 2, \ldots$ satisfy

$$\rho_a(\mathbf{x}_2^n - \mathbf{x}_1^n) \to \rho_a(\mathcal{F}).$$

Then, we find subsequences $\mathbf{x}_1^{n_k}, \mathbf{x}_2^{n_k}$ such that $\mathbf{x}_i^{n_k} \xrightarrow{w} \mathbf{x}_i(a), i = 1, 2$. With Assumption 4.5, by Lemma 4.2, we have $L_1(\mathbf{x}_2^{n_k} - \mathbf{x}_1^{n_k}) \to L_1(\mathbf{x}_2(a) - \mathbf{x}_1(a))$ and $\langle \mathbf{w}, \mathbf{K}(\mathbf{x}_2^{n_k} - \mathbf{x}_1^{n_k}) \rangle \to \langle \mathbf{w}, \mathbf{K}(\mathbf{x}_2(a) - \mathbf{x}_1(a)) \rangle$ for any $\mathbf{w} \in \mathbf{U}$. Hence,

$$\begin{aligned}
\rho_a(\mathcal{F}) &\geq \rho_a(\mathbf{x}_2(a) - \mathbf{x}_1(a)) \\
&= \frac{L_1((\mathbf{x}_2(a) - \mathbf{x}_1(a))/2)^2}{1 + g_{\mathbf{w}_a(\mathbf{x}_2(a) - \mathbf{x}_1(a))}(\mathbf{x}_2(a) - \mathbf{x}_1(a))^2}
\end{aligned}$$

$$= \lim \frac{L_1((\mathbf{x}_2^{n_k} - \mathbf{x}_1^{n_k})/2)^2}{1 + g_{\mathbf{w}_a(\mathbf{x}_2(a)-\mathbf{x}_1(a))}\left(\mathbf{x}_2^{n_k} - \mathbf{x}_1^{n_k}\right)^2}$$

$$\geq \limsup \rho_a(\mathbf{x}_2^{n_k} - \mathbf{x}_1^{n_k}) = \rho_a\left(\mathcal{F}\right).$$

proving the lemma.

To prove Theorem 1, we need the following Lemmas (Lemma 4.19 – Lemma 4.25). We state each Lemma, give a proof and then prove Theorem 4.5.

Lemma 4.19. *Suppose that $\mathcal{F}$ is a bounded closed convex subset of $\mathbf{X}$. For $\mathbf{x}, \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{F}$ and $h \in (0,1)$, Let $\mathbf{x}_h = h\mathbf{x} + (1-h)\,\mathbf{x}_2 - \mathbf{x}_1$ and $\mathbf{x}_0 = \mathbf{x}_2 - \mathbf{x}_1$. Then*

$$\left|\left|g_{\mathbf{w}_a(\mathbf{x}_h)}\left(\mathbf{x}_h\right)\right| - \left|g_{\mathbf{w}_a(\mathbf{x}_0)}\left(\mathbf{x}_0\right)\right|\right| = O\left(h\right).$$

Proof of Lemma 4.19

Since $\left|g_{\mathbf{w}_a(\mathbf{x}_0)}\left(\mathbf{x}_0\right)\right| \geq \left|g_{\mathbf{w}_a(\mathbf{x}_h)}\left(\mathbf{x}_0\right)\right|$, we have

$$\begin{aligned}
\left|g_{\mathbf{w}_a(\mathbf{x}_h)}\left(\mathbf{x}_h\right)\right| - \left|g_{\mathbf{w}_a(\mathbf{x}_0)}\left(\mathbf{x}_0\right)\right| &\leq \left|g_{\mathbf{w}_a(\mathbf{x}_h)}\left(\mathbf{x}_h\right)\right| - \left|g_{\mathbf{w}_a(\mathbf{x}_h)}\left(\mathbf{x}_0\right)\right| \\
&\leq \left|g_{\mathbf{w}_a(\mathbf{x}_h)}\left(\mathbf{x}_h\right) - g_{\mathbf{w}_a(\mathbf{x}_h)}\left(\mathbf{x}_0\right)\right| \\
&= \left|g_{\mathbf{w}_a(\mathbf{x}_h)}\left(\mathbf{x}_h - \mathbf{x}_0\right)\right| \\
&= \left|g_{\mathbf{w}_a(\mathbf{x}_h)}\left(h\tilde{\mathbf{x}}\right)\right| \\
&= \left|\frac{\left\langle \mathbf{w}_a\left(\mathbf{x}_h\right), h\mathbf{K}\tilde{\mathbf{x}}/2\right\rangle}{\left\|\mathbf{Z}\mathbf{w}_a\left(\mathbf{x}_h\right)\right\|}\right| \\
&\leq \frac{h\left\|\mathbf{K}\tilde{\mathbf{x}}\right\|/2}{M} \\
&= O\left(h\right)
\end{aligned}$$

where $M = \inf_{\mathbf{w}\in W_a}\left\|\mathbf{Z}\mathbf{w}\right\| > 0$, and $\tilde{\mathbf{x}} = \mathbf{x} - \mathbf{x}_2$. Similarly, we can show that

$$\left|g_{\mathbf{w}_a(\mathbf{x}_0)}\left(\mathbf{x}_0\right)\right| - \left|g_{\mathbf{w}_a(\mathbf{x}_h)}\left(\mathbf{x}_h\right)\right| = O\left(h\right).$$

This proves the Lemma.

Now we introduce a few new terms. For $\mathbf{u} \in \mathbf{U}$, define

$$I_t^{\mathbf{u}} = \left\{ \mathbf{w} \in W_a : \langle \mathbf{w}, \mathbf{u} \rangle = t \sup_{\mathbf{w}' \in W_a} \langle \mathbf{w}', \mathbf{u} \rangle + (1-t) \inf_{\mathbf{w}' \in W_a} \langle \mathbf{w}', \mathbf{u} \rangle \right\}$$

where $t \in [0,1]$. Now let $\mathbf{u} = \mathbf{u}_{\mathbf{v}} + \mathbf{u}_{\perp}$, where $\mathbf{v}$ is as in (4.6), $\mathbf{u}_{\mathbf{v}} = \alpha \mathbf{v}$ and $\mathbf{u}_{\perp} \in \ker(\mathbf{v}) = \{\mathbf{w} \in \mathbf{U} : \langle \mathbf{w}, \mathbf{v} \rangle = 0\}$. For $\mathbf{w} \in W_a$, we have

$$\begin{aligned} \langle \mathbf{w}, \mathbf{u} \rangle &= \langle \mathbf{w}, \mathbf{u}_{\mathbf{v}} \rangle + \langle \mathbf{w}, \mathbf{u}_{\perp} \rangle = a\alpha + \langle \mathbf{w}, \mathbf{u}_{\perp} \rangle \\ &\in [a\alpha - \|\mathbf{u}_{\perp}\| \theta, a\alpha + \|\mathbf{u}_{\perp}\| \theta] \end{aligned}$$

where $\theta = \sqrt{1-a^2}$. Hence, $\sup_{\mathbf{w}' \in W_a} \langle \mathbf{w}', \mathbf{u} \rangle = a\alpha + \|\mathbf{u}_{\perp}\| \theta$ and $\inf_{\mathbf{w}' \in W_a} \langle \mathbf{w}', \mathbf{u} \rangle = a\alpha - \|\mathbf{u}_{\perp}\| \theta$. For $\mathbf{w} \in I_t^{\mathbf{u}}$,

$$\langle \mathbf{w}, \mathbf{u} \rangle = a\alpha + \langle \mathbf{w}, \mathbf{u}_{\perp} \rangle = a\alpha + (2t-1) \|\mathbf{u}_{\perp}\| \theta.$$

Thus, $\langle \mathbf{w}, \mathbf{u}_{\perp} \rangle = (2t-1) \|\mathbf{u}_{\perp}\| \theta$. For $U, V \subset \mathbf{X}$, define

$$D(U,V) = \max \left\{ \sup_{\mathbf{x} \in U} \inf_{\mathbf{y} \in V} \|\mathbf{x} - \mathbf{y}\|, \sup_{\mathbf{x} \in V} \inf_{\mathbf{y} \in U} \|\mathbf{x} - \mathbf{y}\| \right\}.$$

Now, we have the following Lemma that describes the behavior of D on U.

Lemma 4.20. *For $t \in [0,1]$ and $\mathbf{u} \in \mathbf{U}$, if $\mathbf{u}_{\perp} \neq \mathbf{0}$, then $D\left(I_t^{\mathbf{u}}, I_t^{\mathbf{u}'}\right) \to 0$ as $\mathbf{u}' \to \mathbf{u}$. And the convergence is uniform in t.*

Proof of Lemma 4.20

Let $J_t^{\mathbf{u}} = I_t^{\mathbf{u}} - a\mathbf{v}, J_t^{\mathbf{u}'} = I_t^{\mathbf{u}'} - a\mathbf{v}$. It is not hard to see that $D\left(I_t^{\mathbf{u}}, I_t^{\mathbf{u}'}\right) = D\left(J_t^{\mathbf{u}}, J_t^{\mathbf{u}'}\right)$. It can be shown that $J_t^{\mathbf{u}} = \{\mathbf{w} \in B_\theta(\ker(\mathbf{v})) : \langle \mathbf{w}, \mathbf{u}_{\perp} \rangle = (2t-1) \|\mathbf{u}_{\perp}\| \theta\}$, where $B_\theta(\ker(\mathbf{v})) = \{\mathbf{w} \in \ker(\mathbf{v}) : \|\mathbf{w}\| \leq \theta\}$ and remember that $\theta = \sqrt{1-a^2}$. Now we will show that *diam* $(J_t^{\mathbf{u}}) \to 0$ uniformly in $\mathbf{u}$ if $t \to 1$ or $t \to 0$. Suppose that $\mathbf{w} \in J_t^{\mathbf{u}}$. Let $\mathbf{w}' = (2t-1) \theta \mathbf{u}_{\perp} / \|\mathbf{u}_{\perp}\|$ Then we have $\mathbf{w}' \in J_t^{\mathbf{u}}$. Therefore,

$$\langle \mathbf{w} - \mathbf{w}', \mathbf{w}' \rangle = \frac{(2t-1)\theta (\langle \mathbf{w}, \mathbf{u}_{\perp} \rangle - \langle \mathbf{w}', \mathbf{u}_{\perp} \rangle)}{\|\mathbf{u}_{\perp}\|} = 0.$$

which means that $\|\mathbf{w}\|^2 = \|\mathbf{w}-\mathbf{w}'+\mathbf{w}'\|^2 = \|\mathbf{w}-\mathbf{w}'\|^2 + \|\mathbf{w}'\|^2$. We know that $\|\mathbf{w}'\| = |(2t-1)|\,\theta$ and $\|\mathbf{w}\| \le \theta$. Thus, $\|\mathbf{w}-\mathbf{w}'\|^2 \le \theta^2 - (2t-1)^2\theta^2 = 4t(1-t)\theta^2$ so that

$$\begin{aligned}
diam\left(J_t^{\mathbf{u}}\right) &= \sup_{\mathbf{w}_1,\mathbf{w}_2\in J_t^{\mathbf{u}}} \|\mathbf{w}_1-\mathbf{w}_2\| \\
&\le \sup_{\mathbf{w}_1,\mathbf{w}_2\in J_t^{\mathbf{u}}} \left(\|\mathbf{w}_1-\mathbf{w}'\| + \|\mathbf{w}_2-\mathbf{w}'\|\right) \\
&\le 4\sqrt{t(1-t)}\theta
\end{aligned}$$

Since $4\sqrt{t(1-t)}\theta \to 0$ if $t\to 0$ or $t\to 1$, and it does not depend on $\mathbf{u}$, we have proven the claim that $diam\left(J_t^{\mathbf{u}}\right)\to 0$ uniformly in $\mathbf{u}$.

Now for any $1>\varepsilon>0$, we can find δ such that when $t\ge 1-\delta$ or $t\le\delta$, $diam\left(J_t^{\mathbf{u}}\right) < \frac{\varepsilon}{3}$. When $\|\mathbf{u}'-\mathbf{u}\| < \varepsilon\|\mathbf{u}_\perp\|/6$, we have

$$\begin{aligned}
&\left\|\frac{(2t-1)\,\theta\mathbf{u}_\perp}{\|\mathbf{u}_\perp\|} - \frac{(2t-1)\,\theta\mathbf{u}'_\perp}{\|\mathbf{u}'_\perp\|}\right\| \\
&= \left\|\frac{(2t-1)\,\theta\left(\left(\|\mathbf{u}'_\perp\|-\|\mathbf{u}_\perp\|\right)\mathbf{u}_\perp + \|\mathbf{u}_\perp\|\left(\mathbf{u}_\perp-\mathbf{u}'_\perp\right)\right)}{\|\mathbf{u}_\perp\|\,\|\mathbf{u}'_\perp\|}\right\| \\
&\le \frac{|2t-1|\,\theta}{\|\mathbf{u}'_\perp\|}\left(\left|\|\mathbf{u}'_\perp\|-\|\mathbf{u}_\perp\|\right| + \|\mathbf{u}_\perp-\mathbf{u}'_\perp\|\right) \\
&\le \frac{\left(\left|\|\mathbf{u}'_\perp\|-\|\mathbf{u}_\perp\|\right| + \|\mathbf{u}_\perp-\mathbf{u}'_\perp\|\right)}{\|\mathbf{u}'_\perp\|}
\end{aligned}$$

$$\begin{aligned}
&\le \frac{2\,\|\mathbf{u}_\perp-\mathbf{u}'_\perp\|}{\|\mathbf{u}_\perp\|} \\
&\le \frac{2\,\|\mathbf{u}-\mathbf{u}'\|}{\|\mathbf{u}_\perp\|} < \frac{2\,\|\mathbf{u}_\perp\|\,\varepsilon}{6\,\|\mathbf{u}_\perp\|} = \varepsilon/3.
\end{aligned}$$

Since $(2t-1)\,\theta\mathbf{u}_\perp/\|\mathbf{u}_\perp\| \in J_t^{\mathbf{u}}$ and likewise, $(2t-1)\,\theta\mathbf{u}'_\perp/\|\mathbf{u}'_\perp\| \in J_t^{\mathbf{u}'}$, we have $D\left(J_t^{\mathbf{u}}, J_t^{\mathbf{u}'}\right) < \varepsilon$. Now suppose that $t\in(\delta, 1-\delta)$. For $\mathbf{w}\in J_t^{\mathbf{u}}$, we have $\langle\mathbf{w},\mathbf{u}_\perp\rangle = (2t-1)\,\|\mathbf{u}_\perp\|\,\theta$. First, we consider the case in which $\langle\mathbf{w},\mathbf{u}'_\perp\rangle \le (2t-1)\,\|\mathbf{u}'_\perp\|\,\theta$. Suppose that $\|\mathbf{u}-\mathbf{u}'\| < \delta_1 = \delta\theta\,\|\mathbf{u}_\perp\|\,\varepsilon/4$. Then,

$$\|\mathbf{u}_\perp-\mathbf{u}'_\perp\| < \delta_1, \left|\langle\mathbf{w},\mathbf{u}_\perp\rangle - \langle\mathbf{w},\mathbf{u}'_\perp\rangle\right| < \delta_1,$$

and

$$\left|(2t-1)\,\|\mathbf{u}_\perp\|\,\theta - (2t-1)\,\|\mathbf{u}'_\perp\|\,\theta\right| < \delta_1.$$

Therefore,

$$\begin{aligned}\langle \mathbf{w}, \mathbf{u}'_\perp \rangle &\geq \langle \mathbf{w}, \mathbf{u}_\perp \rangle - \delta_1 = (2t-1)\,\|\mathbf{u}_\perp\|\,\theta - \delta_1 \\ &\geq (2t-1)\,\|\mathbf{u}'_\perp\|\,\theta - 2\delta_1.\end{aligned}$$

Next, let $\mathbf{w}' = p\mathbf{w} + q\theta\mathbf{u}'_\perp / \|\mathbf{u}'_\perp\|$, where $p = \frac{(2-2t)\theta\|\mathbf{u}'_\perp\|}{\theta\|\mathbf{u}'_\perp\| - \langle \mathbf{w}, \mathbf{u}'_\perp\rangle} \leq 1$ and $q = 1-p$. We then have

$$\begin{aligned}\langle \mathbf{w}', \mathbf{u}'_\perp \rangle &= \frac{(2-2t)\,\theta\,\|\mathbf{u}'_\perp\|}{\theta\,\|\mathbf{u}'_\perp\| - \langle \mathbf{w}, \mathbf{u}'_\perp\rangle}\langle \mathbf{w}, \mathbf{u}'_\perp\rangle + \left(1 - \frac{(2-2t)\,\theta\,\|\mathbf{u}'_\perp\|}{\theta\,\|\mathbf{u}'_\perp\| - \langle \mathbf{w}, \mathbf{u}'_\perp\rangle}\right)\theta\,\|\mathbf{u}'_\perp\| \\ &= \frac{(2-2t)\,\theta\,\|\mathbf{u}'_\perp\|}{\theta\,\|\mathbf{u}'_\perp\| - \langle \mathbf{w}, \mathbf{u}'_\perp\rangle}\langle \mathbf{w}, \mathbf{u}'_\perp\rangle + \theta\,\|\mathbf{u}'_\perp\| - \frac{(2-2t)\,\theta\,\|\mathbf{u}'_\perp\|}{\theta\,\|\mathbf{u}'_\perp\| - \langle \mathbf{w}, \mathbf{u}'_\perp\rangle}\theta\,\|\mathbf{u}'_\perp\| \\ &= (2t-2)\,\theta\,\|\mathbf{u}'_\perp\| + \theta\,\|\mathbf{u}'_\perp\| = (2t-1)\,\theta\,\|\mathbf{u}'_\perp\|\end{aligned}$$

and

$$\begin{aligned}&\|\mathbf{w} - \mathbf{w}'\| \\ &= \|\mathbf{w} - (p\mathbf{w} + q\theta\mathbf{u}'_\perp / \|\mathbf{u}'_\perp\|)\| \leq \|q\mathbf{w}\| + q\theta \\ &\leq 2q \\ &= 2\left(1 - \frac{(2-2t)\,\theta\,\|\mathbf{u}'_\perp\|}{\theta\,\|\mathbf{u}'_\perp\| - \langle \mathbf{w}, \mathbf{u}'_\perp\rangle}\right) = 2\left(\frac{(\theta\,\|\mathbf{u}'_\perp\| - \langle \mathbf{w}, \mathbf{u}'_\perp\rangle) - (2-2t)\,\theta\,\|\mathbf{u}'_\perp\|}{\theta\,\|\mathbf{u}'_\perp\| - \langle \mathbf{w}, \mathbf{u}'_\perp\rangle}\right) \\ &= 2\frac{(2t-1)\,\theta\,\|\mathbf{u}'_\perp\| - \langle \mathbf{w}, \mathbf{u}'_\perp\rangle}{\theta\,\|\mathbf{u}'_\perp\| - \langle \mathbf{w}, \mathbf{u}'_\perp\rangle} \\ &\leq 2\frac{2\delta_1}{\theta\,\|\mathbf{u}'_\perp\| - (2t-1)\,\theta\,\|\mathbf{u}'_\perp\|} \\ &= \frac{4\delta_1}{(2-2t)\,\theta\,\|\mathbf{u}'_\perp\|} \leq \frac{4\delta_1}{2\delta\theta\,\|\mathbf{u}'_\perp\|} \leq \frac{2\delta_1}{\delta\theta\,(\|\mathbf{u}_\perp\| - \delta_1)} \leq \frac{2\delta_1}{\delta\theta\,\|\mathbf{u}_\perp\| - \delta_1} \\ &< \varepsilon\end{aligned}$$

For the case of $\langle \mathbf{w}, \mathbf{u}'_\perp\rangle \geq (2t-1)\,\|\mathbf{u}'_\perp\|\,\theta$, we just need to replace the p above by $\frac{2t\|\mathbf{u}'_\perp\|\theta}{\langle \mathbf{w}, \mathbf{u}'_\perp\rangle + \theta\|\mathbf{u}'_\perp\|}$ and $\mathbf{w}'$ by $p\mathbf{w} - q\theta\mathbf{u}'_\perp / \|\mathbf{u}'_\perp\|$ and repeat the discussion. This shows that $\sup_{\mathbf{w}\in J_t^{\mathbf{u}}} \inf_{\mathbf{w}'\in J_t^{\mathbf{u}'}} \|\mathbf{w} - \mathbf{w}'\| \to 0$. Exchanging the role of $\mathbf{u}$ and $\mathbf{u}'$, we can show that $\sup_{\mathbf{w}'\in J_t^{\mathbf{u}'}} \inf_{\mathbf{w}\in J_t^{\mathbf{u}}} \|\mathbf{w} - \mathbf{w}'\| \to 0$. This finishes the proof.

Lemma 4.21. *Define the function* $\varphi_{\mathbf{u}}(t) = dist\left(\mathbf{0}, \mathbf{Z}\left(I_t^{\mathbf{u}}\right)\right)$. *Then,* $\varphi_{\mathbf{u}}$ *is strictly convex.*

Proof of Lemma 4.21

For any $t_1, t_2 \in [0,1]$, suppose that $\mathbf{z}_i \in \mathbf{Z}\left(I_{t_i}^{\mathbf{u}}\right)$, $\|\mathbf{z}_i\| = dist\left(\mathbf{0}, \mathbf{Z}\left(I_{t_i}^{\mathbf{u}}\right)\right)$, $i = 1, 2$. Then, for $p \in (0,1)$, $q = 1-p$, we have $p\mathbf{z}_1 + q\mathbf{z}_2 \in \mathbf{Z}\left(I_{pt_1+qt_2}^{\mathbf{u}}\right)$. Thus $\varphi_{\mathbf{u}}(pt_1 + qt_2) = dist\left(\mathbf{0}, \mathbf{Z}\left(I_{pt_1+qt_2}^{\mathbf{u}}\right)\right) \leq \|p\mathbf{z}_1 + q\mathbf{z}_2\| \leq p\|\mathbf{z}_1\| + q\|\mathbf{z}_2\|$. Since $\mathbf{0}, \mathbf{z}_1, \mathbf{z}_2$ cannot be on the same line, the last inequality is strict. Hence, $\varphi_{\mathbf{u}}$ is strictly convex.

Lemma 4.22. *Let U, V be two closed, bounded and convex subsets of a Hilbert space $\mathbf{H}$, $\mathbf{h} \in \mathbf{H}$ and $dist(\mathbf{h}, U) = l$. Suppose that $D(U, V) < \varepsilon$ for some $0 < \varepsilon < l/2$. Let $\mathbf{h}_1 \in U, \mathbf{h}_2 \in V$ satisfy $\|\mathbf{h}_1 - \mathbf{h}\| = dist(\mathbf{h}, U)$ and $\|\mathbf{h}_2 - \mathbf{h}\| = dist(\mathbf{h}, V)$. Then, $\|\mathbf{h}_1 - \mathbf{h}_2\| < 4\sqrt{l}\sqrt{\varepsilon}$.*

Proof of Lemma 4.22

Clearly, $dist(\mathbf{h}, V) > dist(\mathbf{h}, U) - \varepsilon$. Since $D(U, V) < \varepsilon$, we find $\mathbf{h}' \in V$ such that $\|\mathbf{h}_1 - \mathbf{h}'\| < \varepsilon$. Thus $\|\mathbf{h} - \mathbf{h}'\| < \|\mathbf{h} - \mathbf{h}_1\| + \varepsilon < \|\mathbf{h} - \mathbf{h}_2\| + 2\varepsilon$. Suppose that $\mathbf{h}' - \mathbf{h}_2 = \mathbf{g}_1 + \mathbf{g}_2$, where $\mathbf{g}_1 = p(\mathbf{h} - \mathbf{h}_2)$, and $\mathbf{g}_2 \perp \mathbf{g}_1$. Now assume that $p > 0$. Let $\mathbf{h}_\alpha = (1-\alpha)\mathbf{h}_2 + \alpha\mathbf{h}'$, and

$$
\begin{aligned}
h(\alpha) &= \|\mathbf{h} - \mathbf{h}_\alpha\|^2 = \|\mathbf{h} - [(1-\alpha)\mathbf{h}_2 + \alpha\mathbf{h}']\|^2 \\
&= \left\|(1-\alpha)(\mathbf{h} - \mathbf{h}_2) + \alpha\left(\mathbf{h} - \mathbf{h}'\right)\right\|^2 \\
&= \|(1-\alpha)(\mathbf{h} - \mathbf{h}_2) + \alpha(\mathbf{h} - \mathbf{h}_2) + \alpha(\mathbf{h}_2 - \mathbf{h}')\|^2 \\
&= \|\mathbf{h} - \mathbf{h}_2 + \alpha(\mathbf{h}_2 - \mathbf{h}')\|^2 \\
&= \|\mathbf{h} - \mathbf{h}_2 - \alpha(\mathbf{g}_1 + \mathbf{g}_2)\|^2 = \|\mathbf{h} - \mathbf{h}_2 - \alpha(p(\mathbf{h} - \mathbf{h}_2) + \mathbf{g}_2)\|^2 \\
&= \|(1-\alpha p)(\mathbf{h} - \mathbf{h}_2) - \alpha\mathbf{g}_2\|^2 \\
&= (1-\alpha p)^2\|\mathbf{h} - \mathbf{h}_2\|^2 + \alpha^2\|\mathbf{g}_2\|^2.
\end{aligned}
$$

We have

$$
h'(\alpha) = 2\|\mathbf{g}_2\|^2\alpha - 2p(1-\alpha p).
$$

which gives $h'(0) = -2p < 0$, which contradicts the fact that $\mathbf{h}_2$ achieves the distance between $\mathbf{h}$ and V. This shows that $p \leq 0$. Thus,

$$
\begin{aligned}
\left\|\mathbf{h} - \mathbf{h}'\right\| &= \sqrt{(\|\mathbf{h} - \mathbf{h}_2\| + \|\mathbf{g}_1\|)^2 + \|\mathbf{g}_2\|^2} \\
&\geq \max\left\{\sqrt{\|\mathbf{h} - \mathbf{h}_2\|^2 + \|\mathbf{g}_2\|^2}, \|\mathbf{h} - \mathbf{h}_2\| + \|\mathbf{g}_1\|\right\},
\end{aligned}
$$

which gives $\|\mathbf{h}-\mathbf{h}_2\|+\|\mathbf{g}_1\|<\|\mathbf{h}-\mathbf{h}_2\|+2\varepsilon$. Thus, $\|\mathbf{g}_1\|<2\varepsilon$ and

$$\sqrt{\|\mathbf{h}-\mathbf{h}_2\|^2+\|\mathbf{g}_2\|^2}<\|\mathbf{h}-\mathbf{h}_2\|+2\varepsilon,$$

or

$$\|\mathbf{g}_2\|^2<4\varepsilon^2+4\varepsilon\|\mathbf{h}-\mathbf{h}_2\|$$

Thus,

$$\begin{aligned}\|\mathbf{h}'-\mathbf{h}_2\| &= \sqrt{\|\mathbf{g}_1\|^2+\|\mathbf{g}_2\|^2}\\ &= \sqrt{4\varepsilon^2+4\varepsilon^2+4\varepsilon\|\mathbf{h}-\mathbf{h}_2\|}\\ &= 2\sqrt{2\varepsilon^2+\varepsilon\|\mathbf{h}-\mathbf{h}_2\|}\\ &\leq 3\sqrt{l}\sqrt{\varepsilon}\end{aligned}$$

Now,

$$\begin{aligned}\|\mathbf{h}_1-\mathbf{h}_2\| &\leq \|\mathbf{h}_1-\mathbf{h}'\|+\|\mathbf{h}'-\mathbf{h}_2\|\\ &< \varepsilon+3\sqrt{l}\sqrt{\varepsilon}\\ &\leq 4\sqrt{l}\sqrt{\varepsilon}.\end{aligned}$$

Lemma 4.23. *For every $\mathbf{x}\in\mathcal{F}-\mathcal{F}$, there are at most two choices for $\mathbf{w}_a(\mathbf{x})$. And if there are two choices, say, $\mathbf{w}_1$ and $\mathbf{w}_2$, then, $\mathbf{w}_1(\mathbf{x})$ and $\mathbf{w}_2(\mathbf{x})$ have different signs.*

Proof of Lemma 4.23

Let $\phi_{\mathbf{u}}(t)=a\alpha+(2t-1)\theta\|\mathbf{u}_\perp\|=\langle\mathbf{w},\mathbf{u}\rangle$ for some $\mathbf{w}\in I_t^{\mathbf{u}}$. First, suppose that $\phi_{\mathbf{Kx}}(t_0)=0$, for some $t_0\in(0,1)$. We have $G_a(\mathbf{x})=\sup_{t\in[0,1]}\sup_{\mathbf{w}\in I_t^{\mathbf{Kx}}}|g_{\mathbf{w}}(\mathbf{x})|$. It is easy to see that $\sup_{\mathbf{w}\in I_t^{\mathbf{Kx}}}|g_{\mathbf{w}}(\mathbf{x})|=|\phi_{\mathbf{Kx}}(t)|/\varphi_{\mathbf{Kx}}(t)$. Since $\phi_{\mathbf{kx}}(t)<0$ for $t\in[0,t_0)$ and $\phi_{\mathbf{kx}}(t)>0$ for $t\in(t_0,1]$, we have

$$G_a(\mathbf{x})=\sup_{t\in[0,1]}\frac{|\phi_{\mathbf{Kx}}(t)|}{\varphi_{\mathbf{Kx}}(t)}=\max\left\{\sup_{t\in[0,t_0]}-\frac{\phi_{\mathbf{Kx}}(t)}{\varphi_{\mathbf{Kx}}(t)},\sup_{t\in[t_0,1]}\frac{\phi_{\mathbf{Kx}}(t)}{\varphi_{\mathbf{Kx}}(t)}\right\}$$

Let k_1 be the smallest positive number k such that the line segment $y=k(t-t_0),t\in[t_0,1]$ intersects $y=\varphi_{\mathbf{Kx}}(t)$. Since, by Lemma 4.21, $\varphi_{\mathbf{Kx}}(t)$ is strictly convex, we know that the line segment intersects $\varphi_{\mathbf{Kx}}(t)$ at only one point, say, $(t_1, \varphi_{\mathbf{Kx}}(t_1))$. Thus, elementary calculations give us $\sup_{t\in[t_0,1]} \frac{\phi_{\mathbf{Kx}}(t)}{\varphi_{\mathbf{Kx}}(t)} = \frac{\phi_{\mathbf{Kx}}(t_1)}{\varphi_{\mathbf{Kx}}(t_1)}$. Similarly, we can find $t_2 \in [0, t_0]$ such that $\sup_{t\in[0,t_0]} -\frac{\phi_{\mathbf{Kx}}(t)}{\varphi_{\mathbf{Kx}}(t)} = -\frac{\phi_{\mathbf{Kx}}(t_2)}{\varphi_{\mathbf{Kx}}(t_2)}$. By convexity, we know that for each t, $\varphi_{\mathbf{Kx}}(t) = \|\mathbf{Zw}\|$ for only one $\mathbf{w} \in I_t^{\mathbf{Kx}}$, and that there can be at most two choices for $\mathbf{w}_a(\mathbf{x})$, they are the $\mathbf{w}_1$ and $\mathbf{w}_2$ that satisfy $\varphi_{\mathbf{Kx}}(t_i) = \|\mathbf{Zw}_i\|, i = 1, 2$. Now suppose that $\phi_{\mathbf{Kx}}(t) \geq 0$ or $\phi_{\mathbf{Kx}}(t) \leq 0$ for all $t \in [0, 1]$. Since these two cases can be discussed the same way, we will only do the former, $\phi_{\mathbf{Kx}}(t) \geq 0$.

If $\|(\mathbf{Kx})_\perp\| = 0$, then, $\phi_{\mathbf{Kx}}(t)$ is constant. There is a unique $t' \in [0, 1]$ that minimizes $\varphi_{\mathbf{Kx}}(t)$, and we find $\mathbf{w}' \in I_{t'}^{\mathbf{Kx}}$ such that $\varphi_{\mathbf{Kx}}(t') = \|\mathbf{Zw}'\|$. Then, $\mathbf{w}'$ is the unique choice for $\mathbf{w}_a(\mathbf{x})$. If $\|(\mathbf{Kx})_\perp\| \neq 0$, then, suppose that $t'' \notin (0, 1)$ satisfies $a\alpha + (2t'' - 1)\theta\|\hat{\mathbf{x}}\| = 0$. Now, let k_2 be the smallest positive number k such that the line segment $k(t - t'') = y$, $t \in [0, 1]$ intersects $y = \varphi_{\mathbf{Kx}}(t)$. Again, by the convexity of $\varphi_{\mathbf{Kx}}(t)$, these two intersect at exactly one point, which is denoted by t_3. Suppose that $\mathbf{w}_3$ is the unique element in $I_{t_3}^{\mathbf{Kx}}$ such that $\varphi_{\mathbf{Kx}}(t_3) = \|\mathbf{Zw}_3\|$. Then, $\mathbf{w}_3$ is the unique choice for $\mathbf{w}_a(\mathbf{x})$. This proves the Lemma.

Lemma 4.24. *Let* $\mathbf{x}_1$, $\mathbf{x}_2$; $\mathbf{x}_1^{(n)}$, $\mathbf{x}_2^{(n)}$, $i = 1, 2$; $n = 1, 2, \ldots$ *be elements of* $\mathcal{F}$ *and* $\mathbf{x}_i^{(n)} \to \mathbf{x}_i$, $\mathbf{Kx}_i^{(n)} \to \mathbf{Kx}_i$ *for* $i = 1, 2$. *Then we can find a subsequences of* $\mathbf{x}_i^{(n_k)}$ *so that* $\mathbf{Zw}_a\left(\mathbf{x}_2^{(n_k)} - \mathbf{x}_1^{(n_k)}\right) \to \mathbf{Zw}_a(\mathbf{x}_2 - \mathbf{x}_1)$.

Proof of Lemma 4.24

We only give the proof for the case that a $t_0 \in (0, 1)$ can be found satisfying $\phi_{\mathbf{Kx}}(t_0) = 0$, where $\phi_{\mathbf{K_x}}$ is defined in the proof of the previous lemma. The proof for other cases is similar, in fact simpler. Let $t_0^{(n)}$ be such that $\phi_{\mathbf{Kx}_n}\left(t_0^{(n)}\right) = 0$. It can be easily shown that $\phi_{\mathbf{Kx}_n}(t) \to \phi_{\mathbf{Kx}}(t)$ if $\mathbf{Kx}_n \to \mathbf{Kx}$, and this convergence is uniform in t for $t \in [0, 1]$. This gives $t_0^{(n)} \to t_0$. Let t_1, t_2 be such that $0 \leq t_1 < t_0 < t_2 \leq 1$ and $\frac{-\phi_{\mathbf{Kx}}(t_1)}{\varphi_{\mathbf{Kx}}(t_1)} = \sup_{t\in[0,t_0]} \frac{-\phi_{\mathbf{Kx}}(t)}{\varphi_{\mathbf{Kx}}(t)}$, $\frac{\phi_{\mathbf{Kx}}(t_2)}{\varphi_{\mathbf{Kx}}(t_2)} = \sup_{t\in[t_0,1]} \frac{\phi_{\mathbf{Kx}}(t)}{\varphi_{\mathbf{Kx}}(t)}$, and $\mathbf{w}_1, \mathbf{w}_2$ are the elements in $I_{t_1}^{\mathbf{Kx}}, I_{t_2}^{\mathbf{Kx}}$ respectively satisfying $\|\mathbf{Zw}_1\| = \varphi_{\mathbf{Kx}}(t_1)$ and $\|\mathbf{Zw}_2\| = \varphi_{\mathbf{Kx}}(t_2)$. For n sufficiently large, $t_0^{(n)} \in (0, 1)$, by discarding the initial terms, we can assume that $t_0^{(n)} \in (0, 1)$ for all n. Similar to t_1, t_2 and $\mathbf{w}_1, \mathbf{w}_2$, we define $0 \leq t_1^{(n)} < t_0^{(n)} < t_2^{(n)} \leq 1$ and $\mathbf{w}_1^{(n)} \in I_{t_1^{(n)}}^{\mathbf{Kx}_n}, \mathbf{w}_2^{(n)} \in I_{t_2^{(n)}}^{\mathbf{Kx}_n}$ satisfying $\frac{-\phi_{\mathbf{Kx}}\left(t_1^{(n)}\right)}{\varphi_{\mathbf{Kx}}\left(t_1^{(n)}\right)} = \sup_{t\in\left[0,t_0^{(n)}\right]} \frac{-\phi_{\mathbf{Kx}}(t)}{\varphi_{\mathbf{Kx}}(t)}$, $\frac{\phi_{\mathbf{Kx}}\left(t_2^{(n)}\right)}{\varphi_{\mathbf{Kx}}\left(t_2^{(n)}\right)} = \sup_{t\in\left[t_0^{(n)},1\right]} \frac{\phi_{\mathbf{Kx}}(t)}{\varphi_{\mathbf{Kx}}(t)}$ and $\left\|\mathbf{Zw}_1^{(n)}\right\| = \varphi_{\mathbf{Kx}_n}\left(t_1^{(n)}\right)$, $\left\|\mathbf{Zw}_2^{(n)}\right\| = \varphi_{\mathbf{Kx}_n}\left(t_2^{(n)}\right)$. By Lemma 4.20 it can be shown that $\varphi_{\mathbf{Kx}_n}(t) \to \varphi_{\mathbf{Kx}}(t)$ uniformly in $t \in [0, 1]$. With the uniform (over t) convergence of $\varphi_{\mathbf{Kx}_n}(t)$ and $\phi_{\mathbf{Kx}_n}(t)$ and the strict convexity of $\varphi_{\mathbf{Kx}_n}(t)$ and $\varphi_{\mathbf{Kx}}(t)$, we can show that $t_1^{(n)} \to t_1$ and $t_2^{(n)} \to t_2$. If $\frac{-\phi_{\mathbf{Kx}}(t_1)}{\varphi_{\mathbf{Kx}}(t_1)} \neq \frac{\phi_{\mathbf{Kx}}(t_2)}{\varphi_{\mathbf{Kx}}(t_2)}$, then, there will be only one choice for $\mathbf{w}_a(\mathbf{Kx})$. If $\frac{-\phi_{\mathbf{Kx}}(t_1)}{\varphi_{\mathbf{Kx}}(t_1)} > \frac{\phi_{\mathbf{Kx}}(t_2)}{\varphi_{\mathbf{Kx}}(t_2)}$, then, $\mathbf{w}_1$ is chosen, and eventually, we will also choose $\mathbf{w}_1^{(n)}$ for $\mathbf{w}_a(\mathbf{Kx}_n)$. By Lemma 4.20 we have $D\left(I_{t_1}^{\mathbf{Kx}}, I_{t_1}^{\mathbf{Kx}_n}\right) \to 0$. Since $t_1^{(n)} \to t_1$, it is easy to show that $D\left(I_{t_1}^{\mathbf{Kx}_n}, I_{t_1^{(n)}}^{\mathbf{Kx}_n}\right) \to 0$. Thus, $D\left(I_{t_1}^{\mathbf{Kx}}, I_{t_1^{(n)}}^{\mathbf{Kx}_n}\right) \to 0$. By Lemma 4.22, we know that $\mathbf{Zw}_1^{(n)} \to \mathbf{Zw}_1$. For the case of $\frac{-\phi_{\mathbf{Kx}}(t_1)}{\varphi_{\mathbf{Kx}}(t_1)} < \frac{\phi_{\mathbf{Kx}}(t_2)}{\varphi_{\mathbf{Kx}}(t_2)}$, the proof is similar. Now suppose that $\frac{-\phi_{\mathbf{Kx}}(t_1)}{\varphi_{\mathbf{Kx}}(t_1)} = \frac{\phi_{\mathbf{Kx}}(t_2)}{\varphi_{\mathbf{Kx}}(t_2)}$. Since we have to pick either $\mathbf{w}_1^{(n)}$ or $\mathbf{w}_2^{(n)}$ for $\mathbf{w}_a(\mathbf{Kx}_n)$, we must pick infinitely many from either of the two sequences $\left(\mathbf{w}_1^{(n)}\right)$ or $\left(\mathbf{w}_2^{(n)}\right)$. Without loss of generality, we assume that we pick an infinite number elements from $\left(\mathbf{w}_1^{(n)}\right)$, and form a subsequence $\left(\mathbf{w}_1^{(n_k)}\right)$. Then, we pick $\mathbf{w}_1$ for $\mathbf{w}_a(\mathbf{Kx})$. Again, we can show that $\mathbf{Zw}_1^{(n_k)} \to \mathbf{Zw}_1$.

Lemma 4.25. *Suppose* $\mathbf{w}, \mathbf{w}_1, \mathbf{w}_2 \ldots \in W_a$ *satisfies* $\mathbf{Zw}_n \xrightarrow{w} \mathbf{Zw}$*, then for every subsequence of* $\mathbf{w}_n^{(1)}$ *of* $\mathbf{w}_n$*, there is a subsequence* $\mathbf{w}_{n_k}^{(1)}$ *such that* $\mathbf{w}_{n_k}^{(1)} \xrightarrow{w} \mathbf{w}$*.*

Proof of Lemma 4.25

Suppose $\mathbf{w}_m^{(1)}$ is a subsequence of $\mathbf{w}_n$ such that no subsequence of $\mathbf{w}_m^{(1)}$ converges weakly to $\mathbf{w}$. Then, there exists a subsequence $\mathbf{w}_{m_i}^{(1)}$ which converges weakly to some $\mathbf{w}' \neq \mathbf{w}$. Thus, $\mathbf{Zw}_{m_i}^{(1)} \xrightarrow{w} \mathbf{Zw}'$. But $\mathbf{Zw}_{m_i}^{(1)} \xrightarrow{w} \mathbf{Zw}$ giving $\mathbf{Zw}_1 = \mathbf{Zw}'$. This means $\mathbf{w}_1 = \mathbf{w}'$, giving a contradiction. Hence the result.

With these results, we can now prove Theorem 4.4 .

Proof of Theorem 4.4

For simplicity, let $\mathbf{x}_i = \mathbf{x}_i(a), i = 1, 2$. Suppose that $\mathbf{x} \in \mathcal{F}$. We want to show that $R_{[\mathbf{x}_1,\mathbf{x}_2]}(\hat{L}_0) \geq E_{\mathbf{x}}\left(\hat{L}_0 - L(\mathbf{x})\right)^2$, where $\hat{L}_0 = \hat{L}(c_0\mathbf{w}_0, d_0)$ We have

$$E_{\mathbf{x}}\left(\hat{L}_0 - L(\mathbf{x})\right)^2 = bias\left(\hat{L}_0, \mathbf{x}\right)^2 + \|c_0\mathbf{Zw}_0\|^2$$

and

$$\begin{aligned} R_{[\mathbf{x}_1,\mathbf{x}_2]}(\hat{L}_0) &= E_{\mathbf{x}_i}\left(\hat{L}_0 - L(\mathbf{x}_i)\right)^2 \\ &= bias\left(\hat{L}_0, \mathbf{x}_i\right)^2 + \|c_0\mathbf{Zw}_0\|^2, i = 1, 2. \end{aligned}$$

Therefore, we only need to show that $\left|bias\left(\hat{L}_0, \mathbf{x}\right)\right| \leq \left|bias\left(\hat{L}_0, \mathbf{x}_i\right)\right|, i = 1, 2$. Since $bias\left(\hat{L}_0, \mathbf{x}_1\right)$ and $bias\left(\hat{L}_0, \mathbf{x}_2\right)$ have opposite signs, with out loss of generality, we may assume that $bias\left(\hat{L}_0, \mathbf{x}_2\right)$

and $bias\left(\hat{L}_0, \mathbf{x}\right)$ have the same sign. We will only deal with the case in which $bias\left(\hat{L}_0, \mathbf{x}_2\right) > 0$. The other case can be proven similarly. Let $\mathbf{x}_h = h\mathbf{x} + (1-h)\,\mathbf{x}_2$. If $bias\left(\hat{L}_0, \mathbf{x}\right) > bias\left(\hat{L}_0, \mathbf{x}_2\right)$, let $bias\left(\hat{L}_0, \mathbf{x}\right) - bias\left(\hat{L}_0, \mathbf{x}_2\right) = \Delta$. We have

$$\begin{aligned}
& bias\left(\hat{L}_0, \mathbf{x}_h\right) - bias\left(\hat{L}_0, \mathbf{x}_2\right) \\
& = \langle c_0\mathbf{w}_0, \mathbf{K}\mathbf{x}_h\rangle + d_0 - L\left(\mathbf{x}_h\right) - \left(\langle c_0\mathbf{w}_0, \mathbf{K}\mathbf{x}_2\rangle + d_0 - L\left(\mathbf{x}_2\right)\right) \\
& = \langle c_0\mathbf{w}_0, \mathbf{K}\left(\mathbf{x}_h - \mathbf{x}_2\right)\rangle - L_1\left(\mathbf{x}_h - \mathbf{x}_2\right)
\end{aligned} \tag{4.18}$$

$$\begin{aligned}
& = \langle c_0\mathbf{w}_0, \mathbf{K}\left(h\mathbf{x} + (1-h)\,\mathbf{x}_2 - \mathbf{x}_2\right)\rangle - L_1\left(h\mathbf{x} + (1-h)\,\mathbf{x}_2 - \mathbf{x}_2\right) \\
& = \langle c_0\mathbf{w}_0, \mathbf{K}\left(h\mathbf{x} - h\mathbf{x}_2\right)\rangle - L_1\left(h\mathbf{x} - h\mathbf{x}_2\right) \\
& = h\left(\langle c_0\mathbf{w}_0, \mathbf{K}\left(\mathbf{x} - \mathbf{x}_2\right)\rangle - L_1\left(\mathbf{x} - \mathbf{x}_2\right)\right) \\
& = h\left(\langle c_0\mathbf{w}_0, \mathbf{K}\mathbf{x}\rangle - \langle c_0\mathbf{w}_0, \mathbf{K}\mathbf{x}_2\rangle - L\left(\mathbf{x}\right) + L\left(\mathbf{x}_2\right)\right) \\
& = h\left(bias\left(\hat{L}_0, \mathbf{x}\right) - bias\left(\hat{L}_0, \mathbf{x}_2\right)\right) = h\Delta
\end{aligned}$$

Our technique will be to show that this cannot be true. Let $\psi\left(s,t\right) = \frac{s^2}{1+t^2}$, $\mathbf{s}_h = \mathbf{x}_h - \mathbf{x}_1$, $\mathbf{s}_0 = \mathbf{x}_2 - \mathbf{x}_1$, $\mathbf{s} = \mathbf{x} - \mathbf{x}_2$, and $\mathbf{w}_h = \mathbf{w}_a\left(\mathbf{s}_h\right)$, $\mathbf{w}_0 = \mathbf{w}_a\left(\mathbf{s}_0\right)$. Since $\mathbf{x}_h \to \mathbf{x}_2$, by the continuity of $\mathbf{K}$ on $\mathcal{F}$, we have $\mathbf{K}\mathbf{x}_h \to \mathbf{K}\mathbf{x}_2$, and thus $\mathbf{K}\mathbf{s}_h \to \mathbf{K}\mathbf{s}_0$ Hence, by Lemma 4.24 we can find $\mathbf{w}_{h_k}$ such that $\mathbf{Z}\mathbf{w}_{h_k} \to \mathbf{Z}\mathbf{w}_0$. Further more, Lemma 4.25 tell us that we can find a subsequence of $\mathbf{w}_{h_k}$, which we also write as $\mathbf{w}_{h_k}$ for simplicity, such that $\mathbf{w}_{h_k} \xrightarrow{w} \mathbf{w}_0$. Thus, $g_{\mathbf{w}_{h_k}}\left(\mathbf{s}_{h_k}\right) = \frac{\left\langle \mathbf{w}_{h_k}, \mathbf{K}\mathbf{s}_{h_k}\right\rangle}{\left\|\mathbf{Z}\mathbf{w}_{h_k}\right\|} \to \frac{\langle \mathbf{w}_0, \mathbf{K}\mathbf{s}_0\rangle}{\|\mathbf{Z}\mathbf{w}_0\|} = g_{\mathbf{w}_0}\left(\mathbf{s}_0\right)$, which means that they eventually will have the same sign. With this and Lemma 4.19, we have $g_{\mathbf{w}_{h_k}}\left(\mathbf{s}_{h_k}\right) - g_{\mathbf{w}_0}\left(\mathbf{s}_0\right) = O\left(h_k\right)$. Also,

$$\begin{aligned}
\left(g_{\mathbf{w}_{h_k}} - g_{\mathbf{w}_0}\right)\left(\mathbf{s}_0\right) &= g_{\mathbf{w}_{h_k}}\left(\mathbf{s}_{h_k}\right) - g_{\mathbf{w}_0}\left(\mathbf{s}_0\right) - g_{\mathbf{w}_{h_k}}\left(\mathbf{s}_{h_k}\right) + g_{\mathbf{w}_{h_k}}\left(\mathbf{s}_0\right) \\
&= O\left(h_k\right) - g_{\mathbf{w}_{h_k}}\left(\mathbf{s}_{h_k} - \mathbf{s}_0\right) \\
&= O\left(h_k\right) - h_k g_{\mathbf{w}_{h_k}}\left(\mathbf{s}\right) = O\left(h_k\right)
\end{aligned}$$

Let $\psi^0 = \psi\left(L_1\left(\mathbf{s}_0/2\right), g_{\mathbf{w}_0}\left(\mathbf{s}_0\right)\right)$, $\psi_1\left(s,t\right) = \frac{\partial\psi}{\partial s}\left(s,t\right)$, $\psi_2\left(s,t\right) = \frac{\partial\psi}{\partial t}\left(s,t\right)$,

$$\psi_1^0 = \psi_1\left(L_1\left(\mathbf{s}_0/2\right), g_{\mathbf{w}_0}\left(\mathbf{s}_0\right)\right),$$

and

$$\psi_2^0 = \psi_2 \left(L_1 \left(\mathbf{s}_0 / 2 \right), g_{\mathbf{w}_0} \left(\mathbf{s}_0 \right) \right).$$

We then have

$$\begin{aligned}
&\psi \left(L_1 \left(\mathbf{s}_{h_k}/2 \right), g_{\mathbf{w}_{h_k}} \left(\mathbf{s}_{h_k} \right) \right) \\
&= \psi \left(L_1 \left(\mathbf{s}_0/2 \right) + L_1 \left(h_k \mathbf{s}/2 \right), g_{\mathbf{w}_{h_k}} \left(\mathbf{s}_{h_k} \right) - g_{\mathbf{w}_0} \left(\mathbf{s}_0 \right) + g_{\mathbf{w}_0} \left(\mathbf{s}_0 \right) \right) \\
&= \psi^0 + \psi_1^0 L_1 \left(h_k \mathbf{s}/2 \right) + \psi_2^0 \left(g_{\mathbf{w}_{h_k}} \left(\mathbf{s}_{h_k} \right) - g_{\mathbf{w}_0} \left(\mathbf{s}_0 \right) \right) + o \left(h_k \right) \\
&= \psi^0 + \psi_1^0 L_1 \left(h_k \mathbf{s}/2 \right) + \psi_2^0 \left(g_{\mathbf{w}_{h_k}} \left(\mathbf{s}_{h_k} \right) - g_{\mathbf{w}_0} \left(\mathbf{s}_{h_k} \right) + g_{\mathbf{w}_0} \left(\mathbf{s}_{h_k} \right) - g_{\mathbf{w}_0} \left(\mathbf{s}_0 \right) \right) + o \left(h_k \right) \\
&= \psi^0 + \psi_1^0 L_1 \left(h_k \mathbf{s}/2 \right) + \psi_2^0 \left(\left(g_{\mathbf{w}_{h_k}} - g_{\mathbf{w}_0} \right) \left(\mathbf{s}_{h_k} \right) + g_{\mathbf{w}_0} \left(\mathbf{s}_{h_k} - \mathbf{s}_0 \right) \right) + o \left(h_k \right) \\
&= \psi^0 + \psi_1^0 L_1 \left(h_k \mathbf{s}/2 \right) + \psi_2^0 \left(\left(g_{\mathbf{w}_{h_k}} - g_{\mathbf{w}_0} \right) \left(\mathbf{s}_{h_k} \right) - \left(g_{\mathbf{w}_{h_k}} - g_{\mathbf{w}_0} \right) \left(\mathbf{s}_0 \right) \right. \\
&\left. + \left(g_{\mathbf{w}_{h_k}} - g_{\mathbf{w}_0} \right) \left(\mathbf{s}_0 \right) + g_{\mathbf{w}_0} \left(h_k \mathbf{s} \right) \right) + o \left(h_k \right) \\
&= \psi^0 + \psi_1^0 L_1 \left(h_k \mathbf{s}/2 \right) + \psi_2^0 \left(\left(g_{\mathbf{w}_{h_k}} - g_{\mathbf{w}_0} \right) \left(\mathbf{s}_{h_k} - \mathbf{s}_0 \right) \right. \\
&\left. + \left(g_{\mathbf{w}_{h_k}} - g_{\mathbf{w}_0} \right) \left(\mathbf{s}_0 \right) + g_{\mathbf{w}_0} \left(h_k \mathbf{s} \right) \right) + o \left(h_k \right) \\
&= \psi^0 + \psi_1^0 L_1 \left(h_k \mathbf{s}/2 \right) + \psi_2^0 \left(\left(g_{\mathbf{w}_{h_k}} - g_{\mathbf{w}_0} \right) \left(h_k \mathbf{s} \right) \right. \\
&\left. + \left(g_{\mathbf{w}_{h_k}} - g_{\mathbf{w}_0} \right) \left(\mathbf{s}_0 \right) + g_{\mathbf{w}_0} \left(h_k \mathbf{s} \right) \right) + o \left(h_k \right) \\
&= \psi^0 + \psi_2^0 \left(g_{\mathbf{w}_{h_k}} - g_{\mathbf{w}_0} \right) \left(\mathbf{s}_0 \right) + \psi_1^0 L_1 \left(h_k \mathbf{s}/2 \right) \\
&+ h_k \psi_2^0 \left(g_{\mathbf{w}_{h_k}} - g_{\mathbf{w}_0} \right) \left(\mathbf{s} \right) + \psi_2^0 g_{\mathbf{w}_0} \left(h_k \mathbf{s} \right) + o \left(h_k \right). \\
&= \psi \left(L_1 \left(\mathbf{s}_0/2 \right), g_{\mathbf{w}_{h_k}} \left(\mathbf{s}_0 \right) \right) + \psi_1^0 L_1 \left(h_k \mathbf{s}/2 \right) + \psi_2^0 g_{\mathbf{w}_0} \left(h_k \mathbf{s} \right) + o \left(h_k \right)
\end{aligned}$$

By the definition of $\mathbf{x}_2 (a), \mathbf{x}_1 (a)$, and $\mathbf{w}_0$, we know that

$$\psi \left(L_1 \left(\mathbf{s}_{h_k}/2 \right), g_{\mathbf{w}_{h_k}} \left(\mathbf{s}_{h_k} \right) \right) \leq \psi^0 \leq \psi \left(L_1 \left(\mathbf{s}_0/2 \right), g_{\mathbf{w}_{h_k}} \left(\mathbf{s}_0 \right) \right).$$

Thus, $\psi_1^0 L_1 \left(h_k \mathbf{s}/2 \right) + \psi_2^0 g_{\mathbf{w}_0} \left(h_k \mathbf{s} \right) + o \left(h_k \right) \leq 0$. However,

$$\begin{aligned}
&\psi_1^0 L_1 \left(h_k \mathbf{s}/2 \right) + \psi_2^0 g_{\mathbf{w}_0} \left(h_k \mathbf{s} \right) \\
&= \frac{2 L_1 \left(\mathbf{s}_0/2 \right)}{1 + g_{\mathbf{w}0} \left(\mathbf{s}_0 \right)^2} L_1 \left(h_k \mathbf{s}/2 \right) - \frac{2 L_1 \left(\mathbf{s}_0/2 \right)^2 g_{\mathbf{w}_0} \left(\mathbf{s}_0 \right)}{\left(1 + g_{\mathbf{w}_0} \left(\mathbf{s}_0 \right)^2 \right)^2} g_{\mathbf{w}_0} \left(h_k \mathbf{s} \right)
\end{aligned}$$

$$= \frac{2L_1(\mathbf{s}_0/2)}{1+g_{\mathbf{w}0}(\mathbf{s}_0)^2}\left(L_1(h_k\mathbf{s}/2) - c_0\langle \mathbf{w}_0, h_k\mathbf{Ks}/2\rangle\right).$$

We have $bias\left(\hat{L}_0, \mathbf{x}_2\right) > 0$. Also

$$\left|E_{\mathbf{x}_2}\left(\hat{L}_0\right) - L\left((\mathbf{x}_1+\mathbf{x}_2)/2\right)\right| < \left|L(\mathbf{x}_2) - L\left((\mathbf{x}_1+\mathbf{x}_2)/2\right)\right|$$

and they have the same sign. Therefore, we must have

$$L(\mathbf{x}_2) - L\left((\mathbf{x}_1+\mathbf{x}_2)/2\right) = L_1(\mathbf{s}_0/2) < 0.$$

Hence,

$$\frac{2L_1(\mathbf{s}_0/2)}{1+g_{\mathbf{w}0}(\mathbf{s}_0)^2}\left(L_1(h_k\mathbf{s}/2) - c_0\langle \mathbf{w}_0, h_k\mathbf{Ks}/2\rangle\right) + o(h_k) \le 0,$$

which gives

$$L_1(h_k\mathbf{s}/2) - c_0\langle \mathbf{w}_0, h_k\mathbf{Ks}/2\rangle + \frac{1+g_{\mathbf{w}0}(\mathbf{s}_0)^2}{2L_1(\mathbf{s}_0/2)} o(h_k) \ge 0,$$

or

$$L_1(h_k\mathbf{s}) - c_0\langle \mathbf{w}_0, h_k\mathbf{Ks}\rangle + o(h_k) \ge 0.$$

But by (4.18), we know that $L_1(h_k\mathbf{s}) - c_0\langle \mathbf{w}_0, h_k\mathbf{Ks}\rangle = -h_k\Delta$, giving a contradiction to the last inequality above. Hence proving the desired result.

Proof of Theorem 4.5

Showing that the closedness condition can be removed is done exactly the same as the corresponding part in the proof of Theorem 2 in Donoho (1994). We will now show that the boundedness constraint can also be dropped. Let $\mathcal{F}_k = \{\mathbf{x} \in \mathcal{F} : \|\mathbf{x}\| \le k\}$. It is easy to see that $\rho_a(\mathcal{F}_k) \uparrow \rho_a(\mathcal{F})$. Clearly,

$$\inf_{\mathbf{w}\in W_a, c, d\in\mathbb{R}} R_{\mathcal{F}}\left(\hat{L}\right) \ge \rho_a(\mathcal{F}).$$

We can assume that $\rho_a(\mathcal{F}) < \infty$, for, if not, the result is trivial. Also, for a non-trivial setting, $\rho_a(\mathcal{F}_k) > 0$ for sufficiently large k, so that by ignoring the first few terms, we can assume that $\rho_a(\mathcal{F}_k) > 0$ for all k. Since $\mathcal{F}_k \ne \phi$ for sufficiently large k, we will assume this to be true for all k. By Theorem 4.4, we can find $\mathbf{w}_k \in W_a$ and $c_k, d_k \in \mathbb{R}$ such that the estimator $\hat{L}_k = \hat{L}(c_k\mathbf{w}_k, d_k)$ satisfies $R_{\mathcal{F}_k}\left(\hat{L}_k\right) = \inf_{\mathbf{w}\in W_a, c, d\in\mathbb{R}} R_{\mathcal{F}_k}\left(\hat{L}(c\mathbf{w}, d)\right) = \rho_a(\mathcal{F}_k)$. Now, we will show that $\sup_k |c_k| < \infty$. In fact,

$$\rho_a(\mathcal{F}_k) = \inf_{\mathbf{w}\in W_a, c,d\in\mathbb{R}} R_{\mathcal{F}_k}\left(\hat{L}\right) = R_{\mathcal{F}_k}\left(\hat{L}_k\right)$$
$$= \sup_{\mathbf{x}\in\mathcal{F}_k} bias\left(\hat{L}_k, \mathbf{x}\right)^2 + \|c_k\mathbf{Z}\mathbf{w}_k\|^2$$
$$\geq \|c_k\mathbf{Z}\mathbf{w}_k\|^2,$$

so that

$$\sup_k c_k^2 \|\mathbf{Z}\mathbf{w}_k\|^2 \leq \sup_k \rho_a(\mathcal{F}_k) < \infty.$$

Since $\inf_k \|\mathbf{Z}\mathbf{w}_k\| \geq \inf_{\mathbf{w}\in W_a}\|\mathbf{Z}\mathbf{w}\| > 0$, we have $\sup_k c_k^2 < \infty$. Next, we will show that $\sup_k |d_k| < \infty$. From the above discussion, we observe that

$$\infty > \sup_k \rho_a(\mathcal{F}_k) \geq \sup_k \sup_{\mathbf{x}\in\mathcal{F}_k} bias\left(\hat{L}_k, \mathbf{x}\right)^2$$
$$\geq \sup_k bias\left(\hat{L}_k, \mathbf{x}_0\right)^2$$
$$= \sup_k \left(\langle c_k\mathbf{w}_k, \mathbf{K}\mathbf{x}_0\rangle + d_k - L(\mathbf{x}_0)\right)^2,$$

where $\mathbf{x}_0$ is any vector in $\mathcal{F}_1$. Since $\sup_k |\langle c_k\mathbf{w}_k, \mathbf{K}\mathbf{x}_0\rangle - L(\mathbf{x}_0)| < \infty$, we must have $\sup_k |d_k| < \infty$. Thus, we can find a subsequences which for simplicity, we continue to write as c_k, d_k, and $\mathbf{w}_k$, such that $c_k \to c_0, d_k \to d_0$, and $\mathbf{w}_k \to \mathbf{w}_0 \in W_a$ weakly. We claim that $\hat{L}_0 = \hat{L}(c_k\mathbf{w}_k, d_k)$ is the affine estimator that we are looking for. If fact, we have $\|c_0\mathbf{Z}\mathbf{w}_0\| \leq \liminf_k \|c_k\mathbf{Z}\mathbf{w}_k\| \leq \limsup_k \|c_k\mathbf{Z}\mathbf{w}_k\|$, and

$$bias\left(\hat{L}_0, \mathbf{x}\right) = |\langle c_0\mathbf{w}_0, \mathbf{K}\mathbf{x}\rangle + d_0 - L(\mathbf{x})| = \lim_k |\langle c_k\mathbf{w}_k, \mathbf{K}\mathbf{x}\rangle + d_k - L(\mathbf{x})|$$
$$= \lim_k bias\left(\hat{L}_k, \mathbf{x}\right)$$
$$\leq \liminf \sup_{\mathbf{x}'\in\mathcal{F}_k} bias\left(\hat{L}_k, \mathbf{x}'\right), \forall \mathbf{x}\in\mathcal{F}.$$

Here we used the fact that $\mathbf{x}\in\mathcal{F}_k$ for k large enough. Hence,

$$\sup_{\mathbf{x}\in\mathcal{F}} bias\left(\hat{L}_0, \mathbf{x}\right)^2 \leq \liminf \sup_{\mathbf{x}\in\mathcal{F}_k} bias\left(\hat{L}_k, \mathbf{x}\right)^2.$$

Thus, we have

$$\begin{aligned}\inf_{\mathbf{w}\in W_a,c,d\in\mathbb{R}} R_{\mathcal{F}}\left(\hat{L}\right) &\leq \sup_{\mathbf{x}\in\mathcal{F}} E_{\mathbf{x}}\left(\hat{L}_0 - L(\mathbf{x})\right)^2\\
&= \sup_{\mathbf{x}\in\mathcal{F}} bias\left(\hat{L}_0,\mathbf{x}\right)^2 + \|\mathbf{Z}\mathbf{w}_0\|^2\\
&\leq \liminf \sup_{\mathbf{x}\in\mathcal{F}_k}\left(bias\left(\hat{L}_k,\mathbf{x}\right)^2 + \|\mathbf{Z}\mathbf{w}_k\|^2\right)\\
&= \lim_k \rho_a(\mathcal{F}_k) \leq \inf_{\mathbf{w}\in W_a,c,d\in\mathbb{R}} R_{\mathcal{F}}\left(\hat{L}\right)\end{aligned}$$

proving the desired result.

Proof of Theorem 4.6

Clearly $M = \inf_{\hat{L}\text{ affine}} R_{\mathcal{F}}\left(\hat{L}\right) \leq \lim_{a\to 0}\rho_a(\mathcal{F})$, since $\inf_{\hat{L}\text{ affine}} R_{\mathcal{F}}\left(\hat{L}\right) \leq \rho_a(\mathcal{F})$ for all $a \in (0,1)$. Suppose that $\hat{L}_0 = \hat{L}(\mathbf{w}_0, d_0)$ is an affine estimator such that $R_{\mathcal{F}}\left(\hat{L}_0\right) < \infty$. We will use $\frac{\mathbf{w}_0}{\|\mathbf{w}_0\|}$ as the $\mathbf{v}$ in (4.6). For any $\varepsilon \in (0,1)$, we find an affine estimator $\hat{L}_\varepsilon = \hat{L}(\mathbf{w}_\varepsilon, d_\varepsilon)$ such that $R_{\mathcal{F}}\left(\hat{L}_\varepsilon\right) < M + \varepsilon$. If $\mathbf{w}_\varepsilon \notin \ker(\mathbf{v})$, then, there exists $a_\varepsilon > 0$ and c_ε such that $\mathbf{w}_\varepsilon \in c_\varepsilon W_{a_\varepsilon}$. Then, $\rho_{a_\varepsilon}(\mathcal{F}) \leq R_{\mathcal{F}}\left(\hat{L}_\varepsilon\right) < M + \varepsilon$. If $\mathbf{w}_\varepsilon \in \ker(\mathbf{v})$, then, Let $\hat{L}' = p\hat{L}_0 + q\hat{L}_\varepsilon$ where $0 < p < \varepsilon \Big/ \left(M + \varepsilon + R_{\mathcal{F}}\left(\hat{L}_0\right)\right)$ and $q = 1 - p$. For any $\mathbf{x} \in \mathcal{F}$, we have

$$\begin{aligned}&E_{\mathbf{x}}\left(\hat{L}' - L(\mathbf{x})\right)^2\\
&= E_{\mathbf{x}}\left(p\hat{L}_0 - pL(\mathbf{x}) + q\hat{L}_\varepsilon - qL(\mathbf{x})\right)^2\\
&= E_{\mathbf{x}}\left[\left(p\hat{L}_0 - pL(\mathbf{x})\right)^2 + 2\left(p\hat{L}_0 - pL(\mathbf{x})\right)\left(q\hat{L}_\varepsilon - qL(\mathbf{x})\right) + \left(q\hat{L}_\varepsilon - qL(\mathbf{x})\right)^2\right]\\
&= p^2 E_{\mathbf{x}}\left(\hat{L}_0 - L(\mathbf{x})\right)^2 + q^2 E_{\mathbf{x}}\left(\hat{L}_\varepsilon - L(\mathbf{x})\right)^2\\
&+ 2pqE_{\mathbf{x}}\left[\left(\hat{L}_0 - L(\mathbf{x})\right)\left(\hat{L}_\varepsilon - L(\mathbf{x})\right)\right]\\
&\leq p^2 E_{\mathbf{x}}\left(\hat{L}_0 - L(\mathbf{x})\right)^2 + q^2 E_{\mathbf{x}}\left(\hat{L}_\varepsilon - L(\mathbf{x})\right)^2\\
&+ pq\left(E_{\mathbf{x}}\left(\hat{L}_0 - L(\mathbf{x})\right)^2 + E_{\mathbf{x}}\left(\hat{L}_\varepsilon - L(\mathbf{x})\right)^2\right)\\
&\leq \varepsilon + q^2 E_{\mathbf{x}}\left(\hat{L}_\varepsilon - L(\mathbf{x})\right)^2 + \varepsilon \leq 2\varepsilon + E_{\mathbf{x}}\left(\hat{L}_\varepsilon - L(\mathbf{x})\right)^2 < M + 3\varepsilon.\end{aligned}$$

Since $R_{\mathcal{F}}\left(\hat{L}'\right) \geq \rho_a(\mathcal{F})$ for some $a > 0$, we have $\lim_{a\to 0}\rho_a(\mathcal{F}) \leq M + 3\epsilon$. This finishes the proof.

Proof of Lemma 4.7

Since $\rho_0(\mathbf{x}_2 - \mathbf{x}_1) \leq \rho_a(\mathbf{x}_2 - \mathbf{x}_1)$ for any $0 < a < 1$, it is clear that

$$\lim_{a \to 0} \rho_a (\mathcal{F}) \geq \rho_0 (\mathcal{F})$$

Suppose that there is a positive ε such that $\lim_{a\to 0} \rho_a (\mathcal{F}) > \rho_0 (\mathcal{F}) + \varepsilon$. Then there exist sequences $\mathbf{x}_1^n$, $\mathbf{x}_2^n \in \mathcal{F}$, and $a_n \downarrow 0$ such that $\rho_{a_n} (\mathbf{x}_2^n - \mathbf{x}_1^n) > \rho_0 (\mathcal{F}) + \varepsilon, n = 1, 2, \ldots$. By passing to a subsequence, we can assume that $\mathbf{x}_1^n \xrightarrow{w} \tilde{\mathbf{x}}_1$ and $\mathbf{x}_1^n \xrightarrow{w} \tilde{\mathbf{x}}_2$, where $\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2 \in \mathcal{F}$. Clearly, $\rho_0 (\tilde{\mathbf{x}}_2 - \tilde{\mathbf{x}}_1) \leq \rho_0 (\mathcal{F})$. We can find $\mathbf{w}_0 \in S(\mathbf{U})$ such that $\frac{L_1((\tilde{\mathbf{x}}_2 - \tilde{\mathbf{x}}_1)/2)^2}{1+g_{\mathbf{w}_0}(\tilde{\mathbf{x}}_2 - \tilde{\mathbf{x}}_1)^2} < \rho_0 (\tilde{\mathbf{x}}_2 - \tilde{\mathbf{x}}_1) + \varepsilon/3$. With Assumption 4.5 and Lemma 4.2, we have $L_1 (\mathbf{x}_2^n - \mathbf{x}_1^n)^2 \to L_1 (\tilde{\mathbf{x}}_2 - \tilde{\mathbf{x}}_1)^2$. Also, $\langle \mathbf{w}_0, \mathbf{K} (\mathbf{x}_2^n - \mathbf{x}_1^n) \rangle^2 \to \langle \mathbf{w}_0, \mathbf{K} (\tilde{\mathbf{x}}_2 - \tilde{\mathbf{x}}_1) \rangle^2$ so that we can find N such that for all $n \geq N$,

$$\frac{L_1((\mathbf{x}_2^n - \mathbf{x}_1^n)/2)^2}{1 + g_{\mathbf{w}_0} (\mathbf{x}_2^n - \mathbf{x}_1^n)^2} \leq \frac{L_1((\tilde{\mathbf{x}}_2 - \tilde{\mathbf{x}}_1)/2)^2}{1 + g_{\mathbf{w}_0} (\tilde{\mathbf{x}}_2 - \tilde{\mathbf{x}}_1)^2} + \varepsilon/3.$$

Now, suppose that $\mathbf{w}_0$ is not in $\ker(\mathbf{v})$. We can find N' such that when $n \geq N'$, $c\mathbf{w}_0 \in W_{a_n}$ for some $c \in [-1, 1]$. Thus, for $n \geq \max\{N, N'\}$ we have that

$$\begin{aligned}\rho_{a_n} (\mathbf{x}_2^n - \mathbf{x}_1^n) &\leq \frac{L_1((\mathbf{x}_2^n - \mathbf{x}_1^n)/2)^2}{1 + g_{\mathbf{w}_0} (\mathbf{x}_2^n - \mathbf{x}_1^n)^2} \leq \frac{L_1((\tilde{\mathbf{x}}_2 - \tilde{\mathbf{x}}_1)/2)^2}{1 + g_{\mathbf{w}_0} (\tilde{\mathbf{x}}_2 - \tilde{\mathbf{x}}_1)^2} + \varepsilon/3 \\ &\leq \rho_0 (\tilde{\mathbf{x}}_2 - \tilde{\mathbf{x}}_1) + 2\varepsilon/3,\end{aligned}$$

which is impossible. This shows that $\mathbf{w}_0 \in \ker(\mathbf{v})$. It is not hard to prove that we can find $\mathbf{w}_n \in W_{a_n}$ such that the sequence $\{\mathbf{w}_n\}$ converges to $\mathbf{w}_0$. Since $\mathbf{K}(\mathcal{F} - \mathcal{F})$ is bounded and $\mathbf{Z}$ is a bounded operator, we have $g_{\mathbf{w}_n} (\mathbf{x}_2 - \mathbf{x}_1) \to g_{\mathbf{w}_0} (\mathbf{x}_2 - \mathbf{x}_1)$ for any $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{F}$ and this convergence is uniform. Also, $\{L_1 (\mathbf{x}_2^n - \mathbf{x}_1^n) : n = 1, 2, \ldots\}$ is bounded. Thus there exists N'' such that when $n \geq N''$, we can find $\mathbf{w}_n \in W_{a_n}$ close enough to $\mathbf{w}_0$ such that

$$\frac{L_1((\mathbf{x}_2^n - \mathbf{x}_1^n)/2)^2}{1 + g_{\mathbf{w}_n} (\mathbf{x}_2^n - \mathbf{x}_1^n)^2} \leq \frac{L_1((\mathbf{x}_2^n - \mathbf{x}_1^n)/2)^2}{1 + g_{\mathbf{w}_0} (\mathbf{x}_2^n - \mathbf{x}_1^n)^2} + \varepsilon/3.$$

Hence, for $n \geq \max\{N, N''\}$, we have

$$\begin{aligned}\rho_{a_n} (\mathbf{x}_2^n - \mathbf{x}_1^n) &\leq \frac{L_1((\mathbf{x}_2^n - \mathbf{x}_1^n)/2)^2}{1 + g_{\mathbf{w}_n} (\mathbf{x}_2^n - \mathbf{x}_1^n)^2} \\ &\leq \frac{L_1((\mathbf{x}_2^n - \mathbf{x}_1^n)/2)^2}{1 + g_{\mathbf{w}_0} (\mathbf{x}_2^n - \mathbf{x}_1^n)^2} + \varepsilon/3\end{aligned}$$

$$
\begin{aligned}
&\leq \frac{L_1((\tilde{\mathbf{x}}_2 - \tilde{\mathbf{x}}_1)/2)^2}{1 + g_{\mathbf{w}_0}\left(\tilde{\mathbf{x}}_2 - \tilde{\mathbf{x}}_1\right)^2} + 2\varepsilon/3 \\
&< \rho_a\left(\tilde{\mathbf{x}}_2 - \tilde{\mathbf{x}}_1\right) + \varepsilon \\
&\leq \rho_0\left(\mathcal{F}\right) + \varepsilon,
\end{aligned}
$$

which contradicts the assumption. Thus, $\lim_{a\to 0}\rho_a\left(\mathcal{F}\right) \leq \rho_0\left(\mathcal{F}\right) + \varepsilon$. Since ε is arbitrary, we get

$$
\lim_{a\to 0}\rho_a\left(\mathcal{F}\right) \leq \rho_0\left(\mathcal{F}\right).
$$

completing the proof

Proof of Lemma 4.8

Let $\mathcal{F}_s = \mathcal{F} \cap B_s\left(\mathbf{X}\right)$ where $B_s\left(\mathbf{X}\right) = \left\{\mathbf{x} : \|\mathbf{x}\| \leq s\right\}$. Suppose that $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{F}$. Then, since $\mathcal{F}$ is symmetric, $-\mathbf{x}_1, -\mathbf{x}_2$ are also in $\mathcal{F}$. Let $\mathbf{x}_1' = \left(\mathbf{x}_1 - \mathbf{x}_2\right)/2$, and $\mathbf{x}_2' = \left(\mathbf{x}_2 - \mathbf{x}_1\right)/2$. We see that $\mathbf{x}_1', \mathbf{x}_2' \in \mathcal{F}$, and $\mathbf{x}_2' - \mathbf{x}_1' = \mathbf{x}_2 - \mathbf{x}_1$.

Therefore, $\rho_a\left(\mathbf{x}_2 - \mathbf{x}_1\right) = \rho_a\left(\mathbf{x}_2' - \mathbf{x}_1'\right) \leq \rho_a\left(\mathcal{F}_{\|(\mathbf{x}_2 - \mathbf{x}_1)/2\|}\right)$. Again, we only need to prove that $\lim_{a\to 0}\rho_a\left(\mathcal{F}\right) \leq \rho_0\left(\mathcal{F}\right)$. If not, we find $\delta > 0$ and a sequence a_n such that $a_n \downarrow 0$ and

$$
\lim_{n\to\infty}\rho_{a_n}\left(\mathcal{F}\right) = \tau\left(\mathcal{F}\right) \geq \rho_0\left(\mathcal{F}\right) + \delta. \tag{4.19}
$$

Let$M > \sqrt{\frac{64\|\mathbf{Z}\|^2\rho_0(\mathcal{F})}{\delta} + 32\left\|\mathbf{Z}\right\|^2}$, and $M' > \chi\left(M\right)$. By Lemma 4.7,

$$
\lim_{n\to\infty}\rho_{a_n}\left(\mathcal{F}_{M'}\right) = \rho_0\left(\mathcal{F}_{M'}\right).
$$

Hence, for $\varepsilon \in (0, \delta/4)$we find m' such that

$$
\begin{aligned}
\rho_{a_{m'}}\left(\mathcal{F}_{M'}\right) &< \rho_0\left(\mathcal{F}_{M'}\right) + \varepsilon \\
&\leq \rho_0\left(\mathcal{F}\right) + \varepsilon
\end{aligned} \tag{4.20}
$$

and $a_{m'} < 1/4$. Equation (4.19) gives us $\rho_{a_{m'}}(\mathcal{F}) \geq \rho_0(\mathcal{F}) + \delta$. Thus, there exists a pair $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{F}$ such that

$$\begin{aligned} \rho_{a_{m'}}(\mathbf{x}_2 - \mathbf{x}_1) &> \rho_{a_{m'}}(\mathcal{F}) - \varepsilon \\ &\geq \rho_0(\mathcal{F}) + \delta - \varepsilon. \end{aligned} \tag{4.21}$$

Let $\mathbf{x}_1' = (\mathbf{x}_1 - \mathbf{x}_2)/2$, and $\mathbf{x}_2' = (\mathbf{x}_2 - \mathbf{x}_1)/2$. If $\|(\mathbf{x}_2 - \mathbf{x}_1)/2\| \leq M'$, then $\mathbf{x}_1', \mathbf{x}_2' \in \mathcal{F}_{M'}$, and this gives

$$\begin{aligned} \rho_0(\mathcal{F}) + \delta - \varepsilon &< \rho_{a_{m'}}(\mathbf{x}_2 - \mathbf{x}_1) \\ &= \rho_{a_{m'}}(\mathbf{x}_2' - \mathbf{x}_1') \\ &\leq \rho_{a_{m'}}(\mathcal{F}_{M'}) \\ &< \rho_0(\mathcal{F}) + \varepsilon, \end{aligned}$$

which is a contradiction. Hence, $\|(\mathbf{x}_2 - \mathbf{x}_1)/2\| > M'$.

Now, let $\hat{\mathbf{x}}_i = \frac{M'}{\|(\mathbf{x}_2 - \mathbf{x}_1)/2\|}\mathbf{x}_i', i = 1, 2$. We have $\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2 \in \mathcal{F}_{M'}$. Let $\mathbf{w}_0 = \frac{\mathbf{K}(\hat{\mathbf{x}}_2 - \hat{\mathbf{x}}_1)}{\|\mathbf{K}(\hat{\mathbf{x}}_2 - \hat{\mathbf{x}}_1)\|}$. Then, either $c\mathbf{w}_0 \in W_{a_{m'}}$ for some $c \in [-1, 1]$ or there exists $\mathbf{w}' \in W_{a_{m'}}$ such that $\|\mathbf{w}' - c\mathbf{w}_0\| < 2a_{m'}$ for $c = 1$ or -1. In either case, we can find a $\mathbf{w}'$ such that $\|\mathbf{w}' - c\mathbf{w}_0\| / \|c\mathbf{w}_0\| < 2a_{m'}$, with $0 \neq c \in [-1, 1]$. Thus,

$$\begin{aligned} |g_{\mathbf{w}'}(\hat{\mathbf{x}}_2 - \hat{\mathbf{x}}_1)| &= \left| \frac{\langle \mathbf{w}', \mathbf{K}(\hat{\mathbf{x}}_2 - \hat{\mathbf{x}}_1)/2 \rangle}{\|\mathbf{Z}\mathbf{w}'\|} \right| \\ &= \left| \frac{\langle \mathbf{w}' - c\mathbf{w}_0 + c\mathbf{w}_0, \mathbf{K}(\hat{\mathbf{x}}_2 - \hat{\mathbf{x}}_1)/2 \rangle}{\|\mathbf{Z}(\mathbf{w}' - c\mathbf{w}_0 + c\mathbf{w}_0)\|} \right| \\ &\geq \frac{|\langle \mathbf{w}' - c\mathbf{w}_0, \mathbf{K}(\hat{\mathbf{x}}_2 - \hat{\mathbf{x}}_1)/2 \rangle + \langle c\mathbf{w}_0, \mathbf{K}(\hat{\mathbf{x}}_2 - \hat{\mathbf{x}}_1)/2 \rangle|}{\|\mathbf{Z}(\mathbf{w}' - c\mathbf{w}_0)\| + \|c\mathbf{Z}\mathbf{w}_0\|} \\ &\geq \frac{|\langle c\mathbf{w}_0, \mathbf{K}(\hat{\mathbf{x}}_2 - \hat{\mathbf{x}}_1)/2 \rangle| - |\langle \mathbf{w}' - c\mathbf{w}_0, \mathbf{K}(\hat{\mathbf{x}}_2 - \hat{\mathbf{x}}_1)/2 \rangle|}{\|\mathbf{Z}(\mathbf{w}' - c\mathbf{w}_0)\| + \|c\mathbf{Z}\mathbf{w}_0\|} \\ &\geq \frac{|c| \|\mathbf{K}(\hat{\mathbf{x}}_2 - \hat{\mathbf{x}}_1)\|/2 - |c| a_{m'} \|\mathbf{K}(\hat{\mathbf{x}}_2 - \hat{\mathbf{x}}_1)\|}{\|\mathbf{Z}(\mathbf{w}' - c\mathbf{w}_0)\| + \|c\mathbf{Z}\mathbf{w}_0\|} \\ &> \frac{c \|\mathbf{K}(\hat{\mathbf{x}}_2 - \hat{\mathbf{x}}_1)\|/4}{2c\|\mathbf{Z}\|} \\ &= \frac{\|\mathbf{K}(\hat{\mathbf{x}}_2 - \hat{\mathbf{x}}_1)\|}{8\|\mathbf{Z}\|} \\ &\geq \frac{M}{4\|\mathbf{Z}\|}, \end{aligned}$$

and

$$|g_{\mathbf{w}'}(\mathbf{x}_2' - \mathbf{x}_1')| = \frac{\|(\mathbf{x}_2 - \mathbf{x}_1)/2\|}{M'} |g_{\mathbf{w}'}(\hat{\mathbf{x}}_2 - \hat{\mathbf{x}}_1)| \geq |g_{\mathbf{w}'}(\hat{\mathbf{x}}_2 - \hat{\mathbf{x}}_1)|$$

It can be shown that, $\mathbf{w}_{a_{m'}}(\hat{\mathbf{x}}_2 - \hat{\mathbf{x}}_1)$ and $\mathbf{w}_{a_{m'}}(\mathbf{x}_2' - \mathbf{x}_1')$ can be chosen to be the same, which is denoted $\tilde{\mathbf{w}}$. Therefore,

$$|g_{\tilde{\mathbf{w}}}(\hat{\mathbf{x}}_2 - \hat{\mathbf{x}}_1)| \geq |g_{\mathbf{w}'}(\hat{\mathbf{x}}_2 - \hat{\mathbf{x}}_1)| \geq \frac{M}{4\|\mathbf{Z}\|}$$

and

$$|g_{\tilde{\mathbf{w}}}(\mathbf{x}_2' - \mathbf{x}_1')| \geq |g_{\mathbf{w}'}(\mathbf{x}_2' - \mathbf{x}_1')| \geq \frac{M}{4\|\mathbf{Z}\|},$$

giving

$$\begin{aligned}
&\left|\rho_{a_{m'}}(\mathbf{x}_2' - \mathbf{x}_1') - \frac{[L_1((\mathbf{x}_2' - \mathbf{x}_1')/2)]^2}{g_{\tilde{\mathbf{w}}}(\mathbf{x}_2' - \mathbf{x}_1')^2}\right| \\
&= \left|\frac{[L_1((\mathbf{x}_2' - \mathbf{x}_1')/2)]^2}{1 + g_{\tilde{\mathbf{w}}}(\mathbf{x}_2' - \mathbf{x}_1')^2} - \frac{[L_1((\mathbf{x}_2' - \mathbf{x}_1')/2)]^2}{[g_{\tilde{\mathbf{w}}}(\mathbf{x}_2' - \mathbf{x}_1')]^2}\right| \\
&= \left|\frac{-[L_1((\mathbf{x}_2' - \mathbf{x}_1')/2)]^2}{\left(1 + [g_{\tilde{\mathbf{w}}}(\mathbf{x}_2' - \mathbf{x}_1')]^2\right)\left([g_{\tilde{\mathbf{w}}}(\mathbf{x}_2' - \mathbf{x}_1')]^2\right)}\right| \\
&= \rho_{a_{m'}}(\mathbf{x}_2' - \mathbf{x}_1') \frac{1}{[g_{\tilde{\mathbf{w}}}(\mathbf{x}_2' - \mathbf{x}_1')]^2} \\
&\leq \frac{16\|\mathbf{Z}\|^2}{M^2} \rho_{a_{m'}}(\mathbf{x}_2' - \mathbf{x}_1')
\end{aligned}$$

Similarly,

$$\left|\rho_{a_{m'}}(\hat{\mathbf{x}}_2 - \hat{\mathbf{x}}_1) - \frac{[L_1((\hat{\mathbf{x}}_2 - \hat{\mathbf{x}}_1)/2)]^2}{[g_{\tilde{\mathbf{w}}}(\hat{\mathbf{x}}_2 - \hat{\mathbf{x}}_1)]^2}\right| \leq \frac{16\|\mathbf{Z}\|^2}{M^2} \rho_{a_{m'}}(\hat{\mathbf{x}}_2 - \hat{\mathbf{x}}_1).$$

However, noticing that

$$\begin{aligned}
\frac{[L_1((\mathbf{x}_2' - \mathbf{x}_1')/2)]^2}{[g_{\tilde{\mathbf{w}}}(\mathbf{x}_2' - \mathbf{x}_1')]^2} &= \frac{[L_1((\mathbf{x}_2' - \mathbf{x}_1')/2)]^2}{\langle \tilde{\mathbf{w}}, \mathbf{x}_2' - \mathbf{x}_1'/2\rangle^2 / \|\mathbf{Z}\tilde{\mathbf{w}}\|^2} \\
&= \frac{[L_1((\hat{\mathbf{x}}_2 - \hat{\mathbf{x}}_1)/2)]^2}{\langle \tilde{\mathbf{w}}, (\hat{\mathbf{x}}_2 - \hat{\mathbf{x}}_1)/2\rangle^2 / \|\mathbf{Z}\tilde{\mathbf{w}}\|} \\
&= \frac{[L_1((\hat{\mathbf{x}}_2 - \hat{\mathbf{x}}_1)/2)]^2}{g_{\tilde{\mathbf{w}}}(\hat{\mathbf{x}}_2 - \hat{\mathbf{x}}_1)^2}
\end{aligned}$$

we have

$$\begin{aligned}
&\left|\rho_{a_{m'}}\left(\mathbf{x}_2'-\mathbf{x}_1'\right)-\rho_{a_{m'}}\left(\hat{\mathbf{x}}_2-\hat{\mathbf{x}}_1\right)\right| \\
&=\left|\rho_{a_{m'}}\left(\mathbf{x}_2'-\mathbf{x}_1'\right)-\frac{\left[L_1((\mathbf{x}_2'-\mathbf{x}_1')/2)\right]^2}{\left[g_{\tilde{\mathbf{w}}}\left(\mathbf{x}_2'-\mathbf{x}_1'\right)\right]^2}+\frac{\left[L_1\left(\left(\hat{\mathbf{x}}_2-\hat{\mathbf{x}}_1\right)/2\right)\right]^2}{\left[g_{\tilde{\mathbf{w}}}\left(\hat{\mathbf{x}}_2-\hat{\mathbf{x}}_1\right)\right]^2}-\rho_{a_{m'}}\left(\hat{\mathbf{x}}_2-\hat{\mathbf{x}}_1\right)\right| \\
&\leq\left|\rho_{a_{m'}}\left(\mathbf{x}_2'-\mathbf{x}_1'\right)-\frac{\left[L_1((\mathbf{x}_2'-\mathbf{x}_1')/2)\right]^2}{\left[g_{\tilde{\mathbf{w}}}\left(\mathbf{x}_2'-\mathbf{x}_1'\right)\right]^2}\right|+\left|\frac{\left[L_1\left(\left(\hat{\mathbf{x}}_2-\hat{\mathbf{x}}_1\right)/2\right)\right]^2}{\left[g_{\tilde{\mathbf{w}}}\left(\hat{\mathbf{x}}_2-\hat{\mathbf{x}}_1\right)\right]^2}-\rho_{a_{m'}}\left(\hat{\mathbf{x}}_2-\hat{\mathbf{x}}_1\right)\right| \\
&\leq\frac{16\left\|\mathbf{Z}\right\|^2}{M^2}\rho_{a_{m'}}\left(\mathbf{x}_2'-\mathbf{x}_1'\right)+\frac{16\left\|\mathbf{Z}\right\|^2}{M^2}\rho_{a_{m'}}\left(\hat{\mathbf{x}}_2-\hat{\mathbf{x}}_1\right) \\
&=\frac{16\left\|\mathbf{Z}\right\|^2}{M^2}\left(\rho_{a_{m'}}\left(\mathbf{x}_2'-\mathbf{x}_1'\right)+\rho_{a_{m'}}\left(\hat{\mathbf{x}}_2-\hat{\mathbf{x}}_1\right)\right) \\
&\leq\frac{16\left\|\mathbf{Z}\right\|^2}{M^2}\left(\rho_{a_{m'}}\left(\mathbf{x}_2'-\mathbf{x}_1'\right)+\rho_{a_{m'}}\left(\mathcal{F}_{M'}\right)\right).
\end{aligned}$$

Thus,

$$\begin{aligned}
\rho_{a_{m'}}\left(\mathcal{F}_{M'}\right)&\geq\rho_{a_{m'}}\left(\hat{\mathbf{x}}_2-\hat{\mathbf{x}}_1\right) \\
&\geq\rho_{a_{m'}}\left(\mathbf{x}_2'-\mathbf{x}_1'\right)-\left|\rho_{a_{m'}}\left(\mathbf{x}_2'-\mathbf{x}_1'\right)-\rho_{a_{m'}}\left(\hat{\mathbf{x}}_2-\hat{\mathbf{x}}_1\right)\right| \\
&\geq\rho_{a_{m'}}\left(\mathbf{x}_2'-\mathbf{x}_1'\right)-\frac{16\left\|\mathbf{Z}\right\|^2}{M^2}\left(\rho_{a_{m'}}\left(\mathbf{x}_2'-\mathbf{x}_1'\right)+\rho_{a_{m'}}\left(\mathcal{F}_{M'}\right)\right),
\end{aligned}$$

or

$$\rho_{a_{m'}}\left(\mathcal{F}_{M'}\right)+\frac{16\left\|\mathbf{Z}\right\|^2}{M^2}\rho_{a_{m'}}\left(\mathcal{F}_{M'}\right)\geq\rho_{a_{m'}}\left(\mathbf{x}_2'-\mathbf{x}_1'\right)-\frac{16\left\|\mathbf{Z}\right\|^2}{M^2}\rho_{a_{m'}}\left(\mathbf{x}_2'-\mathbf{x}_1'\right),$$

giving

$$\left(1+\frac{16\left\|\mathbf{Z}\right\|^2}{M^2}\right)\rho_{a_{m'}}\left(\mathcal{F}_{M'}\right)\geq\left(1-\frac{16\left\|\mathbf{Z}\right\|^2}{M^2}\right)\rho_{a_{m'}}\left(\mathbf{x}_2'-\mathbf{x}_1'\right).$$

By (4.20) and (4.21), we have.

$$\left(1+\frac{16\left\|\mathbf{Z}\right\|^2}{M^2}\right)\left(\rho_0\left(\mathcal{F}\right)+\varepsilon\right)\geq\left(1-\frac{16\left\|\mathbf{Z}\right\|^2}{M^2}\right)\left(\rho_0\left(\mathcal{F}\right)+\delta-\varepsilon\right).$$

And this gives

$$\frac{64\left\|\mathbf{Z}\right\|^2\rho_0\left(\mathcal{F}\right)}{\delta}+32\left\|\mathbf{Z}\right\|^2\geq M^2$$

proving the Lemma

4.4.2 Proofs for Section 4.3

For convenience, in the proofs of this section, we shall use $K, K_1, \dots$ and $M, M_1, \dots$ as generic constants which may wary from line to line. They may depend on the fixed numbers such as m, p, or C, but not on the function f or the number of sampling points n.

For the proof of Lemma 4.10, we need the following results.

Lemma 4.26. *Let $f \in \mathbf{W}_{[0,T]}(m,p,C)$, and $0 \le k < m$. If $\left|f^{(k)}(t)\right| \ge A$, for $t \in [t_1, t_2]$ with $t_2 - t_1 = B$, then, we can find $\tau_1, \tau_2 \in [t_1, t_2]$, with $\tau_2 - \tau_1 \ge \lambda_k B$, such that $|f(t)| \ge \mu_k A B^k$. Here, λ_k, μ_k are constants depend only on k.*

Proof

The result for $k = 0$ is apparent. Now let $k = 1$. By the continuity of $f'(t)$, $|f'(t)| \ge A$ for $t \in [t_1, t_2]$ implies that either $f'(t) \ge A$ or $f'(t) \le -A$. We will only provide the proof for the former case, since the proof for the latter case is similar. Suppose that $t' = (t_2 + t_1)/2$. If $f(t') \ge 0$, then, let $t_1^{(1)} = 3t_2/4 + t_1/4, t_2^{(1)} = t_2$. We have

$$\frac{f\left(t_1^{(1)}\right) - f(t')}{t_1^{(1)} - t'} \ge A,$$

so that,

$$\begin{aligned} f\left(t_1^{(1)}\right) &\ge A\left(t_1^{(1)} - t'\right) + f(t') \ge A\left(t_1^{(1)} - t'\right) \\ &= \frac{1}{4}A(t_2 - t_1) \ge \frac{1}{4}AB. \end{aligned}$$

Thus, for $t \in \left[t_1^{(1)}, t_2^{(1)}\right]$, $f(t) \ge AB/4$. If $f(t') \le 0$, then, let $t_1^{(1)} = t_1, t_2 = 3t_1/4 + t_2/4$. Then we have $f(t) \le -AB/4$ for $t \in \left[t_1^{(1)}, t_2^{(1)}\right]$. The proof of the lemma can be completed by an induction on k.

Proof of Lemma 4.10

First, we prove the lemma for $p > 1$ and $k = m-1$. Suppose that $\left\|f^{(m-1)}\right\|_\infty = \left|f^{(m-1)}(t')\right| = 2A$. Since the discussions for $f^{(m-1)}(t') = 2A$ and that for $f^{(m-1)}(t') = -2A$ are almost the same, we only give a proof for the case $f^{(m-1)}(t') = 2A$. Let $\delta = \left(\frac{A}{C}\right)^{\frac{p}{p-1}}$ and $\delta' = \min\{\delta, T/2\}$. Then, we have either $[t', t'+\delta'] \subseteq [0,T]$ or $[t'-\delta', t'] \subseteq [0,T]$. Without loss of generality, we may assume that $[t', t'+\delta'] \subseteq [0,T]$. We claim that $f^{(m-1)}(t) \geq A$ for $t \in [t', t'+\delta']$. If not, we can find a $t'' \in [t', t'+\delta']$ such that $f^{(m-1)}(t'') < A$. Then we have

$$\begin{aligned}\int_{t'}^{t'+\delta'} \left|f^{(m)}(t)\right| dt &\geq \left|\int_{t'}^{t''} f^{(m)}(t)\, dt\right| \\ &= \left|f^{(m-1)}(t'') - f^{(m-1)}(t')\right| > A,\end{aligned}$$

giving,

$$\begin{aligned}\left\|f^{(m)}\right\|_p^p &\geq \int_{t'}^{t'+\delta'} \left|f^{(m)}(t)\right|^p dt \\ &\geq \delta' \left(\frac{1}{\delta'}\int_{t'}^{t'+\delta'} \left|f^{(m)}(t)\right| dt\right)^p \\ &> \delta'^{1-p}A^p \geq \delta^{1-p}A^p = \left(\frac{A}{C}\right)^{-p} A^p = C^p,\end{aligned}$$

which is impossible. Hence, by Lemma 4.26 we find $\tau_1, \tau_2 \in [t', t'+\delta]$ with $\tau_2 - \tau_1 = \delta'' \geq \lambda_{m-1}\delta'$, and for $t \in [\tau_1, \tau_2]$, we have $|f(t)| \geq \mu_{m-1}A\delta'^{m-1}$. Now,

$$\begin{aligned}\|f\|_r &\geq \left(\lambda_{m-1}\delta' \left(\mu_{m-1}A\delta'^{m-1}\right)^r\right)^{1/r} \\ &= \left(\lambda_{m-1}\delta'\right)^{1/r} \mu_{m-1}A\delta'^{m-1} \\ &= KA\delta'^{(1/r+m-1)}\end{aligned}$$

If $\frac{1}{2}\left\|f^{(m-1)}\right\|_\infty \geq C\left(\frac{T}{2}\right)^{\frac{p-1}{p}}$, then, $\delta' = \frac{T}{2}$ and

$$\|f\|_r \geq KA\left(\frac{T}{2}\right)^{(1/r+m-1)}$$

or

$$\|f\|_r \geq K\left\|f^{(m-1)}\right\|_\infty,$$

which means that $\frac{1}{2}\left\|f^{(m-1)}\right\|_\infty \leq C\left(\frac{T}{2}\right)^{\frac{p-1}{p}}$ for $\|f\|_r$ small enough. If this is the case, we have

$\delta' = \delta$ and

$$\begin{aligned}
\|f\|_r &\geq KA\delta^{(1/r+m-1)} \\
&= KA\left(\frac{A}{C}\right)^{\frac{p}{p-1}(1/r+m-1)} \\
&= K(2A)^{1+\frac{p}{p-1}(1/r+m-1)}
\end{aligned}$$

Therefore

$$\begin{aligned}
\left\|f^{(m-1)}\right\|_\infty &\leq K\|f\|_r^{\frac{1}{1+\frac{p}{p-1}\left(\frac{1}{r}+m-1\right)}} \\
&= K\|f\|_r^{\alpha_{m-1}}.
\end{aligned}$$

For $p = 1$ and $k = m-1$, we have $\alpha_{m-1} = 0$. Suppose that $\left\|f^{(m-1)}\right\|_\infty = \left|f^{(m-1)}(t')\right| = A > C$. Again, we can assume that $f^{(m-1)}(t') = A$. Then, we have $f^{(m-1)}(t) \geq A - C$ for any $t \in [0,T]$. For, if $f^{(m-1)}(t'') < A - C$ for some $t'' \in [0,T]$, then

$$\begin{aligned}
\left\|f^{(m)}\right\|_1 = \int_0^T \left|f^{(m)}(t)\right| dt &\geq \left|\int_{t'}^{t''} \left|f^{(m)}(t)\right| dt\right| \\
&\geq \left|\int_{t'}^{t''} f^{(m)}(t)\, dt\right| > C
\end{aligned}$$

which is impossible. Again, by applying Lemma 4.26 the same way as before we can show that $\left(\left\|f^{(m-1)}\right\|_\infty - C\right) \vee 0 \to 0$ as $\|f\|_r \to 0$. Therefore, $\left\|f^{(m-1)}\right\|_\infty = O(1)$.

Now, suppose that we have proven the Lemma for k, $1 < k \leq m-1$. Let $\left\|f^{(k-1)}\right\|_\infty = \left|f^{(k-1)}(t')\right| = 2A$. As before, we assume that $f^{(k-1)}(t') = 2A$. Let $\delta = A/\left\|f^{(k)}\right\|_\infty$ and $\delta' = \min\{\delta, T/2\}$. Then, either $[t', t'+\delta'] \subset [0,T]$ or $[t'-\delta', t'] \subset [0,T]$. Without loss of generality, we may assume the former. For every $t \in [t', t'+\delta']$, we have $f^{(k-1)}(t) \geq A$. Now, by applying Lemma 4.26, we can find $[\tau_1, \tau_2] \subset [t', t'+\delta']$ with $\tau_2 - \tau_1 = \delta'' \geq \lambda_{k-1}\delta'$ and for $t \in [\tau_1, \tau_2]$, $|f(t)| \geq \mu_{k-1}A\delta'^{k-1}$. Next,

$$\begin{aligned}
\|f\|_r &\geq \left(\int_{\tau_1}^{\tau_2} |f(t)|^r\, dt\right)^{1/r} \geq \left(\delta''\left(\mu_{k-1}A\delta'^{k-1}\right)^r\right)^{1/r} \\
&= KA\delta'^{r^{-1}+k-1}.
\end{aligned}$$

If $\left\|f^{(k-1)}\right\|_\infty > T\left\|f^{(k)}\right\|_\infty$, we have $\delta' = T/2$. Then,

$$\|f\|_r \geq KA\left(\frac{T}{2}\right)^{r^{-1}+k-1}$$

or

$$\left\|f^{(k-1)}\right\|_\infty = 2A \leq K\|f\|_r\left(\frac{T}{2}\right)^{-\left(r^{-1}+k-1\right)}. \tag{4.22}$$

If $\|f\|_r \leq 1$ then (4.22) gives

$$\left\|f^{(k-1)}\right\|_\infty \leq K\|f\|_r^{\alpha_{k-1}}$$

since $\alpha_{k-1} \leq 1$. If (4.22) is violated, then $\left\|f^{(k-1)}\right\|_\infty \leq t\left\|f^{(k)}\right\|_\infty$, which means that $\delta' = \delta$. Therefore

$$\begin{aligned}\|f\|_r \geq KA\delta^{r^{-1}+k-1} &= KA\left(A\Big/\left\|f^{(k)}\right\|_\infty\right)^{r^{-1}+k-1}\\ &= KA^{r^{-1}+k}\left\|f^{(k)}\right\|_\infty^{1-r^{-1}-k}.\end{aligned} \tag{4.23}$$

By the assumption, we can find ε and K_1 such that when $\|f\|_r \leq \varepsilon$, we have $\left\|f^{(k)}\right\|_\infty \leq K_1\|f\|_r^{\alpha_k}$. Then by (4.23), when $\|f\|_r \leq \varepsilon$ we have

$$\begin{aligned}\left\|f^{(k-1)}\right\|_\infty = 2A &\leq K\|f\|_r^{\frac{1}{r^{-1}+k}}\left\|f^{(k)}\right\|_\infty^{\frac{r^{-1}+k-1}{r^{-1}+k}}\\ &\leq K\|f\|_r^{\frac{1}{r^{-1}+k}}\left(K_1\|f\|_r^{\alpha_k}\right)^{\frac{r^{-1}+k-1}{r^{-1}+k}}\\ &= K\|f\|_r^{\alpha_{k-1}}.\end{aligned}$$

This proves the Lemma.

The following lemmas are needed for proving Lemma 4.11. Assume that $f \in \mathscr{L}_2[0,T]$ and $F(x) = \int_0^x f(t)\,dt$ for $x \in [0,T]$.

Lemma 4.27.

$$\begin{aligned}\int_0^y\int_0^u\int_0^{T-t} f(x+t)f(x)\,dxdtdu = \frac{1}{2}\int_{1-y}^{T}(F(T)-F(x))^2\,dx+\\ +\frac{1}{2}\int_0^{T-y}(F(x+y)-F(x))^2\,dx+\frac{1}{2}\int_0^y F^2(x)\,dx.\end{aligned} \tag{4.24}$$

Proof

Let $F(x) = \int_0^x f(t)\,dt$. We have

$$\begin{aligned}
&\int_0^u \int_0^{T-t} f(x+t) f(x)\,dxdt \\
&= \int_0^{T-u} \int_0^u f(x+t) f(x)\,dtdx + \int_{T-u}^T \int_0^{T-x} f(x+t) f(x)\,dtdx \\
&= \int_0^{T-u} (F(x+u) - F(x)) f(x)\,dx + \int_{T-u}^T (F(T) - F(x)) f(x)\,dx \\
&= \int_0^{T-u} F(x+u) f(x)\,dx + F(T)(F(T) - F(T-u)) - \int_0^T F(x) f(x)\,dx.
\end{aligned}$$

Thus,

$$\begin{aligned}
&\int_0^y \int_0^u \int_0^{T-t} f(x+t) f(x)\,dxdtdu \\
&= \int_0^y \left(\int_0^{T-u} F(x+u) f(x)\,dx + F(T)(F(T) - F(T-u)) \right. \\
&\left. - \int_0^T F(x) f(x)\,dx \right) du \\
&= \int_0^y \int_0^{T-u} F(x+u) f(x)\,dxdu + \int_0^y F(T)(F(T) - F(T-u))\,du \\
&- \int_0^y \int_0^T F(x) f(x)\,dxdu \\
&= \int_0^{T-y} \int_0^y F(x+u) f(x)\,dudx + \int_{T-y}^T \int_0^{T-x} F(x+u) f(x)\,dudx + \\
&+ \int_0^y F(T)(F(T) - F(T-u))\,du - \frac{1}{2} y (F(T))^2 \\
&= \int_0^{T-y} (G(x+y) - G(x)) f(x)\,dx + \int_{T-y}^T (G(T) - G(x)) f(x)\,dx + \\
&+ (G(T-y) - G(T)) F(T) + \frac{1}{2} y (F(T))^2 \\
&= \int_0^{T-y} G(x+y) f(x)\,dx - \int_0^T G(x) f(x)\,dx + G(T)(F(T) - F(T-y)) + \\
&+ (G(T-y) - G(T)) F(T) + \frac{1}{2} y (F(T))^2 \\
&= \int_0^{T-y} G(x+y) f(x)\,dx - \int_0^T G(x) f(x)\,dx + F(T) G(T-y) \\
&- F(T-y) G(T) + \frac{1}{2} y (F(T))^2
\end{aligned}$$

$$
\begin{aligned}
&= \int_0^{T-y} G(x+y)\, dF(x) - \int_0^T G(x)\, dF(x) + F(T)\, G(T-y) \\
&- F(T-y)\, G(T) + \frac{1}{2} y\, (F(T))^2 \\
&= F(T)\, G(T-y) - \int_0^{T-y} F(x)\, F(x+y)\, dx - G(T)\, F(T) \\
&+ \int_0^T (F(x))^2\, dx + \frac{1}{2} y\, (F(T))^2 \\
&= -F(T) \int_{T-y}^T F(x) - \int_0^{T-y} F(x)\, F(x+y)\, dx + \int_0^T F(x)^2\, dx + \frac{1}{2} y F(T)^2 \\
&= -\int_{T-y}^T F(T)\, F(x)\, dx - \int_0^{T-y} F(x)\, F(x+y)\, dx + \frac{1}{2} \int_0^{T-y} F(x)^2\, dx \\
&+ \frac{1}{2} \int_{T-y}^T F(x)^2\, dx + \frac{1}{2} \int_0^{T-y} F(x+y)^2\, dx + \frac{1}{2} \int_0^y F(x)^2\, dx + \frac{1}{2} \int_{T-y}^T F(T)^2\, dx \\
&= \frac{1}{2} \int_{T-y}^T \left(F(T)^2 - 2F(T)\, F(x) + F(x)^2 \right) dx \\
&+ \frac{1}{2} \int_0^{T-y} \left(F(x)^2 - 2F(x)\, F(x+y) + F(x+y)^2 \right) dx + \frac{1}{2} \int_0^y F(x)^2\, dx \\
&= \frac{1}{2} \int_{T-y}^T (F(T) - F(x))^2\, dx + \frac{1}{2} \int_0^{T-y} (F(x+y) - F(x))^2\, dx + \frac{1}{2} \int_0^y F(x)^2\, dx
\end{aligned}
$$

Lemma 4.28. *There exists a constant A_c such that for any $0 < \delta < T$, and $s \in [0,T]$ such that $[s, s+\delta] \subset [0,T]$, we have*

$$
\int_0^T \int_0^T f(u)\, f(v)\, |u-v|^{2H-2}\, dudv \geq A_c \delta^{2H-2} \left| F(s+\delta) - F(s) \right|^2 .
$$

Proof

It is enough to prove the lemma for bounded f, since such functions are dense in$\mathscr{L}_2[0,T]$, and hence in$\mathscr{L}_\phi^2[0,T]$. Let $g(t) = \int_0^{T-t} f(x+t)\, f(x)\, dx$, $G(t) = \int_0^t g(s)\, ds$, and $Q(t) = \int_0^t G(s)\, ds$ for $t \in [0,T]$. First, notice that

$$
\begin{aligned}
&\int_0^T \int_0^T f(u)\, f(v)\, |u-v|^{2H-2}\, dudv \\
&= \int_0^T \int_0^v f(u)\, f(v)\, (v-u)^{2H-2}\, dudv + \int_0^T \int_v^T f(u)\, f(v)\, (u-v)^{2H-2}\, dudv \\
&= 2 \int_0^T x^{2H-2} g(x)\, dx = 2 \int_0^T x^{2H-2} dG(x)
\end{aligned}
$$

$$
\begin{aligned}
&= 2x^{2H-2}G(x)\,\big|_0^T - 2\int_0^T (2H-2)\,x^{2H-3}G(x)\,dx \\
&= 2G(T) - 2(2H-2)\int_0^T x^{2H-3}dQ(x) \\
&= 2G(T) - 2(2H-2)\left(x^{2H-3}Q(x)\,\big|_0^T - \int_0^T (2H-3)\,x^{2H-4}Q(x)\,dx\right) \\
&= 2G(T) - 2(2H-2)\,Q(T) + 2(2H-2)(2H-3)\int_0^T x^{2H-4}Q(x)\,dx
\end{aligned}
$$

It can be shown that $G(T) > 0$ and by Lemma 4.27 $Q(T)$ is also positive. So we only need to show that there exists K such that $\int_0^T x^{2H-4}Q(x)\,dx \geq K\delta^{2H-2}\left|F(s+\delta) - F(s)\right|^2$. By Lemma 4.27 we have

$$
\begin{aligned}
&\int_0^T x^{2H-4}Q(x)\,dx \\
&\geq \int_0^T x^{2H-4}\int_0^{T-x} \left(F(y+x) - F(y)\right)^2 dydx
\end{aligned}
$$

Without loss of generality, we assume that $F(s+\delta) > F(s)$. Let $h_0 = F(s+\delta) - F(s)$, $\delta_0 = \delta$, $s_0 = s$ and $t_0 = s+\delta$. Beginning at $k = 0$, we repeat the process below recursively until certain condition (inequality (4.25)) is met. Define

$$
\eta_1 = \sup\left\{F(t) - F(s_k) : s_k \leq t \leq \frac{\lambda-1}{\lambda}s_k + \frac{1}{\lambda}t_k\right\}
$$

and

$$
\eta_2 = \inf\left\{F(t) - F(s_k) : \frac{1}{\lambda}s_k + \frac{\lambda-1}{\lambda}t_k \leq t \leq t_k\right\}
$$

where $\lambda = 3^{\frac{1}{1-H}}$. If

$$
\eta_2 - \eta_1 \geq \frac{1}{3}h_k, \tag{4.25}
$$

is not satisfied, we have either $\eta_1 > \frac{1}{3}h_k$ or $\eta_2 < \frac{2}{3}h_k$. Without loss of generality, we assume that the former is true. Then we find $\tau \in \left(s_k, \frac{\lambda-1}{\lambda}s_k + \frac{1}{\lambda}t_k\right]$ such that $F(\tau) - F(s_k) > \frac{1}{3}h_k$. Now let $s_{k+1} = s_k$, $t_{k+1} = \tau$, $\delta_{k+1} = t_{k+1} - s_{k+1}$, and $h_{k+1} = F(t_{k+1}) - F(s_{k+1})$.

We claim that the condition (4.25) will be met after a finite number of repetitions, because if not, than for any positive integer m, we have $h_m \geq \left(\frac{1}{3}\right)^m h_0$ and $\delta_m \leq \frac{1}{\lambda^m}\delta$. Thus $\frac{F(t_m)-F(s_m)}{t_m-t_s} = \frac{1}{3^m}\Big/\frac{1}{\lambda^m}$ which goes to infinity, contradicting the assumption that f is bounded. So, there is a k

such that condition (4.25) is not met. We have

$$\begin{aligned}
&\int_0^T x^{2H-4}\int_0^{T-x}\left(F\left(y+x\right)-F\left(y\right)\right)^2 dydx\\
&\geq \int_0^{\frac{1}{\lambda}\delta_k} x^{2H-4}\int_0^{\frac{\lambda-1}{\lambda}\delta_k}\left(F\left(y+x\right)-F\left(y\right)\right)^2 dydx\\
&\geq \int_0^{\frac{1}{\lambda}\delta_k} x^{2H-4}\frac{\lambda}{\lambda-1}\delta_k^{-1}\left(\int_0^{\frac{\lambda-1}{\lambda}\delta_k}\left|F\left(y+x\right)-F\left(y\right)\right| dy\right)^2 dx\\
&\geq \frac{\lambda}{\lambda-1}\delta_k^{-1}\int_0^{\frac{1}{\lambda}\delta_k} x^{2H-4}\left(\int_0^{\frac{\lambda-1}{\lambda}\delta_k}\left(F\left(y+x\right)-F\left(y\right)\right) dy\right)^2 dx\\
&= \frac{\lambda}{\lambda-1}\delta_k^{-1}\int_0^{\frac{1}{\lambda}\delta_k} x^{2H-4}\left(\int_0^{\frac{\lambda-1}{\lambda}\delta_k}\int_y^{x+y} f\left(t\right) dtdy\right)^2 dx\\
&= \frac{\lambda}{\lambda-1}\delta_k^{-1}\int_0^{\frac{1}{\lambda}\delta_k} x^{2H-4}\left(\int_0^{\frac{\lambda-1}{\lambda}\delta_k}\int_0^{x} f\left(u+v\right) dudv\right)^2 dx\\
&= \frac{\lambda}{\lambda-1}\delta_k^{-1}\int_0^{\frac{1}{\lambda}\delta_k} x^{2H-4}\left(\int_0^{x}\int_0^{\frac{\lambda-1}{\lambda}\delta_k} f\left(u+v\right) dvdu\right)^2 dx\\
&= \frac{\lambda}{\lambda-1}\delta_k^{-1}\int_0^{\frac{1}{\lambda}\delta_k} x^{2H-4}\left(\int_0^{x}\left(F\left(u+\frac{\lambda-1}{\lambda}\delta_k\right)-F\left(u\right)\right) du\right)^2 dx\\
&\geq \frac{\lambda}{\lambda-1}\delta_k^{-1}\int_0^{\frac{1}{\lambda}\delta_k} x^{2H-4}\left(\int_0^{x}\frac{1}{3}h_k du\right)^2 dx\\
&= \frac{\lambda}{\lambda-1}\delta_k^{-1}\frac{1}{9}h_k^2\int_0^{\frac{1}{\lambda}\delta_k} x^{2H-2}dx\\
&\geq \frac{1}{2H-1}\frac{\lambda}{\lambda-1}\left(\frac{1}{\lambda}\right)^{2H-1}\frac{1}{9}\delta^{2H-2}h_0^2.
\end{aligned}$$

This finishes the proof.

Proof of Lemma 4.11

Suppose that $f^{(k)}\left(t_0\right)=2\varepsilon$ with $\varepsilon<1$, and $\left\|f^{(k)}\right\|_\infty=\left|f^{(k)}\left(t'\right)\right|=2A\geq 2\varepsilon$. Again, we assume that $f^{(k)}\left(t'\right)=2A$, since the discussion for the case of $f^{(k)}\left(t'\right)=-2A$ is almost the same. Like in the proof of Lemma 4.10, we can find $[\tau_1,\tau_2]\subset[0,T]$ with $\tau_2-\tau_1=l\geq\lambda_{k-1}\delta'$, $\delta'=\min\{\delta,T/2\}$ where

$$\delta=\begin{cases}\left(\frac{A}{C}\right)^{\frac{p}{p-1}} & \text{if} \quad k=m-1,\\ A\big/\left\|f^{(k+1)}\right\|_\infty & \text{if} \quad k<m-1,\end{cases}$$

and $|f(t)| \geq \mu_k A \delta'^k$ for $t \in [\tau_1, \tau_2]$. If we take

$$g(t) = \begin{cases} l^{-1/2} & \text{if } t \in [\tau_1, \tau_2], \\ 0 & \text{o/w.} \end{cases}$$

We have $\|g\|_2 = 1$ and

$$\begin{aligned} |\langle g, f \rangle| &\geq l \left(l^{-1/2} \right) \mu_k A \delta'^k \\ &= l^{1/2} \mu_k A \delta'^k \end{aligned}$$

We also have

$$\begin{aligned} \|\mathbf{Z}_\sigma g\|^2 &= \sigma^2 H (2H-1) \int_0^T \int_0^T g(u) g(v) |u-v|^{2H-2} dudv \\ &= \sigma^2 H (2H-1) l^{-1} \int_{\tau_1}^{\tau_2} \int_{\tau_1}^{\tau_2} |u-v|^{2H-2} dudv \\ &= \sigma^2 H (2H-1) l^{-1} \int_0^l \int_0^l |u-v|^{2H-2} dudv \\ &= \sigma^2 H (2H-1) l^{-1} \int_0^l \left(\int_0^v |u-v|^{2H-2} du + \int_v^l |u-v|^{2H-2} du \right) dv \\ &= \sigma^2 H (2H-1) l^{-1} \int_0^l \left(\int_0^v (v-u)^{2H-2} du + \int_v^l (u-v)^{2H-2} du \right) dv \\ &= \sigma^2 H (2H-1) l^{-1} \int_0^l \left(\int_0^v (v-u)^{2H-2} d(u-v) \right. \\ &\left. + \int_v^l (u-v)^{2H-2} d(u-v) \right) dv \\ &= \sigma^2 H (2H-1) l^{-1} \int_0^l \left(- \left(\frac{1}{2H-1} (v-u)^{2H-1} |_0^v \right) \right. \\ &\left. + \frac{1}{2H-1} (u-v)^{2H-1} |_v^l \right) dv \\ &= \sigma^2 H (2H-1) l^{-1} \int_0^l \left(\frac{1}{2H-1} v^{2H-1} + \frac{1}{2H-1} (l-v)^{2H-1} \right) dv \\ &= \sigma^2 H l^{-1} \left(\int_0^l v^{2H-1} dv + \int_0^l (l-v)^{2H-1} dv \right) \\ &= \sigma^2 H l^{-1} \left(\frac{1}{2H} v^{2H} |_0^l - \frac{1}{2H} (l-v)^{2H} |_0^l \right) \\ &= \frac{1}{2} \sigma^2 l^{-1} \left(l^{2H} + l^{2H} \right) = \sigma^2 l^{2H-1}. \end{aligned}$$

Then,

$$\frac{|\langle g,f\rangle|/2}{\|\mathbf{Z}_\sigma g\|} \geq \frac{l^{1/2}\mu_k A\delta'^k}{2\sigma l^{H-1/2}} = \frac{1}{2\sigma} l^{1-H}\mu_k A\delta'^k$$
$$\geq \frac{1}{2\sigma}\left(\lambda_{k-1}\delta'\right)^{1-H}\mu_k A\delta'^k$$
$$= KA\delta'^{1-H+k}/\sigma, \tag{4.26}$$

If $\delta' = T/2$, then, (4.26) gives

$$\frac{|\langle g,f\rangle|/2}{\|\mathbf{Z}_\sigma g\|} \geq KA/\sigma \geq K\varepsilon/\sigma \geq K\varepsilon^{\gamma_k}/\sigma.$$

If $k = m-1$ and $\delta' = (A/C)^{\frac{p}{p-1}}$, then (4.26) gives

$$\frac{|\langle g,f\rangle|/2}{\|\mathbf{Z}_\sigma g\|} \geq KA\,(A/C)^{\frac{p}{p-1}(1-H+(m-1))}/\sigma = KA^{1+\frac{p}{p-1}(m-H)}/\sigma$$
$$= KA^{1+\frac{m-H}{1-p^{-1}}}/\sigma = KA^{\frac{1-H+m-p^{-1}}{1-p^{-1}}}/\sigma$$
$$= KA^{\gamma_{m-1}}/\sigma \geq K\varepsilon^{\gamma_{m-1}}/\sigma.$$

If $k < m-1$ and $\delta' = A\big/\left\|f^{(k+1)}\right\|_\infty$, because $f^{(k)} \in \mathbf{W}_{[0,T]}\,(m-k,p,C)$, we apply Lemma 4.10 and get positive ε_0 and K' such that

$$\left\|f^{(k+1)}\right\|_\infty \leq K'\left\|f^{(k)}\right\|_\infty^{\frac{m-k-1-p^{-1}}{m-k-p^{-1}}} \leq K'A^{\frac{m-k-1-p^{-1}}{m-k-p^{-1}}}.$$

when $\varepsilon \leq \varepsilon_0$. Thus,

$$\frac{|\langle g,f\rangle|/2}{\|\mathbf{Z}_\sigma g\|} \geq KA\left(A\Big/\left\|f^{(k+1)}\right\|_\infty\right)^{1-H+k}/\sigma$$
$$= KA^{2-H+k}\left\|f^{(k+1)}\right\|_\infty^{-(1-H+k)}/\sigma$$
$$\geq KA^{2-H+k}\left(K'\left\|f^{(k)}\right\|_\infty^{\frac{m-k-1-p^{-1}}{m-k-p^{-1}}}\right)^{-(1-H+k)}/\sigma$$
$$= KA^{-(1-H+k)\frac{m-k-1-p^{-1}}{m-k-p^{-1}}+2-H+k}/\sigma = KA^{1+(1-H+k)\left(1-\frac{m-k-1-p^{-1}}{m-k-p^{-1}}\right)}/\sigma$$
$$= KA^{1+(1-H+k)\frac{1}{m-k-p^{-1}}}/\sigma = KA^{1+\frac{1-H+k}{m-k-p^{-1}}}/\sigma$$
$$= KA^{\frac{m-p^{-1}+1-H}{m-k-p^{-1}}}/\sigma = KA^{\gamma_k}/\sigma \geq K\varepsilon^{\gamma_k}/\sigma.$$

Up to now, we have proven that there is a constant K such that when ε is small enough,

$$\upsilon\left(\varepsilon;\mathbf{W}_{[0,T]}\left(m,p,\frac{1}{2}C\right)\right)\geq K\varepsilon^{\gamma_k}/\sigma.$$

Next we will complete the other part of the proof. Let $f\in\mathbf{W}_{[0,T]}(m,p,C)$ satisfying $f^{(i)}(0)=f^{(i)}(T)=0$ for $0\leq i\leq m$ and $f^{(k)}(t_0)/2=1$. Let

$$f_\delta(t)=\begin{cases}\delta^{m-1/p}f\left(\frac{t-t_0}{\delta}+t_0\right), & \text{if } t\in I_\delta=(t_0-t_0\delta,t_0+(T-t_0)\delta)\\ 0, & \text{otherwise}\end{cases}$$

where $1\geq\delta>0$. We have $\left\|f_\delta^{(m)}\right\|_p=\left\|f^{(m)}\right\|_p$. so that $f_\delta\in\mathbf{W}_{[0,T]}(m,p,C)$. For any bounded function g in $\mathscr{L}_2[0,T]$, we have

$$\begin{aligned}|\langle g,f_\delta\rangle|&=\left|\int_{I_\delta}f_\delta(t)\,g(t)\,dt\right|=\left|\int_{I_\delta}f_\delta(t)\,dG(t)\right|\\&=\left|\int_{I_\delta}f_\delta'(t)\,G(t)\,dt\right|\leq\|f_\delta'\|_1\left\|\mathbf{1}_{I_\delta}G\right\|_\infty,\end{aligned}$$

where $G(t)=\int_{t_0-t_0\delta}^{t}f_\delta(s)\,ds$ for $t\in[t_0-t_0\delta,t_0+(T-t_0)\delta]$. Suppose that t_1 and $t_2\in I_\delta$, $t_1<t_2$ satisfy that $\left\|\mathbf{1}_{I_\delta}G\right\|_\infty=|G(t_2)-G(t_1)|$. By Lemma 4.28, we have

$$\begin{aligned}\int_0^T\int_0^T g(u)\,g(v)\,|u-v|^{2H-2}\,dudv&\geq A_c(t_2-t_1)^{2H-2}\,|G(t_2)-G(t_1)|^2\\&\geq A_c\delta^{2H-2}\,|G(t_2)-G(t_1)|^2.\end{aligned}$$

Thus,

$$\begin{aligned}\frac{|\langle g,f_\delta\rangle/2|}{\|\mathbf{Z}_1g\|}&\leq\frac{\|f_\delta'\|_1\,|G(t_2)-G(t_1)|}{A_c^{1/2}\delta^{H-1}\,|G(t_2)-G(t_1)|}\\&=\frac{\|f_\delta'\|_1}{A_c^{1/2}\delta^{H-1}}\\&\leq A_c^{-1/2}\,\|f'\|_1\,\delta^{m-p^{-1}+1-H}.\end{aligned}$$

Notice that the right hand side of the above inequality is free of g, so we have

$$G_0(f_\delta)\leq A_c^{-1/2}\,\|f\|_1\,\delta^{m-p^{-1}+1-H}/\sigma$$

Let

$$\varepsilon = \frac{d^k}{dt^k} f_\delta (t_0) /2 = \delta^{m-k-p^{-1}} f^{(k)} (t_0) /2.$$

Since $f_\delta \in \mathbf{W}_{[0,T]} \left(m, p, \frac{C}{2}\right) - \mathbf{W}_{[0,T]} \left(m, p, \frac{C}{2}\right)$, we have

$$\begin{aligned} \upsilon \left(\varepsilon; \mathbf{Z}_\sigma, \mathbf{W}_{[0,T]} \left(m, p, \frac{C}{2}\right)\right) &\leq G_0 (f_\delta) \\ &\leq A_c^{-1/2} \|f\|_1 \delta^{m-p^{-1}+1-H} /\sigma \\ &= K \varepsilon^{\frac{m-p^{-1}+1-H}{m-k-p^{-1}}} /\sigma \end{aligned}$$

proving the Lemma.

Proof of Theorem 4.12

By Lemma 4.11 we know that there exists $\varepsilon_0 > 0$ and constants A and B such that $A\varepsilon^{\gamma_k}/\sigma \leq \upsilon (\varepsilon; \mathbf{Z}_\sigma) \leq B\varepsilon^{\gamma_k}/\sigma$ when $0 < \varepsilon \leq \varepsilon_0$. By (4.13) and Corollary 4.9, we have

$$\begin{aligned} \inf_{\hat{L} \text{ affine}} R_{\mathbf{W}_{[0,T]}(m,p,C)} \left(\hat{L}\right) &= \sup_{\varepsilon > 0} \frac{\varepsilon^2}{1 + \upsilon (\varepsilon; \mathbf{Z}_\alpha)^2} \\ &= \max \left\{ \sup_{\varepsilon_0 > \varepsilon > 0} \frac{\varepsilon^2}{1 + \upsilon (\varepsilon; \mathbf{Z}_\alpha)^2}, \sup_{\varepsilon \geq \varepsilon_0} \frac{\varepsilon^2}{1 + \upsilon (\varepsilon; \mathbf{Z}_\alpha)^2} \right\} \\ &\leq \max \left\{ \sup_{\varepsilon_0 > \varepsilon > 0} \frac{\varepsilon^2}{1 + \upsilon (\varepsilon; \mathbf{Z}_\alpha)^2}, \sup_{\varepsilon \geq \varepsilon_0} \frac{\varepsilon^2}{\upsilon (\varepsilon; \mathbf{Z}_\alpha)^2} \right\} \\ &= \max \left\{ \sup_{\varepsilon_0 > \varepsilon > 0} \frac{\varepsilon^2}{1 + \upsilon (\varepsilon; \mathbf{Z}_\alpha)^2}, \frac{(\varepsilon_0)^2 \sigma^2}{\upsilon ((\varepsilon_0); \mathbf{Z}_1)^2} \right\} . \end{aligned}$$

The last step follows due to the convexity of $\upsilon (\varepsilon)$. We now have

$$\sup_{\varepsilon_0 > \varepsilon > 0} \frac{\varepsilon^2}{1 + B\varepsilon^{2\gamma_k}/\sigma^2} \leq \sup_{\varepsilon_0 > \varepsilon > 0} \frac{\varepsilon^2}{1 + \upsilon (\varepsilon; \mathbf{Z}_\alpha)^2} \leq \sup_{\varepsilon_0 > \varepsilon > 0} \frac{\varepsilon^2}{1 + A\varepsilon^{2\gamma_k}/\sigma^2}.$$

Straight forward calculation gives that for σ sufficiently small, we have

$$B^{1/\gamma_k} \gamma_k^{-1} (\gamma_k - 1)^{1-1/\gamma_k} \sigma^{2/\gamma_k} \leq \sup_{1/2 > \varepsilon > 0} \frac{\varepsilon^2}{1 + \upsilon (\varepsilon; \mathbf{Z}_\alpha)^2} \leq A^{1/\gamma_k} \gamma_k^{-1} (\gamma_k - 1)^{1-1/\gamma_k} \sigma^{2/\gamma_k}$$

This yields,

$$\sup_{1/2 > \varepsilon > 0} \frac{\varepsilon^2}{1 + \upsilon (\varepsilon; \mathbf{Z}_\alpha)^2} \asymp \sigma^{2/\gamma_k}.$$

Since

$$\frac{(\varepsilon_0)^2 \sigma^2}{\upsilon((\varepsilon_0); \mathbf{Z}_1)^2} = o\left(\sigma^{2/\gamma_k}\right),$$

we have (4.14)

Some preparation is needed for the proof of Theorem 4.13. Like in Duncan et al. (2000), we define $\varepsilon : \mathscr{L}_\phi^2 \to \mathscr{L}_1(\Omega, \mathscr{F}, P)$ as

$$\varepsilon(f) := \exp\left\{\int_0^T f(t)\, dZ_t - \frac{1}{2}|f|_\phi^2\right\},$$

and

$$\mathcal{E} = \left\{\sum_{k=1}^n a_k \varepsilon(f_k), n \in \mathbb{N}, a_k \in \mathbb{R}, f_k \in \mathscr{L}_\phi^2 \text{ for } k \in \{1, \ldots, n\}\right\}.$$

As in Duncan et al. (2000), for $f \in \mathscr{L}_\phi^2$, we have $\varepsilon(f) \in \mathscr{L}_p(\Omega, \mathscr{F}, P)$ for each $p \geq 1$. Following the same lines as in the proof of Theorem 3.1 in Duncan et al. (2000), we can prove the following lemma.

Lemma 4.29. *$\mathcal{E}$ is a dense set of $\mathscr{L}_p(\Omega, \mathscr{F}, P)$ for each $p \geq 1$.*

Proof of Theorem 4.13

For simplicity, we assume that $\sigma = 1$. By Corollary 4.9 we have $\inf_{\hat{L} \text{ affine}} R_{\mathcal{F}}\left(\hat{L}\right) = \rho_0(\mathcal{F})$. Suppose that $[\mathbf{x}_1, \mathbf{x}_2]$ is a one-dimensional subproblem such that

$$\inf_{\hat{L} \text{ affine}} R_{[\mathbf{x}_1, \mathbf{x}_2]}\left(\hat{L}\right) > \inf_{\hat{L} \text{ affine}} R_{\mathcal{F}}\left(\hat{L}\right) + \varepsilon.$$

We will show that

$$\frac{\inf_{\hat{L} \text{ affine}} R_{[\mathbf{x}_1, \mathbf{x}_2]}\left(\hat{L}\right)}{\inf_{\hat{T} \text{ measurable}} R_{[\mathbf{x}_1, \mathbf{x}_2]}\left(\hat{T}\right)} \leq 1.25.$$

Let $\hat{T} \in \mathscr{L}_1(\Omega, \mathscr{F}, P)$. Without loss of generality, we can assume that $\hat{T}$ is bounded. By Lemma 4.29, $\mathcal{E}$ is dense in $\mathscr{L}_4(\Omega, \mathscr{F}, P)$. Thus, we can find $X = \sum_{i=1}^r a_i \varepsilon(f_i)$ such that $E\left[\Delta^4\right] \leq \varepsilon^2$ and $E\left[\Delta^2\right] \leq \varepsilon$ where $\Delta = \hat{T} - X$. By Theorem 30.7 in Samko et al. (1993), we can find functions $g_1(t), g_2(t)$ on $[0, T]$ such that $\mathbf{x}_i(t) = \int_0^T \phi(t, u)\, g_i(u)\, du, i = 1, 2$. For $f \in [\mathbf{x}_1, \mathbf{x}_2]$, suppose that $f = \theta \mathbf{x}_1 + (1 - \theta)\mathbf{x}_2$. We have $f(t) = \int_0^T \phi(t, u)\, g(u)\, du$ where $g = \theta g_1 + (1 - \theta) g_2$. By Theorem

3.3 in Duncan et al. (2000)

$$
\begin{aligned}
&E\left\{\left|\Delta^2\left(Z_\cdot+\int_0^\cdot f(s)\,ds\right)\right|\right\}\\
&=E\left\{\Delta^2\left(Z_\cdot\right)e^{\int_0^T g(s)dZ_s-\frac{1}{2}|g|_\phi^2}\right\}\\
&\le\left(E\left(\Delta^4\right)\right)^{1/2}\left(E\left(\exp 2\left(\int_0^T g(s)\,dZ_s-\frac{1}{2}|g|_\phi^2\right)\right)\right)^{1/2}\\
&\le\varepsilon e^{\frac{1}{2}|g|_\phi^2}\\
&\le\varepsilon\exp\left(\frac{1}{2}\max\left\{|g_1|_\phi^2,|g_2|_\phi^2\right\}\right). \qquad (4.27)
\end{aligned}
$$

Let $X_i=\int_0^T f_i(t)\,dY(t)\,, i=1,\ldots,r$. Now, consider the minimax risk of estimating $L(f)$ by functions of $X_1,\ldots,X_r$, knowing that $f\in[\mathbf{x}_1,\mathbf{x}_2]$. Without loss of generality, we can assume that the f_i's are linearly independent. Under such assumptions, it is easy to show that $\mathbf{X}=(X_1,\ldots,X_r)'$ is a Gaussian vector with positive definite covariance matrix

$$
\Sigma=\begin{pmatrix}
|f_1|_\phi^2 & \langle f_1,f_2\rangle_\phi & \ldots & \langle f_1,f_r\rangle_\phi\\
\langle f_2,f_1\rangle_\phi & |f_2|_\phi^2 & \ldots & \langle f_2,f_r\rangle_\phi\\
\ldots & \ldots & \ldots & \ldots\\
\langle f_r,f_1\rangle_\phi & \langle f_r,f_2\rangle_\phi & \ldots & |f_r|_\phi^2
\end{pmatrix}
$$

and mean vector

$$
\mu_f=E_f(\mathbf{X})=\left(\int_0^T f_1(t)\,f(t)\,dt,\ldots,\int_0^T f_r(t)\,f(t)\,dt\right)'=(\theta\mu_1+(1-\theta)\,\mu_2)\,,
$$

where $\mu_i=E_{\mathbf{x}_i}(\mathbf{X})\,, i=1,2$. It is easy to see that $L(f)$ is a linear function of θ, and the estimation of $L(f)$ is equivalent to the estimation of θ. Since a sufficient statistic for θ is $S=\left(\mu_1^T-\mu_2^T\right)\Sigma^{-1}\mathbf{X}$, which is distributed

$$
N\left(\left(\mu_1^T-\mu_2^T\right)\Sigma^{-1}\mu_f,\left(\mu_1^T-\mu_2^T\right)\Sigma^{-1}(\mu_1-\mu_2)\right).
$$

Thus,

$$
\inf_{\eta\text{ measurable}}R_{[\mathbf{x}_1,\mathbf{x}_2]}(\eta(\mathbf{X}))=\inf_{\kappa\text{ measurable}}\sup_{\theta\in[0,1]}E_\theta\left\{(\kappa(S)-L(f))^2\right\}
$$

But, we know that

$$\frac{\inf_{\kappa \text{ affine}} \sup_{\theta\in[0,1]} E_\theta \left\{ \left(\kappa\left(S\right) - L\left(f\right)\right)^2 \right\}}{\inf_{\kappa \text{ measurable}} \sup_{\theta\in[0,1]} E_\theta \left\{ \left(\kappa\left(S\right) - L\left(f\right)\right)^2 \right\}} \leq 1.25,$$

and since $\kappa\left(S\right)$ is an affine estimator if κ is affine, we have

$$\frac{\inf_{\hat{L} \text{ affine}} R_{[\mathbf{x}_1,\mathbf{x}_2]}\left(\hat{L}\right)}{\inf_{\eta \text{ measurable}} R_{[\mathbf{x}_1,\mathbf{x}_2]}\left(\eta\left(\mathbf{X}\right)\right)} \leq 1.25. \tag{4.28}$$

Clearly, for $f \in [\mathbf{x}_1, \mathbf{x}_2]$,

$$E_f\left(\left(X\left(Z. + \int_0^{\cdot} f\left(s\right) ds\right) - L\left(f\right)\right)^2\right) \geq \inf_{\eta \text{ measurable}} R_{[\mathbf{x}_1,\mathbf{x}_2]}\left(\eta\left(\mathbf{X}\right)\right). \tag{4.29}$$

By (4.27) we have

$$\begin{aligned}
&\left(E\left(\left(X\left(Z. + \int_0^{\cdot} f\left(s\right) ds\right) - L\left(f\right)\right)^2\right)\right)^{1/2} \\
&\leq \left(E\left(\left(\hat{T}\left(Z. + \int_0^{\cdot} f\left(s\right) ds\right) - L\left(f\right)\right)^2\right)\right)^{1/2} + \left(E\left\{\left|\Delta^2\left(Z. + \int_0^{\cdot} f\left(s\right) ds\right)\right|\right\}\right)^{1/2} \\
&\leq \left(E\left(\left(\hat{T}\left(Z. + \int_0^{\cdot} f\left(s\right) ds\right) - L\left(f\right)\right)^2\right)\right)^{1/2} + \left(\varepsilon \exp\left(\frac{1}{2}\max\left\{\left|g_1\right|_\phi^2, \left|g_2\right|_\phi^2\right\}\right)\right)^{1/2} \\
&= \left(E\left(\left(\hat{T}\left(Z. + \int_0^{\cdot} f\left(s\right) ds\right) - L\left(f\right)\right)^2\right)\right)^{1/2} + O\left(\varepsilon^{1/2}\right).
\end{aligned}$$

This gives us

$$\begin{aligned}
&\inf_{\eta \text{ measurable}} R_{[\mathbf{x}_1,\mathbf{x}_2]}\left(\eta\left(\mathbf{X}\right)\right) \\
&\leq E\left(\left(\hat{T}\left(Z. + \int_0^{\cdot} f\left(s\right) ds\right) - L\left(f\right)\right)^2\right) + O\left(\varepsilon^{1/2}\right).
\end{aligned} \tag{4.30}$$

By (4.28), (4.29), and (4.30) we have

$$\frac{\inf_{\hat{L} \text{ affine}} R_{[\mathbf{x}_1,\mathbf{x}_2]}\left(\hat{L}\right)}{E\left(\left(\hat{T}\left(Z. + \int_0^{\cdot} f\left(s\right) ds\right) - L\left(f\right)\right)^2\right) + O\left(\varepsilon^{1/2}\right)} \leq 1.25.$$

Taking infimum over $f \in [\mathbf{x}_1, \mathbf{x}_2]$ we have

$$\frac{\inf_{\hat{L} \text{ affine}} R_{[\mathbf{x}_1,\mathbf{x}_2]}\left(\hat{L}\right)}{\inf_{\hat{T} \text{ measurable}} R_{[\mathbf{x}_1,\mathbf{x}_2]}\left(\hat{T}\right) + O\left(\varepsilon^{1/2}\right)} \leq 1.25.$$

Since the ε in the above inequality is arbitrary, the proof is completed.

For simplicity, we work with the cases in which $n = 2lm$ for some $l \geq 1$. The result can be easily generalized to the general cases. Let $\|\mathbf{K}_n f\|_2 = M$, we have

$$\min_{1\leq i\leq l}\left\{f^2\left(t_{2(j-1)l+i}\right)\right\} \leq \frac{1}{l}\Sigma_{i=1}^{l} f^2\left(t_{2(j-1)l+i}\right) \leq \frac{nM^2}{l} = 2mM^2, \tag{4.31}$$

where $j = 1, \ldots, m$. Let ξ_j, $j = 1, \ldots, m$ be such that $\min_{1\leq i\leq l}\left\{f^2\left(t_{2(j-1)l+i}\right)\right\} = f^2(\xi_j)$. We have $|\xi_i - \xi_j| \geq T/2m$ for any $1 \leq i, j \leq m$, $i \neq j$. Let $P_k(f)$ be the Lagrangian interpolation polynomial for $(\xi_i, f(\xi_i)), \ldots, (\xi_k, f(\xi_k))$, $k = 2, \ldots, m$. By (4.31) it can be shown that there exists a K independent of f such that

$$\begin{aligned}(P_k(f))^{(k-1)}(t) &\leq K \max\{|f(\xi_i)| : i = 1, \ldots, k\} \\ &\leq KM. \end{aligned} \tag{4.32}$$

Notice that we can find $\zeta_k \in (0, T), k = 1, \ldots m-1$ such that

$$f^{(k)}(\zeta_k) = (P_{k+1}(f))^{(k)}(\zeta_k).$$

With this observation and also by combining (4.32) and the fact that $\left\|f^{(m)}\right\|_p \leq C$, we have

$$\begin{aligned}\left|\left\|f^{(m-1)}\right\|_\infty - f^{(m-1)}(\zeta_k)\right| &\leq \int_0^T \left|f^{(m)}\right| dt \\ &\leq T^{1-p^{-1}}\left(\int_0^T \left|f^{(m)}\right|^p dt\right)^{1/p} \\ &= CT^{1-p^{-1}}.\end{aligned}$$

Thus

$$\left\|f^{(m-1)}\right\|_{\infty} \leq \left(P_m\left(f\right)\right)^{(m-1)}\left(\zeta_{m-1}\right) + CT^{1-p^{-1}}$$
$$= KM + CT^{1-p^{-1}}$$

Now, (4.17) can be readily shown by induction.

For the proof of Theorem 4.16, we need the following lemmas. Let $h = T/n$

Lemma 4.30. $\left|\sqrt{T}\left\|\mathbf{K}_n f\right\|_2 - \|f\|_2\right| = \left\|\mathbf{K}_n f\right\|_2 O_1\left(h\right) + O_2\left(h\right)$

Proof

For $t \in \left[t_i - h/2, t_i + h/2\right]$, $i = 1, \ldots, n$

$$f\left(t\right) = f\left(t_i\right) + \left(t - t_i\right) f'\left(\xi\right).$$

So,

$$\left|f\left(t\right) - f\left(t_i\right)\right| \leq \left|\left(t - t_i\right) f'\left(\xi\right)\right| \leq \frac{h}{2}\left(KM\left\|\mathbf{K}_n f\right\|_2 + CT^{m-1-p^{-1}}\right)$$

where K is as in Lemma 4.15.

$$\int_{t_i-h/2}^{t_i+h/2}\left|f\left(t\right) - f\left(t_i\right)\right|^2 dt \leq \left(\frac{h}{2}\left(KM\left\|\mathbf{K}_n f\right\|_2 + CT^{m-1-p^{-1}}\right)\right)^2 h.$$

Since

$$\left\|\mathbf{K}_n f\right\|_2 = \left(\frac{1}{n}\sum_{i=1}^{n} f^2\left(t_i\right)\right)^{1/2}$$
$$= \left(\frac{h}{T}\sum_{i=1}^{n} f^2\left(t_i\right)\right)^{1/2}$$
$$= \sqrt{\frac{1}{T}}\left(\sum_{i=1}^{n}\int_{t_i-h/2}^{t_i+h/2} f^2\left(t_i\right) dt\right)^{1/2}$$
$$= \sqrt{\frac{1}{T}}\left(\int_0^T f_n^2\left(t\right) dt\right)^{1/2} = \sqrt{\frac{1}{T}}\left\|f_n\right\|_2,$$

where function f_n is defined on $[0,T)$ as $f_n = f(t_i)$ for $t \in [t_i - h/2, t_i + h/2), i = 1, \ldots, n$. Thus,

$$\begin{aligned}\left|\sqrt{T}\,\|\mathbf{K}_n f\|_2 - \|f\|_2\right| &= \left|\|f_n\|_2 - \|f\|_2\right| \le \|f_n - f\|_2 \\ &= \left(\Sigma_{i=1}^n \int_{t_i-h/2}^{t_i+h/2} |f(t) - f(t_i)|^2\, dt\right)^{1/2} \\ &\le \left(n\left(\frac{h}{2}\left(K\,\|\mathbf{K}_n f\|_2 + CT^{m-2} + T^{m-1}\right)\right)^2 h\right)^{1/2} \\ &= \|\mathbf{K}_n f\|_2\, O_1(h) + O_2(h)\end{aligned}$$

Now we have the following result.

Lemma 4.31. *Suppose that $f \in \mathcal{F}$ with $\|f\|_2 \le M$ and $L(f)/2 = \varepsilon \ge Bn^{\frac{H-1}{\gamma_k}}$ for some positive number B. Then there exists B' independent of f and n such that for n sufficiently large and ε sufficiently small, we have $G_0\left(f; \tilde{\mathbf{Z}}_n, \mathbf{K}_n, \mathcal{F}\right) \ge B'\varepsilon^{\gamma_k} / n^{H-1}$.*

Proof

For any $f \in \mathcal{F}_M$, and $|L(f/2)| = \varepsilon$, we have

$$G_0\left(f; \tilde{\mathbf{Z}}_n, \mathbf{K}_n, \mathcal{F}\right) \ge \frac{\langle \mathbf{K}_n g, \mathbf{K}_n f/2\rangle}{\left\|\tilde{\mathbf{Z}}_n \mathbf{K}_n g\right\|}$$

where g is as in the proof of Lemma 4.11. Let $1 \le i_1 < i_2 \le n$ be the smallest and largest integer i such that $t_i \in [\tau_1, \tau_2]$ respectively. We know that

$$\lambda_k \delta' \le \tau_2 - \tau_1 \le \delta' = \min\{\delta, T/2\} \tag{4.33}$$

where

$$\delta = \begin{cases} (A/C)^{\frac{p}{p-1}} & \text{if } k = m-1, \\ A/\left\|f^{(k+1)}\right\|_\infty & \text{if } k < m-1, \end{cases}$$

with $A = \left\|f^{(k)}\right\|_\infty /2 \ge L(f)/2 = \varepsilon$. Then, we have

$$\langle \mathbf{K}_n g, \mathbf{K}_n f/2\rangle = \frac{1}{2n}\sum_{i=i_1}^{i_2} l^{-1/2} f(t_i)$$

where $l = \tau_2 - \tau_1$. Now we will show that $[\tau_1, \tau_2]$ becomes dense with t_i's as n goes to infinity. If

$\delta' = T/2$, then clearly the number of t_i's in the interval $[\tau_1, \tau_2]$ will go to infinity. Now assume that $\delta' = \delta$. First, assume that $k = m-1$. We then have $\delta' = (A/C)^{\frac{p}{p-1}} \geq (\varepsilon/C)^{\frac{p}{p-1}} \geq Kn^{\frac{H-1}{\gamma_k}\frac{p}{p-1}} = Kn^{\frac{(H-1)}{m-p^{-1}+1-H}}$. Since $\frac{(H-1)}{m-p^{-1}+1-H}$ is larger than -1, the t_i's will also become dense in $[\tau_1, \tau_2]$ as n goes to infinity. Next, if $k < m-1$, we have $\delta' = A/\left\|f^{(k+1)}\right\|_\infty = \left\|f^{(k)}\right\|_\infty /2\left\|f^{(k+1)}\right\|_\infty$. By Lemma 4.10, again we have that δ' is bigger than $Kn^{\frac{(H-1)}{m-p^{-1}+1-H}}$ for some constant K and sufficiently large n. Thus, again, the t_i's will become dense in $[\tau_1, \tau_2]$ as n goes to infinity.

We have,

$$
\begin{aligned}
&T\left\langle \mathbf{K}_n g, \mathbf{K}_n f/2\right\rangle - \langle g, f/2\rangle \\
&= \frac{T}{2n}\sum_{i=i_1}^{i_2} l^{-1/2} f(t_i) - \frac{1}{2}\int_{\tau_1}^{\tau_2} l^{-1/2} f(t)\,dt \\
&= \frac{T}{2n}\sum_{i=i_1}^{i_2} l^{-1/2} f(t_i) - \frac{1}{2} l^{-1/2}\left(\int_{\tau_1}^{t_{i_1}} f(t)\,dt + \sum_{i=i_1}^{i_2-1}\int_{t_i}^{t_{i+1}} f(t)\,dt + \int_{t_{i_2}}^{\tau_2} f(t)\,dt\right) \\
&= \frac{1}{2l^{1/2}}\sum_{i=i_1}^{i_2-1}\int_{t_i}^{t_{i+1}} f(t_i)\,dt + \frac{T}{2nl^{1/2}} f(t_{i_2}) \\
&- \frac{1}{2l^{1/2}}\left(\int_{\tau_1}^{t_{i_1}} f(t)\,dt + \sum_{i=i_1}^{i_2-1}\int_{t_i}^{t_{i+1}} f(t)\,dt + \int_{t_{i_2}}^{\tau_2} f(t)\,dt\right) \\
&= \frac{1}{2l^{1/2}}\sum_{i=i_1}^{i_2-1}\int_{t_i}^{t_{i+1}} \left(f(t_i) - f(t)\right)dt + \frac{T}{2nl^{1/2}} f(t_{i_2}) \\
&- \frac{1}{2l^{1/2}}\left(\int_{\tau_1}^{t_{i_1}} f(t)\,dt + \int_{t_{i_2}}^{\tau_2} f(t)\,dt\right).
\end{aligned}
$$

So, that

$$
\begin{aligned}
&\left|T\left\langle \mathbf{K}_n g, \mathbf{K}_n f/2\right\rangle - \langle g, f/2\rangle\right| \\
&\leq \frac{1}{2l^{1/2}}\left(\sum_{i=i_1}^{i_2-1}\int_{t_i}^{t_{i+1}} \left|f(t_i) - f(t)\right| dt + \frac{T}{n}\left|f(t_{i_2})\right| + \int_{\tau_1}^{t_{i_1}} |f(t)|\,dt + \int_{t_{i_2}}^{\tau_2} |f(t)|\,dt\right) \\
&\leq \frac{1}{2l^{1/2}}\left(l\frac{T}{n}\left\|f'\right\|_\infty + \frac{3T}{n}\|f\|_\infty\right) = l^{1/2}\frac{T}{2n}\left\|f'\right\|_\infty + \frac{3T}{2l^{1/2}n}\|f\|_\infty. \qquad (4.34)
\end{aligned}
$$

From Lemma 4.10 we know that $\lim_{\varepsilon\to 0}\sup\left\{\|f\|_\infty : \|f\|_2 = \varepsilon\right\} = 0$. From this and the fact that $\sup\left\{\|f\|_\infty : \|f\|_2 = \varepsilon\right\}$ is a concave function of ε, we know that

$$\sup\left\{\|f\|_\infty : \|f\|_2 = \varepsilon\right\}$$

is finite for every ε. And since $\|f\|_2 \leq M$, we have

$$\|f\|_\infty \leq M_1 = \sup\left\{\|f\|_\infty : \|f\|_2 = M\right\} < \infty.$$

Similarly, we have

$$\|f'\|_\infty \leq M_2 = \sup\left\{\|f'\|_\infty : \|f\|_2 = M\right\} < \infty.$$

It is shown in the proof of Lemma 4.11 that $|\langle g, f\rangle| \geq l^{1/2}\mu_k A\delta'^k$. Hence, (4.34) gives

$$\begin{aligned}
&\left|\frac{T\left\langle \mathbf{K}_n g, \mathbf{K}_n f/2\right\rangle - \langle g, f/2\rangle}{\langle g, f/2\rangle}\right| \\
&\leq \frac{l^{1/2}\frac{T}{n}M_2 + \frac{3T}{l^{1/2}n}M_1}{l^{1/2}\mu_k A\delta'^k} \\
&= \frac{TM_2}{n\mu_k A\delta'^k} + \frac{3TM_1}{ln\mu_k A\delta'^k} \\
&\leq \frac{TM_2}{n\mu_k A\delta'^k} + \frac{3TM_1}{\lambda_k n\mu_k A\delta'^{k+1}} \qquad (4.35)
\end{aligned}$$

If $k = m-1$, we have

$$\delta = (A/C)^{\frac{p}{p-1}}$$

So that, (4.35) becomes

$$\begin{aligned}
&\left|\frac{T\left\langle \mathbf{K}_n g, \mathbf{K}_n f/2\right\rangle - \langle g, f/2\rangle}{\langle g, f/2\rangle}\right| \\
&\leq \max\left\{\frac{TM_2}{n\mu_{m-1}A\,(T/2)^{m-1}}, \frac{TM_2}{n\mu_{m-1}A\left((A/C)^{\frac{p}{p-1}}\right)^{m-1}}\right\} \\
&+ \max\left\{\frac{3TM_1}{\lambda_{m-1}n\mu_{m-1}A\,(T/2)^m}, \frac{3TM_1}{\lambda_{m-1}n\mu_{m-1}A\left((A/C)^{\frac{p}{p-1}}\right)^m}\right\} \\
&\leq \max\left\{K_2 n^{\frac{1-H}{\gamma_{m-1}}-1}, K_3 n^{\frac{1-H}{\gamma_{m-1}}\left(\frac{(m-1)p}{p-1}+1\right)-1}\right\} + \max\left\{K_4 n^{\frac{1-H}{\gamma_{m-1}}-1}, K_5 n^{\frac{1-H}{\gamma_{m-1}}\left(\frac{mp}{p-1}+1\right)-1}\right\} \\
&= O\left(n^{\frac{1-H}{\gamma_{m-1}}\left(\frac{mp}{p-1}+1\right)-1}\right) \\
&= O\left(n^\beta\right)
\end{aligned}$$

where $\beta = \frac{Hp^{-1}-Hm}{m-p^{-1}+1-H}$. If $k < m-1$, we have

$$\delta = A \Big/ \left\|f^{(k+1)}\right\|_\infty ,$$

For ε sufficiently small, by (4.35) and Lemma 4.10 we have

$$\begin{aligned}
&\left|\frac{T\left\langle \mathbf{K}_n g, \mathbf{K}_n f/2\right\rangle - \langle g, f/2\rangle}{\langle g, f/2\rangle}\right| \\
&\le \frac{TM_2}{n\mu_k A\delta'^k} + \frac{3TM_1}{\lambda_k n\mu_k A\delta'^{k+1}} \\
&\le \max\left\{K_2 n^{\frac{1-H}{\gamma_k}-1}, \frac{TM_2}{n\mu_k A\left(A\Big/\left\|f^{(k+1)}\right\|_\infty\right)^k}\right\} \\
&+ \max\left\{K_4 n^{\frac{1-H}{\gamma_k}-1}, \frac{3TM_1}{\lambda_k n\mu_k A\left(A\Big/\left\|f^{(k+1)}\right\|_\infty\right)^{k+1}}\right\} && (4.36)\\
&\le \max\left\{K_2 n^{\frac{1-H}{\gamma_k}-1}, K_3' n^{\left(\frac{1-H}{\gamma_k}\right)\left(\frac{k}{m-k-p^{-1}}+1\right)-1}\right\} \\
&+ \max\left\{K_4 n^{\frac{1-H}{\gamma_k}-1}, K_5' n^{\left(\frac{1-H}{\gamma_k}\right)\left(\frac{k+1}{m-k-p^{-1}}+1\right)-1}\right\} \\
&= O\left(n^{\left(\frac{1-H}{\gamma_k}\right)\left(\frac{k+1}{m-k-p^{-1}}+1\right)-1}\right) \\
&= O\left(n^\beta\right) && (4.37)
\end{aligned}$$

Since we assume that either $m > 1$ or $p > 1$, β is always negative. Now notice that

$$\begin{aligned}
\left\|\tilde{\mathbf{Z}}_n \mathbf{K}_n g\right\|^2 &= \frac{1}{n^2} l^{-1} \sum_{j=i_1}^{i_2} \sum_{i=i_1}^{i_2} R(i-j) \\
&= \frac{1}{n^2} n'^2 \left(n'\right)^{2H-2} l^{-1} \sum_{j=i_1}^{i_2} \sum_{i=i_1}^{i_2} \frac{1}{n'^2} R(i-j) \left(\frac{1}{n'}\right)^{2H-2} .
\end{aligned}$$

where $n' = i_2 - i_1 + 1$. Let

$$S = \sum_{j=i_1}^{i_2} \sum_{i=i_1}^{i_2} \frac{1}{n'^2} R(i-j) \left(\frac{1}{n'}\right)^{2H-2} ,$$

and we can see that

$$S \to C_1 \int_0^1 \int_0^1 |u-v|^{2H-2}\, dudv \tag{4.38}$$

as n' goes to infinity. In fact, we define the function

$$f_{n'}(u,v) = \begin{cases} R(i-j)\left(\frac{1}{n'}\right)^{2H-2} & \text{if } (u,v) \in \left(\frac{i-1}{n'}, \frac{i}{n'}\right] \times \left(\frac{j-1}{n'}, \frac{j}{n'}\right] \text{ for } i = 1, \ldots, n'; \\ 0 & \text{otherwise.} \end{cases}$$

Then, we can see that $f_{n'}(u,v) \to C_1 |u-v|^{2H-2}$ a.e. where on $[0,1]\times[0,1]$, and $S = \int_0^1 \int_0^1 f_{n'}(u,v)\, dudv$. Further more, Since $R(k) \sim C_1 |k|^{2H-2}$, we can find K such that $R(k) \le K|k|^{2H-2}$ for $k > 0$. Now, letting $K' = \max\left\{2^{2-2H}K, R(0)\right\}$, we have $R(k) \le K'(|k|+1)^{2H-2}$ for $k \ge 0$. Thus,

$$R(i-j)\left(\frac{1}{n'}\right)^{2H-2} \le K'\left(\frac{|i-j|+1}{n'}\right)^{2H-2} \le K'|u-v|^{2H-2}$$

for $(u,v) \in \left(\frac{i-1}{n'}, \frac{i}{n'}\right] \times \left(\frac{j-1}{n'}, \frac{j}{n'}\right]$, which means that the function $f_n(u,v)$ is dominated by $K'|u-v|^{2H-2}$. By dominant convergence theorem (4.38) follows. Now, letting $\sigma_n = \frac{C_1^{1/2} T^{1-H} n^{H-1}}{(H(2H-1))^{1/2}}$, we have

$$\begin{aligned}
&\frac{\left|T^2 \left\|\tilde{\mathbf{Z}}_n \mathbf{K}_n g\right\|^2 - \left\|\mathbf{Z}_{\sigma_n} g\right\|^2\right|}{\left\|\mathbf{Z}_{\sigma_n} g\right\|^2} \\
&= \frac{\left|T^2 n^{-2} n'^2 n'^{2H-2} l^{-1} S - \sigma_n^2 H(2H-1) l^{-1} \int_0^l \int_0^l |u-v|^{2H-2}\, dudv\right|}{\sigma_n^2 l^{2H-1}} \\
&= \frac{n^{2H-2} T^{2-2H} \left(\frac{Tn'}{n}\right)^{2H} l^{-1} S - \sigma_n^2 H(2H-1) l^{-1} l^{2H-2} l^2 \int_0^l \int_0^l \left|\frac{u-v}{l}\right|^{2H-2} d\frac{u}{l} d\frac{v}{l}}{\sigma_n^2 l^{2H-1}} \\
&= \frac{n^{2-2H} T^{2-2H} \left(\frac{n'T}{n}\right)^{2H} l^{-1} S - \sigma_n^2 H(2H-1) l^{2H-1} \int_0^1 \int_0^1 |s-t|^{2H-2}\, dsdt}{\sigma_n^2 l^{2H-1}} \\
&= \frac{n^{2-2H} T^{2-2H} l^{2H-1} \left(\left(\frac{n'T}{nl}\right)^{2H} S - C_1 \int_0^1 \int_0^1 |s-t|^{2H-2}\, dsdt\right)}{\sigma_n^2 l^{2H-1}} \\
&= \frac{n^{2-2H} T^{2-2H} \left(\left(\frac{n'T}{nl}\right)^{2H} S - C_1 \int_0^1 \int_0^1 |s-t|^{2H-2}\, dsdt\right)}{\sigma_n^2} \\
&= C_1^{-1} (H(2H-1))^{-1} \left(\left(\frac{n'T}{nl}\right)^{2H} S - C_1 \int_0^1 \int_0^1 |s-t|^{2H-2}\, dsdt\right).
\end{aligned}$$

Hence, using (4.38) and $\frac{n'T}{nl} \to 1$ we have

$$\left| \frac{T^2 \left\| \tilde{\mathbf{Z}}_n g_n \right\|^2 - \left\| \mathbf{Z}_{\sigma_n} g \right\|^2}{\left\| \mathbf{Z}_{\sigma_n} g \right\|^2} \right| \to 0.$$

Since

$$\begin{aligned} \left| \frac{T^2 \left\| \tilde{\mathbf{Z}}_n g_n \right\|^2 - \left\| \mathbf{Z}_{\sigma_n} g \right\|^2}{\left\| \mathbf{Z}_{\sigma_n} g \right\|^2} \right| &= \left| \frac{\left(T \left\| \tilde{\mathbf{Z}}_n g_n \right\| - \left\| \mathbf{Z}_{\sigma_n} g \right\| \right) \left(T \left\| \tilde{\mathbf{Z}}_n g_n \right\| + \left\| \mathbf{Z}_{\sigma_n} g \right\| \right)}{\left\| \mathbf{Z}_{\sigma_n} g \right\|^2} \right| \\ &\geq \left| \frac{T \left\| \tilde{\mathbf{Z}}_n g_n \right\| - \left\| \mathbf{Z}_{\sigma_n} g \right\|}{\left\| \mathbf{Z}_{\sigma_n} g \right\|} \right|, \end{aligned}$$

we have

$$\left| \frac{T \left\| \tilde{\mathbf{Z}}_n g_n \right\| - \left\| \mathbf{Z}_{\sigma_n} g \right\|}{\left\| \mathbf{Z}_{\sigma_n} g \right\|} \right| \to 0,$$

which also gives

$$\frac{T \left\| \tilde{\mathbf{Z}}_n g_n \right\|}{\left\| \mathbf{Z}_{\sigma_n} g \right\|} \to 1.$$

Hence,

$$\begin{aligned} &\frac{\left| \frac{\langle \mathbf{K}_n g, \mathbf{K}_n f/2 \rangle}{\left\| \tilde{\mathbf{Z}}_n \mathbf{K}_n g \right\|} - \frac{\langle g, f/2 \rangle}{\left\| \mathbf{Z}_{\sigma_n} g \right\|} \right|}{\left| \frac{\langle g, f/2 \rangle}{\left\| \mathbf{Z}_{\sigma_n} g \right\|} \right|} \\ &= \frac{\left| \frac{T \langle \mathbf{K}_n g, \mathbf{K}_n f/2 \rangle}{T \left\| \tilde{\mathbf{Z}}_n \mathbf{K}_n g \right\|} - \frac{\langle g, f/2 \rangle}{\left\| \mathbf{Z}_{\sigma_n} g \right\|} \right|}{\left| \frac{\langle g, f/2 \rangle}{\left\| \mathbf{Z}_{\sigma_n} g \right\|} \right|} = \frac{\left| \frac{T \langle \mathbf{K}_n g, \mathbf{K}_n f/2 \rangle \left\| \mathbf{Z}_{\sigma_n} g \right\| - T \left\| \tilde{\mathbf{Z}}_n \mathbf{K}_n g \right\| \langle g, f/2 \rangle}{T \left\| \tilde{\mathbf{Z}}_n \mathbf{K}_n g \right\| \left\| \mathbf{Z}_{\sigma_n} g \right\|} \right|}{\left| \frac{\langle g, f/2 \rangle}{\left\| \mathbf{Z}_{\sigma_n} g \right\|} \right|} \\ &= \left| \frac{T \langle \mathbf{K}_n g, \mathbf{K}_n f/2 \rangle \left\| \mathbf{Z}_{\sigma_n} g \right\| - T \left\| \tilde{\mathbf{Z}}_n \mathbf{K}_n g \right\| \langle g, f/2 \rangle}{T \left\| \tilde{\mathbf{Z}}_n \mathbf{K}_n g \right\| \langle g, f/2 \rangle} \right| \\ &= \left| \frac{\begin{array}{c} \left(T \langle \mathbf{K}_n g, \mathbf{K}_n f/2 \rangle \left\| \mathbf{Z}_{\sigma_n} g \right\| - \langle g, f/2 \rangle \left\| \mathbf{Z}_{\sigma_n} g \right\| \right) \\ + \left(\langle g, f/2 \rangle \left\| \mathbf{Z}_{\sigma_n} g \right\| - T \left\| \tilde{\mathbf{Z}}_n \mathbf{K}_n g \right\| \langle g, f/2 \rangle \right) \end{array}}{T \left\| \tilde{\mathbf{Z}}_n \mathbf{K}_n g \right\| \langle g, f/2 \rangle} \right| \end{aligned}$$

$$= \left| \frac{\|\mathbf{Z}_{\sigma_n} g\| \left(T \langle \mathbf{K}_n g, \mathbf{K}_n f/2\rangle - \langle g, f/2\rangle\right) + \langle g, f/2\rangle \left(\|\mathbf{Z}_{\sigma_n} g\| - T \left\|\tilde{\mathbf{Z}}_n \mathbf{K}_n g\right\|\right)}{\left\|\tilde{\mathbf{Z}}_n \mathbf{K}_n g\right\| \langle g, f/2\rangle} \right|$$

$$= \left| \frac{\|\mathbf{Z}_{\sigma_n} g\|}{\left\|\tilde{\mathbf{Z}}_n \mathbf{K}_n g\right\|} \frac{T \langle \mathbf{K}_n g, \mathbf{K}_n f/2\rangle - \langle g, f/2\rangle}{\langle g, f/2\rangle} + \frac{\|\mathbf{Z}_{\sigma_n} g\| - T \left\|\tilde{\mathbf{Z}}_n \mathbf{K}_n g\right\|}{\left\|\tilde{\mathbf{Z}}_n \mathbf{K}_n g\right\|} \right| \to 0.$$

In the proof of Lemma 4.11 we have shown that

$$\frac{\langle g, f/2\rangle}{\|\mathbf{Z}_{\sigma_n} g\|} \geq K_1 \varepsilon^\gamma / \sigma_n,$$

for some positive K_1 independent of f , and ε sufficiently small. Thus, for n sufficiently large and ε sufficiently small we have

$$G_0\left(f; \tilde{\mathbf{Z}}_n, \mathbf{K}_n, \mathcal{F}\right) \geq \frac{\langle \mathbf{K}_n g, \mathbf{K}_n f/2\rangle}{\left\|\tilde{\mathbf{Z}}_n \mathbf{K}_n g\right\|} \geq K'' \varepsilon^\gamma / \sigma_n = B' \varepsilon^\gamma / n^{H-1} .$$

The following lemma is the discrete version of Lemma 4.28. Since its proof is similar to that of Lemma 4.28 except that integrations are replaced by summation, we omitted it here.

Lemma 4.32. *For* $\mathbf{x} = (x_1, \ldots, x_n)^T \in \mathbb{R}^n$*, and* $1 \leq i_1 \leq i_2 \leq n$*, we can find a positive* A_d *independent of* n *and* $\mathbf{x}$ *such that*

$$\sum_{j=1}^{n} \sum_{i=1}^{n} x_i x_j R(i-j) \geq A_d (i_2 - i_1 + 1)^{2H-2} \left| \sum_{i=i_1}^{i_2} x_i \right|^2$$

Proof of Theorem 4.16

By Corollary 4.9 we have

$$\inf_{\hat{L} \text{ affine}} R_{\mathbf{W}_{[0,T]}(m,p,C)}\left(\hat{L}; \mathbf{K}_n, \tilde{\mathbf{Z}}_n\right) = \rho_0\left(\mathcal{F}; \mathbf{K}_n, \tilde{\mathbf{Z}}_n\right).$$

By Lemma 4.15, there exists M such that if $\|f\|_2 > M$, then $\|\mathbf{K}_n f\|_2 > M' > 0$. Hence,

$$\rho_0\left(\mathcal{F}; \mathbf{K}_n, \tilde{\mathbf{Z}}_n\right) = \sup_{f \in \mathcal{F} - \mathcal{F}} \rho_0\left(f; \mathcal{F}, \tilde{\mathbf{Z}}_n, L, \mathbf{K}_n\right)$$

$$= \max\{\rho_1, \rho_2, \rho_3\}$$

where $\rho_i = \sup \{\rho_0(f) : f \in \mathcal{F}_i\}$, $i = 1, 2, 3$ with

$$\mathcal{F}_1 = \{f \in \mathcal{F} - \mathcal{F} : \|f\|_2 > M\},$$

$$\mathcal{F}_2 = \left\{f \in \mathcal{F} - \mathcal{F} : |L(f)/2| < Kn^{\frac{H-1}{\gamma}}\right\},$$

and

$$\mathcal{F}_3 = \left\{f \in \mathcal{F} - \mathcal{F} : \|f\|_2 \le M, |L(f)/2| \ge Kn^{\frac{H-1}{\gamma}}\right\}.$$

By Lemma 4.31 we have

$$\begin{aligned}
\rho_3 &\le \sup_{\varepsilon \ge Kn^{\frac{H-1}{\gamma}}, f \in \mathcal{F}_3} \frac{\varepsilon^2}{1 + G_0(f)^2} \\
&\le \sup_{\varepsilon \ge Kn^{\frac{H-1}{\gamma}}} \frac{\varepsilon^2}{1 + \left(B_1 \frac{\varepsilon^\gamma}{n^{H-1}}\right)^2} \\
&\le \sup_{\varepsilon} \frac{\varepsilon^2}{1 + \left(B_1 \frac{\varepsilon^\gamma}{n^{H-1}}\right)^2} = O\left(n^{(2H-2)/\gamma}\right)
\end{aligned}$$

and

$$\begin{aligned}
\rho_1 &\le \sup_{f \in \mathcal{F}_1} \frac{(L(f)/2)^2}{1 + \frac{\|\mathbf{K}_n f/2\|_2^2}{\|\tilde{\mathbf{Z}}_n\|^2}}. \\
&= \sup_{f \in \mathcal{F}_1} \frac{\frac{(L(f))^2}{\|\mathbf{K}_n f\|_2^2}}{\frac{4}{\|\mathbf{K}_n f\|_2^2} + \frac{1}{\|\tilde{\mathbf{Z}}_n\|^2}}
\end{aligned}$$

By Lemma 4.15, we can find K such that $\left\|f^{(k)}\right\|_\infty \le K \|\mathbf{K}_n f\|_2$. Hence

$$\frac{|L(f)|}{\|\mathbf{K}_n f\|_2} = \frac{\left|f^{(k)}(t_0)\right|}{\|\mathbf{K}_n f\|_2} \le \frac{\left\|f^{(k)}\right\|_\infty}{\|\mathbf{K}_n f\|_2} \le K.$$

Therefore,

$$\rho_1 \le \sup_{f \in \mathcal{F}_2} \frac{K^2}{\frac{1}{\|\tilde{\mathbf{Z}}_n\|^2}} = K^2 \left\|\tilde{\mathbf{Z}}_n\right\|^2 = O\left(n^{2H-2}\right).$$

Also,

$$\rho_2 \le \left(Kn^{\frac{H-1}{\gamma}}\right)^2 = O\left(n^{\frac{2H-2}{\gamma}}\right).$$

Thus, $\max\{\rho_1, \rho_2, \rho_3\} = O\left(n^{\frac{2H-2}{\gamma}}\right)$. This prove one side of the inequality.

Now, let f_δ be as defined in the proof of Lemma 4.11, and let

$$i_1 = \min\{1 \le i \le n : t_i \ge t_0 - t_0\delta\}$$

and

$$i_2 = \max\{1 \le i \le n : t_i \le t_0 + (1 - t_0)\,\delta\}.$$

For any $\mathbf{x} \in \mathbb{R}^n$, we have

$$\begin{aligned}
\langle \mathbf{x}, \mathbf{K}_n f_\delta \rangle &= \frac{1}{n} \sum_{i_1}^{i_2} f_\delta(t_i)\, x_i \\
&= \frac{1}{n} \left(f_\delta(t_{i_2}) \left(\sum_{i=i_1}^{i_2} x_i \right) - \sum_{i=i_1}^{i_2-1} \left(f_\delta(t_{i+1}) - f_\delta(t_i) \right) \left(\sum_{j=i_1}^{i} x_i \right) \right) \\
&\le \frac{1}{n} \left(\sum_{i=i_1}^{i_2-1} \left| f_\delta(t_{i+1}) - f_\delta(t_i) \right| \right) \sup_{i=i_1}^{i_2-1} \left| \sum_{j=i_1}^{i} x_i \right|
\end{aligned}$$

Suppose that $i_1 \le j_1 \le j_2 \le i_2 - 1$, and $\left| \sum_{i=j_1}^{j_i} x_i \right| = \sup_{i=i_1}^{i_2-1} \left| \sum_{j=i_1}^{i} x_i \right|$. By Lemma 4.32 we have

$$\begin{aligned}
\left\| \tilde{\mathbf{Z}}_n \mathbf{x} \right\|^2 &= \frac{1}{n^2} \sum_{j=1}^{n} \sum_{i=1}^{n} x_i x_j R(i-j) \\
&\ge \frac{1}{n^2} A_d \left(j_2 - j_1 + 1 \right)^{2H-2} \left| \sum_{i=j_1}^{j_2} x_i \right|^2 \\
&\ge \frac{1}{n^2} A_d \left(i_2 - i_1 \right)^{2H-2} \sup_{i=i_1}^{i_2-1} \left| \sum_{j=i_1}^{i} x_i \right|^2
\end{aligned}$$

So,

$$G_0(f_\delta) = \sup_{\mathbf{x} \in \mathbb{R}^n} \frac{\langle \mathbf{x}, \mathbf{K}_n f_\delta / 2 \rangle}{\left\| \tilde{\mathbf{Z}}_n \mathbf{x} \right\|} \le A_d^{1/2} \sum_{i=i_1}^{i_2-1} \left| f_\delta(t_{i+1}) - f_\delta(t_i) \right| \left(i_2 - i_1 \right)^{1-H}. \tag{4.39}$$

Since

$$\begin{aligned}
&\sum_{i=i_1}^{i_2-1}\left|f_\delta\left(t_{i+1}\right)-f_\delta\left(t_i\right)\right|-\int_{t_0-t_0\delta}^{t_0+(1-t_0)\delta}\left|f_\delta'\left(t\right)\right|dt\\
&=\frac{T}{n}\sum_{i=i_1}^{i_2-1}f_\delta'\left(\xi_i\right)-\int_{t_0-t_0\delta}^{t_0+(1-t_0)\delta}f_\delta'\left(t\right)dt\\
&\le\frac{T}{n}\int_{t_0-t_0\delta}^{t_0+(1-t_0)\delta}\left\|f_\delta''\right\|_\infty dt=\frac{T}{n}\delta\left\|f_\delta''\right\|_\infty\\
&=\frac{T}{n}\delta^{m-1-p^{-1}}\left\|f''\right\|_\infty,
\end{aligned}$$

by (4.39)

$$\begin{aligned}
G_0\left(f_\delta\right)&\le A_d^{-1/2}\left(\left\|f_\delta'\right\|_1+\frac{T}{n}\delta^{m-1-p^{-1}}\left\|f''\right\|_\infty\right)\left(i_2-i_1\right)^{1-H}\\
&=A_d^{-1/2}\left(\delta^{m-p^{-1}}\left\|f'\right\|_1+\frac{T}{n}\delta^{m-1-p^{-1}}\left\|f''\right\|_\infty\right)\left(i_2-i_1\right)^{1-H}.
\end{aligned}$$

Now let $\delta=n^{\frac{H-1}{m-p^{-1}+1-H}}$. Since $i_2-i_1\sim n\delta$, we have

$$G_0\left(f_\delta\right)=O\left(1\right).$$

Thus,

$$\begin{aligned}
\inf_{\hat{L}\text{ affine}}R_{\mathbf{W}_{[0,T]}(m,p,C/2)}\left(\hat{L};\mathbf{K}_n,\tilde{\mathbf{Z}}_n\right)&=\rho_0\left(\mathbf{W}_{[0,T]}\left(m,p,C/2\right)\right)\\
&\ge\frac{\left(L\left(f_\delta^{(k)}/2\right)\right)^2}{1+\left(G_0\left(f\right)\right)^2}\\
&=\frac{L\left(f_\delta^{(k)}/2\right)}{1+\left(G_0\left(f\right)\right)^2}\\
&=\frac{\left(\delta^{m-k-p^{-1}}\right)^2}{1+\left(G_0\left(f\right)\right)^2}\\
&=\frac{n^{2\left(\frac{(H-1)\left(m-k-p^{-1}\right)}{m-p^{-1}+1-H}\right)}}{1+\left(G_0\left(f\right)\right)^2}\\
&=O\left(n^{\frac{2H-2}{\gamma_k}}\right).
\end{aligned}$$

Chapter 5

Estimation of survival functions based on doubly censored data

5.1 Introduction

The estimation of the distribution function has been of special interest in biometry, reliability, and medical follow-up studies. In these fields, incomplete data are frequently encountered. One example is the case of doubly censored data. In such a case, the random variable of interest (X) could be censored either from the right by a right censoring variable (U) or from the left by a left censoring variable (V). Thus, the observations have the form (Z, δ) where $Z = \max\{\min\{X, U\}, V\}$ and

$$\delta = \begin{cases} 1, \text{ if } V < X \leq U \\ 2, \text{ if } X > U \\ 3, \text{ if } X \leq V. \end{cases}$$

We use $(X_i, U_i, V_i, Z_i, \delta_i)$, $i = 1, \ldots, n$ to denote a sample of size n. In the literature (Yu and Li, 2001; Gu and Zhang, 1993; Zhang and Li, 1996; Turnbull, 1974b), self-consistent estimators are frequently used for the estimation of the distribution function. Let S_X, S_U, S_V be the survival function of the random variables X, U, and V respectively. Define $Q^{(i)}(t) = \Pr\{Z > t, \delta = i\}$, $i = 1, 2, 3$ and

$Q^{(0)}(t) = \sum_{i=1}^{3} Q^{(i)}(t)$. We have

$$\begin{aligned} Q^{(0)}(t) &= \Pr\{Z > t\} = \Pr\{V > t\} + \Pr\{V \leq t, U > t, X > t\} \\ &= S_V(t) + S_X(t)\left(S_U(t) - S_V(t)\right), \end{aligned}$$

$$\begin{aligned} Q^{(2)}(t) &= \Pr\{Z > t, \delta = 2\} \\ &= -\int_{t<u} S_X(u)\, dS_U(u), \end{aligned}$$

and

$$\begin{aligned} Q^{(3)}(t) &= \Pr\{Z > t, \delta = 3\} \\ &= -\int_{t<u} \left(1 - S_X(u)\right) dS_V(u). \end{aligned}$$

Using the above relationships, we can write

$$\begin{aligned} & Q^{(0)}(t) - \int_{u\leq t} \frac{S_X(t)}{S_X(u)} dQ^{(2)}(u) + \int_{t<u} \frac{1 - S_X(t)}{1 - S_X(u)} dQ^{(3)}(u) \\ &= Q^{(0)}(t) - \int_{u\leq t} S_X(t)\, dS_U(t) + \int_{t<u} \left(1 - S_X(t)\right) dS_V(u) \\ &= S_V(t) + S_X(t)\left(S_U(t) - S_V(t)\right) + S_X(t) F_U(t) - \left(1 - S_X(t)\right) S_V(t) \\ &= S_X(t). \end{aligned} \tag{5.1}$$

Equation (5.1) is usually referred to as the self-consistency equation. After replacing the $Q^{(i)}$ functions by their respective empirical versions as estimators, the following equation can be established

$$S_n(t) = Q_n^{(0)}(t) - \int_{u\leq t} \frac{S_n(t)}{S_n(u)} dQ_n^{(2)}(u) + \int_{t<u} \frac{1 - S_n(t)}{1 - S_n(u)} dQ_n^{(3)}(u). \tag{5.2}$$

A solution $S_n(\cdot)$ to (5.2) is called a self-consistent estimator for S_X. The existence of self-consistent estimators, as well as their strong consistency and asymptotic normality has been widely discussed in the literature. Chang and Yang (1987), Samuelsen (1989), Gu and Zhang (1993), and Yu and Li (2001) proved the consistency and/or asymptotic normality of self-consistent estimators, each under different assumptions.

In addition to the self-consistent estimators, a popular technique for estimating the survival function for doubly censored data is the nonparametric likelihood method. To define the nonparametric likelihood, we proceed as follows. Let S be the survival function. Then it can be seen that the joint distribution of the data can be written as a functional of S as

$$L_1(S) = C\prod_{i=1}^{n}\left(S(Z_i-) - S(Z_i)\right)^{I(\delta_i=1)} S(Z_i)^{I(\delta_i=2)} \left(1 - S(Z_i)\right)^{I(\delta_i=3)}$$

where C is a term that does not depend on S. Then, the nonparametric maximum likelihood estimator (NPMLE) for S_X is the survival function that maximizes

$$L(S) = \prod_{i=1}^{n}\left(S(Z_i-) - S(Z_i)\right)^{I(\delta_i=1)} S(Z_i)^{I(\delta_i=2)} \left(1 - S(Z_i)\right)^{I(\delta_i=3)}. \tag{5.3}$$

Mykland and Ren (1996) proposed an algorithm for the computation of a self-consistent estimator and the NPMLE of the survival function S. It has been shown that for doubly censored data, the NPMLE is also a self-consistent estimator, but not necessarily vice versa.

The above mentioned estimators have a few restrictions. The previous researchers only discussed self-consistency equations like equation (5.2) above, in which empirical processes are used as estimators for the survival function or partial survival functions ($Q^{(i)}$, $i = 1, 2, 3$) based on the observed Z and δ. There are many occasions where other types of estimators are preferred. For example, if it is reasonable to assume that S_X is smooth, then it is preferred to obtain a smooth estimator such as a kernel type estimator for S_X. In fact, it has been shown by Falk (1983) and Lemdani and Ould-Said (2001) that when the underlying S_X is smooth, under certain regularity conditions, kernel type smooth estimators are superior to the empirical estimators in a MSE sense.

To examine the use of smooth estimators of $Q^{(i)}$, we discuss general self-consistency equations that are not restricted to using empirical processes as estimators for the $Q^{(i)}$'s. In particular, we consider equations of the form

$$r(t) = \mu^{(0)}((t,\infty]) + \int_{u\le t}\frac{r(t)}{r(u)}d\mu^{(2)}(u) - \int_{t<u}\frac{1-r(t)}{1-r(u)}d\mu^{(3)}(u), \quad t\in[-\infty,\infty] \tag{5.4}$$

where the $\mu^{(i)}$'s are positive measures on $[-\infty,\infty]$ for $i = 0,\ldots,3$ with $\mu^{(0)} = \sum \mu^{(i)}$ and

$$\mu^{(0)}([-\infty,\infty]) = 1. \tag{5.5}$$

Note that here, we used the extended real line. This is for convenience in our discussions. Any survival function S defined on $(-\infty, \infty)$ can be extended to $[-\infty, \infty]$ by defining $S(-\infty) = 1$ and $S(\infty) = 0$. In the sequel, we will refer to (5.4) as the general self-consistency equation and (5.2) as the discrete self-consistency equation. The $\mu^{(i)}$'s in (5.4) could be measures generated by some estimator of the $Q^{(i)}$ functions. When this is the case, the solution r for (5.4) can be considered as an estimator for S_X. The measures $\mu^{(i)}$'s above can be considered as distributions of general data.

Using the same idea, we define a generalized nonparametric log likelihood functional based on such general data. We then show that the generalized nonparametric maximum likelihood estimator (GNPMLE hereafter), if exists, also satisfies the general self-consistency equation (5.4), which extends the existing results concerning the relationship between self-consistent estimators and NPMLEs. In general, a self-consistent estimator may not be the GNPMLE. We discuss conditions under which a generalized self-consistent estimator is also the GNPMLE. These results further leads us to the proof the existence of the GNPMLE.

Section 5.2 gives the development of the self consistent estimators and the GNPMLE. The proofs of results in Section 5.2 is given in Section 5.4.

5.2 The Main Results

Let

$$\begin{aligned}\Theta = \{G : G &\text{ is a nonincreasing function from } [-\infty, +\infty] \text{ to } [0,1] \\ &\text{with } G(-\infty) = 1 \text{ and } G(\infty) = 0\}\end{aligned}$$

and

$$\begin{aligned}\Theta^1 = \{G : G &\text{ is a right continuous, nonincreasing function from } [-\infty, +\infty] \text{ to } [0,1] \\ &\text{with } G(-\infty) = 1 \text{ and } G(\infty) = 0 \text{ and } \lim_{t\to\infty} G(t) = 0\Big\}.\end{aligned}$$

Functions in Θ^1 associate with probability measures defined on $(-\infty, \infty)$. These are considered "proper" survival functions.

Now we will define the generalized nonparametric log likelihood for functions in Θ. For

$S \in \Theta$, we can write it as $\tilde{S}+S^{\perp}+\Delta_S$, where $\tilde{S}+S^{\perp}$ is right continuous with $\tilde{S}$ absolutely continuous with respect to $\mu^{(1)}$ and $S^{\perp}$ singular with respect to $\mu^{(1)}$. Δ_S is a left continuous step function. Here $\tilde{S}$, $S^{\perp}$, and Δ_S are chosen so that all of them are nonincreasing and $\tilde{S}(\infty) = S^{\perp}(\infty) = \Delta_S(\infty) = 0$. Let $\mu_{\tilde{S}}$ be the probability measure defined on $[-\infty, \infty]$ related to $\tilde{S}$ through

$$\tilde{S}(t) = \int_{(t,\infty]} d\mu_{\tilde{S}}(u)$$

and f the Radon-Nikodym derivative of $\mu_{\tilde{S}}$ with respect to $\mu^{(1)}$. Define the vector $\mu = \left(\mu^{(1)}, \mu^{(2)}, \mu^{(3)}\right)$. Consider the likelihood functional defined by (5.3). We have

$$\begin{aligned}
\log L(S) = &\sum_{i=1}^{n} \left(I(\delta_i = 1) \log\left(S(Z_i-) - S(Z_i)\right)\right. \\
&\left.+ I(\delta_i = 2) \log S(Z_i) + I(\delta_i = 3) \log\left(1 - S(Z_i)\right)\right) \\
&= -n \int_{-\infty}^{\infty} \left(\log f(t)\, dQ_n^{(1)}(t) + \log S(t)\, dQ_n^{(2)}(t) + \log\left(1 - S(t)\right) dQ_n^{(3)}(t)\right)
\end{aligned}$$

where f defined by $f(t) = n\left(S(t-) - S(t)\right)$. Mimicking this, we can generalize the above log likelihood to define the following generalized log-likelihood functional:

$$l(S; \mu) = \int_{[-\infty,\infty]} \left(\log f(x)\, d\mu^{(1)}(x) + \log S(x)\, d\mu^{(2)}(x) + \log\left(1 - S(x)\right) d\mu^{(3)}(x)\right). \tag{5.6}$$

When there is no risk of confusion, we will suppress the dependence of l on μ.

If there exists an S that maximizes l in Θ, we call it the GNPMLE. It has been shown by Mykland and Ren (1996) that the classical NPMLE is also a self-consistent estimator (but not vice-versa). Now the following questions arise.

(1) For any $\mu = \left(\mu^{(1)}, \mu^{(2)}, \mu^{(3)}\right)$ satisfying (5.5), does a GNPMLE always exist in Θ^1?

(2) Is the GNPMLE unique?

(3) Is the GNPMLE also a solution to the self-consistent equation (5.4)?

If the answers to the (1) and (3) were both yes, then this would also guarantees the existence of a solution for in (5.4) Θ^1, since the GNPMLE itself will be a solution for (5.4). Unfortunately, the answer to (1) is ‘no’, as shown in the following example.

Example 5.1. Define

$$h^{(1)}(t) = \begin{cases} 1/5, \text{ if } t < 1/2, \\ 0, \text{ if } t \geq 1/2, \end{cases}$$

$$h^{(2)}(t) = \begin{cases} 2/5, \text{ if } t < 0, \\ 1/5 - 1/5t, \text{ if } 0 \leq t < 1, \\ 0, \text{ if } t \geq 1, \end{cases}$$

and

$$h^{(3)}(t) = \begin{cases} 2/5, \text{ if } t < 0, \\ 2/5 - 1/5t \text{ if } 0 \leq t < 1, \\ 0, \text{ if } t \geq 1. \end{cases} .$$

Let $\mu^{(i)}(E) = -\int_E dh^{(i)}$ for $i = 1, 2, 3$. It can be verified that $\sum \mu^{(i)}([-\infty, \infty]) = 1$. Let $\mu = \left(\mu^{(1)}, \mu^{(2)}, \mu^{(3)}\right)$. Now we will find the GNPMLE – the function $S \in \Theta$ that maximizes $l(S; \mu)$. For any $S \in \Theta$,

$$\begin{aligned} l(S) &= \int_{[-\infty,\infty]} \left(\log f(x)\, d\mu^{(1)}(x) + \log S(x)\, d\mu^{(2)}(x) + \log(1 - S(x))\, d\mu^{(3)}(x)\right) \\ &= -\int_{[-\infty,\infty]} \left(\log f(x)\, dh^{(1)}(x) + \log S(x)\, dh^{(2)}(x) + \log(1 - S(x))\, dh^{(3)}(x)\right) \\ &= \frac{1}{5}\log f(1/2) + \frac{1}{5}\log S(0) + \frac{1}{5}\int_{(0,1)} \log S(x)\, dx + \frac{1}{5}(1 - S(1)) \\ &+ \frac{1}{5}\int_{(0,1)} \log(1 - S(x))\, dx. \end{aligned}$$

We realize that the S that maximizes l has to be constant on the intervals $(0, 1/2)$ and $(1/2, 1)$, since if not, we can replace S by

$$\bar{S}(x) = \begin{cases} S(x), \text{ if } x \notin (0,1), \\ 2\int_{(0,1/2)} S(x)\, dx \text{ if } x \in (0, 1/2), \\ 2\int_{(1/2,1)} S(x)\, dx \text{ if } x \in [1/2, 1). \end{cases}$$

Then, we have

$$
\begin{aligned}
l\left(\bar{S}\right) - l\left(S\right) = &\frac{1}{5}\log f\left(1/2\right) - \frac{1}{5}\log\left(5\left(\bar{S}\left(1/2-\right) - \bar{S}\left(1/2\right)\right)\right) \\
&+ \frac{1}{5}\left(\frac{1}{2}\log\left(2\int_{(0,1/2)} S\left(x\right)dx\right) - \int_{(0,1/2)} \log S\left(x\right)dx\right) \\
&+ \frac{1}{5}\left(\frac{1}{2}\log\left(2\int_{(1/2,1)} S\left(x\right)dx\right) - \int_{(1/2,1)} \log S\left(x\right)dx\right) \\
&> 0
\end{aligned}
$$

by the convexity of the logarithm function. Thus, we assume that

$$
S\left(x\right) = \begin{cases} 1 \text{ if } x = 0, \\ a \text{ if } x \in \left(0, 1/2\right), \\ b \text{ if } x \in \left[1/2, 1\right), \\ 0 \text{ if } x = 1. \end{cases}
$$

Then

$$
l\left(S\right) = \frac{1}{5}\log\left(5\left(a-b\right)\right) + \frac{1}{10}\left(\log a + \log b\right) + \frac{1}{10}\left(\log\left(1-a\right) + \log\left(1-b\right)\right).
$$

By straight forward calculations, we can show $a = \left(3+\sqrt{3}\right)/6$ and $b = \left(3-\sqrt{3}\right)/6$ maximize l. Thus,

$$
S\left(t\right) = \begin{cases} 1, \text{ if } t \le 0, \\ \left(3+\sqrt{3}\right)/6, \text{ if } 0 < t < 1/2, \\ \left(3-\sqrt{3}\right)/6, \text{ if } 1/2 \le t < 1, \\ 0, \text{ if } t \ge 1. \end{cases}
$$

Note that $S \notin \Theta^1$. However, S can be approximated arbitrarily close by continuous functions. Thus, we can find a sequence $\{S_n\} \subset \Theta^1$ such that $S_n \to S$ and $l\left(S_n\right) \to l\left(S\right)$. But S is the unique function in Θ that maximizes l. This shows that for this example, the maximum log likelihood can be approached by functions in Θ^1, but cannot be attained by functions in Θ^1.

Subsequently in this chapter, we will show that the GNPMLE always exists in Θ. As shown by the above example, it is possible for GNPMLE not to exist in Θ^1. In other words, it is possible that the GNPMLE is not a proper survival function. To deal with this situation, we will give conditions under which the GNPMLE exists in Θ^1. These conditions turn out to be very reasonable.

As to the second question above, which is about the uniqueness of the GNPMLE, the answer is yes. The GNPMLE is unique in Θ in the $\mu^{(0)}$-a.s. sense. This can be proved by the convexity of l. Suppose that S_1 and S_2 both maximizes l in Θ. Define $S = \frac{S_1+S_2}{2}$, then

$$\begin{aligned}
l(S) &= \int_{[-\infty,\infty]} \left(\log \frac{f_1(x)+f_2(x)}{2} d\mu^{(1)}(x) + \log \frac{S_1(x)+S_2(x)}{2} d\mu^{(2)}(x) \right. \\
&\quad \left. + \log\left(1 - \frac{S_1(x)+S_2(x)}{2}\right) d\mu^{(3)}(x) \right) \\
&\geq \frac{1}{2} \left(\int_{[-\infty,\infty]} \left(\log f_1(x)\, d\mu^{(1)}(x) + \log S_1(x)\, d\mu^{(2)}(x) + \log(1 - S_1(x))\, d\mu^{(3)}(x) \right) \right. \\
&\quad \left. + \int_{[-\infty,\infty]} \left(\log f_2(x)\, d\mu^{(1)}(x) + \log S_2(x)\, d\mu^{(2)}(x) + \log(1 - S_2(x))\, d\mu^{(3)}(x) \right) \right) \\
&= \frac{1}{2} \left(l(S_1) + l(S_2) \right).
\end{aligned}$$

The above inequality is strict if $S_1 \neq S_2$ μ_0-a.s. Since S_1 and S_2 are GNPMLEs, we must have $l(S) = l(S_1) = l(S_2)$, thus, $S_1 = S_2$ $\mu^{(0)}$-a.s.

The following theorem shows that the GNPMLE satisfies the self-consistency equation (5.4). This is a generalization of the well known result in the literature which states that NPMLEs are also self-consistent.

Theorem 5.2. *Assume that* $\mu = \left(\mu^{(1)}, \mu^{(2)}, \mu^{(3)}\right)$ *satisfies condition* (5.5). *If* $S \in \Theta$ *maximizes* $l(\cdot;\mu)$, *then* S *is a solution to* (5.4).

The reverse of the above theorem is not always true. A solution to the self-consistency equation (5.4) may not maximize l. However, the following theorem shows that if a solution to (5.4) satisfies an extra condition, then it is the GNPMLE.

Theorem 5.3. *If* $S \in \Theta$ *is a solution to* (5.4), *then it is the GNPMLE if and only if it satisfies*

$$1 - \int_{(-\infty,t)} \frac{1}{S} d\mu^{(2)} - \int_{[t,\infty]} \frac{1}{1-S} d\mu^{(3)} \geq 0$$

for any $t \in [-\infty, \infty]$.

Note that Mykland and Ren (1996) proved a similar result for the discrete case. Thus, the above theorem can be considered extension of results in the literature on self-consistency equation based on the empirical processes. The next theorem shows the existence of the GNPMLE in Θ. Its proof is based on the previous two theorems.

Theorem 5.4. *The log-likelihood function defined by* (5.6) *is maximized by an S in Θ.*

As shown by Example 5.1, it is possible that the GNPMLE is not in Θ^1. But if the measures $\mu^{(i)}$'s are generated by continuous partial survival functions with compact support, such as the smoothed empirical partial survival functions (obtained by using a smooth kernel with compact support), then it is easy to see that the GNPMLE always exists in Θ^1. First of all, under such conditions, the GNPMLE S can be always chosen so that $\lim_{t\to-\infty} S(t) = 1$ and $\lim_{t\to\infty} S(t) = 0$. Then, if S is not right-continuous, we can always define S_1 by $S_1(t) = \lim_{t'\downarrow t} S(t')$. Clearly $l(S_1) = l(S)$. In the next theorem, we give other sufficient conditions under which the GNPMLE exists in Θ^1. Here, as in the definition of the log likelihood functional, we assume that $S = \tilde{S} + S^{\perp} + \Delta_S$, where $\tilde{S}$ can be written as $\int_{(t,\infty]} f d\mu^{(1)}$.

Theorem 5.5. *If $\mu^{(0)}(\{-\infty,\infty\}) = 0$ and there exists a constant K such that*
(i) for any $t \in [-\infty,\infty)$, either there exists $t' > t$ such that $\mu^{(0)}((t,t')) = 0$ or

$$\limsup_{t'\downarrow t} \frac{\left(\mu^{(2)} + \mu^{(3)}\right)((t,t'))}{\mu^{(1)}((t,t'))} \leq K$$

and
(ii) either $\mu^{(0)}(t'',\infty) = 0$ for some $t'' > 0$ or $\limsup_{t\to\infty} \frac{\left(\mu^{(2)}+\mu^{(3)}\right)((t,\infty))}{\mu^{(1)}((t,\infty))} \leq K$, then we can find $S \in \Theta^1$ such that for any $T \in \Theta$, $l(T) \leq l(S)$.

In particular, if the measures $\mu^{(i)}$'s are all atomic and the set of all atoms of these measures contains no limit point, then the condition in the statement of the above theorem is satisfies. Thus, these conditions are satisfied if the measures $\mu^{(i)}$ are generated by the empirical partial survival functions $Q_n^{(i)}$'s. If the measures $\mu^{(i)}$'s are continuous measures, then informally speaking, the conditions (i) and (ii) requires that $\mu^{(1)}$ is not dominated by $\mu^{(2)} + \mu^{(3)}$, which is understandable since if $\mu^{(2)} + \mu^{(3)}$ dominates $\mu^{(1)}$, that means the density of censored observation is much higher than that of the uncensored observations, which will pose difficulty to the estimation of the survival function S_X.

5.3 Conclusion

In this chapter, we examined some theoretical issues regarding self-consistent and generalized non parametric estimators of the survival function S_X using doubly censored data. It remains to explore the empirical properties of the proposed estimators using a detailed simulation study. In addition, it may be worth while to examine the large sample properties of the proposed estimators and compare them to the existing self-consistent estimators.

5.4 Proofs

The following lemmas are needed for the proof of Theorem 5.2.

Lemma 5.6. *For any positive measure μ defined on $[-\infty, \infty]$ with Radon-Nikodym derivative g with respect to a measure σ, there exists a version of g, such that $x \in Supp\,(g)$ if and only if $\mu\,((a, x]) > 0$ for any $a < x$.*

Proof. Let A be the subset of $Supp\,(g)$ such that for any $x \in A$, we find $a_x < x$ such that $\mu\,((a_x, x]) = 0$. Let $B = \cup_{x \in A}\,(a_x, x)$. Then B is an open set. Thus, B is the disjoint union of at most a countable number of intervals. For each one of these intervals $I = (a, b)$, define $\tau = \sup\{t > b,\ \mu\,(t, b) = 0\}$. If $t \neq a$, we must have $t \in (a_x, x)$ for some $x \in A$, thus, $\mu\,(a_x, x) > 0$, controdiction. So, $\tau = a$, which means $\mu\,(I) = 0$. This shows that $\mu\,(B) = 0$. Clearly, any $x \in A$ is either an end point or an inner point of an interval in B. This shows that $A \setminus B$ is at most countable. Thus, $\mu\,(A) = 0$. Now let

$$\tilde{g}\,(x) = \begin{cases} 0, \text{ if } x \in A, \\ g\,(x) \text{ otherwise.} \end{cases}$$

Then $\tilde{g} = g$ $\mu-$a.e. and $\tilde{g}$ satisfies the condition in the lemma. □

Lemma 5.7. *Suppose that $S \in \Theta$ maximizes $l\,(\cdot; \mu)$. Then there exists a constant C and a version of f such that*

$$\frac{1}{f\,(t)} = C - \int_{u<t} \frac{1}{S\,(u)} d\mu^{(2)}\,(u) - \int_{u \geq t} \frac{1}{1 - S\,(u)} d\mu^{(3)}\,(u) \tag{5.7}$$

if $t \in Supp\,(f)$;

$$C - \int_{u<t} \frac{1}{S\,(u)} d\mu^{(2)}\,(u) - \int_{u \geq t} \frac{1}{1 - S\,(u)} d\mu^{(3)}\,(u) = 0 \tag{5.8}$$

if $t \notin Supp(f)$ and $S(a) - S(t) > 0$ for any $a \in [-\infty, t)$; and

$$C - \int_{u \le t} \frac{1}{S(u)} d\mu^{(2)}(u) - \int_{u > t} \frac{1}{1 - S(u)} d\mu^{(3)}(u) = 0 \tag{5.9}$$

if S is not right continuous at t.

Proof. We will only give the proof for (5.7) and (5.8). The proof for (5.9) is similar and is left to the reader. Let f be chosen according to Lemma 5.6. If for some t_0,

$$S(t) = \begin{cases} 1, \text{ if } t < t_0, \\ 0, \text{ if } t \ge t_0. \end{cases}$$

then the proof is trivial. So we assume that for some $\delta \in (0,1)$, $\{x : S(x) = \delta\} \neq \emptyset$. Let $t_0 = \sup\{x : S(x) = \delta\}$. For any t such that $S(a) - S(t) > 0$ for any $a < t$, with no loss of generality, we assume that $t < t_0$. We will only provide the proof for the case in which $\tilde{\mu}_S((a, t_0)) > 0$ for any $a < t_0$. The proof for the other case is very similar and we will omit it. Find sequences $a_n \uparrow t$, $b_n \uparrow t_0$ and $\varepsilon_n \to 0$. First, assume that $t \in Supp(f)$. Define

$$f_n = \begin{cases} f(x), \text{ if } x \in [-\infty, \infty] \setminus ((a_n, t] \cup (b_n, t_0)), \\ (1 + c_n) f(x), \text{ if } x \in (a_n, t], \\ (1 - d_n) f(x), \text{ if } x \in (b_n, t_0). \end{cases}$$

Here $c_n = \frac{\varepsilon_n}{\tilde{\mu}_S((a_n, t])}$ and $d_n = \frac{\varepsilon_n}{\tilde{\mu}_S((b_n, t_0))}$. Let measure $\tilde{\mu}_n$ be define by $\tilde{\mu}_n(E) = \int_E f_n(x) d\mu^{(1)}(x)$, then $\tilde{\mu}_n([-\infty, \infty]) = \tilde{\mu}_S([-\infty, \infty])$. Define $\mu_n = \tilde{\mu}_n + \mu_S^{\perp}$ and $S_n = G_{\mu_n} + \Delta_S$. Then $S_n \in \Theta$. We have

$$\begin{aligned} & l(S_n) \\ &= \int_{[-\infty,\infty]} \left(\log f_n(x) d\mu^{(1)}(x) + \log S_n(x) d\mu^{(2)}(x) + \log(1 - S_n(x)) d\mu^{(3)}(x)\right) \\ &= \left(\int_{[-\infty, a_n]} + \int_{(a_n, t]} + \int_{(t, b_n]} + \int_{(b_n, t_0)} + \int_{[t_0, \infty]}\right) \left(\log f_n(x) d\mu^{(1)}(x) + \right. \\ & \left. + \log S_n(x) d\mu^{(2)}(x) + \log(1 - S_n(x)) d\mu^{(3)}(x)\right) \end{aligned}$$

and

$$
\begin{aligned}
&l\left(S_n\right)-l\left(S\right)\\
&=\left(\int_{\left(a_n,t\right]}+\int_{\left(t,b_n\right]}+\int_{\left(b_n,t_0\right)}\right)\left(\log\frac{f_n\left(x\right)}{f\left(x\right)}d\mu^{(1)}\left(x\right)\right.\\
&\left.+\log\frac{S_n\left(x\right)}{S\left(x\right)}d\mu^{(2)}\left(x\right)+\log\frac{1-S_n\left(x\right)}{1-S\left(x\right)}d\mu^{(3)}\left(x\right)\right)\\
&=\int_{\left(a_n,t\right]}\log\left(1+c_n\right)d\mu^{(1)}\left(x\right)+\int_{\left(b_n,t\right)}\log\left(1-d_n\right)d\mu^{(1)}\left(x\right)\\
&+\int_{\left(a_n,t\right)}\left(\log\frac{S_n\left(x\right)}{S\left(x\right)}d\mu^{(2)}\left(x\right)+\log\frac{1-S_n\left(x\right)}{1-S\left(x\right)}d\mu^{(3)}\left(x\right)\right)\\
&=\left(c_n+o\left(c_n\right)\right)\mu^{(1)}\left(\left(a_n,t\right]\right)-\left(d_n+o\left(d_n\right)\right)\mu^{(1)}\left(\left(b_n,t\right)\right)\\
&+\left(\mu^{(2)}+\mu^{(3)}\right)\left(\left(a_n,t\right)\cup\left(b_n,t_0\right)\right)O\left(\varepsilon_n\right)\\
&+\int_{\left[t,b_n\right]}\left(\log\frac{S\left(x\right)-\varepsilon_n}{S\left(x\right)}d\mu^{(2)}\left(x\right)+\log\frac{1-S\left(x\right)+\varepsilon_n}{1-S\left(x\right)}d\mu^{(3)}\left(x\right)\right)\\
&=\frac{\varepsilon_n}{\bar{f}_{\left(a_n,t\right]}}+o\left(\frac{\varepsilon_n}{\bar{f}_{\left(a_n,t\right]}}\right)-\left(\frac{\varepsilon_n}{\bar{f}_{\left(b_n,t_0\right)}}\right)+o\left(\frac{\varepsilon_n}{\bar{f}_{\left(b_n,t_0\right)}}\right)\\
&+\left(\mu^{(2)}+\mu^{(3)}\right)\left(\left(a_n,t\right)\cup\left(b_n,t_0\right)\right)O\left(\varepsilon_n\right)\\
&+\int_{\left[t,b_n\right]}\left(\log\left(1-\frac{\varepsilon_n}{S\left(x\right)}\right)d\mu^{(2)}\left(x\right)+\log\left(1+\frac{\varepsilon_n}{1-S\left(x\right)}\right)d\mu^{(3)}\right)\\
&=\varepsilon_n\left(\frac{1}{\bar{f}_{\left(a_n,t\right]}}-\frac{1}{\bar{f}_{\left(b_n,t_0\right)}}+\int_{\left[t,b_n\right]}\left(\frac{1}{1-S\left(x\right)}d\mu^{(3)}-\frac{1}{S\left(x\right)}d\mu^{(2)}\right)\right)\\
&+o\left(\varepsilon_n\right).
\end{aligned}
$$

Since $l\left(S_n\right)-l\left(S\right)<0$, we must have

$$
\lim_{n\to\infty}\left(\frac{1}{\bar{f}_{\left(a_n,t\right]}}-\frac{1}{\bar{f}_{\left(b_n,t_0\right)}}+\int_{\left[t,b_n\right]}\left(\frac{1}{1-S\left(x\right)}d\mu^{(3)}\left(x\right)-\frac{1}{S\left(x\right)}d\mu^{(2)}\left(x\right)\right)\right)=0
$$

Thus, we proved that

$$
\begin{aligned}
\lim_{n\to\infty}\frac{1}{\bar{f}_{\left(a_n,t\right]}}&=C'-\int_{\left[t,t_0\right)}\left(\frac{1}{1-S\left(x\right)}d\mu^{(3)}\left(x\right)-\frac{1}{S\left(x\right)}d\mu^{(2)}\left(x\right)\right)\\
&=C'-\int_{\left[t,t_0\right)}\frac{1}{1-S\left(x\right)}d\mu^{(3)}\left(x\right)+\int_{\left[t,t_0\right)}\frac{1}{S\left(x\right)}d\mu^{(2)}\left(x\right)
\end{aligned}
$$

$$= C' - \int_{[t,\infty]} \frac{1}{1-S(x)} d\mu^{(3)}(x) + \int_{[t_0,\infty]} \frac{1}{1-S(x)} d\mu^{(3)}(x) -$$
$$- \int_{[-\infty,t)} \frac{1}{S(u)} d\mu^{(2)}(u) + \int_{[-\infty,t_0)} \frac{1}{S(u)} d\mu^{(2)}(u)$$
$$= C - \int_{[-\infty,t)} \frac{1}{S(u)} d\mu^{(2)}(u) - \int_{[t,\infty]} \frac{1}{1-S(x)} d\mu^{(3)}(x).$$

Now define

$$\tilde{f}(t) = \begin{cases} f(t) \text{ if } t \notin E(f), \\ \lim_{a \uparrow t} \bar{f}_{(a,t]} \text{ otherwise.} \end{cases}$$

We will prove that $\tilde{f} = f$ $\mu^{(1)}$−a.s. We just need to show that

$$\int_{(a,b]} \left(f - \tilde{f}\right) d\mu^{(1)} = 0$$

for any interval $(a,b]$. In fact, we can assume that $b \in Supp(f)$, since if not, $\int_{(a,b]} \left(f - \tilde{f}\right) d\mu^{(1)} = \int_{(a,t']} \left(f - \tilde{f}\right) d\mu^{(1)}$ where $t' = \sup((a,b] \cap Supp(f))$. Easy to see that $t' \in Supp(f)$. Define $A = (a,b] \cap Supp(f)$. Let $\tilde{f}_k = \tilde{f} \wedge k$, then $\tilde{f}_k$ is bounded and left continuous on A. If A contains one single point b, then the proof is trivial. Otherwise, let $b_0 = b$. We can find $a_0 \in (a,b)$, such that $\left|1 - \frac{\tilde{f}_k(b_0)\mu^{(1)}(a_0,b_0]}{\int_{(a_0,b_0]} \tilde{f}_k d\mu^{(1)}}\right| < \varepsilon$ by the left continuity of f'_k and also

$$\frac{\bar{f}_{(a_0,b_0]} - \tilde{f}_k(b_0)}{\tilde{f}_k(b_0)} \begin{cases} \in (-\varepsilon, \varepsilon) \text{ if } \tilde{f}_k(b_0) < k \\ > -\varepsilon \text{ if } \tilde{f}_k(b_0) = k \end{cases}$$

by the definition of $\tilde{f}_k$. Now for an ordinal number α, assuming that we have constructed intervals of the form $(a_\beta, b_\beta]$ with $a_\beta \in (a, b_\beta)$, $b_\beta \in A$ for all $\beta < \alpha$, and for each $\beta > 0$, $b_\beta = \sup\{A \setminus \cup_{\beta' < \beta}(a_\beta, b_\beta]\}$ (easy to see that $b_\beta \in A$.) Now, If $A \setminus \cup_{\beta' < \alpha}(a_{\beta'}, b_{\beta'}] \neq \emptyset$, then let

$$b_\alpha = \sup\{A \setminus \cup_{\beta' < \alpha}(a_{\beta'}, b_{\beta'}]\} \in A, \tag{5.10}$$

we can find $a_\alpha \in A$ such that

$$\left|1 - \frac{\int_{(a_\alpha,b_\alpha]} \tilde{f}_k d\mu^{(1)}}{\tilde{f}_k(b_\alpha)\mu^{(1)}(a_\alpha, b_\alpha]}\right| < \varepsilon\ . \tag{5.11}$$

and and
$$\frac{\bar{f}_{(a_\alpha,b_\alpha]}-\tilde{f}_k(b_\alpha)}{\bar{f}_{(a_\alpha,b_\alpha]}}\begin{cases}\in(-\varepsilon,\varepsilon)\ \text{ if } \tilde{f}_k(b_\alpha)<k\\ >-\varepsilon \text{ if } \tilde{f}_k(b_\alpha)=k\end{cases} \tag{5.12}$$

Clearly, this construction has to stop after at most countable number of intervals. Thus, we have a (possibly finite) sequence of intervals $(a_\alpha, b_\alpha]$, for $0 \le \alpha < \alpha_0$ such that $a_\alpha \in (a, b_\alpha), b_\alpha \in A$ and (5.10), (5.11) and (5.12) are satisfied. Moreover, $A \setminus \cup_{\beta'<\alpha}(a_{\beta'}, b_{\beta'}] = \emptyset$. From the construction of these intervals, it can be proved that $A \subset \cup_{\alpha<\alpha_0}(a_\alpha, b_\alpha]$. Thus,

$$\begin{aligned}\left|\int_{(a,b]}\left(f-\tilde{f}_k\right)d\mu^{(1)}\right| &= \left|\sum_\alpha\int_{(a_\alpha,b_\alpha]}\left(f-\tilde{f}_k\right)d\mu^{(1)}\right|\\ &\le \left|\left(\sum_{\alpha\in\mathcal{I}_k}+\sum_{\alpha\in\mathcal{J}_k}\right)\left(\bar{f}_{(a_\alpha,b_\alpha]}-\tilde{f}_k(b_\alpha)\right)\mu^{(1)}((a_\alpha,b_\alpha])\right|\\ &+\left|\left(\sum_\alpha\right)\left(\tilde{f}_k(b_\alpha)\mu^{(1)}((a_\alpha,b_\alpha])-\int_{(a_\alpha,b_\alpha]}\tilde{f}_k d\mu^{(1)}\right)\right|\\ &\le \varepsilon\sum_\alpha\left(\bar{f}_{(a_\alpha,b_\alpha]}+\tilde{f}_k(b_\alpha)\right)\mu^{(1)}((a_\alpha,b_\alpha])+\\ &+(1+\varepsilon)\sum_{\alpha\in\mathcal{I}_k}\left(\bar{f}_{(a_\alpha,b_\alpha]}+\tilde{f}_k(b_\alpha)\right)\mu^{(1)}((a_\alpha,b_\alpha])\end{aligned} \tag{5.13}$$

where $\mathcal{I}_k = \left\{0 \le \alpha < \alpha_0 : \tilde{f}_k(b_\alpha) < k\right\}$ and $\mathcal{J}_k = \left\{0 \le \alpha < \alpha_0 : \tilde{f}_k(b_\alpha) = k\right\}$. Since f is integrable, for large enough k we have

$$\sum_{\alpha\in\mathcal{I}_k}\int_{(a_\alpha,b_\alpha]}fd\mu^{(1)} = \sum_{\alpha\in\mathcal{I}_k}\bar{f}_{(a_\alpha,b_\alpha]}\mu^{(1)}((a_\alpha,b_\alpha]) < \varepsilon.$$

Thus, by (5.13) we have
$$\left|\int_{(a,b]}\left(f-\tilde{f}_k\right)d\mu^{(1)}\right| \le O(\varepsilon).$$

Since ε is arbitrary, this shows that $\int_{(a,b]}\left(f-\tilde{f}_k\right)d\mu^{(1)} = 0$.

If $t \notin Supp(f)$, Define

$$f_n = \begin{cases} f(x), \text{ if } x \in [-\infty,\infty]\setminus(b_n,t_0),\\ (1-d_n)f(x), \text{ if } x\in(b_n,t_0).\end{cases}$$

and

$$S_n(x) = \begin{cases} S(x) \text{ if } x \in [-\infty, \infty] \setminus (a_n, t_0), \\ S(a_n) - \frac{(S(x) - S(a_n))}{S(t) - S(a_n)} \varepsilon_n \text{ if } x \in [a_n, t), \\ S(x) - \varepsilon_n \text{ if } x \in [t, b_n], \\ S(x) - \varepsilon_n + d_n \tilde{\mu}_S((b_n, x]) \text{ if } x \in (b_n, t_0). \end{cases}$$

Then, f_n is the Radon-Nickodym derivative of the right continuous part of S_n with respect to $\mu^{(1)}$. Thus,

$$\begin{aligned} l(S_n) &= \int_{[-\infty,\infty]} \Big(\log f_n(x)\, d\mu^{(1)}(x) + \log S_n(x)\, d\mu^{(2)}(x) + \log(1 - S_n(x))\, d\mu^{(3)}(x) \Big) \\ &= \left(\int_{[-\infty,a_n]} + \int_{(a_n,t]} + \int_{(t,b_n]} + \int_{(b_n,t_0)} + \int_{[t_0,\infty]} \right) \Big(\log f_n(x)\, d\mu^{(1)}(x) + \\ &+ \log S_n(x)\, d\mu^{(2)}(x) + \log(1 - S_n(x))\, d\mu^{(3)}(x) \Big) \end{aligned}$$

and

$$\begin{aligned} & l(S_n) - l(S) \\ &= \left(\int_{(a_n,t]} + \int_{(t,b_n]} + \int_{(b_n,t_0)} \right) \left(\log \frac{f_n(x)}{f(x)} d\mu^{(1)}(x) \right. \\ &\left. + \log \frac{S_n(x)}{S(x)} d\mu^{(2)}(x) + \log \frac{1 - S_n(x)}{1 - S(x)} d\mu^{(3)}(x) \right). \end{aligned}$$

For n big enough, $f(x) = f_n(x) = 0$ for $x \in (a_n, t)$, we have

$$\begin{aligned} & l(S_n) - l(S) \\ &= \int_{(b_n,t)} \log(1 - d_n)\, d\mu^{(1)}(x) + \int_{(a_n,t_0)} \left(\log \frac{S_n(x)}{S(x)} d\mu^{(2)}(x) \right. \\ &\left. + \log \frac{1 - S_n(x)}{1 - S(x)} d\mu^{(3)}(x) \right) \\ &= -(d_n + o(d_n))\, \mu^{(1)}((b_n, t)) + \left(\mu^{(2)} + \mu^{(3)} \right) ((a_n, t) \cup (b_n, t_0))\, O(\varepsilon_n) \\ &= \varepsilon_n \left(-\frac{1}{\bar{f}_{(b_n,t_0)}} + \int_{[t,b_n]} \left(\frac{1}{1 - S(x)} d\mu^{(3)}(x) - \frac{1}{S(x)} d\mu^{(2)}(x) \right) \right) + o(\varepsilon_n). \end{aligned}$$

Be a similar discussion as in the previous case,

$$\lim_{n\to i} -\frac{1}{\bar{f}_{(b_n,t_0)}} + \int_{[t,b_n]} \left(\frac{1}{1-S(x)} d\mu^{(3)}(x) - \frac{1}{S(x)} d\mu^{(2)}(x) \right) = 0.$$

Thus,

$$C - \int_{u<t} \frac{1}{S(u)} d\mu^{(2)}(u) - \int_{u\geq t} \frac{1}{1-S(u)} d\mu^{(3)}(u) = 0.$$

This finishes the proof. Using these results, now we can prove Theorem 5.2.

Proof of Theorem 5.2

By Lemma 5.7, for $x \in Supp(f)$, we have.

$$\frac{1}{f(t)} = C - \int_{u<t} \frac{1}{S(u)} d\mu^{(2)}(u) - \int_{u\geq t} \frac{1}{1-S(u)} d\mu^{(3)}(u).$$

Thus, for $t \in [-\infty, \infty]$,

$$d\mu^{(1)}(t) = Cf(t)\, d\mu^{(1)}(t) - \int_{u<t} \frac{f(t)\, d\mu^{(1)}(t)}{S(u)} d\mu^{(2)}(u) - \int_{u\geq t} \frac{f(t)\, d\mu^{(1)}(t)}{1-S(u)} d\mu^{(3)}(u)$$

or

$$d\mu^{(1)}(t) = Cd\mu_{\tilde{S}}(t) - \int_{u<t} \frac{d\mu_{\tilde{S}}(t)}{S(u)} d\mu^{2}(u) - \int_{u\geq t} \frac{d\mu_{\tilde{S}}(t)}{1-S(u)} d\mu^{(3)}(u) \tag{5.14}$$

By the same lemma we also have

$$0 = Cd\mu^{\perp}_{\tilde{S}+S^{\perp}}(t) - \int_{u<t} \frac{d\mu^{\perp}_{\tilde{S}+S^{\perp}}(t)}{S(u)} d\mu^{(2)}(u) - \int_{u\geq t} \frac{d\mu^{\perp}_{\tilde{S}+S^{\perp}}(t)}{1-S(u)} d\mu^{(3)}(u) \tag{5.15}$$

and

$$0 = -Cd\Delta_S + \int_{u\leq t} \frac{d\Delta_S(t)}{S(u)} d\mu^{(2)}(u) + \int_{u>t} \frac{d\Delta_S(t)}{1-S(u)} d\mu^{(3)}(u) \tag{5.16}$$

for $t \in [-\infty, \infty]$. Combining (5.14) and (5.15) we have

$$\begin{aligned} Cd\left(\tilde{S} + S^{\perp}\right) = -d\mu^{(0)}(t) &+ \int_{u<t} \frac{d\left(\tilde{S} + S^{\perp}\right)}{S(u)} d\mu^{(2)}(u) + d\mu^{(2)}(t) \\ &+ \int_{u\geq t} \frac{d\left(\tilde{S} + S^{\perp}\right)}{1-S(u)} d\mu^{(3)}(u) + d\mu^{(3)}(t) \end{aligned}$$

$$= -d\mu^{(0)}(t) + d_- \left(\int_{u \le t} \frac{S(t)}{S(u)} d\mu^{(2)}(u) \right)$$
$$- d_- \left(\int_{u > t} \frac{1 - S(t)}{1 - S(u)} d\mu^{(3)}(u) \right).$$

Thus,

$$\begin{aligned}
CS(t) &= C \left(1 + \int_{u \le t} d\left(\tilde{S} + S^{\perp}\right) + \int_{u<t} d\Delta_s(u) \right) \\
&= C + \left(- \int_{u \le t} d\mu^{(0)}(u) + \int_{y \le t} d_- \left(\int_{u \le y} \frac{S(y)}{S(u)} d\mu^{(2)}(u) \right) \right. \\
&- \int_{y \le t} d_- \left(\int_{u > y} \frac{1 - S(y)}{1 - S(u)} d\mu^{(3)}(u) \right) \Bigg) \\
&+ \int_{y<t} \left(\int_{u \le y} \frac{d\Delta_S(y)}{S(u)} d\mu^{(2)}(u) + \int_{u>y} \frac{d\Delta_S(y)}{1 - S(u)} d\mu^{(3)}(u) \right) \\
&= C + \left(- \int_{u \le t} d\mu^{(0)}(u) + \int_{y \le t} d_- \left(\int_{u \le y} \frac{S(y)}{S(u)} d\mu^{(2)}(u) \right) \right. \\
&- \int_{y \le t} d_- \left(\int_{u > y} \frac{1 - S(y)}{1 - S(u)} d\mu^{(3)}(u) \right) \Bigg) \\
&+ \int_{y<t} \left(d_+ \left(\int_{u \le y} \frac{S(y)}{S(u)} d\mu^{(2)}(u) \right) - d_+ \left(\int_{u>y} \frac{1 - S(y)}{1 - S(u)} d\mu^{(3)}(u) \right) \right) \\
&= C - \int_{u \le t} d\mu^{(0)}(u) + \int_{u \le t} \frac{S(t)}{S(u)} d\mu^{(2)}(u) - \int_{u>t} \frac{1 - S(t)}{1 - S(u)} d\mu^{(3)}(u) \\
&= C - 1 + S^{(0)}(t) + \int_{u \le t} \frac{S(t)}{S(u)} d\mu^{(2)}(u) - \int_{u>t} \frac{1 - S(t)}{1 - S(u)} d\mu^{(3)}(u).
\end{aligned}$$

At $t = \infty$, we have

$$0 = C - 1.$$

Thus,

$$C = 1.$$

This finishes the proof.

Proof of Theorem 5.3

Suppose that S satisfies the conditions in the statement of the theorem. Assume that l is maximized by $T \in \Theta$ and $T \neq S$. White $S = \tilde{S} + S^{\perp} + \Delta_S$ as we did in earlier discussion. Similarly, let $T = \tilde{T} + T^{\perp} + \Delta_T$. Assume that $\tilde{S} = \int_{(t,\infty]} f d\mu^{(1)}$ and $\tilde{T} = \int_{(t,\infty]} g d\mu^{1}$. Define $H = T - S$ and

$h = g - f$. Also define

$$\tilde{H} = \tilde{T} - \tilde{S} = \int_{(t,\infty]} (f' - f)\, d\mu^1 = \int_{(t,\infty]} h d\mu^1,$$

$$H_1 = T^{\perp} - S^{\perp},$$

and

$$H_2 = \Delta_T - \Delta_S.$$

Let $S_\varepsilon = S + \varepsilon H$ where $\varepsilon \in (0,1)$. We have

$$\begin{aligned} l(S_\varepsilon) - l(S) &= \int_{[-\infty,\infty]} \Big(\log f_\varepsilon(x)\, d\mu^{(1)}(x) + \log S_\varepsilon(x)\, d\mu^{(2)}(x) + \log(1 - S_\varepsilon(x))\, d\mu^{(3)}(x)\Big) \\ &- \int_{[-\infty,\infty]} \Big(\log f(x)\, d\mu^{(1)}(x) + \log S(x)\, d\mu^{(2)}(x) + \log(1 - S(x))\, d\mu^{(3)}(x)\Big) \\ &= \int_{[-\infty,\infty]} \left(\log\left(1 + \frac{\varepsilon h(x)}{f(x)}\right) d\mu^{(1)}(x) + \log\left(1 + \frac{\varepsilon H(x)}{S(x)}\right) d\mu^{(2)}(x)\right. \\ &\left. + \log\left(1 - \frac{\varepsilon H(x)}{1 - S(x)}\right) d\mu^{(3)}(x)\, .\right) \end{aligned}$$

by Lemma 5.7 and Theorem 5.2, we have that for any $t \in Supp(f)$, $f(t) > 1$. Also, $\int_{[-\infty,\infty]} \frac{1}{S} d\mu^2 \leq 1$ and $\int_{[-\infty,\infty]} \frac{1}{1-S} d\mu^3 \leq 1$. Thus, it can be shown that

$$\begin{aligned} &\left(\int \log\left(\frac{f + \varepsilon h}{f}\right) d\mu^1 + \log\left(\frac{S + \varepsilon H}{S}\right) d\mu^2 + \log\left(\frac{1 - S - \varepsilon H}{1 - S}\right) d\mu^3\right) \\ &= \int \left(\frac{\varepsilon h}{f} d\mu^1 + \frac{\varepsilon H}{S} d\mu^2 - \frac{\varepsilon H}{1 - S} d\mu^3\right) + o(\varepsilon), \end{aligned}$$

which gives that

$$\lim_{\varepsilon \to 0_+} \frac{l(S_\varepsilon) - l(S)}{\varepsilon} = \int_{[-\infty,\infty]} \left(\frac{h}{f} d\mu^1 + \frac{H}{S} d\mu^2 - \frac{H}{1 - S} d\mu^3\right).$$

We have

$$\begin{aligned} &\int \left(\frac{h}{f} d\mu^1 + \frac{H}{S} d\mu^2 - \frac{H}{1 - S} d\mu^3\right) \\ &= \int \frac{h}{f} d\mu^1 + \int \frac{H}{S} d\mu^2 - \int \frac{H}{1 - S} d\mu^3 \end{aligned}$$

$$
\begin{aligned}
&= \int_{[-\infty,\infty]} \frac{h}{f} d\mu^1 + \int_{[-\infty,\infty]} \frac{\tilde{H} + H_1 + H_2}{S} d\mu^2 - \int_{[-\infty,\infty]} \frac{\tilde{H} + H_1 + H_2}{1-S} d\mu^3 \\
&= \left(\int_{[-\infty,\infty]} \frac{h}{f} d\mu^1 + \int_{(-\infty,\infty]} \frac{\tilde{H}}{S} d\mu^2 - \int_{(-\infty,\infty]} \frac{\tilde{H}}{1-S} d\mu^3 \right) \\
&+ \int_{(-\infty,\infty]} \frac{H_1}{S} d\mu^2 - \int_{(-\infty,\infty]} \frac{H_1}{1-S} d\mu^3 + \int_{(-\infty,\infty]} \frac{H_2}{S} d\mu^2 - \int_{(-\infty,\infty]} \frac{H_2}{1-S} d\mu^3 \\
&= \left(\int_{[-\infty,\infty]} \frac{h}{f} d\mu^1 + \int_{(-\infty,\infty]} \frac{\tilde{H}}{S} d\mu^2 - \int_{(-\infty,\infty]} \frac{\tilde{H}}{1-S} d\mu^3 \right) \\
&+ \int_{(-\infty,\infty]} \frac{H_1}{S} d\mu^2 - \int_{[-\infty,\infty]} \frac{H_1}{1-S} d\mu^3 + \int_{(-\infty,\infty]} \frac{H_2}{S} d\mu^2 - \int_{[-\infty,\infty]} \frac{H_2}{1-S} d\mu^3 \\
&= \int_{[-\infty,\infty]} \frac{h}{f} d\mu^1 + \tilde{H}(\infty) \int_{(-\infty,\infty]} \frac{1}{S} d\mu^{(2)} - \int_{(-\infty,\infty]} \int_{(-\infty,t)} \frac{1}{S} d\mu^{(2)} d\tilde{H}(t) \\
&- \tilde{H}(-\infty) \int_{(-\infty,\infty]} \frac{1}{1-S} d\mu^{(3)} - \int_{(-\infty,\infty]} \int_{[t,\infty]} \frac{1}{1-S} d\mu^{(3)} d\tilde{H}(t) \\
&+ H_1(\infty) \int_{(-\infty,\infty]} \frac{1}{S} d\mu^{(2)} - \int_{(-\infty,\infty]} \int_{(-\infty,t)} \frac{1}{S} d\mu^{(2)} dH_1(t) \\
&- H_1(-\infty) \int_{(-\infty,\infty]} \frac{1}{1-S} d\mu^{(3)} - \int_{(-\infty,\infty]} \int_{[t,\infty]} \frac{1}{1-S} d\mu^{(3)} dH_1(t) \\
&+ H_2(\infty) \int_{(-\infty,\infty]} \frac{1}{S} d\mu^{(2)} - \int_{[-\infty,\infty)} \int_{(-\infty,t]} \frac{1}{S} d\mu^{(2)} dH_2 \\
&- H_2(-\infty) \int_{(-\infty,\infty]} \frac{1}{1-S} d\mu^{(3)} - \int_{[-\infty,\infty)} \int_{(t,\infty]} \frac{1}{1-S} d\mu^{(3)} dH_2 \\
&= \int_{[-\infty,\infty]} \frac{h}{f} d\mu^1 - \int_{(-\infty,\infty]} \int_{(-\infty,t)} \frac{1}{S} d\mu^{(2)} d\tilde{H}(t) - \int_{(-\infty,\infty]} \int_{[t,\infty]} \frac{1}{1-S} d\mu^{(3)} d\tilde{H}(t) \\
&- \int_{(-\infty,\infty]} \int_{(-\infty,t)} \frac{1}{S} d\mu^{(2)} dH_1(t) - \int_{(-\infty,\infty]} \int_{[t,\infty]} \frac{1}{1-S} d\mu^{(3)} dH_1(t) \\
&- \int_{[-\infty,\infty)} \int_{(-\infty,t]} \frac{1}{S} d\mu^{(2)} dH_2 - \int_{[-\infty,\infty)} \int_{(t,\infty]} \frac{1}{1-S} d\mu^{(3)} dH_2 \\
&= \int_{[-\infty,\infty]} \frac{h}{f} d\mu^1 - \int_{(-\infty,\infty]} \int_{(-\infty,t)} \frac{1}{S} d\mu^{(2)} d\tilde{H}(t) - \int_{(-\infty,\infty]} \int_{[t,\infty]} \frac{1}{1-S} d\mu^{(3)} d\tilde{H}(t) \\
&- \int_{(-\infty,\infty]} \int_{(-\infty,t)} \frac{1}{S} d\mu^{(2)} dH_1(t) - \int_{(-\infty,\infty]} \int_{[t,\infty]} \frac{1}{1-S} d\mu^{(3)} dH_1(t) \\
&- \int_{[-\infty,\infty)} \int_{(-\infty,t]} \frac{1}{S} d\mu^{(2)} dH_2 - \int_{[-\infty,\infty)} \int_{(t,\infty]} \frac{1}{1-S} d\mu^{(3)} dH_2 \\
&= \int_{[-\infty,\infty]} \frac{h}{f} d\mu^1 + \int_{[-\infty,\infty]} \left(1 - \int_{(-\infty,t)} \frac{1}{S} d\mu^{(2)} - \int_{[t,\infty]} \frac{1}{1-S} d\mu^{(3)} \right) d\tilde{H}(t) \\
&+ \int_{[-\infty,\infty]} \left(1 - \int_{(-\infty,t)} \frac{1}{S} d\mu^{(2)} - \int_{[t,\infty]} \frac{1}{1-S} d\mu^{(3)} \right) dH_1(t) \\
&+ \int_{[-\infty,\infty]} \left(1 - \int_{(-\infty,t]} \frac{1}{S} d\mu^{(2)} - \int_{(t,\infty]} \frac{1}{1-S} d\mu^{(3)} \right) dH_2
\end{aligned}
$$

$$= \int_{[-\infty,\infty]} \frac{h}{f} d\mu^1 - \int_{[-\infty,\infty]} h \left(1 - \int_{(-\infty,t)} \frac{1}{S} d\mu^{(2)} - \int_{[t,\infty]} \frac{1}{1-S} d\mu^{(3)}\right) d\mu_1$$
$$+ \int_{[-\infty,\infty]} \left(1 - \int_{(-\infty,t)} \frac{1}{S} d\mu^{(2)} - \int_{[t,\infty]} \frac{1}{1-S} d\mu^{(3)}\right) dH_1(t)$$
$$+ \int_{[-\infty,\infty]} \left(1 - \int_{(-\infty,t]} \frac{1}{S} d\mu^{(2)} - \int_{(t,\infty]} \frac{1}{1-S} d\mu^{(3)}\right) dH_2(t).$$

By Lemma 5.7 and earlier discussion, we have

$$\int \left(\frac{h}{f} d\mu^1 + \frac{H}{S} d\mu^2 - \frac{H}{1-S} d\mu^3\right)$$
$$= \int_{[-\infty,\infty]} \left(1 - \int_{(-\infty,t)} \frac{1}{S} d\mu^{(2)} - \int_{[t,\infty]} \frac{1}{1-S} d\mu^{(3)}\right) dH_1(t)$$
$$+ \int_{[-\infty,\infty]} \left(1 - \int_{(-\infty,t)} \frac{1}{S} d\mu^{(2)} - \int_{[t,\infty]} \frac{1}{1-S} d\mu^{(3)}\right) dH_2(t)$$

Let $H_1 = H_1^0 + H_1^{\perp}$, where H_1^0 is absolutely continuous with respect to S_1 and $H_1^{\perp}$ is singular with respect to S_1. Since $H_1 = H_1^0 + H_1^{\perp} = S_1' - S_1$, it is easy to see that $H_1^{\perp}$ is non-increasing. Thus, by Lemma 5.7, we have

$$\int_{[-\infty,\infty]} \left(1 - \int_{(-\infty,t)} \frac{1}{S} d\mu^{(2)} - \int_{[t,\infty]} \frac{1}{1-S} d\mu^{(3)}\right) dH_1(t)$$
$$= \int_{[-\infty,\infty]} \left(1 - \int_{(-\infty,t)} \frac{1}{S} d\mu^{(2)} - \int_{[t,\infty]} \frac{1}{1-S} d\mu^{(3)}\right) dH_1^0(t)$$
$$+ \int_{[-\infty,\infty]} \left(1 - \int_{(-\infty,t)} \frac{1}{S} d\mu^{(2)} - \int_{[t,\infty]} \frac{1}{1-S} d\mu^{(3)}\right) dH_1^{\perp}(t)$$
$$= \int_{[-\infty,\infty]} \left(1 - \int_{(-\infty,t)} \frac{1}{S} d\mu^{(2)} - \int_{[t,\infty]} \frac{1}{1-S} d\mu^{(3)}\right) dH_1^{\perp}(t) \leq 0. \tag{5.17}$$

Define set $A = \{t : S_2(t+) - S_2(t) > 0\}$ and set $B = A \cup \{t : T_2(t+) - T_2(t) > 0\}$. Then by Lemma 5.7, $1 - \int_{(-\infty,t)} \frac{1}{S} d\mu^{(2)} - \int_{[t,\infty]} \frac{1}{1-S} d\mu^{(3)} = 0$ for any $t \in A$. It is easy to see that for any $t \in B \setminus A$, we have $H_2(t+) - H_2(t) < 0$. Thus,

$$\int_{[-\infty,\infty]} \left(1 - \int_{(-\infty,t)} \frac{1}{S} d\mu^{(2)} - \int_{[t,\infty]} \frac{1}{1-S} d\mu^{(3)}\right) dH_2(t) \leq 0. \tag{5.18}$$

Combining (5.17) and (5.18) we have

$$\int \left(\frac{h}{f} d\mu^1 + \frac{H}{S} d\mu^2 - \frac{H}{1-S} d\mu^3 \right) \leq 0. \tag{5.19}$$

Since l is concave, we have for some $\varepsilon_0 \in (0,1)$, $l\left(S_{\varepsilon_0}\right) = l\left(\left(1-\varepsilon_0\right) S + \varepsilon_0 T\right) > \left(1-\varepsilon_0\right) l\left(S\right) + \varepsilon_0 l\left(T\right)$. Thus, $\frac{l\left(S_{\varepsilon_0}\right) - l(S)}{\varepsilon_0} > \left(l\left(T\right) - l\left(S\right)\right) \geq 0$. For any $\varepsilon \in (0, \varepsilon_0)$,

$$\begin{aligned} S_\varepsilon &= \varepsilon T + (1-\varepsilon) S = \varepsilon \frac{S_{\varepsilon_0} - (1-\varepsilon_0) S}{\varepsilon_0} + (1-\varepsilon) S \\ &= \frac{\varepsilon}{\varepsilon_0} S_{\varepsilon_0} - \frac{\varepsilon(1-\varepsilon_0) S}{\varepsilon_0} + (1-\varepsilon) S \\ &= \frac{\varepsilon}{\varepsilon_0} S_{\varepsilon_0} + \frac{\varepsilon_0 - \varepsilon}{\varepsilon_0} S \end{aligned}$$

Again, by the fact that l is concave, we have

$$l\left(S_\varepsilon\right) > \frac{\varepsilon}{\varepsilon_0} l\left(S_{\varepsilon_0}\right) + \frac{\varepsilon_0 - \varepsilon}{\varepsilon_0} l\left(S\right)$$

and

$$\frac{l\left(S_\varepsilon\right) - l\left(S\right)}{\varepsilon} > \frac{\left(l\left(S_{\varepsilon_0}\right) - l\left(S\right)\right)}{\varepsilon_0} > 0.$$

Thus, $\lim_{\varepsilon \to 0_+} \frac{l(S_\varepsilon) - l(S)}{\varepsilon} \geq \frac{\left(l\left(S_{\varepsilon_0}\right) - l(S)\right)}{\varepsilon_0} > 0$, which contradicts (5.19), finishing the proof.

Proof of Theorem 5.4

First, find sequences of atomic measures $\mu_n^{(1)}$, $\mu_n^{(2)}$, $\mu_n^{(3)}$ satisfying

1. $\mu_n^{(i)}$, $i = 1, 2, 3$ consists of atoms of measure $1/n$,
2. Let $\mu_n^{(0)} = \sum \mu_n^{(i)}$, then $\mu_n^{(0)}\left[-\infty, \infty\right] = 1$,
3. For $i = 1, 2, 3$, $\mu_n^{(i)}$ converges to $\mu^{(i)}$ in distribution.

By results in the literate, there exists functions S_n that maximizes $l\left(S; \mu_n\right)$, where $\mu_n = \left(\mu_n^{(1)}, \mu_n^{(2)}, \mu_n^{(3)}\right)$. Thus, S_n satisfies the self-consistency equation

$$S_n\left(t\right) = \mu_n^{(0)}\left(\left(t, \infty\right]\right) + \int_{u \leq t} \frac{S_n\left(t\right)}{S_n\left(u\right)} d\mu_n^{(2)}\left(u\right) - \int_{t < u} \frac{1 - S_n\left(t\right)}{1 - S_n\left(u\right)} d\mu_n^{(3)}\left(u\right), \quad t \in \left[-\infty, \infty\right] \tag{5.20}$$

and also,

$$1-\int_{(-\infty,t)}\frac{1}{S_n}d\mu_n^{(2)}-\int_{[t,\infty]}\frac{1}{1-S_n}d\mu_n^{(3)}\geq 0$$

for any $t\in[-\infty,\infty]$. Since S_n is bounded and monotone, we can find a subsequence of S_{n_k} such that S_{n_k} converges to some $S\in\Theta$ uniformly. Thus, by bounded convergence theorem, we have

$$\begin{aligned}&\lim_{n\to\infty}\left(1-\int_{(-\infty,t)}\frac{1}{S_n}d\mu_n^{(2)}-\int_{[t,\infty]}\frac{1}{1-S_n}d\mu_n^{(3)}\right)\\&=1-\int_{(-\infty,t)}\frac{1}{S}d\mu^{(2)}-\int_{[t,\infty]}\frac{1}{1-S}d\mu^{(3)}\geq 0\end{aligned}$$

for any $t\in[-\infty,\infty]$. By taking the limit on both sides of (5.20), we have

$$S(t)=\mu^{(0)}((t,\infty])+\int_{u\leq t}\frac{S(t)}{S(u)}d\mu_n^{(2)}(u)-\int_{t<u}\frac{1-S(t)}{1-S(u)}d\mu^{(3)}(u),\quad t\in[-\infty,\infty].$$

Thus, by Theorem 5.3, $S(t)$ is the GNPMLE. This finishes the proof.

Proof of Theorem 5.5

First we will show that there exists an $S\in\Theta$ that is right continuous on $[-\infty,\infty)$ and maximizes l. Suppose that $S\in\Theta$ is not right continuous. First assume that S is not right continuous at $t_0\in[-\infty,\infty)$. If there exists a t' such that $\mu^{(i)}((t_0,t'))=0$, $i=1,2,3$, then Let

$$\tilde{S}_1=\begin{cases}S(t_0)+n(t'-t_0)(S(t')-S(t_0)) & \text{if } t\in[t_0,t'],\\ S(t) & \text{other size.}\end{cases}$$

Then, clearly, $l\left(\tilde{S}_1\right)=l(S)$ and $\tilde{S}_1$ are right continuous at t_0. This shows that points of right discontinuity like t_0 above can be removed without changing the likelihood. Thus, we assume that S has no points of right discontinuity of this type. Now assume that S is not right continuous at t_0 and $\limsup_{t\downarrow t_0}\frac{\left(\mu^{(2)}+\mu^{(3)}\right)((t_0,t))}{\mu^{(1)}((t_0,t))}\leq K_1$ for some constant K_1. We can find a sequence t_n, such that $t_n\downarrow t$, and $\frac{\left(\mu^{(2)}+\mu^{(3)}\right)((t_0,t_n))}{\mu^{(1)}((t_0,t_n))}\leq K_1+1=K$. Now, define

$$f_n(t)=\begin{cases}(c_n+1)f(t) & \text{if } t\in(t_0,t_n),\\ f(t) & \text{other wise}\end{cases}$$

where $c_n=\frac{S(t_0)-S(t_n)}{\mu_{\tilde{S}}((t_0,t_n))}$. Let $\tilde{S}_n=\int_{(t,\infty]}fd\mu^{(1)}$, and Δ_1 is a pure jump left-continuous function with all the same jumps as Δ_S except at t_0, at which point Δ_1 does not have a jump. Now define $S_n = \tilde{S}_n + S^{\perp} + \Delta_1$. S_n has the same value as S everywhere except inside the interval (t_0, t_n). The construction of the function S_n guarantees that compared with S_n, it is right-continuous at t_0. We have

$$\begin{aligned}
l(S_n) - l(S) &= \int_{(t_0,t_n)} \Big((\log f_n - \log f)\, \mu^{(1)} + (\log S_n - \log S)\, d\mu^{(2)} \\
&\quad + (\log(1-S_n) - \log(1-S))\, d\mu^{(3)} \Big) \\
&= \log(c_n+1)\, \mu^{(1)}((t_0,t_n)) + \int_{(t_0,t_n)} \Big((\log S_n - \log S)\, d\mu^{(2)} \\
&\quad + (\log(1-S_n) - \log(1-S))\, d\mu^{(3)} \Big) \\
&\le \log(c_n+1)\, \mu^{(1)}((t_0,t_n)) + M \left(\mu^{(2)} + \mu^{(3)} \right) ((t_0,t_n)) \\
&\le (\log(c_n+1) - KM)\, \mu^{(1)}((t_0,t_n))
\end{aligned}$$

where $M = \max\left\{ \frac{S(t_0)-S(t_n)}{S(t_0)}, \frac{S(t_0)-S(t_n)}{1-S(t_0)} \right\}$. Since $\mu_{\tilde{S}}((t_0,t_n)) \to 0$ and $S(t_0) - S(t_n) \to d > 0$, we have $c_n \to \infty$. Thus, for some n_0 large enough, we have $l(S_{n_0}) - l(S) > 0$. Thus, S cannot be the GNPMLE.

By a similar discussion as above, we can also show that S can be chosen such that $\lim_{t\to\infty} S(t) = 0$.

Bibliography

Akaike, H. (1970), Statistical predictor identification, *Ann. Inst. Statist. Math.*, 22, pp. 203–217.

Andersen, T. G. and Bollerslev, T. (1997), Heterogeneous information arrivals and return volatility dynamics: uncovering the long run in high frequency returns, *J. Finance*, 52, pp. 975–1005.

Blum, J. R. and Susarla, V. (1980), maximal deviation theory density and failure estimates based on censored data, *Multivariate Analysis V*, pp. 213–222.

Burke, M. D. and L., H. (1984), Density and failure rate estimation in a competing risks model, *Sankhya Series A*, 46, pp. 135–154.

Cai, T. T. and Low, M. G. (2003), A note on nonparametric estimation of linear functionals, *Ann. Statist.*, 31, pp. 1140–1153.

Cai, T. T. and Low, M. G. (2004), Minimax estimation of linear functionals over nonconvex parameter spaces, *Ann. Statist.*, 32, pp. 552–576.

Chang, M. N. and Yang, G. L. (1987), Strong consistency of a nonparametric estimator of the survival function with doubly censored data, *Ann. Statist.*, 15, pp. 1536–1547.

Csorgo, M., E., G., and L., H. (1991), Central limit theorems for l_p distances of kernel estimators of densities under random censorship, *Ann. Statist.*, 19, pp. 1813–1831.

Deo, R. S. (1997), Nonparametric regression with long-memory errors, *Statistics and Probability Letters*, 33, pp. 89–94.

Donoho, D. L. (1994), Statistical estimation and optimal recovery, *Ann. Statist.*, 22, pp. 238–270.

Donoho, D. L. and Johnstone, L. M. (1998), Minimax estimation via wavelet shrinkage, *Ann. Statist.*, 26, pp. 879–921.

Donoho, D. L. and Liu, R. C. (1991), Geometrizng rates of convergence, iii, *Ann. Statist.*, 19, pp. 668–701.

Duncan, T. E., Hu, Y., and Pasik-Duncan, B. (2000), Stochastic calculus for fractional brownian motion. i. theory, *SIAM J. Control Optim.*, 38, pp. 582–612.

Eubank, R. L. (1998), *Spline smoothing and non parametric regression*, Dekker, New York.

Falk, M. (1983), Relative efficiency and deficiency of kernel type estimators of smooth distribution functions, *Statist. Neerlandica*, 37, pp. 73–83.

Falk, M. (1984), Relative deficiency of kernel type estimators of quantiles, *Ann. Statist.*, 12, pp. 261–268.

Fan, J. (1993), Local linear regression smoothers and their minimax efficiencies, *Ann. Statist.*, 21, pp. 196–216.

Fan, J. and Li, R. (2001), Variable selection via nonconcave penalized likelihood and its oracle properties, *Journal of the American Statistical Association*, 96, pp. 1348–1360.

Ferguson, T. S. (1973), A bayesian analysis of some nonparametric problems, *Ann. Statist.*, 2, pp. 615–629.

Foldes, A. and Retjo, L. (1981), Strong uniform consistency of nonparametric survival curve estimators from randomly censored data, *Ann. Statist.*, 9, pp. 122–129.

Gu, M. G. and Zhang, C.-H. (1993), Asymptotic properties of self-consistent estimators based on doubly censored data, *Ann. Statist.*, 21, pp. 611–624.

Harwell, M. (1995), *Microbond tests for ribbon fibers, M.S. thesis, Department of Chemical Engineering*, Clemson University.

Ibragimov, I. A. and Khasminskii, R. Z. (1984), On nonparametric estimation of the value of a linear functional in gaussian white noise, *Theory of Probability And Its Applications*, 29, pp. 18–32.

Ihaka, R. and Gentleman, R. (1996), R: A language for data analysis and graphics, *Journal of Computational and Graphical Statistics*, 5, pp. 299–314 (http://www.r–project.org).

Johnstone, I. M. (1999), Wavelet shrinkage for correlated data and inverse problems: adaptivity results, *Statistica Sinica*, 9, pp. 51–83.

Kaplan, E. and Meier, P. (1958), Nonparametric estimation from incomplete observations, *Journal of the American Statistical Association*, 53, pp. 457–481.

Kuhn, J. W. and Padgett, W. J. (1997), Local bandwidth selection for kernel density estimation from right-censored data based on asymptotic mean absolute error, *Nonlinear Analysis*, 30, pp. 4375–4384.

Kulasekera, K., Williams, C. L., and Manatunga, A. (2001), Smooth estimation of the reliability function, *Lifetime Data Analysis*, pp. 415–433.

Kvam, P. H., Tiwari, R. C., and Zalkikar, J. N. (2000), Nonparametric bayes estimation of contamination levels using observations from the residual distribution, *Journal of the American Statistical Association*, 95, pp. 1119–1126.

Lehmann, E. L. and Caeslla, G. (1998), *Theory of point estimation 2nd ed*, Springer, New York.

Lemdani, M. and Ould-Said, E. (2001), Relative deficiency of the kaplan-meier estimator with respect to a smoothed estimator, *Math. Methods Statist.*, 10, pp. 215–231.

Mallows, C. L. (1973), Some comments on c_p, *Technometrics*, 15, pp. 661–675.

Memin, J., Mishura, Y., and Valkeila, E. (2001), Inequalities for the moments of wiener integrals with respect to a fractional brownian motion, *Statistics and Probability Letters*, 51, pp. 197–206.

Mykland, P. A. and Ren, J.-J. (1996), Algorithms for computing self-consistent and maximum likelihood estimators with doubly censored data, *Ann. Statist.*, 24, pp. 1740–1764.

Padgett, w. J. (1986), A kernel type-estimator of a quantile function from right censored data, *Journal of the American Statistical Association*, 81, pp. 215–222.

Padgett, W. J. and McNichols, D. T. (1984), Nonparametric estimation from accelerated life test with random censorship, *Reliability Theory and Models*, pp. 155–167.

Painter, S. (1996), Existence of non-gaussian scaling behavior in heterogeneous sedimentary formations, *water resourse res.*, 32, pp. 1183–1195.

Park, C., Hernandez-Campos, F., Marron, J. S., and Smith, F. D. (2005), Long-range dependence in a changing internet traffic mix, *Computer Networks*, 48, pp. 401–422.

Phadia, E. G. (1980), A note on empirical bayes estimation of a distribution function based on censored data, *Ann. Statist.*, 8, pp. 226–229.

Rao, C. R. and Wu, Y. (1989), A strongly consistent procedure for model selection in a regression problem, *Biometrika*, 76, pp. 369–374.

Sacks, J. and Ylvisaker, D. (1981), Asymptotically optimum kernels for density estimation at a point, *Ann. Statist.*, 9, pp. 334–346.

Samko, S. G., Kilbas, A. A., and Marichev, O. I. (1993), *Fractional integrals and derivatives theory and applications*, Gordon and Breach Science Publishers.

Samuelsen, S. O. (1989), Asymptotic theory for nonparametric estimators from doubly censored data, *Scandinavian Journal of Statistics*, 16, pp. 1–21.

Schwarz, G. (1978), Estimating the dimensions of a model, *Ann. Statist.*, 6, pp. 461–464.

Shao, J. (1997), An asymptotic theory for linear model selection, *Statistica Sinica*, 7, pp. 221–264.

Shi, P. and Tsai, C. (2002), Regression model selection —a residual likelihood approach, *Journal of the Royal Statistical Society: Series B*, 64.

Susarla, v. and Van Ryzin, J. (1976), Nonparametric bayesian estimation of survival curves from incomplete observations, *Journal of the American Statistical Association*, 61, pp. 897–902.

Susarla, V. and Van Ryzin, J. (1978), Large sample thoery for a bayesian nonparametric survival curve estimator based on censored samples, *Ann. Statist.*, 6, pp. 755–768.

Tiwari, R. c. (1988), Nonparametric bayes estimation of a distribution under nomination sampling, *IEEE Trans. In Rel.*, 37, pp. 558–561.

Turcotte, D. L. (1994), Fractal theory and the estimatioin of extreme floods, *J. Res. Nat. Inst. Stand. Tech.*, 99, pp. 377–389.

Turnbull, B. W. (1974a), Nonparametric estimation of a survivorship function with boubly censored data, *Journal of the American Statistical Association*, 69, pp. 169–173.

Turnbull, B. W. (1974b), Nonparametric estimation of a survivorship function with doubly censored data, *Journal of the American Statistical Association*, 69, pp. 169–173.

Uzunogullari, U. and Wang, J. L. (1992), A comparison of hazard rate estimators for left truncated and right censored data, *Biometrika*, 79, pp. 297–310.

Wang, Y. (1996), Function estimation via wavelet shrinkage for long-memory data, *Ann. Statist.*, 24, pp. 466–484.

Wang, Y. (1997), Minimax estimation via wavelets for indirect long-memory data, *Journal of statistical planning and inference*, 64, pp. 45–55.

Whittle, P. (1960), Bounds for the moments of linear and quadratic forms in independent variables, *Theory of Probability And Its Applications*, 5(3), pp. 302–305.

Wornell, G. W. and Oppenheim, A. V. (1992), Estimation of fractal signals from noisy measurements using wavelets, *IEEE Trans. Signal Proc.*, 40, pp. 611–623.

Yu, Q. Q. and Li, L. X. (2001), Asymptotic properties of self-consistent estimators with doubly-censored data, *Acta Mathematica Sinica, English Series*, 17, pp. 581–594.

Zhang, C.-h. and Li, X. (1996), Linear regression with doubly censored data, *Ann. Statist.*, 24, pp. 2720–2743.

Zhao, L. H. (1997), Minimax linear estimation in a white noise problem, *Ann. Statist.*, 25, pp. 745–755.

Zheng, X. and Loh, W. (1997), A consistent variable selection criterion for linear models with high-dimensional covariates, *Statistica Sinica*, 7, pp. 311–325.