Concentration inequalities for Poisson point processes with applications to non-parametric statistics

Inauguraldissertation
zur Erlangung des akademischen Grades
eines Doktors der Naturwissenschaften
der Universität Mannheim

vorgelegt von
Diplom-Mathematiker Martin Kroll
aus Euskirchen

Mannheim, 2017

Dekan: Professor Dr. Heinz Jürgen Müller, Universität Mannheim
Referent: Professor Dr. Martin Schlather, Universität Mannheim
Korreferent: Professor Dr. Jan Johannes, Ruprecht-Karls-Universität Heidelberg

Tag der mündlichen Prüfung: 22. März 2017

Abstract

In the first part of this thesis we derive new concentration inequalities for maxima of empirical processes associated with independent but not necessarily identically distributed Poisson point processes. The proofs are based on a careful application of Ledoux's entropy method.

In the second part of the thesis, we show potential applications of the concentration results derived in the first part to non-parametric statistics: we consider intensity estimation for Poisson point processes from direct (Chapter 3) and indirect (Chapter 4) observations and non-parametric Poisson regression (Chapter 5). For all the considered models we develop a minimax theory (upper and lower bounds) under abstract smoothness assumptions on the unknown functional parameter. We study projection estimators in terms of trigonometric basis functions. The performance of these estimators crucially depends on the choice of a dimension parameter. For all our applications, we propose a fully data-driven selection of the dimension parameter based on model selection. The resulting adaptive estimators either attain optimal rates of convergence or are suboptimal only by a logarithmic factor.

Zusammenfassung

Im ersten Teil der vorliegenden Arbeit leiten wir neue Konzentrationsungleichungen für Maxima von empirischen Prozessen assoziiert zu unabhängigen, aber nicht notwendigerweise identisch verteilten Poissonschen Punktprozessen her. Die Beweise basieren auf einer Anwendung von Ledoux' Entropie-Methode.

Im zweiten Teil der Arbeit behandeln wir mögliche Anwendungen der Konzentrationsresultate aus dem ersten Teil in der nichtparametrischen Statistik: Wir betrachten Intensitätsschätzung für Poissonsche Punktprozesse ausgehend von direkten (Kapitel 3) und indirekten (Kapitel 4) Beobachtungen sowie nichtparametrische Poisson-Regression (Kapitel 5). Für alle betrachteten Modelle entwickeln wir eine Minimax-Theorie (obere und untere Schranken) unter abstrakten Glattheitsannahmen an den unbekannten funktionalen Parameter. Wir betrachten Projektionsschätzer basierend auf trigonometrischen Basisfunktionen. Die Güte dieser Schätzer hängt entscheidend von der Wahl eines Dimensionsparameters ab. Für alle betrachteten Anwendungen schlagen wir, basierend auf Modellwahl, eine rein datengetriebene Wahl des Dimensionsparameters vor. Die daraus resultierenden adaptiven Schätzer nehmen entweder die optimale Konvergenzrate an oder sind suboptimal um lediglich einen logarithmischen Faktor.

Contents

Introduction

> Das Neue ergibt sich aus dem Alten, aber auch das Alte verändert sich fortwährend im Lichte des Neuen und nimmt Merkmale an, die auf keiner früheren Stufe sichtbar waren.
>
> *(Arnold Hauser)*

Poisson point processes (PPPs) are of fundamental importance in probability theory and statistics, both from a theoretical and an applied point of view. For instance, they serve as elementary building blocks for complex point process models which are used in stochastic geometry [Sto+13], and a wide range of applications including, amongst others, extreme value theory [Res87], finance [BH09], forestry [PS00], and queueing theory [Bré81].

The distribution of a PPP is completely determined by its so-called *intensity measure*. Thus, from a statistical point of view, the (non-parametric) estimation of the intensity measure from observed realizations of the point process is of central importance. The theoretical analysis of adaptive non-parametric estimators, however, is often essentially based on the availability of appropriate concentration inequalities. Hence, besides being of independent interest, the derivation of such concentration inequalities is of fundamental importance for non-parametric statistics, and turns out to be a hard challenge in probability theory.

This thesis establishes novel concentration inequalities for PPPs and discusses potential applications of such inequalities to non-parametric estimation. Accordingly, the thesis is divided into two main parts: the first part recaps basic point process terminology and provides concentration inequalities for maxima of empirical processes associated with independent but not necessarily identically distributed PPPs. The second part is devoted to applications of these concentration results to non-parametric estimation in models where the observations are either independent realizations of point processes or closely related to such observations: intensity estimation from direct and indirect observations as well as estimation of the regression function in a Poisson regression model will be studied. In the sequel, we will give a short summary of the topics and methodology the reader can expect from the respective parts of this work.

Part I: Concentration inequalities

Concentration inequalities belong to the main tools in probability theory and statistics. In particular, classical results like the inequalities due to Markov, Hoeffding, Bernstein and Bennett are exhaustively used. The theoretical analysis of many estimation procedures in non-parametric statistics, however, is based on more elaborate concentration results that have been derived during the last decades. The recent monograph [BLM16] provides a comprehensive introduction into this topic.

The following result by Cirel'son, Ibragimov and Sudakov [CIS76] is regarded as one of the starting points in the modern development of concentration inequalities. The following formulation is taken from [BLM16] (cf. Theorem 5.6 therein).

Theorem 1. *Let $\boldsymbol{X} = (X_1, \dots, X_n)$ be a vector of n independent standard normal random variables. Let $f : \mathbb{R}^n \to \mathbb{R}$ denote a Lipschitz function with Lipschitz constant L. Then, for all*

$t > 0$,

$$\mathbb{P}(f(\boldsymbol{X}) - \mathbb{E}f(\boldsymbol{X}) \geq t) \leq e^{-\frac{t^2}{2L^2}}.$$

The original proof of Theorem 1 is based on stochastic calculus, an alternative one can be found in [BLM16]. Further concentration results were obtained by using martingale arguments [Yur76], [SS87], [McD89].

It turns out that in statistical applications one is often interested in concentration inequalities for maxima of empirical processes associated with a sequence of random variables in non-Gaussian frameworks. Ground-breaking results in this setup have been developed in a series of papers by Talagrand in the mid-1990s [Tal95]; [Tal96]. The following formulation of the *Talagrand inequality* is taken from [Mas00] (cf. Theorem 1 therein).

THEOREM 2 (Talagrand). *Consider n independent and identically distributed random variables $X_1, \ldots, X_n$ with values in some measurable space $(\mathbb{X}, \mathscr{X})$. Let $\mathcal{S}$ be some countable family of real-valued measurable functions on $(\mathbb{X}, \mathscr{X})$, such that $\|s\|_\infty \leq b < \infty$ for every $s \in \mathcal{S}$. Let $Z = \sup_{s\in\mathcal{S}} \sum_{i=1}^n s(X_i)$ and $v = \mathbb{E}[\sup_{s\in\mathcal{S}} \sum_{i=1}^n s^2(X_i)]$. Then for every positive number x,*

$$\mathbb{P}(Z \geq \mathbb{E}Z + x) \leq K \exp\left[-\frac{1}{K'}\frac{x}{b}\log\left(1 + \frac{xb}{v}\right)\right]$$

and

$$\mathbb{P}(Z \geq \mathbb{E}Z + x) \leq K \exp\left[-\frac{x^2}{2(c_1 v + c_2 bx)}\right] \tag{1}$$

where K, K', c_1 and c_2 are universal positive constants. Moreover, the same inequalities hold when replacing Z by $-Z$.

The variance factor v in the statement of Theorem 2 is called the *weak variance* (cf. [BLM16], p. 314). Talagrand's original proof is essentially based on geometric arguments and rather involved. Ledoux [Led96] proposed the *entropy method* as a different and more accessible approach to regain Talagrand's results but did not exactly recover the statement of Theorem 2. Instead, he proved a version of Theorem 2 with $v = \mathbb{E}[\sup_{s\in\mathcal{S}} \sum_{i=1}^n s^2(X_i)]$ replaced with

$$v = \mathbb{E}\left[\sup_{s\in\mathcal{S}} \sum_{i=1}^n s^2(X_i)\right] + \frac{4}{21} b\mathbb{E}[Z].$$

In addition, Ledoux was able to obtain reasonably sized constants in the statement of Talagrand's inequality. Based on an adaption of Gross's logarithmic Sobolev inequality in the Gaussian case to the non-Gaussian setup, Massart [Mas00] gave a version of (1) in a framework where the random variables $X_1, \ldots, X_n$ are independent but eventually not identically distributed. In this case, he was able to show that (1) holds with $K = 1$, $c_1 = 8$, and $c_2 = 2.5$. Massart also remarked that from a statistical point of view one is more interested in a version of the bound (1) with $v = \mathbb{E}[\sup_{s\in\mathcal{S}} \sum_{i=1}^n s^2(X_i)]$ replaced by

$$v = \sup_{s\in\mathcal{S}} \mathbb{E}[\sum_{i=1}^n s^2(X_i)]$$

which is usually called the *wimpy variance.* In [Mas00] such a version was shown, however, the correctness of a version with $c_1 = 1$ was only conjectured. This result was finally proven by Klein and Rio in [KR05].

THEOREM 3 ([KR05], Theorem 2.1). *Let $X_1, \ldots, X_n$ be a sequence of independent random variables with values in some Polish space $\mathbb{X}$ and let $\mathcal{S}$ be a countable class of measurable functions with* *values in* $[-1,1]^n$*. Suppose that* $\mathbb{E}[s^k(X_k)] = 0$ *for any* $s = (s^1, \ldots, s^n) \in \mathcal{S}$ *and any* $k \in \{1, \ldots, n\}$*. Put* $S_n(s) = s^1(X_1) + \ldots + s^n(X_n)$ *for* $s \in \mathcal{S}$*,* $Z = \sup_{s \in \mathcal{S}} S_n(s)$ *and define* $L_Z(t) = \log \mathbb{E}[e^{tZ}]$ *as the logarithm of the moment-generating function of* Z*. Then, for any positive* t*,*

a) $$L_Z(t) \leq t\mathbb{E}Z + \frac{t}{2}\left(2\mathbb{E}Z + V_n\right)\left(\exp((e^{2t}-1)/2) - 1\right).$$

Setting $\upsilon := 2\mathbb{E}Z + V_n$ *with* $V_n := \sup_{s \in \mathcal{S}} \operatorname{Var} S_n(s)$*, we obtain that, for any non-negative* x*,*

b) $$\mathbb{P}\left(Z \geq \mathbb{E}Z + x\right) \leq \exp\left(-\frac{x}{4}\log(1 + 2\log(1 + x/\upsilon))\right),$$

and

c) $$\mathbb{P}\left(Z \geq \mathbb{E}Z + x\right) \leq \exp\left(-\frac{x^2}{\upsilon + \sqrt{\upsilon^2 + 3\upsilon x} + (3x/2)}\right) \leq \exp\left(-\frac{x^2}{2\upsilon + 3x}\right).$$

Before introducing our point process setup, let us sketch two of the main tools that are commonly used for the derivation of concentration results. Although in our setup the occurring terms will be more complicated, these two techniques determine the structure of our later approach, in particular of the proof of Theorem 2.1 given in Chapter 2 below.

Ledoux's entropy method

The following lemma contains the key argument of the *entropy method*. Its formulation is taken from [Kle03] (cf. p. 16, Lemme 1 therein).

LEMMA 4. *Let* $X_1, \ldots, X_n$ *be independent random variables with values in a Polish space* $\mathbb{X}$*. Let* $\mathscr{F}_n$ *be the* σ*-field generated by* $X_1, \ldots, X_n$ *and* $\mathscr{F}_n^k$ *the* σ*-field generated by* $X_1, \ldots, X_{k-1}, X_{k+1}, \ldots, X_n$*. Let* $\mathbb{E}_n^k$ *denote the conditional expectation operator associated to* $\mathscr{F}_n^k$ *and let* f *be a strictly positive* $\mathscr{F}_n$*-measurable function with* $\mathbb{E}[f \log f] < \infty$*. Then it holds that*

$$\mathbb{E}[f \log f] - \mathbb{E}[f]\log \mathbb{E}[f] \leq \sum_{k=1}^{n} \mathbb{E}[f \log(f/\mathbb{E}_n^k f)]. \tag{2}$$

If Z denotes the random variable of interest (for instance, Z might be defined as in Theorem 3), applying the above lemma to the function $f(t) = e^{tZ}$ yields that the left-hand side of (2) is equal to $tF'(t) - F(t)\log F(t)$ where $F(t) = \mathbb{E}[\exp(tZ)]$ is the moment-generating function of Z. If one is able to bound the term on the right-hand side by some term of the form $F(t)V(t)$, division by $F(t)$ yields that

$$tL_Z'(t) - L_Z(t) \leq V(t) \tag{3}$$

where $L_Z(t) = \log F(t)$. Now, Herbst's argument can be used to deduce from (3) an upper bound for the logarithm of the moment-generating function.

Herbst's argument

The starting point of Herbst's argument is the observation that with L_Z as above we obtain from (3)

$$\frac{L_Z'(t)}{t} - \frac{L_Z(t)}{t^2} \leq \widetilde{V}(t)$$

for $t > 0$ and $\widetilde{V}(t) = \frac{V(t)}{t^2}$. One observes that the left-hand side of the last inequality is equal to the derivative of $\frac{L_Z(t)}{t}$. Thus, for every positive $\varepsilon > 0$, we get by integration

$$\frac{L_Z(t)}{t} - \frac{L_Z(\varepsilon)}{\varepsilon} \leq \int_{\varepsilon}^{t} \widetilde{V}(s)ds.$$

Taking the limit $\varepsilon \to 0$ on the left-hand side yields $\lim_{\varepsilon \to 0} L_Z(t)/t - L_Z(\varepsilon)/\varepsilon = L_Z(t)/t - \mathbb{E}Z$ and if one is able to find a reasonable expression for the integral on the right-hand side, one can obtain a bound on the logarithm of the moment-generating function of the form

$$L_Z(t) \leq t\mathbb{E}Z + \check{V}(t)$$

for some suitable function $\check{V}$. Such a bound can usually by used to obtain upper bounds for tail probabilities via Markov's inequality.

The main contribution of the first part of this thesis is to establish an analogue of Theorem 3 and a variant for left-hand side deviations from the mean (inspired by Theorem 2.3 in [KR05]) in a framework where the random variables are replaced with PPPs.

More precisely, we will consider independent (but eventually not identically distributed) PPPs $N_1, \ldots, N_n$ with finite intensity measures on a Polish space $\mathbb{X}$. As in the statement of Theorem 3, we start with a countable set of measurable functions $\mathcal{S}$ from $\mathbb{X}$ to $[-1, 1]$. For $s = (s^1, \ldots, s^n) \in \mathcal{S}$, we define

$$I^k(s) := \int_{\mathbb{X}} s^k(x)(dN_k(x) - d\Lambda_k(x)) \quad \text{and} \quad S_n(s) := I^1(s) + \ldots + I^n(s).$$

All the $I^k(s)$ are, exactly as the $s^k(X_k)$ in the statement of Theorem 3, centred random variables and we aim for concentration inequalities for the quantity $Z := \sup_{s \in \mathcal{S}} S_n(s)$ in the flavour of Theorem 3. Let us already mention that our results derived in Chapter 2 cannot be immediately deduced from Theorem 3 by interpreting PPPs as random variables in the space of locally finite $\mathbb{N}_0$-valued random measures equipped with an appropriate topology.

The following diagram illustrates how the first part of this thesis relates to and builds upon prior work.

$$\begin{array}{ccc} \textbf{[Mas00]} & \xrightarrow{\text{PPP}} & \textbf{[RB03]} \\ \downarrow & & \downarrow \\ \textbf{[KR05]} & \xrightarrow{\text{PPP}} & \textit{Chapter 2} \end{array}$$

The arrows labelled 'PPP' indicate that the work on the right-hand side transfers the results on the left-hand side of the arrow to the setup with PPPs. Vertical arrows indicate an improvement of results concerning the numerical constants involved. Not surprisingly, we will borrow ideas from both [RB03] and [KR05] to obtain our results.

Results from [KR05] have been exploited at various places in the literature as a starting point for further concentration results that can then be used in statistical applications. Following this guideline, we will also obtain a further concentration result (Proposition 2.13 together with the following Remark 2.15 in Chapter 2) that turns out to be useful for our statistical applications in Chapters 3–5 in the second part of this thesis.

Part II: Applications to non-parametric estimation

The second part of this thesis provides examples of how the concentration results derived in first part can be used to obtain theoretical results concerning the performance of adaptive non-parametric estimators. We will consider three different non-parametric statistical models which are treated in Chapters 3–5, respectively. The structure of the individual chapters will be essentially the same: first, inference of the unknown functional parameter of interest from the respective observations will be studied from a minimax point of view under mean integrated squared error loss. Under mild technical assumptions on the unknown infinite-dimensional parameter minimax upper and lower bounds will be determined. As the method of choice we focus on orthonormal series estimators in terms of the ordinary trigonometric basis (Chapters 3 and 5) or its complex-valued variant (Chapter 4). Such orthonormal series estimators of some functional parameter $\lambda \in \mathbb{L}^2(\mathbb{X})$

(here, $\mathbb{L}^2(\mathbb{X})$ denotes the set of square-integrable functions on $\mathbb{X}$ with respect to some pre-specified measure; in our applications we will exclusively consider the Lebesgue measure and $\mathbb{X}$ will be a bounded subset of $\mathbb{R}$) take on the form

$$\widehat{\lambda}_k(\cdot) := \sum_{0\leq |j|\leq k} \widehat{[\lambda]}_j \varphi_j(\cdot), \tag{3}$$

and are motivated by the $\mathbb{L}^2$-convergent representation $\lambda(\cdot) = \sum_{j\in\mathbb{Z}}[\lambda_j]\varphi_j$ in terms of some orthonormal basis $\{\varphi_j\}_{j\in\mathbb{Z}}$ in $\mathbb{L}^2(\mathbb{X}, dx)$ where the (generalized) Fourier coefficients $[\lambda]_j$ are given by

$$[\lambda]_j := \langle \lambda, \varphi_j \rangle_{\mathbb{L}^2}.$$

Certainly, the $\widehat{[\lambda]}_j$ in (3) should be (reasonable) estimators of the true $[\lambda]_j$. The quantity $k \in \mathbb{N}_0$ in (3) is a dimension parameter that has to be chosen by the statistician.

As the performance criterion for potential estimators $\widetilde{\lambda}$ of the unknown λ based on the respective observations we consider the mean integrated squared error $\mathbb{E}[\|\widetilde{\lambda} - \lambda\|^2]$ where $\mathbb{E}$ denotes the expectation operator associated with the distribution of the observations and expectation is taken under the true parameter λ. The minimax point of view consists in considering the worst case scenario over some class Λ of potential candidates of λ, that is in studying the *maximum risk*

$$\sup_{\lambda\in\Lambda} \mathbb{E}[\|\widetilde{\lambda} - \lambda\|^2].$$

Usually, the definition of the class Λ imposes structural pre-assumptions upon the function λ, for instance that λ belongs to some Sobolev ellipsoid, an ellipsoid of (generalized) analytic functions, or some Besov space. An estimator $\widehat{\lambda}$ is called *minimax optimal* if

$$\sup_{\lambda\in\Lambda} \mathbb{E}[\|\widehat{\lambda} - \lambda\|^2] = \inf_{\widetilde{\lambda}} \sup_{\lambda\in\Lambda} \mathbb{E}[\|\widetilde{\lambda} - \lambda\|^2],$$

and the quantity on the right-hand side is called the *minimax risk*. An estimator $\widehat{\lambda}$ is called *rate optimal* if

$$\sup_{\lambda\in\Lambda} \mathbb{E}[\|\widehat{\lambda} - \lambda\|^2] \lesssim \inf_{\widetilde{\lambda}} \sup_{\lambda\in\Lambda} \mathbb{E}[\|\widetilde{\lambda} - \lambda\|^2]$$

which by definition means that $\sup_{\lambda\in\Lambda} \mathbb{E}[\|\widehat{\lambda} - \lambda\|^2] \leq C \inf_{\widetilde{\lambda}} \sup_{\lambda\in\Lambda} \mathbb{E}[\|\widetilde{\lambda} - \lambda\|^2]$ for some constant C that does not depend on the sample size of the observations. In this thesis, we content ourselves throughout with the derivation of *rate optimal* estimators. It will turn out that the maximum risk of the estimator in (3) crucially depends on the correct specification of the dimension parameter k: the optimal choice k_n^* of this parameter in the minimax sense usually depends on the *a priori* knowledge of the class Λ. More precisely, its optimal value is such that the optimal compromise in the trade-off between bias and variance terms is achieved.

Since assuming the membership of λ to some *a priori* specified class Λ is not feasible in practice, there is need for a fully data-driven choice of the dimension parameter k which does not depend on any structural pre-assumptions on the parameter λ. Such an estimator is called *adaptive*. There are several approaches for data-driven selection procedures of so-called smoothing parameters, for instance *cross-validation* [AC10] or *Lepski's method* [Lep91]. Another approach to fully data-driven estimation is aggregation (cf., for instance, [BTW07], [LM09], [RT12]). In this thesis, we will exclusively use the *model selection* approach to adaptive estimation which has been introduced in the 1990s in a series of papers (see [BM97], [BM98], [BBM99], and [Mas07] for comprehensive treatments of this approach). In the following, let us give a sketch of this model selection approach.

Adaptive estimation via model selection in a nutshell

In this paragraph, we stick by the terminology and standard notation commonly used in papers dealing with model selection in non-parametric statistics. For $n \in \mathbb{N}$ (denoting the number of observations in our estimation frameworks later on), let us denote with $\mathcal{M}_n$ a finite set of *admissible* 'models'. Take note that the cardinality of the set of models is allowed to vary with the sample size n. Every model $\mathfrak{m} \in \mathcal{M}_n$ is assumed to be associated with a linear subspace $\mathcal{S}_\mathfrak{m}$ of $\mathbb{L}^2(\mathbb{X})$ and an estimator $\widehat{\lambda}_\mathfrak{m} \in \mathcal{S}_\mathfrak{m}$. Furthermore, assume that there is a linear subspace $\mathcal{S}_n \subseteq \mathbb{L}^2(\mathbb{X})$ such that $\mathcal{S}_\mathfrak{m} \subseteq \mathcal{S}_n$ for all $\mathfrak{m} \in \mathcal{M}_n$. The task of model selection is to choose from the collection $(\widehat{\lambda}_\mathfrak{m})_{\mathfrak{m}\in\mathcal{M}_n}$ an estimator in a completely data-driven way. For this purpose, two further ingredients are necessary, namely

(i) an empirical contrast function $\Upsilon_n : \mathcal{S}_n \to \mathbb{R}$, and

(ii) a penalty function $\textsc{Pen} : \mathcal{M}_n \to \mathbb{R}, \mathfrak{m} \mapsto \textsc{Pen}_\mathfrak{m}$.

Note that the dependence of the contrast function on the given observations is suppressed in our notation. The penalty term is often of the form

$$\textsc{Pen}_\mathfrak{m} = \frac{\kappa \mathbb{D}_\mathfrak{m} L_\mathfrak{m}}{n}$$

with a numerical constant κ, $\mathbb{D}_\mathfrak{m}$ the 'dimension' of the model $\mathfrak{m}$ and a weight factor $L_\mathfrak{m} \geq 1$. The penalty terms considered in this thesis will contain a random variable as proportionality factor instead of a deterministic κ. Moreover, different kind of weight factors $L_\mathfrak{m}$ will be considered. In Chapter 3, we consider the choice $L_\mathfrak{m} \equiv 1$ and in Chapters 4 and 5 the choice $L_\mathfrak{m} \equiv \log n$. These two choices are the standard ones used in the research literature (cf. [BBM99], p. 58). In Chapter 4, we will also consider a more elaborate choice of the penalty that is inspired by the definition of the penalty in [JS13a].

Based on the definition of contrast and penalty, a fully data-driven estimator from the collection $\{\widehat{\lambda}_\mathfrak{m}\}_{\mathfrak{m}\in\mathcal{M}_n}$ is chosen as $\widehat{\lambda}_{\widehat{\mathfrak{m}}}$ where

$$\widehat{\mathfrak{m}} := \operatorname*{argmin}_{\mathfrak{m}\in\mathcal{M}_n} \{\Upsilon(\widehat{\lambda}_\mathfrak{m}) + \textsc{Pen}_\mathfrak{m}\},$$

and one chooses an arbitrary minimizing model if there is no unique minimizer. Typically, under some conditions, one can derive for the data-driven estimator $\widehat{\lambda}_\mathfrak{m}$ so-called 'oracle inequalities' of the form

$$\mathbb{E}[\|\widehat{\lambda}_{\widehat{\mathfrak{m}}} - \lambda_\mathfrak{m}\|^2] \lesssim \inf_{\mathfrak{m}\in\mathcal{M}_n} [\|\lambda - \lambda_\mathfrak{m}\|^2 + \textsc{Pen}_\mathfrak{m}] + \text{'terms of lower order'} \tag{4}$$

where $\lambda_\mathfrak{m}$ denotes the projection of the function λ onto the linear space $\mathcal{S}_\mathfrak{m}$. Obviously, in order to make the first term on the right-hand side small, one should choose the class of models $\mathcal{M}_n$ as extensive as possible. However, in order to show that some of the terms arising in the proof of (4) are indeed 'of lower order', one usually has to impose some restrictions on the size of $\mathcal{M}_n$. The standard condition usually postulated in the literature is the existence of a universal constant C such that $\sum_{\mathfrak{m}\in\mathcal{M}_n} \exp(-L_\mathfrak{m}\mathbb{D}_\mathfrak{m}) \leq C < \infty$. Our definitions of the considered adaptive estimators are such that a similar condition is in fact satisfied. Often one can establish for the remainder terms the parametric rate n^{-1} as an upper bound. For this purpose, concentration inequalities are used: in Gaussian regression frameworks one can use for instance the classical inequalities due to [CIS76]. In density estimation setups, arguments are based on Talagrand's inequality and consequences of it. In our applications we mainly build our arguments on the concentration inequalities derived in the first part of the thesis.

The abstract model selection paradigm sketched above has been applied in great variety of non-parametric estimation problems. The following list provides some exemplary applications and is far away from being exhaustive:

- density estimation in mixed Poisson models [CGC15],
- adaptive estimation of the spectrum of a stationary Gaussian sequence [Com01],
- adaptive estimation of the transition density of hidden Markov chains [Lac08],
- circular deconvolution [JS13a],
- adaptive functional linear regression [CJ12],
- estimation of the jump size density for mixed compound Poisson processes [Com+15],
- non-parametric estimation of covariance functions [Big+10],
- optimal adaptive estimation of the relative density [CL15],
- non-parametric adaptive estimation of the drift for a jump diffusion process [Sch14].

Let us briefly sketch how the representation in the second part of this thesis fits into the general framework of adaptive estimation via model selection. For $n \in \mathbb{N}$ being the number of observations, the collection $\mathcal{M}_n$ of models will be given by the set $\mathcal{M}_n = \{0, \ldots, N_n\}$ for some $N_n \leq n$. For all $k \in \mathcal{M}_n$, the estimator $\widehat{\lambda}_k$ from Equation (3) is the associated orthonormal series estimator on the linear subspace $\mathcal{S}_k = \operatorname{span}(\{\varphi_j : 0 \leq |j| \leq k\}) \subseteq \mathcal{S}_{N_n}$ (recall that we denote with $\{\varphi_j\}_{j \in \mathbb{Z}}$ an orthonormal basis of the space $\mathbb{L}^2(\mathbb{X}) \ni \lambda$). In this specific situation, the data-driven choice $\widehat{k}$ of k can be written as

$$\widehat{k} := \operatorname*{argmin}_{0 \leq k \leq N_n} \{\Upsilon(\widehat{\lambda}_k) + \textsc{Pen}_k\}.$$

For instance, the choice of the penalty that we will use in Chapter 3 is proportional to $\frac{2k+1}{n}$ (the proportionality factor being a random variable ≥ 1) which fits into the general setup by setting $\mathbb{D}_{\mathfrak{m}} = 2k + 1$ and $L_{\mathfrak{m}} = 1$. In this case, we will obtain a result of the form

$$\mathbb{E}[\|\widehat{\lambda}_{\widehat{k}} - \lambda\|^2] \lesssim \min_{0 \leq k \leq N_n} \max \left\{ \|\lambda_k - \lambda\|^2, \frac{2k+1}{n} \right\} + \text{'terms of lower order'}, \tag{5}$$

and this bound even holds uniformly over the considered classes of potential parameters λ.

Here, by definition $\lambda_k = \sum_{0 \leq |j| \leq k} [\lambda]_j \varphi_j$ and the term $\|\lambda_k - \lambda\|^2$ corresponds to the squared bias. Finding the minimum on the right-hand side of (5) can be viewed as looking for the best compromise between squared bias and penalty. If the penalty term can be chosen proportional to the variance of $\widehat{\lambda}_k$ (which holds true in the setup of Chapter 3), finding the best compromise between squared bias and penalty is equivalent to finding the best compromise between squared bias and variance. Thus, the estimator $\widehat{\lambda}_{\widehat{k}}$ will be minimax optimal over a class Λ of functions if $k_n^* \leq N_n$ (as above, k_n^* denotes the optimal choice of the dimension parameter from a minimax point of view). For that reason, one would like to choose the quantity N_n as large as possible. However, for too large values of N_n, it might be infeasible to control the remainder terms that lead to the 'terms of lower order' in (5). For all our statistical models we will exploit the concentration results tailored to the PPP framework considered in the first part of the thesis in order to control the remainder terms of lower order. For the adaptive inverse intensity estimation in case of Cox observations in Chapter 4 and the Poisson regression model investigated in Chapter 5, we will have to exploit well-known concentration results for random variables in addition.

For the rest of this introduction, let us give a brief overview of the statistical models that we will consider in more detail later. Moreover, we provide some motivational background and give references to related work.

Intensity estimation from direct observations

In the first application, we aim at estimating non-parametrically the intensity function λ of a PPP on some pre-specified compact interval $I \subsetneq \mathbb{R}$. We will consider the unit interval $I = [0, 1]$ without loss of generality. Here, the observations are given by an i.i.d. sample $N_1, \ldots, N_n$ from the Poisson process of interest. Using the representation of point processes as $\mathbb{N}_0$-valued random measures the observations take on the form

$$N_i = \sum_j \delta_{x_{ij}}$$

where $\delta_\bullet$ denotes the Dirac measure with mass concentrated at $\bullet$. Since one has access to observations with the target intensity, we will refer to this kind of observations as *direct* observations.

Intensity estimation in parametric and non-parametric models has been dealt with in a wide range of monographs and research papers. For general treatments of the subject, we refer to [Kar91] as an introduction to the statistics of point processes, [Kut98] for examples of intensity estimation in different parametric and non-parametric models, and [MW04] for estimation in general spatial models. Early approaches to non-parametric intensity estimation include kernel [Rud82]; [Kut98] and histogram estimators [Rud82]. In addition, the paper [Rud82] already discusses adaptive estimation of the intensity. Baraud and Birgé [BB09] consider a Hellinger type loss function and propose a histogram estimator for intensity estimation. Other contributions focus on non-linear wavelet thresholding techniques, see, for instance, the articles [Kol99]; [WN07]; [RBR10]; [San14] and [Big+13]. The paper [Big+13] proposes a non-linear hard thresholding estimator for intensity estimation from *noisy* observations. The article [Bir07] proposes a model selection approach based on hypothesis testing for adaptive intensity estimation. Moreover, there exist other approaches to non-parametric intensity estimation in more specific models. Let us mention the paper [GN00] that proposes a minimum complexity estimator in the Aalen model and [PW04] that uses a wavelet approach to estimation in a multiplicative intensity model, without making a claim to be exhaustive. The paper most closely related to our presentation is [RB03] where intensity estimation from one single direct observation on the interval $[0, T]$ is considered and asymptotics as $T \to \infty$ are studied. The analysis of the adaptive estimator in that paper is also based on the use of concentration inequalities but our analysis is rather inspired by the one in [JS13a] in a circular deconvolution model.

Intensity estimation from indirect observations in a circular model

The second statistical model that we consider is closely related to the first one. As in this model, we are interested in estimating the intensity function λ of some PPP, now with state space $I = [0, 1)$. In contrast to the previous model, we are now not able to observe direct realizations of the point process with the target intensity but instead observe the i.i.d. sample $N_1, \ldots, N_n$ where

$$N_i = \sum_j x_{ij} + \varepsilon_{ij} - \lfloor x_{ij} + \varepsilon_{ij} \rfloor \tag{6}$$

where the ε_{ij} are additive errors. Here, the hidden point processes $\widetilde{N}_i = \sum_j \delta_{x_{ij}}$ are PPPs with intensity function $\lambda \in \mathbb{L}^2([0, 1), dx)$ which is the functional parameter of interest. This leads to a *statistical inverse problem* which is closely related to (circular) deconvolution problems [JS13a]; [CL10]; [CL11].

At this point, some comments seem to be necessary. The first one concerns the additive errors ε_{ij} in (6). In our investigation we will assume that the ε_{ij} are stationary in the sense that $\varepsilon_{ij} \sim f$ for some unknown error density f. Note that different dependency structures concerning the additive errors ε_{ij} lead to different kinds of point process observations. We will focus on the following two cases:

1. the errors ε_{ij} are i.i.d. $\sim f$. In this case the observed point processes N_i are again Poisson. We will refer to this case in Chapter 4 as model 1 or the model with Poisson observations.

2. The error does only depend on the index i, that is $\varepsilon_{ij} = \varepsilon_i$ for $i = 1, \ldots, n$ and arbitrary j. This means that all the points of the hidden point process $\widetilde{N}_i$ are shifted by the same amount ε_i modulo 1. In this case, the observed point processes stem from a Cox process. We will

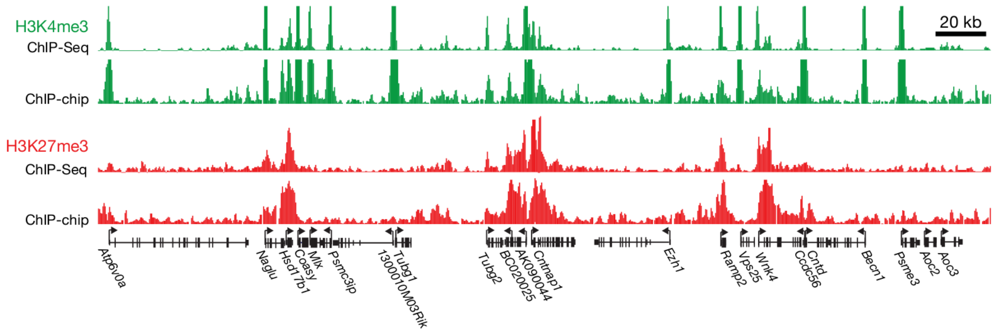


Figure 1.: ChipSeq Data are useful to determine the evolutionary state of a cell. The figure is taken from [Mik+07].

refer to this model as model 2 or the model with Cox process observations. This model has already been studied in the recent article [Big+13].

The second comment concerns the knowledge of the distribution of the additive errors. The conservative assumption is to assume that the error distribution is known [Big+13]. However, if the error distribution is not known in advance, which is obviously more realistic, then the model has to be further specified to guarantee even identifiability. For instance, one can assume that the error distribution belongs to some certain parametric class (this approach has been considered in a blind convolution framework in [SVB10]). Another option, which is the one we will consider in this work, is to assume the availability of a second i.i.d. sample $Y_1, \ldots, Y_m \sim f$. The availability of a second sample from the error distribution has already been assumed in density deconvolution setups, for instance in [DH93], [Joh09] or [CL11]. Not surprisingly, the rate of convergence in this setup will depend on both sample sizes n and m.

The third and last comment deals with the, on a first view, cryptical form of the contaminated observations in (6). Obviously, via the standard identification of $[0, 1)$ with the circle of perimeter 1, we can interpret the observations in (6) as contaminated point processes on the circle. The main technical advantage is that we can make use of the *complex trigonometric basis* in this setup and exploit its convenient behaviour with respect to deconvolution. From a practical point of view, circular models are capable of modelling periodic intensities that appear in various applications. We refer the interested reader for a detailed list of examples to the paper [HWMZ03].

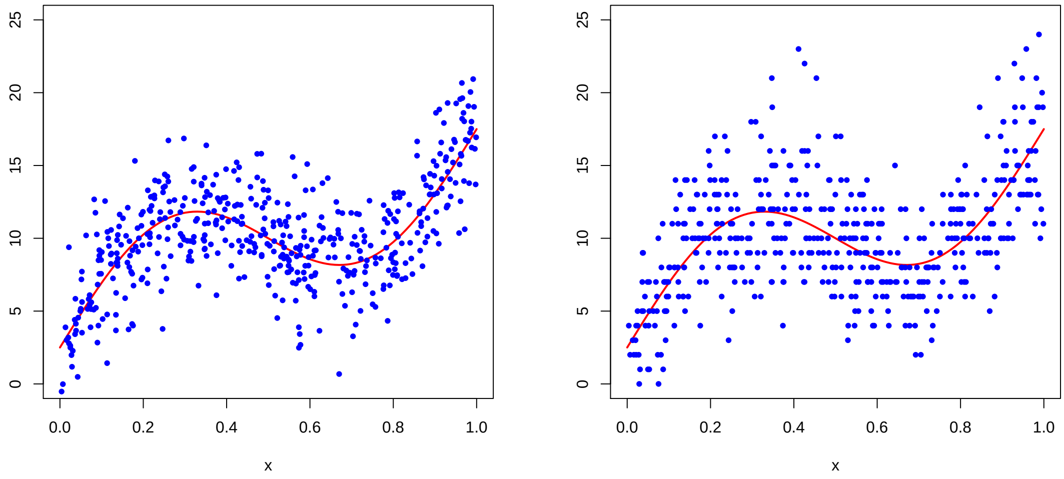
A recent motivation for dealing with this circular model in the case of Cox process observations comes from genomics, in particular the analysis of so-called ChIP-seq data (Figure 1). The analysis of such data is useful to determine the state of a cell in its evolution from totipotent stemcells to terminal differentiation [Mik+07]. We refer to [Big+13] for further details concerning the motivation from genomics of the statistical model and to [Mik+07] where the biological background is explained in detail.

The considered models 1 and 2 can be seen as special instances of *Poisson inverse problems* [AB06]. Besides [Big+13], [CJ02] is another paper falling under the umbrella of Poisson inverse problems where a wavelet shrinkage approach for intensity estimation of tomographic data is studied.

Nonparametric Poisson regression

Chapter 5 is concerned with a regression model for count data. We assume the availability of an i.i.d. sample of observations $(X_1, Y_1), \ldots, (X_n, Y_n)$ satisfying

$$Y_i | X_i \sim \mathcal{P}(T\lambda(X_i)), \qquad i = 1, \ldots, n. \tag{7}$$



(a) Gaussian homoscedastic regression

(b) Poisson heteroscedastic regression

Figure 2.: Comparison of homoscedastic Gaussian regression and heteroscedastic Poisson regression. The red curve is the unknown regression function, the blue points are the observations. In the Gaussian case, the observations follow Equation (8) for normally distributed noise ε_i; in the Poisson case, the observations obey model (7).

Here, $\mathcal{P}(\alpha)$ denotes the Poisson distribution with parameter $\alpha \geq 0$, $T > 0$, and the functional parameter of interest is $\lambda : [0,1] \to [0,\infty)$. In this thesis, we will exclusively consider the random design case: the explanatory variables $X_1, \ldots, X_n$ form an i.i.d. sample where $X_i \sim f$ for some probability density function f on $[0,1]$.

Regression models for count data are widely used in the natural and social sciences as well as in economics [CT98]; [Win08]. The standard approach to regression for count data is a *generalized linear model* [Str13] of the form

$$\mathbb{E}[Y] = \exp(\beta x)$$

with link function $g(x) = \log(x)$. We refer the reader to [Win08] for a detailed treatment of this model. Besides purely parametric approaches there exist also semi-parametric approaches to Poisson regression problems (see, for instance, Chapter 12 in [CT98]). However, purely non-parametric approaches seem to be rare. One approach is to use the Anscombe transform [Ans48] of the data and treat the data as if they were Gaussian. Recent work has considered the regression model (7) in a high-dimensional framework using the LASSO and the group LASSO [IPR16]. Applications of a related regression model in a geostatistical context are provided in [DTM98]. However, this paper makes use of a fully parametric approach and suggests MCMC techniques for fitting a model to given data. The paper [CP02] introduces a semi-parametric Bayesian model for count data regression and applies it as a prognostic model for early breast cancer data.

Note that one characteristic feature of the regression model defined through (7) is that it naturally contains heteroscedastic noise (see Figure 2). Besides work on regression under the assumption of homoscedasticity [Bar00], there exists already research that considers model selection techniques in regression frameworks containing heteroscedasticity [Sau13]. However, in [Sau13] the observations are of the form

$$Y_i = r(X_i) + \sigma(X_i)\varepsilon_i \tag{8}$$

where r is the unknown regression function to be estimated, the residuals ε_i have zero mean and variance one, and the function σ models the unknown heteroscedastic noise level. Note that this model does not contain our model (7). Besides the paper [IPR16] mentioned above, there does not seem to exist another contribution that considers non-parametric Poisson regression via the model selection approach. In the recent paper [KYS13], the authors consider a model selection approach in a parametric model via a bias-corrected AIC criterion.

The investigation of an adaptive estimator for Poisson regression under integrated squared error following the guidelines sketched in the paragraph on model selection above will include concentration results both for general random variables and those tailored to PPPs as derived in the first part of this thesis. Our construction of the adaptive estimator is split into two steps: the first approach is based on the assumption that an upper bound for $\|\lambda\|_\infty$ is known in advance. This upper bound is used in the definition of the penalty. In order to dispense with the *a priori* knowledge of an upper bound for $\|\lambda\|_\infty$, we replace the upper bound in the definition of the penalty with an estimator of $\|\lambda\|_\infty$. We follow an approach sketched in [BM97] which was used in [Com01] for the adaptive estimation of the spectral density of a stationary Gaussian sequence. The estimator of $\|\lambda\|_\infty$ is defined as the plug-in estimator $\|\widetilde{\lambda}\|_\infty$ where $\widetilde{\lambda}$ is an appropriately defined projection estimator of λ in terms of an orthonormal basis of piecewise polynomials. The resulting adaptive estimator of λ attains optimal rates of convergence up to a logarithmic factor.

Some of the results derived in this thesis have already been published in the following preprints:

[Kro16] Kroll, M. Concentration inequalities for Poisson point processes with application to adaptive intensity estimation. *arXiv preprint* (2016). arXiv: `1612.07901` (this paper is based on Chapters 2 and 3)

[Kro17] Kroll, M. Nonparametric intensity estimation from indirect point process observations under unknown error distribution. *arXiv preprint* (2017). arXiv: `1703.05619` (this paper is based on Chapter 4)

Acknowledgements

First of all, I am indebted to my supervisor Prof. Dr. Martin Schlather for the excellent supervision of my dissertation. His enthusiasm for mathematics and his curiosity about new topics provided a large part to my motivation.

I also would like to express my deep gratitude to Prof. Dr. Jan Johannes for reading so many intermediate versions of my papers and providing lots of helpful comments and remarks. Since he was already the supervisor of my diploma thesis, it is a honour to have him as second supervisor again. Indeed, without his encouragement, I would have barely started a PhD project years after my diploma. I am also thankful to him for inviting me to a research stay at the ENSAI in Rennes in March 2015 and the opportunity to give a talk at a session organized by him at the CFE-CMStatistics in Sevilla in December 2016.

I also thank all my colleagues at the statistics group, the Institute of Mathematics, and the Research Training Group RTG 1953 'Statistical Modeling of Complex Systems and Processes'. I also thank Anja Gilliar for taking care of all the administrative issues during the last two-and-a-half years which allowed me to concentrate on my scientific work.

Finally I sincerely thank my mother for constant support through my whole life.

Financial support by the Deutsche Forschungsgemeinschaft (DFG) through the Research Training Group RTG 1953 'Statistical Modeling of Complex Systems and Processes' is also gratefully acknowledged.

Part I.

Concentration inequalities for Poisson point processes

1. Preliminaries on point processes

In this chapter, we provide the fundamental concepts and results from point process theory needed in this thesis. We mainly follow the representation in [Kal76] and state the definitions and results for point processes with a *locally compact second countable Hausdorff* (LCCB) space as state space. This assumption concerning the topology will be satisfied by all state spaces considered in the applications in the second part of this thesis.

1.1. Random measures and point processes

For an arbitrary topological space $\mathbb{X}$, we denote its σ-field of Borel sets with $\mathscr{B} = \mathscr{B}(\mathbb{X})$. In addition, we denote with $\mathscr{B}'$ the subset of $\mathscr{B}$ containing all *topologically bounded* (that is, *relatively compact*) sets in $\mathscr{B}$.

Definition 1.1. Let $\mathbb{X}$ be a LCCB space. A measure μ on $(\mathbb{X}, \mathscr{B})$ is called *locally finite* (or *Radon*) if $\mu(B) < \infty$ for all $B \in \mathscr{B}'$. Let $\mathfrak{M} = \mathfrak{M}(\mathbb{X})$ be the set of all *locally finite* measures on $\mathbb{X}$ and $\mathfrak{N} = \mathfrak{N}(\mathbb{X}) \subseteq \mathfrak{M}$ be the subset of $\mathbb{N}_0$-valued locally finite measures. Furthermore, let $\mathscr{M} = \mathscr{M}(\mathcal{X})$ and $\mathscr{N} = \mathscr{N}(\mathcal{X})$ be the σ-fields in $\mathfrak{M}$ and $\mathfrak{N}$ which are generated by the mappings $\mu \mapsto \mu(B)$ for $B \in \mathscr{B}'$, respectively.

Remark 1.2. $\mathscr{N} \subseteq \mathscr{M}$ (cf. Lemma 1.5 in [Kal76]).

Definition 1.3. Let $\mathbb{X}$ be LCCB space. A *random measure* with state space $\mathbb{X}$ is a measurable mapping from some probability space $(\Omega, \mathscr{A}, \mathbb{P})$ to $(\mathfrak{M}, \mathscr{M})$. A *point process* with state space $\mathbb{X}$ is a measurable mapping from some probability space $(\Omega, \mathscr{A}, \mathbb{P})$ to $(\mathfrak{N}, \mathscr{N})$.

The σ-fields $\mathscr{M}$ and $\mathscr{N}$ can be equivalently defined as the Borel σ-fields corresponding to the so-called *vague topology* on the sets $\mathfrak{M}$ and $\mathfrak{N}$. For this, denote with $\mathcal{F} = \mathcal{F}(\mathbb{X})$ the class of all $\mathscr{B}$-measurable functions $f : \mathbb{X} \to [0, \infty)$, and with $\mathcal{F}_c = \mathcal{F}_c(\mathbb{X})$ the subclass of all continuous functions in $\mathcal{F}$ with compact support. Then, by definition, the vague topology is the topology generated by the base consisting of all finite intersections of subsets of $\mathfrak{M}$ (resp. $\mathfrak{N}$) of the form $\{\xi : s < \xi f < t\}$ with $f \in \mathcal{F}_c$, $s, t \in \mathbb{R}$ and $\xi f = \int_{\mathbb{X}} f d\xi$. Thus, a sequence of measures $\xi_i \in \mathfrak{M}$ tends to ξ if and only if $\xi_i f \to \xi f$ for all $f \in \mathcal{F}_c$.

The following theorem (together with the subsequent remark) will be exploited tacitly several times in the proofs of Chapter 2.

Theorem 1.4 ([Kal76], A 7.7.). *The spaces $\mathfrak{M}$ and $\mathfrak{N}$ equipped with the vague topology are Polish.*

Remark 1.5. The statement of Theorem 1.4 still holds true for state spaces that are not LCCB but only Polish. In this more general case, the *vague* topology has to be replaced with the so-called $w^{\#}$-topology ('*weak-hash*'-topology). In the case of locally compact $\mathbb{X}$, the notions of vague and $w^{\#}$-convergence coincide (see Appendix A2.6 in [DVJ03]).

1.2. The L-transform

By definition, the distribution of a random measure (or point process) ξ is the probability distribution $\mathbb{P}^{\xi}$ on $(\mathfrak{M}, \mathscr{M})$ (or $(\mathfrak{N}, \mathscr{N})$) given by

$$\mathbb{P}^{\xi}(M) = \mathbb{P}(\xi^{-1}(M)) = \mathbb{P}(\xi \in M), \quad M \in \mathscr{M} \text{ (or } M \in \mathscr{N}).$$

Theorem 1.7 below states equivalent conditions for equality in distribution of random measures. One of these equivalent conditions is stated in terms of the L-transform, which we define now.

Definition 1.6. Let ξ be a random measure with state space $\mathbb{X}$. The mapping

$$L_\xi : \mathcal{F} \to \mathbb{R} \cup \{\infty\}, \quad f \mapsto L_\xi(f) := \mathbb{E}[e^{-\xi f}]$$

is called the *L-transform* of ξ.

The L-transform uniquely determines the distribution of a random measure:

Theorem 1.7 (cf. [Kal76], Theorem 3.1). *Let ξ and η be random measures with state space $\mathbb{X}$. Then, the following assertions are equivalent:*

(i) $\xi \stackrel{d}{=} \eta$,

(ii) $\xi f \stackrel{d}{=} \eta f$ *for all* $f \in \mathcal{F}_c$,

(iii) $L_\xi(f) = L_\eta(f)$ *for all* $f \in \mathcal{F}_c$,

(iv) $(\xi(B_1), \ldots, \xi(B_k)) \stackrel{d}{=} (\eta(B_1), \ldots, \eta(B_k))$ *for all* $k \in \mathbb{N}$ *and* $B_1, \ldots, B_k \in \mathscr{B}'$.

Definition 1.8. A point process N with state space $\mathbb{X}$ satisfying

$$L_N(f) = e^{-\mu(1-e^{-f})}, \qquad f \in \mathcal{F},$$

for some $\mu \in \mathfrak{M}$ is called *Poisson point process* (PPP) with intensity measure Λ.

By Theorem 1.7, the distribution of a PPP is uniquely determined by its L-transform. For a proof of existence, we refer the reader to Chapter 1 of [Kal76]. Let us mention the following alternative characterization of PPPs (cf. [Ser09], Chapter 3, Definition 16) which is more intuitive than the one given by the L-transform:

Proposition 1.9. *A point process N on $\mathbb{X}$ is a Poisson point process with locally finite intensity measure μ if and only if the following two conditions are satisfied:*

(i) for $n \in \mathbb{N}$ *and* disjoint $B_1, \ldots, B_n \in \mathscr{B}'$, *the random variables* $N(B_1), \ldots, N(B_n)$ *are independent,*

(ii) for each $B \in \mathscr{B}'$, *the random variable* $N(B)$ *follows a Poisson distribution with parameter* $\mu(B)$.

In Chapter 4, we will encounter Cox processes which are a natural generalization of PPPs.

Definition 1.10. Let η be a random measure with state space $\mathbb{X}$. A point process N with state space $\mathbb{X}$ is called *Cox process* with *directing measure* η if

$$L_N(f) = \mathbb{E}[e^{-\eta(1-e^{-f})}] = L_\eta(1 - e^{-f}).$$

A Cox process is uniquely determined by its directing measure η (cf. Corollary 3.2 in [Kal76]). Since Cox processes arise from PPPs by mixing, the existence of such processes can be shown by means of a general existence theorem for mixtures of random measures (cf. Lemma 1.7 in [Kal76]).

1.3. Infinite divisibility

In the proofs of Chapter 2, we will exploit the fact that PPPs are *infinitely divisible*. Recall that a random variable X is called *infinitely divisible* if for each $n \in \mathbb{N}$, there exist i.i.d. random variables $X_1, \ldots, X_n$ such that $X \stackrel{d}{=} X_1 + \ldots + X_n$. The definition for the case of random measures and point processes is totally analogous.

Definition 1.11. A random measure ξ with state space $\mathbb{X}$ is called *infinitely divisible* if for each $n \in \mathbb{N}$ there exist i.i.d. random measures $\xi_1, \dots, \xi_n$ such that

$$\xi \overset{d}{=} \xi_1 + \ldots + \xi_n.$$

Analogously, a point process N is said to be *infinitely divisible* if for each $n \in \mathbb{N}$ there exist i.i.d. point processes $N_1, \dots, N_n$ such that

$$N \overset{d}{=} N_1 + \ldots + N_n. \tag{1.1}$$

Remark 1.12. There exist point processes N which are infinitely divisible as random measures but not as point processes. The simplest examples of this type are provided by deterministic elements of $\mathfrak{N}$.

For a full characterization of infinitely divisible random measures and point processes we refer to Chapters 6 and 7 of [Kal76]. For our purposes, it is sufficient to note that PPPs are infinitely divisible. More precisely, if N is a PPP with intensity measure μ, then equation (1.1) is satisfied for $N_1, \dots, N_n$ being i.i.d. PPPs with intensity $\frac{\mu}{n}$, respectively.

1.4. Campbell's theorem

Definition 1.13. Let N be a point process with state space $\mathbb{X}$. The mapping $\mu : \mathscr{B} \to \mathbb{R} \cup \{\infty\}, B \mapsto \mathbb{E}[N(B)]$ is called the *mean measure* of N.

Note that for Poisson processes the intensity measure and the mean measure coincide. The following theorem will be frequently used in this thesis.

Theorem 1.14 (cf. [Ser09], Chapter 3, Theorem 24). *Let N be a point process on the state space $\mathbb{X}$ with mean measure μ. Then, for any measurable $f : \mathbb{X} \to \mathbb{C}$, it holds*

$$\mathbb{E}\left[\int_{\mathbb{X}} f(x) dN(x)\right] = \int_{\mathbb{X}} f(x)\mu(dx) \tag{1.2}$$

provided that the integral on the right-hand side exists. If, in addition, N is a Poisson process, then

$$\operatorname{Var}\left(\int_{\mathbb{X}} f(x) dN(x)\right) = \int_{\mathbb{X}} |f(x)|^2 \mu(dx)$$

provided that the integral on the right-hand side exists.

Equation (1.2) is usually referred to as Campbell's theorem or compensation formula.

Remark 1.15. In [Ser09], Theorem 1.14 is stated for real-valued functions only. The statement for complex-valued functions follows by decomposition into real and imaginary part.

2. Concentration inequalities for Poisson processes

In this chapter, we derive concentration inequalities for suprema of empirical processes associated with Poisson point processes. These results will be used in the second part of this thesis for the derivation of upper risk bounds of adaptive non-parametric estimators in different models but might also be of independent interest. Although it would be sufficient to derive concentration inequalities for right-hand side deviations from the mean in view of our intended applications, we also state and prove concentration inequalities for left-hand side deviations.

The main technical prerequisites needed in this chapter are the finiteness of the intensity measures and the assumption that the common state space of the point processes is Polish. More precisely, we use the following notations: $N_1, \ldots, N_n$ is a sequence of independent PPPs with finite intensity measures (denoted with $\Lambda_1, \ldots, \Lambda_n$) on some Polish space $\mathbb{X}$ equipped with the σ-field $\mathscr{B}$ generated by the open sets in $\mathbb{X}$. Note that, thanks to the assumption that $\mathbb{X}$ is Polish, the space $\mathfrak{N}$ of $\mathbb{N}_0$-valued locally finite measures equipped with an appropriate topology (see Chapter 1) is itself Polish (see Theorem 1.4 and Remark 1.5).

In this framework, let $\mathcal{S}$ be a countable class of measurable functions from the space $\mathbb{X}$ into $[-1,1]^n$. For $s = (s^1, \ldots, s^n) \in \mathcal{S}$ and $k \in \{1, \ldots, n\}$, we define

$$I^k(s) := \int_{\mathbb{X}} s^k(x)(dN_k(x) - d\Lambda_k(x)) \quad \text{and} \quad S_n(s) := I^1(s) + \ldots + I^n(s). \tag{2.1}$$

The principal aim of this chapter is to establish concentration inequalities for the random variable $Z := \sup_{s \in \mathcal{S}} S_n(s)$.

2.1. Concentration inequalities for right-hand side deviations

The following theorem is the first main result of this chapter.

Theorem 2.1. *Let $N_1, \ldots, N_n$ be independent PPPs on a Polish space $\mathbb{X}$ with* finite *intensity measures $\Lambda_1, \ldots, \Lambda_n$, and $\mathcal{S}$ be a countable class of measurable functions from $\mathbb{X}$ to $[-1,1]^n$. For $s \in \mathcal{S}$, define $S_n(s)$ as in* (2.1) *and consider $Z := \sup_{s \in \mathcal{S}} S_n(s)$. Let $L(t) = L_Z(t) := \log \mathbb{E}[\exp(tZ)]$ denote the logarithm of the moment-generating function of Z and $V_n := \sup_{s \in \mathcal{S}} \operatorname{Var}(S_n(s))$. Then, for any non-negative t,*

a) $$L_Z(t) \leq t\mathbb{E}Z + \frac{t}{2}(2\mathbb{E}Z + V_n)(\exp((e^{2t}-1)/2) - 1).$$

Setting $\upsilon := 2\mathbb{E}Z + V_n$, we obtain that, for any non-negative x,

b) $$\mathbb{P}(Z \geq \mathbb{E}Z + x) \leq \exp\left(-\frac{x}{4}\log(1 + 2\log(1 + x/\upsilon))\right),$$

and

c) $$\mathbb{P}(Z \geq \mathbb{E}Z + x) \leq \exp\left(-\frac{x^2}{\upsilon + \sqrt{\upsilon^2 + 3\upsilon x} + (3x/2)}\right) \leq \exp\left(-\frac{x^2}{2\upsilon + 3x}\right).$$

Remark 2.2. We emphasize that Theorem 2.1 cannot be immediately deduced from Theorem 1.1 in [KR05]. For instance, if $s^k \equiv 1$ the stochastic integral $\int_{\mathbb{X}} s^k(x) dN_k(x)$ is an unbounded function of N_k (interpreted as a random variable in an appropriately defined state space) but obviously $s^k \equiv 1$ fits into the framework of Theorem 2.1.

Remark 2.3. The bounds obtained in Theorem 2.1 translate literally (that is, even with exact coincidence of the numerical constants involved) the ones obtained in Theorem 1.1 in [KR05] to the setup with PPPs. This observation is in accordance with the one made in the article [RB03] where the derived concentration inequalities translate literally previous results for the random variable setup due to [Mas00].

2.1.1. Notation and preparatory results

In this section, we introduce some notation and state preliminary results. The proof of Theorem 2.1, based on these results, is given in Section 2.1.2. The key property used to prove Theorem 2.1 is the *infinite divisibility* of the PPPs $N_1, \ldots, N_n$: for every $k \in \{1, \ldots, n\}$ and $\ell \in \mathbb{N}$, there exist i.i.d. PPPs N_{kj} such that

$$N_k \stackrel{d}{=} \sum_{j=1}^{\ell} N_{kj}. \tag{2.2}$$

The common intensity measure of the N_{kj} in this representation is Λ_k/ℓ. Throughout this chapter, the dependence of N_{kj}, Λ_{kj}, and derived quantities on ℓ is often suppressed for the sake of convenience.

Define $\Lambda := \sup_{k=1,\ldots,n} \Lambda_k(\mathbb{X})$ and $\Delta = \Delta(\ell) := \Lambda/\ell$. For $s \in \mathcal{S}$, let $I^{kj}(s) := \int_{\mathbb{X}} s^k(x)(dN_{kj}(x) - d\Lambda_{kj}(x))$. We define the random variable $X_{kj} := N_{kj}(\mathbb{X})$, that is, X_{kj} is the total number of points of the point process N_{kj}, and the event Ω_{kj} via $\Omega_{kj} := \{X_{kj} \leq 1\}$.

Remark 2.4. A natural interpretation of the proof of Theorem 2.1 given below is to consider the result being obtained in a setup with a *triangular array* of point processes

$$\begin{array}{l} N_1, \ldots, N_n \\ N_{11}, N_{12}, N_{21}, N_{22}, \ldots, N_{n1}, N_{n2} \\ N_{11}, N_{12}, N_{13}, N_{21}, N_{22}, N_{23}, \ldots, N_{n1}, N_{n2}, N_{n3} \\ \vdots \end{array}$$

where the point processes in each row are independent and the intensity measures of the single point processes in a row tend to zero when the row index tends to infinity. All asymptotic considerations will be obtained under the equivalent regimes $\ell \to \infty$ and $\Delta \to 0$, respectively.

Lemma 2.5. $\mathbb{P}(\Omega_{kj}^{\mathsf{c}}) \leq \Delta^2/2$.

Proof. The function $h : \mathbb{N}_0 \to \mathbb{R}, n \mapsto n^2 - n$ is non-negative and non-decreasing. Since $\Omega_{kj}^{\mathsf{c}} = \{X_{kj} \geq 2\}$ the claim estimate follows from Markov's inequality. □

Let us define the σ-fields

$$\mathscr{F}_n := \sigma(\{N_{11}, \ldots, N_{n\ell}\}) \qquad \text{and} \qquad \mathscr{F}_n^{kj} := \sigma\left(\{N_{11}, \ldots, N_{n\ell}\} \setminus \{N_{kj}\}\right).$$

Further, let $\mathbb{E}_n^{kj}[\,\cdot\,] := \mathbb{E}[\,\cdot\,|\mathscr{F}_n^{kj}]$, $\mathbb{P}_n^{kj}(A) := \mathbb{E}_n^{kj}[\mathbb{1}_A]$, $f = f(t) := \exp(tZ)$, and $f_{kj} = f_{kj}(t) := \mathbb{E}_n^{kj}[f]$. It will turn out to be sufficient to prove the results of this chapter under the following finiteness assumption.

Assumption 2.6. *$\mathcal{S} = \{s_1, \ldots, s_m\}$ is a* finite *set of measurable functions.*

Under the validity of Assumption 2.6, let τ denote the minimal value of i such that $Z = S_n(s_i)$.

Lemma 2.7. *Let Assumption 2.6 hold. Then, for any non-negative t,*

a) $f/f_{kj} \leq \exp(tI^{kj}(s_\tau))$, and in addition *b)* $\exp(-2(1+\Delta)t)(1-e^{(2+3\Delta)t}\exp(\Delta(e^{2t}-1)/2)\cdot\Delta/\sqrt{2})\le f/f_{kj}$ *on* Ω_{kj}.

PROOF. In order to prove statement a), set $S_n^{kj}(s) := S_n(s) - I^{kj}(s)$ and $Z_{kj} := \sup_{s\in\mathcal{S}} S_n^{kj}(s)$. Moreover, define τ_{kj} to be the minimal i such that $S_n^{kj}(s_i) = Z_{kj}$. Then, Z_{kj} is $\mathscr{F}_n^{kj}$-measurable, and we have

$$\exp(t(Z_{kj}+X_{kj}+\Delta)) \ge f \ge \exp(tZ_{kj})\cdot\exp(tI^{kj}(s_{\tau_{kj}})). \tag{2.3}$$

The random variable τ_{kj} is $\mathscr{F}_n^{kj}$-measurable which implies $\mathbb{E}_n^{kj}[I^{kj}(s_{\tau_{kj}})] = 0$. Thus, by Jensen's inequality, we obtain from the second estimate in (2.3) that

$$f_{kj} \ge \exp(tZ_{kj})\cdot\mathbb{E}_n^{kj}[\exp(tI^{kj}(s_{\tau_{kj}}))] \ge \exp(tZ_{kj}) \ge \exp(tS_n^{kj}(s_\tau)),$$

and consequently $f_{kj} \ge f\cdot\exp(-tI^{kj}(s_\tau))$ which implies statement a).

For the proof of b), we retain the notation introduced in the proof of statement a). From the left-hand side inequality in (2.3), we obtain

$$\begin{aligned} f_{kj} &\le e^{t(Z_{kj}+\Delta)}\cdot\mathbb{E}[e^{tX_{kj}}\mathbb{1}_{\Omega_{kj}}] + e^{t(Z_{kj}+\Delta)}\cdot\mathbb{E}[e^{tX_{kj}}\mathbb{1}_{\Omega_{kj}^{\mathsf{c}}}] \\ &\le e^{t(Z_{kj}+1+\Delta)} + e^{t(Z_{kj}+\Delta)}\cdot\mathbb{E}[e^{2tX_{kj}}]^{1/2}\mathbb{P}(\Omega_{kj}^{\mathsf{c}})^{1/2}. \end{aligned}$$

Multiplication with $\mathbb{1}_{\Omega_{kj}}$ on both sides, using the estimate $\mathbb{P}(\Omega_{kj}^{\mathsf{c}})^{1/2} \le \Delta/\sqrt{2}$ from Lemma 2.5, and recalling the formula for the moment-generating function of a Poisson distributed random variable yields

$$f_{kj}\mathbb{1}_{\Omega_{kj}} \le e^{t(Z_{kj}+1+\Delta)}\mathbb{1}_{\Omega_{kj}} + e^{t(Z_{kj}+\Delta)}\cdot\exp(\Delta(e^{2t}-1)/2)\cdot\Delta/\sqrt{2}\cdot\mathbb{1}_{\Omega_{kj}},$$

from which we conclude by exploiting the right-hand side inequality of (2.3) and the definition of Ω_{kj} that

$$f_{kj}\mathbb{1}_{\Omega_{kj}} \le fe^{2(1+\Delta)t}\mathbb{1}_{\Omega_{kj}} + fe^{(1+2\Delta)t}\cdot\exp(\Delta(e^{2t}-1)/2)\cdot\Delta/\sqrt{2}\cdot\mathbb{1}_{\Omega_{kj}},$$

and hence by elementary transformations

$$(1 - f/f_{kj}\cdot e^{(1+2\Delta)t}\exp(\Delta(e^{2t}-1)/2)\cdot\Delta/\sqrt{2})\cdot\mathbb{1}_{\Omega_{kj}} \le f/f_{kj}\cdot e^{2(1+\Delta)t}\mathbb{1}_{\Omega_{kj}}.$$

Now, by the statement of assertion a) and the definition of Ω_{kj}

$$(1 - e^{(2+3\Delta)t}\exp(\Delta(e^{2t}-1)/2)\cdot\Delta/\sqrt{2})\cdot\mathbb{1}_{\Omega_{kj}} \le f/f_{kj}\cdot e^{2(1+\Delta)t}\cdot\mathbb{1}_{\Omega_{kj}},$$

which yields the claim assertion after division by $e^{2(1+\Delta)t}$. □

In the sequel, we use the abbreviation $c(t,\ell) := 1 - e^{(2+3\Delta)t}\exp(\Delta(e^{2t}-1)/2)\cdot\Delta/\sqrt{2}$. Note that $c(t,\ell)\le 1$ and, for any fixed non-negative t, $c(t,\ell)\to 1$ as $\ell\to\infty$. In particular, $c(t,\ell)\in[1/2,1]$, for sufficiently large ℓ, say $\ell \ge \ell_0 = \ell_0(t)$. Under the validity of Assumption 2.6, we consider for $k\in\{1,\dots,n\}$ and $j\in\{1,\dots,\ell\}$ the positive and $\mathscr{F}_n^{kj}$-measurable random variables h_{kj} defined by

$$h_{kj} = h_{kj}(t) := \sum_{i=1}^{m}\mathbb{P}_n^{kj}(\tau=i)\exp(tS_n^{kj}(s_i)) = \mathbb{E}_n^{kj}[\exp(tS_n^{kj}(s_\tau))]. \tag{2.4}$$

From now on, we denote by C a numerical constant *independent of* ℓ (but surely depending on the fixed value of t considered) whose value may change depending on the context. The following lemma summarizes estimates which are used for the rest of this section. Since all the estimates are easy to obtain, we omit its proof.

LEMMA 2.8. *Let Assumption 2.6 hold,* $\eta(x) = 1-\exp(-x) - e^{2(1+\Delta)t-\log c(t,\ell)}x$ *for* $\ell\ge\ell_0$ *and* S_n^{kj} *be defined as in the proof of Lemma 2.7. Then, the estimate* $\mathbb{E}[X]\le C$ *holds true, where* X *can be* *replaced by any of the following random variables:*

a) h_{kj}^4,

b) $(f_{kj} - f)^4$,

c) $(f \log(f/f_{kj}))^4$,

d) $(f\eta(tI^{kj}(s_\tau)))^4$,

e) $(I^{kj}(s))^4$,

f) $(I^{kj}(s))^2$,

g) $\exp(tS_n^{kj}(s_\tau))$, *and*

h) $\exp(4tS_n^{kj}(s_\tau))$.

The constant C can be chosen to be independent of k and j, and in statements d)–h), it can in addition be chosen independent of s and s_τ, respectively.

Lemma 2.9. *Let Assumption 2.6 hold, and let h_{kj} be defined as in* (2.4)*. Then, for all $\ell \geq \ell_0$, we have*

$$\sum_{k=1}^{n}\sum_{j=1}^{\ell}\mathbb{E}[(f-h_{kj})\mathbb{1}_{\Omega_{kj}}] \leq e^{2(1+\Delta)t-\log c(t,\ell)}\mathbb{E}\,[f]\log\mathbb{E}[f] + C\cdot\ell^{-1/2}.$$

Proof. We begin the proof with the observation that

$$\mathbb{E}[(f-h_{kj})\mathbb{1}_{\Omega_{kj}}] = \mathbb{E}[f-h_{kj}] + \mathbb{E}[(h_{kj}-f)\mathbb{1}_{\Omega_{kj}^{\mathsf{c}}}] \leq \mathbb{E}[f-h_{kj}] + \mathbb{E}[h_{kj}\mathbb{1}_{\Omega_{kj}^{\mathsf{c}}}] \tag{2.5}$$

where the last estimate is due to the fact that f is non-negative. We have the decomposition

$$\begin{aligned}\mathbb{E}[f-h_{kj}] &= \mathbb{E}[f(1-\exp(-tI^{kj}(s_\tau)) - te^{2(1+\Delta)t-\log c(t,\ell)}I^{kj}(s_\tau)] + te^{2(1+\Delta)t-\log c(t,\ell)}\mathbb{E}[fI^{kj}(s_\tau)]\\ &= \mathbb{E}[f\eta(tI^{kj}(s_\tau))\mathbb{1}_{\Omega_{kj}}] + \mathbb{E}[f\eta(tI^{kj}(s_\tau))\mathbb{1}_{\Omega_{kj}^{\mathsf{c}}}] + te^{2(1+\Delta)t-\log c(t,\ell)}\mathbb{E}[fI^{kj}(s_\tau)],\end{aligned}$$

where the function η is defined via $\eta(x) = 1-\exp(-x) - e^{2(1+\Delta)t-\log c(t,\ell)}x$. Note that η is non-increasing on the interval $[-2(1+\Delta)t + \log c(t,\ell), \infty)$. This fact in combination with Lemma 2.7 implies that

$$\mathbb{E}[f\eta(tI^{kj}(s_\tau))\mathbb{1}_{\Omega_{kj}}] \leq \mathbb{E}[(f - f_{kj} - e^{2(1+\Delta)t-\log c(t,\ell)}f\log(f/f_{kj}))\mathbb{1}_{\Omega_{kj}}].$$

By the identities $\mathbb{1}_{\Omega_{kj}} = 1 - \mathbb{1}_{\Omega_{kj}^{\mathsf{c}}}$ and $\mathbb{E}[f - f_{kj}] = 0$, we thus obtain

$$\begin{aligned}\mathbb{E}[f\eta(tI^{kj}(s_\tau))\mathbb{1}_{\Omega_{kj}}] \leq\ & \mathbb{E}[(f_{kj}-f)\mathbb{1}_{\Omega_{kj}^{\mathsf{c}}}] + e^{2(1+\Delta)t-\log c(t,\ell)}\mathbb{E}[f\log(f/f_{kj})\mathbb{1}_{\Omega_{kj}^{\mathsf{c}}}]\\ & - e^{2(1+\Delta)t-\log c(t,\ell)}\mathbb{E}[f\log(f/f_{kj})].\end{aligned}$$

Using Hölder's inequality and Lemma 2.8, we obtain the estimate

$$\mathbb{E}[(f_{kj}-f)\mathbb{1}_{\Omega_{kj}^{\mathsf{c}}}] \leq \mathbb{E}[(f_{kj}-f)^4]^{1/4}\cdot\mathbb{P}(\Omega_{kj}^{\mathsf{c}})^{3/4} \leq C\cdot\ell^{-3/2},$$

and by the same argument $\mathbb{E}[f\log(f/f_{kj})\mathbb{1}_{\Omega_{kj}^{\mathsf{c}}}] \leq C\cdot\ell^{-3/2}$, $\mathbb{E}[f\eta(tI^{kj}(s_\tau))\mathbb{1}_{\Omega_{kj}^{\mathsf{c}}}] \leq C\cdot\ell^{-3/2}$, and $\mathbb{E}[h_{kj}\mathbb{1}_{\Omega_{kj}^{\mathsf{c}}}] \leq C\cdot\ell^{-3/2}$. Putting the obtained estimates into (2.5), we obtain

$$\mathbb{E}[(f-h_{kj})\mathbb{1}_{\Omega_{kj}}] \leq e^{2(1+\Delta)t-\log c(t,\ell)}(t\mathbb{E}[fI^{kj}(s_\tau)] - \mathbb{E}[f\log(f/f_{kj})]) + C\cdot\ell^{-3/2}.$$

Summation over k and j yields

$$\sum_{k=1}^{n}\sum_{j=1}^{\ell}\mathbb{E}[(f-h_{kj})\mathbb{1}_{\Omega_{kj}}]\leq e^{2(1+\Delta)t-\log c(t,\ell)}\left(t\mathbb{E}[fZ]-\sum_{k=1}^{n}\sum_{j=1}^{\ell}\mathbb{E}[f\log(f/f_{kj})]\right)+C\cdot\ell^{-1/2}.$$

By application of Proposition 4.1 from [Led96], we have

$$-\sum_{k=1}^{n}\sum_{j=1}^{\ell}\mathbb{E}[f\log(f/f_{kj})]\leq-\mathbb{E}[f\log f]+\mathbb{E}\left[f\right]\cdot\log\mathbb{E}[f],$$

and thus

$$\sum_{k=1}^{n}\sum_{j=1}^{\ell}\mathbb{E}[(f-h_{kj})\mathbb{1}_{\Omega_{kj}}]\leq e^{2(1+\Delta)t-\log c(t,\ell)}\mathbb{E}\left[f\right]\cdot\log\mathbb{E}[f]+C\cdot\ell^{-1/2}.$$

□

Lemma 2.10. *Consider the function r defined through $r(t,x):=x\log x+(1+t)(1-x)$. Then, for any $s\in\mathcal{S}$ and $t\geq 0$,*

$$\mathbb{E}[r((1+\Delta)t,\exp(tI^{kj}(s)))\,\mathbb{1}_{\Omega_{kj}}]\leq Ct^2\ell^{-3/2}+\frac{t^2}{2}\mathbb{E}[(I^{kj}(s))^2].$$

Proof. For fixed non-negative t consider the functions η,δ defined through $\eta(x)=r((1+\Delta)t,e^{tx})=e^{tx}tx+(1+(1+\Delta)t)(1-e^{tx})$ and $\delta(x)=\eta(x)-x\eta'(0)-\frac{(tx)^2}{2}$, respectively. We have $\delta(0)=0$ and $\delta'(x)=t^2(x-(1+\Delta))(e^{tx}-1)$. Thus, the sign of $\delta'(x)$ coincides with the one of $x(x-(1+\Delta))$. This implies that $\delta(x)\leq\delta(0)=0$ for all $x\leq 1+\Delta$, and hence $\eta(x)\leq x\eta'(0)+(tx)^2/2$. Since the estimate $I^{kj}(s)\leq 1+\Delta$ holds on Ω_{kj}, by the preceding arguments we obtain

$$r((1+\Delta)t,e^{tI^{kj}(s)})\mathbb{1}_{\Omega_{kj}}\leq(-(1+\Delta)t^2I^{kj}(s)+(tI^{kj}(s))^2/2)\mathbb{1}_{\Omega_{kj}}.$$

Taking expectations on both sides yields

$$\mathbb{E}[r((1+\Delta)t,\exp(tI^{kj}(s)))\mathbb{1}_{\Omega_{kj}}]\leq\mathbb{E}[(-(1+\Delta)t^2I^{kj}(s)+(tI^{kj}(s))^2/2)\mathbb{1}_{\Omega_{kj}}].$$

Therefrom, by means of the relation $\mathbb{1}_{\Omega_{kj}}\leq 1$, we obtain

$$\mathbb{E}[r((1+\Delta)t,\exp(tI^{kj}(s)))\mathbb{1}_{\Omega_{kj}}]\leq-(1+\Delta)t^2\mathbb{E}[I^{kj}(s)\mathbb{1}_{\Omega_{kj}}]+\frac{t^2}{2}\mathbb{E}[(I^{kj}(s))^2].$$

The identity $\mathbb{1}_{\Omega_{kj}}=1-\mathbb{1}_{\Omega_{kj}^c}$, Hölder's inequality and Lemma 2.8 imply that

$$\mathbb{E}[r((1+\Delta)t,\exp(tI^{kj}(s)))\mathbb{1}_{\Omega_{kj}}]\leq Ct^2\ell^{-3/2}+\frac{t^2}{2}\mathbb{E}[(I^{kj}(s))^2],$$

(recall that $\mathbb{E}[I^{kj}(s)]=0$ for all $s\in\mathcal{S}$) which finishes the proof of the lemma. □

Remark 2.11. There is a clear correspondence between some of the auxiliary results proved above and the auxiliary results in [KR05]. Lemmata 3.1, 3.2, and 3.3 in that paper correspond to our Lemmata 2.7, 2.9, and 2.10, respectively. Both, results and proofs turn out to be more intricate in the PPP setup considered here.

2.1.2. Proof of Theorem 2.1

First note that it is sufficient to prove statements a)–c) of Theorem 2.1 for the case of finite $\mathcal{S}$. Based on this, the case of countable $\mathcal{S}$ follows using the monotone convergence theorem. Thus, we assume from now on without loss of generality that $\mathcal{S} = \{s_1, \ldots, s_m\}$, and the preceding auxiliary results from Section 2.1.1 (which were mostly obtained under the validity of Assumption 2.6) are available. For fixed t and $\ell \geq \ell_0 = \ell_0(t)$ (here, $\ell_0(t)$ is defined as in the Section 2.1.1), let us represent the PPP N_k as in (2.2) as the superposition of ℓ i.i.d. PPPs N_{kj} with intensity measures Λ_k/ℓ, respectively. Then, application of Proposition 4.1 from [Led96] and the decomposition $1 = \mathbb{1}_{\Omega_{kj}} + \mathbb{1}_{\Omega_{kj}^{\mathsf{c}}}$ yield

$$\begin{aligned}\mathbb{E}\left[f \log f\right] - \mathbb{E}[f] \log \mathbb{E}[f] &\leq \sum_{k=1}^{n} \sum_{j=1}^{\ell} \mathbb{E}[f \log(f/f_{kj})] \\ &= \underbrace{\sum_{k=1}^{n} \sum_{j=1}^{\ell} \mathbb{E}[f \log(f/f_{kj})\, \mathbb{1}_{\Omega_{kj}}]}_{=:\,\square} + \underbrace{\sum_{k=1}^{n} \sum_{j=1}^{\ell} \mathbb{E}[f \log(f/f_{kj})\, \mathbb{1}_{\Omega_{kj}^{\mathsf{c}}}]}_{=:\,\blacksquare},\end{aligned} \tag{2.6}$$

and we investigate the two terms separately.
Examination of $\square$: For $k \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, \ell\}$, consider the strictly positive random variables g_{kj} defined through

$$g_{kj} = g_{kj}(t) := \sum_{i=1}^{m} \mathbb{P}_n^{kj}(\tau = i) \exp\left(t S_n(s_i)\right).$$

We have the elementary decomposition

$$\mathbb{E}[f \log(f/f_{kj}) \mathbb{1}_{\Omega_{kj}}] = \mathbb{E}[g_{kj} \log(f/f_{kj}) \mathbb{1}_{\Omega_{kj}}] + \mathbb{E}[(f - g_{kj}) \log(f/f_{kj}) \mathbb{1}_{\Omega_{kj}}]. \tag{2.7}$$

Note that $\mathbb{E}_n^{kj}[f/f_{kj}] = 1$, and thus

$$\mathbb{E}[g_{kj} \log(f/f_{kj}) \mathbb{1}_{\Omega_{kj}}] \leq \sup\{\mathbb{E}[g_{kj} h \mathbb{1}_{\Omega_{kj}}] : h \text{ is } \mathscr{F}_n\text{-measurable with } \mathbb{E}_n^{kj}[e^h] \leq 1\}.$$

Thus, due to the duality formula for the relative entropy (see p. 83 in [Led96] or Proposition 2.12 in [Mas07]), we obtain

$$\mathbb{E}[g_{kj} \log(f/f_{kj}) \mathbb{1}_{\Omega_{kj}}] \leq \mathbb{E}[g_{kj} \mathbb{1}_{\Omega_{kj}} \log(g_{kj} \mathbb{1}_{\Omega_{kj}})] - \mathbb{E}[g_{kj} \mathbb{1}_{\Omega_{kj}} \log \mathbb{E}_n^{kj}[g_{kj} \mathbb{1}_{\Omega_{kj}}]].$$

Putting this estimate into equation (2.7) yields

$$\begin{aligned}\mathbb{E}[f \log\left(f/f_{kj}\right) \mathbb{1}_{\Omega_{kj}}] \leq \mathbb{E}[g_{kj} \mathbb{1}_{\Omega_{kj}} \log(g_{kj} \mathbb{1}_{\Omega_{kj}})] - \mathbb{E}[g_{kj} \mathbb{1}_{\Omega_{kj}} \log \mathbb{E}_n^{kj}[g_{kj} \mathbb{1}_{\Omega_{kj}}]] \\ + \mathbb{E}[(f - g_{kj}) \log(f/f_{kj}) \mathbb{1}_{\Omega_{kj}}],\end{aligned}$$

and by summation over k and j we obtain

$$\begin{aligned}\square \leq &\sum_{k=1}^{n} \sum_{j=1}^{\ell} \mathbb{E}[g_{kj} \mathbb{1}_{\Omega_{kj}} \log(g_{kj} \mathbb{1}_{\Omega_{kj}})] - \sum_{k=1}^{n} \sum_{j=1}^{\ell} \mathbb{E}[g_{kj} \mathbb{1}_{\Omega_{kj}} \log \mathbb{E}_n^{kj}[g_{kj} \mathbb{1}_{\Omega_{kj}}]] \\ &+ \sum_{k=1}^{n} \sum_{j=1}^{\ell} \mathbb{E}[(f - g_{kj}) \log(f/f_{kj})\, \mathbb{1}_{\Omega_{kj}}].\end{aligned} \tag{2.8}$$

Lemma 2.7, combined with the facts that $f - g_{kj} \geq 0$ and $t I^{kj}(s_\tau) \mathbb{1}_{\Omega_{kj}} \leq (1 + \Delta) t \mathbb{1}_{\Omega_{kj}}$, implies

$$\mathbb{E}[(f - g_{kj}) \log(f/f_{kj}) \mathbb{1}_{\Omega_{kj}}] \leq (1 + \Delta) t \mathbb{E}[(f - g_{kj}) \mathbb{1}_{\Omega_{kj}}]. \tag{2.9}$$

For $k \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, \ell\}$, consider the positive and $\mathscr{F}_n^{kj}$-measurable random variables h_{kj} defined as in Equation (2.4). By the variational definition of relative entropy (see Equation (1.5) in [Led96] or Proposition 2.12 in [Mas07]), we obtain

$$\begin{aligned}\mathbb{E}_n^{kj}[g_{kj}\mathbb{1}_{\Omega_{kj}}\log(g_{kj}\mathbb{1}_{\Omega_{kj}})] - \mathbb{E}_n^{kj}[g_{kj}\mathbb{1}_{\Omega_{kj}}\log\mathbb{E}_n^{kj}[g_{kj}\mathbb{1}_{\Omega_{kj}}]] \\ \leq \mathbb{E}_n^{kj}[(g_{kj}\log(g_{kj}/h_{kj}) - g_{kj} + h_{kj})\mathbb{1}_{\Omega_{kj}}].\end{aligned}$$

By taking expectations on both sides of the last estimate, and combining the result with (2.9) we obtain from (2.8) that

$$\begin{aligned}\square &\leq \sum_{k=1}^{n}\sum_{j=1}^{\ell}\mathbb{E}[(g_{kj}\log(g_{kj}/h_{kj}) + (1+(1+\Delta)t)(h_{kj}-g_{kj}))\mathbb{1}_{\Omega_{kj}}] \\ &\quad + (1+\Delta)t\sum_{k=1}^{n}\sum_{j=1}^{\ell}\mathbb{E}[(f-h_{kj})\,\mathbb{1}_{\Omega_{kj}}] =: \square_1 + \square_2.\end{aligned}$$

In order to bound $\square_1$ from above, introduce the function r defined via

$$r(t,x) = x\log x + (1+t)(1-x).$$

By the definition of g_{kj} and h_{kj} we have

$$g_{kj}\log(g_{kj}/h_{kj}) + (1+(1+\Delta)t)(h_{kj}-g_{kj}) = h_{kj}r((1+\Delta)t, g_{kj}/h_{kj}),$$

and the convexity of r with respect to x yields

$$h_{kj}r((1+\Delta)t, g_{kj}/h_{kj}) \leq \sum_{i=1}^{m}\mathbb{P}_n^{kj}(\tau=i)\exp(tS_n^{kj}(s_i))r((1+\Delta)t, \exp(tI^{kj}(s_i))).$$

Hence, multiplication with $\mathbb{1}_{\Omega_{kj}}$ and application of the $\mathbb{E}_n^{kj}$ operator yield

$$\mathbb{E}_n^{kj}[h_{kj}r((1+\Delta)t, g_{kj}/h_{kj})\mathbb{1}_{\Omega_{kj}}] \leq \sum_{i=1}^{m}\mathbb{P}_n^{kj}(\tau=i)\exp(tS_n^{kj}(s_i))\mathbb{E}[r((1+\Delta)t, \exp(tI^{kj}(s_i)))\mathbb{1}_{\Omega_{kj}}].$$

The expectation on the right-hand side can be bounded by means of Lemma 2.10, and we obtain

$$\begin{aligned}&\mathbb{E}_n^{kj}[h_{kj}r((1+\Delta)t, g_{kj}/h_{kj})\mathbb{1}_{\Omega_{kj}}] \\ &\quad \leq Ct^2\ell^{-3/2}\mathbb{E}_n^{kj}[\exp(tS_n^{kj}(s_\tau))] + \frac{t^2}{2}\mathbb{E}_n^{kj}\left[\sum_{i=1}^{m}\mathbb{1}_{\{\tau=i\}}\exp(tS_n^{kj}(s_i))\mathbb{E}[(I^{kj}(s_i))^2]\right].\end{aligned} \tag{2.10}$$

In order to further bound the second term on the right-hand side of the last estimate, we consider the decomposition

$$\begin{aligned}\mathbb{E}_n^{kj}\left[\sum_{i=1}^{m}\mathbb{1}_{\{\tau=i\}}\exp(tS_n^{kj}(s_i))\mathbb{E}[(I^{kj}(s_i))^2]\right] &= \mathbb{E}_n^{kj}\left[\sum_{i=1}^{m}\mathbb{1}_{\{\tau=i\}}\exp(tS_n^{kj}(s_i))\mathbb{1}_{\Omega_{kj}}\mathbb{E}[(I^{kj}(s_i))^2]\right] \\ &\quad + \mathbb{E}_n^{kj}\left[\sum_{i=1}^{m}\mathbb{1}_{\{\tau=i\}}\exp(tS_n^{kj}(s_i))\mathbb{1}_{\Omega_{kj}^c}\mathbb{E}[(I^{kj}(s_i))^2]\right],\end{aligned} \tag{2.11}$$

and we bound the two terms on the right-hand side of (2.11) separately. In order to treat the first one, note that on Ω_{kj} we have $\exp(tS_n^{kj}(s_i)) \leq \exp(2t(1+\Delta) + tS_n(s_i))$, from which we conclude that

$$\mathbb{E}_n^{kj}\left[\sum_{i=1}^{m}\mathbb{1}_{\{\tau=i\}}\exp(tS_n^{kj}(s_i))\mathbb{1}_{\Omega_{kj}}\mathbb{E}[(I^{kj}(s_i))^2]\right]$$

$$\leq e^{2(1+\Delta)t}\mathbb{E}_n^{kj}\left[\sum_{i=1}^m \mathbb{1}_{\{\tau=i\}}\exp(tS_n(s_i))\mathbb{E}[(I^{kj}(s_i))^2]\right]. \quad (2.12)$$

For the second term on the right-hand side of (2.11), we have by Lemma 2.8 that

$$\mathbb{E}_n^{kj}\left[\sum_{i=1}^m \mathbb{1}_{\{\tau=i\}}\exp(tS_n^{kj}(s_i))\mathbb{1}_{\Omega_{kj}^{\mathsf{c}}}\mathbb{E}[(I^{kj}(s_i))^2]\right] \leq C\cdot\mathbb{E}_n^{kj}[\exp(tS_n^{kj}(s_\tau))\mathbb{1}_{\Omega_{kj}^{\mathsf{c}}}],$$

and thus by putting this last estimate and (2.12) into (2.11) we obtain

$$\begin{aligned}&\mathbb{E}_n^{kj}\left[\sum_{i=1}^m \mathbb{1}_{\{\tau=i\}}\exp(tS_n^{kj}(s_i))\mathbb{E}[(I^{kj}(s_i))^2]\right]\\ &\quad\leq e^{2(1+\Delta)t}\mathbb{E}_n^{kj}\left[\sum_{i=1}^m \mathbb{1}_{\{\tau=i\}}\exp(tS_n(s_i))\mathbb{E}[(I^{kj}(s_i))^2]\right] + C\cdot\mathbb{E}_n^{kj}[\exp(tS_n^{kj}(s_\tau))\mathbb{1}_{\Omega_{kj}^{\mathsf{c}}}].\end{aligned}$$

By taking expectations on both sides of (2.10) and summation over k and j, we obtain by means of the derived estimates in combination with Lemma 2.8 that

$$\begin{aligned}\square_1 &\leq Ct^2\ell^{-1/2} + \frac{t^2}{2}e^{2(1+\Delta)t}\mathbb{E}\left[\sum_{i=1}^m \mathbb{1}_{\{\tau=i\}}\exp(tS_n(s_i))\sum_{k=1}^n\sum_{j=1}^\ell \mathbb{E}[(I^{kj}(s_i))^2]\right]\\ &\quad+\frac{Ct^2}{2}\sum_{k=1}^n\sum_{j=1}^\ell \mathbb{E}[\exp(tS_n^{kj}(s_\tau))\mathbb{1}_{\Omega_{kj}^{\mathsf{c}}}].\end{aligned}$$

Since $\sum_{k=1}^n\sum_{j=1}^\ell \mathbb{E}[(I^{kj}(s_i))^2] \leq V_n$ and $\mathbb{E}[\exp(tS_n^{kj}(s_\tau))\mathbb{1}_{\Omega_{kj}^{\mathsf{c}}}] \leq C\cdot\ell^{-3/2}$ (the last estimate follows from Hölder's inequality and Lemma 2.8), we obtain

$$\square_1 \leq Ct^2\ell^{-1/2} + \frac{t^2}{2}e^{2(1+\Delta)t}V_n\mathbb{E}[f].$$

A suitable bound for $\square_2$ follows directly from Lemma 2.9. By combining the derived estimates for $\square_1$ and $\square_2$, we obtain

$$\square \leq Ct^2\ell^{-1/2} + \frac{t^2}{2}e^{2(1+\Delta)t}V_n\mathbb{E}[f] + (1+\Delta)te^{2(1+\Delta)t-\log c(t,l)}\mathbb{E}\left[f\right]\log\mathbb{E}[f] + C\cdot\ell^{-1/2}. \quad (2.13)$$

Examination of $\blacksquare$: By Hölder's inequality, Lemmata 2.5, 2.7, and 2.8, we have

$$\blacksquare \leq \sum_{k=1}^n\sum_{j=1}^\ell \mathbb{E}[(tfI^{kj}(s_\tau))^4]^{1/4}\mathbb{P}(\Omega_{kj}^{\mathsf{c}})^{3/4} \leq C\cdot\ell^{-1/2}. \quad (2.14)$$

We now merge the examinations of the terms $\square$ and $\blacksquare$. More precisely, by combining (2.6) with (2.13) and (2.14) and letting ℓ tend towards infinity we obtain that

$$tL'(t) - (te^{2t}+1)L(t) \leq \frac{V_n}{2}t^2e^{2t}.$$

As in [KR05], setting $\gamma(t) = t^{-2}\exp((1-e^{2t})/2)$ one can derive by means of *Herbst's argument* that

$$t\gamma(t)L(t) \leq \mathbb{E}Z + \frac{V_n}{2}(1-\exp((1-e^{2t})/2)),$$

which implies assertion a). In order to prove statement b), we apply the generalized Markov inequality for the function $x\mapsto\exp(tx)$ and apply assertion a) with $t=\frac{1}{2}\log(1+2\log(1+x/\upsilon))$.

For the proof of statement c), we need the following lemma, a proof of which can be found in [KR05].

Lemma 2.12 ([KR05], Lemma 3.4). *Under the assumptions of Theorem 2.1, we have for any* $t \in (0, \frac{2}{3})$ *that*

$$L(t) \leq t\mathbb{E}Z + (2\mathbb{E}Z + V_n) \cdot \frac{t^2}{2-3t}.$$

By Lemma 2.12 and the generalized Markov inequality, we obtain

$$\mathbb{P}(Z \geq \mathbb{E}Z + x) \leq \exp\left(\frac{\upsilon t^2}{2-3t} - tx\right).$$

The first inequality in assertion c) follows therefrom by the fact that the Legendre transform of $t \mapsto \frac{\upsilon t^2}{2-3t}$ is given by $x \mapsto \frac{4}{9}(\upsilon + \frac{3x}{2} - \sqrt{\upsilon^2 + 3x\upsilon})$, and the second inequality is due to elementary calculus. □

2.2. Intermezzo: A useful consequence of Theorem 2.1

In this section, we state and prove a consequence of Theorem 2.1 which turns out to be useful for the statistical applications in the second part of this thesis. As will become clear from the proof, it can be regarded as an integrated version of statement c) from Theorem 2.1.

Proposition 2.13. *Let* $N_1, \ldots, N_n$ *be independent PPPs on a Polish space* $\mathbb{X}$ *with finite intensity measures* $\Lambda_1, \ldots, \Lambda_n$*. Set* $\nu_n(r) = \frac{1}{n}\sum_{k=1}^n \left\{\int_{\mathbb{X}} r(x)dN_k(x) - \int_{\mathbb{X}} r(x)d\Lambda_k(x)\right\}$ *for* r *contained in a countable class* $\mathcal{R}$ *of complex-valued measurable functions.*
Then, there exist constants $c_1, c_2 = \frac{1}{6}$*, and* c_3 *such that for any* $\varepsilon > 0$

$$\mathbb{E}\left[\left(\sup_{r\in\mathcal{R}} |\nu_n(r)|^2 - c(\varepsilon)H^2\right)_+\right] \leq c_1 \left\{\frac{\upsilon}{n}\exp\left(-c_2\varepsilon\frac{nH^2}{\upsilon}\right) + \frac{M_1^2}{C^2(\varepsilon)n^2}\exp\left(-c_3 C(\varepsilon)\sqrt{\varepsilon}\frac{nH}{M_1}\right)\right\}$$

where $C(\varepsilon) = (\sqrt{1+\varepsilon} - 1) \wedge 1$*,* $c(\varepsilon) = 4(1+2\varepsilon)$ *and* M_1*,* H *and* υ *are such that*

$$\sup_{r\in\mathcal{R}} \|r\|_\infty \leq M_1, \quad \mathbb{E}\left[\sup_{r\in\mathcal{R}} |\nu_n(r)|\right] \leq H, \text{ and } \sup_{r\in\mathcal{R}} \operatorname{Var}\left(\int_{\mathbb{X}} r(x)dN_k(x)\right) \leq \upsilon\, \forall k.$$

Remark 2.14. Analogues of Proposition 2.13 in a setup with random variables instead of point processes have been used in the context of adaptive non-parametric estimation at various places, see, for instance, [CRT06], [Lac08] and [JS13a]. The proof given below follows along the lines of the proof given in [Cha13] to a great extent (with slight modifications concerning the numerical constants).

Remark 2.15. As a by-product of the proof of Proposition 2.13, we obtain that in case that the class $\mathcal{R}$ consists of *real-valued* functions only, one can replace the constant $c(\varepsilon) = 4(1+2\varepsilon)$ with $c(\varepsilon) = 2(1+2\varepsilon)$.

Proof of Proposition 2.13. For $r \in \mathcal{R}$ and $k \in \{1, \ldots, n\}$ define functions $s_r^k : \mathbb{X} \to \mathbb{C}$ via

$$s_r^k(x) := \frac{r(x)}{M_1}.$$

Hence, for all $r \in \mathcal{R}$ and and $k \in \{1, \ldots, n\}$, we have $|s_r^k(x)| \leq 1$ and we can apply statement c) of Theorem 2.1 for both $\mathcal{S} = \{(\Re s_r^1, \ldots, \Re s_r^n) : r \in \mathcal{R}\}$ and $\mathcal{S} = \{(\Im s_r^1, \ldots, \Im s_r^n) : r \in \mathcal{R}\}$ (the quantity Z corresponds to $\frac{n}{M_1}\sup_{r\in\mathcal{R}} \Re\nu_n(r)$ and $\frac{n}{M_1}\sup_{r\in\mathcal{R}} \Im\nu_n(r)$, respectively). In the sequel, we will only give estimates for the *real part* since the corresponding estimates for the *imaginery* *part* are identical. Application of Theorem 2.1 c) yields for any $x > 0$ that

$$\mathbb{P}\left(\frac{n}{M_1}\sup_{r\in\mathcal{R}}\Re\nu_n(r) \geq \frac{n}{M_1}\mathbb{E}\left[\sup_{r\in\mathcal{R}}\Re\nu_n(r)\right] + x\right) \leq \exp\left(-\frac{x^2}{2\upsilon_{\Re} + 3x}\right)$$

with $\upsilon_{\Re} = 2\mathbb{E}Z + V_n$ where $V_n = \sup_{r\in\mathcal{R}} \operatorname{Var}(S_n(\Re s_r))$, and S_n is defined as in the statement of Theorem 2.1. Specializing with $x = ny/M_1$ and using the fact that $\sup_{r\in\mathcal{R}}\Re\nu_n(r) \leq \sup_{r\in\mathcal{R}}|\nu_n(r)|$ yield that for any $y > 0$ we have

$$\begin{aligned}\mathbb{P}\left(\sup_{r\in\mathcal{R}}\Re\nu_n(r) \geq H + y\right) &\leq \mathbb{P}\left(\sup_{r\in\mathcal{R}}\Re\nu_n(r) \geq \mathbb{E}\left[\sup_{r\in\mathcal{R}}\Re\nu_n(r)\right] + y\right)\\ &\leq \exp\left(-\frac{n^2y^2}{2M_1^2\upsilon_{\Re} + 3M_1ny}\right).\end{aligned}$$

Note that on the one hand we have $\mathbb{E}Z \leq nH/M_1$, and on the other hand

$$\begin{aligned}V_n &= \sup_{r\in\mathcal{R}}\operatorname{Var}\left(\frac{1}{M_1}\sum_{k=1}^{n}\int_{\mathbb{X}}\Re r(x)(dN_k(x) - d\Lambda_k(x))\right)\\ &= \frac{1}{M_1^2}\sup_{r\in\mathcal{R}}\operatorname{Var}\left(\sum_{k=1}^{n}\int_{\mathbb{X}}\Re r(x)(dN_k(x) - d\Lambda_k(x))\right)\\ &\leq \frac{n\upsilon}{M_1^2}\end{aligned}$$

which in combination imply $\upsilon_{\Re} \leq 2nH/M_1 + n\upsilon/M_1^2$. We have

$$\mathbb{P}\left(\sup_{r\in\mathcal{R}}\Re\nu_n(r) \geq H + y\right) \leq \exp\left(-\frac{ny^2}{2(2M_1H + \upsilon) + 3M_1y}\right)$$

which is used to obtain

$$\begin{aligned}\mathbb{P}\left(\sup_{r\in\mathcal{R}}|\Re\nu_n(r)| \geq H + y\right) &\leq \mathbb{P}\left(\sup_{r\in\mathcal{R}}\Re\nu_n(r) \geq H + y\right) + \mathbb{P}\left(\sup_{r\in\mathcal{R}}-\Re\nu_n(r) \geq H + y\right)\\ &= \mathbb{P}\left(\sup_{r\in\mathcal{R}}\Re\nu_n(r) \geq H + y\right) + \mathbb{P}\left(\sup_{r\in\mathcal{R}}\Re\nu_n(-r) \geq H + y\right)\\ &\leq 2\cdot\exp\left(-\frac{ny^2}{2(2M_1H + \upsilon) + 3M_1y}\right).\end{aligned}$$

Below, we will apply this estimate for $y = \mu + \eta H$ for μ, η to be specified. This choice of y yields

$$\begin{aligned}\frac{y^2}{2(\upsilon + 2M_1H) + 3M_1y} &= \frac{\mu^2 + \eta^2H^2 + 2\eta\mu H}{2\upsilon + 4HM_1 + 3M_1\mu + 3M_1\eta H}\\ &\geq \frac{\mu^2 + 2\eta\mu H}{2\upsilon + 3\mu M_1 + M_1H(4 + 3\eta)}\\ &=: \frac{a + b}{c + d + e}.\end{aligned} \tag{2.15}$$

For arbitrary $a, b, c, d, e > 0$ we have the estimate

$$\frac{a+b}{c+d+e} \geq \frac{a+b}{3(c\vee d\vee e)} = \frac{1}{3}\left(\frac{a+b}{c}\wedge\frac{a+b}{d}\wedge\frac{a+b}{e}\right) \geq \frac{1}{3}\left(\frac{a}{c}\wedge\frac{a}{d}\wedge\frac{b}{e}\right).$$

For a, b, c, d, e as defined in (2.15), this estimate implies

$$\frac{y^2}{2(\upsilon + 2M_1 H) + 3M_1 y} \geq \frac{1}{3}\left[\frac{\mu^2}{2\upsilon} \wedge \frac{2\mu}{M_1}\left(\frac{1}{6} \wedge \frac{\eta}{4+3\eta}\right)\right].$$

For any $\eta \geq 0$, we obtain

$$\frac{1}{6} \wedge \frac{\eta}{4+3\eta} \geq \frac{\eta \wedge 1}{7}$$

due the trivial estimate $1/6 \geq (\eta \wedge 1)/7$ combined with

$$\frac{\eta}{4+3\eta} - \frac{\eta\wedge 1}{7} = \begin{cases} \frac{7\eta - 4 - 3\eta}{7(4+3\eta)} = \frac{4(\eta-1)}{7(4+3\eta)} \geq 0, & \text{if } \eta \geq 1, \\ \frac{3\eta - 3\eta^2}{7(4+3\eta)} \geq 0, & \text{if } 1 \geq \eta \geq 0. \end{cases}$$

Thus, we have

$$\frac{y^2}{2(\upsilon + 2M_1 H) + 3M_1 y} \geq \frac{1}{3}\left[\frac{\mu^2}{2\upsilon} \wedge \frac{2(\eta\wedge 1)}{7}\frac{\mu}{M_1}\right]$$

which in turn implies

$$\mathbb{P}\left(\sup_{r\in\mathcal{R}} |\Re \nu_n(r)| \geq \mu + (\eta+1)H\right) \leq 2\exp\left(-\frac{n}{3}\left[\frac{\mu^2}{2\upsilon} \wedge \frac{2(\eta\wedge 1)}{7}\frac{\mu}{M_1}\right]\right). \tag{2.16}$$

After these preliminaries, we start the proof of the claim assertion by means of the estimate

$$\begin{aligned}
\mathbb{E}\left[\left(\sup_{r\in\mathcal{R}} |\nu_n(r)|^2 - 4(1+2\varepsilon)H^2\right)_+\right] &= \int_0^\infty \mathbb{P}\left(\sup_{r\in\mathcal{R}} |\nu_n(r)|^2 \geq 4(1+2\varepsilon)H^2 + t\right) dt \\
&= \int_0^\infty \mathbb{P}\left(\sup_{r\in\mathcal{R}} |\nu_n(r)| \geq \sqrt{4(1+\varepsilon)H^2 + 4(\varepsilon H^2 + t/4)}\right) dt \\
&\leq \int_0^\infty \mathbb{P}\left(\sup_{r\in\mathcal{R}} |\nu_n(r)| \geq \sqrt{2(1+\varepsilon)}H + \sqrt{2(\varepsilon H^2 + t/4)}\right) dt
\end{aligned}$$

where the last line is due to the estimate $\sqrt{a} + \sqrt{b} \leq \sqrt{2a+2b}$. From this we conclude

$$\begin{aligned}
\mathbb{E}\left[\left(\sup_{r\in\mathcal{R}} |\nu_n(r)|^2 - 4(1+2\varepsilon)H^2\right)_+\right] &\leq \int_0^\infty \mathbb{P}\left(\sup_{r\in\mathcal{R}} |\Re\nu_n(r)| \geq \sqrt{1+\varepsilon}H + \sqrt{\varepsilon H^2 + t/4}\right) dt \\
&\quad + \int_0^\infty \mathbb{P}\left(\sup_{r\in\mathcal{R}} |\Im\nu_n(r)| \geq \sqrt{1+\varepsilon}H + \sqrt{\varepsilon H^2 + t/4}\right) dt.
\end{aligned}$$

We apply (2.16) with $\eta = \sqrt{1+\varepsilon} - 1$ and $\mu = \sqrt{\varepsilon H^2 + t/4}$ to both terms and obtain

$$\begin{aligned}
&\mathbb{E}\left[\left(\sup_{r\in\mathcal{R}} |\nu_n(r)|^2 - 4(1+2\varepsilon)H^2\right)_+\right] \\
&\qquad \leq 4\int_0^\infty \exp\left(-\frac{n}{3}\left\{\frac{\varepsilon H^2 + t/4}{2\upsilon} \wedge \frac{2(\eta\wedge 1)}{7}\frac{\sqrt{\varepsilon H^2 + t/4}}{M_1}\right\}\right) dt \\
&\qquad \leq 4\int_0^\infty \exp\left(-\frac{n}{3}\frac{\varepsilon H^2 + t/4}{2\upsilon}\right) dt + 4\int_0^\infty \exp\left(-\frac{n}{3}\frac{2(\eta\wedge 1)}{7}\frac{\sqrt{\varepsilon H^2 + t/4}}{M_1}\right) dt.
\end{aligned}$$

Using the estimate $\sqrt{a} + \sqrt{b} \leq \sqrt{2a+2b}$ once again implies

$$\mathbb{E}\left[\left(\sup_{r\in\mathcal{R}} |\nu_n(r)|^2 - 4(1+2\varepsilon)H^2\right)_+\right] \leq 4\exp\left(-\frac{n\varepsilon H^2}{6\upsilon}\right)\int_0^\infty \exp\left(-\frac{nt}{24\upsilon}\right) dt$$

$$
\begin{aligned}
&\qquad + 4\int_0^\infty \exp\left(-\frac{2n(\eta\wedge 1)}{21\sqrt{2}M_1}(\sqrt{\varepsilon}H + \sqrt{t/4})\right) dt\\
&\le 4\exp\left(-\frac{n\varepsilon H^2}{6\upsilon}\right)\int_0^\infty \exp\left(-\frac{nt}{24\upsilon}\right) dt\\
&\qquad + 4\exp\left(-\frac{2n(\eta\wedge 1)}{21\sqrt{2}M_1}\sqrt{\varepsilon}H\right)\int_0^\infty \exp\left(-\frac{n(\eta\wedge 1)}{21\sqrt{2}M_1}\sqrt{t}\right) dt\\
&= 4\left\{\exp\left(-\frac{n\varepsilon H^2}{6\upsilon}\right)\frac{24\upsilon}{n} + \exp\left(-\frac{2n(\eta\wedge 1)}{21\sqrt{2}M_1}\sqrt{\varepsilon}H\right)\cdot\frac{42^2M_1^2}{n^2(\eta\wedge 1)^2}\right\}\\
&= 48\left\{\exp\left(-\frac{n\varepsilon H^2}{6\upsilon}\right)\frac{2\upsilon}{n} + \exp\left(-\frac{2n(\eta\wedge 1)}{21\sqrt{2}M_1}\sqrt{\varepsilon}H\right)\cdot\frac{147M_1^2}{n^2(\eta\wedge 1)^2}\right\}.
\end{aligned}
$$

□

2.3. Concentration inequalities for left-hand side deviations

The following theorem is the second main result of this chapter and complements Theorem 2.1 by providing concentration inequalities for left-hand side deviations of Z from its mean.

THEOREM 2.16. *Under the assumptions of Theorem 2.1, for any non-negative t,*

a) $$L_Z(-t) \le -t\mathbb{E}Z + \frac{\upsilon}{9}(e^{3t} - 3t - 1).$$

Consequently, for any non-negative x, we have

b) $$\mathbb{P}(Z \le \mathbb{E}Z - x) \le \exp\left(-\frac{\upsilon}{9}h\left(\frac{3x}{\upsilon}\right)\right),$$

where $h(x) = (1+x)\log(1+x) - x$, and

c) $$\mathbb{P}(Z \le \mathbb{E}Z - x) \le \exp\left(-\frac{x^2}{\upsilon + \sqrt{\upsilon^2 + 2\upsilon x} + x}\right) \le \exp\left(-\frac{x^2}{2\upsilon + 2x}\right).$$

REMARK 2.17. As in the case of right-hand side deviations, the concentration inequalities in Theorem 2.16 translate literally the results in the random variable framework due to [KR05].

2.3.1. Notation and preliminary results

We maintain a large part of the notation introduced in Section 2.1.1 for the proof of Theorem 2.1. In particular, we use again the representation $N_k \overset{d}{=} \sum_{j=1}^{\ell} N_{kj}$ of the PPPs N_k as the superposition of independent PPPs N_{kj} with intensity Λ_k/ℓ and use the shorthand notations $\Lambda := \sup_{k=1,\dots,n}\Lambda_k(\mathbb{X})$ and $\Delta = \Delta(\ell) := \Lambda/\ell$. Besides, we retain the definition $\Omega_{kj} := \{X_{kj} \le 1\}$ where $X_{kj} := N_{kj}(\mathbb{X})$ and the definition of the I_{kj}. Let us further assume that Assumption 2.6 holds, that is, $\mathcal{S} = \{s_1, \dots, s_m\}$ is finite. Define now

$$L_i(t) := \sum_{k=1}^{n}\sum_{j=1}^{\ell} \log\mathbb{E}[\exp(-tI^{kj}(s_i))], \qquad i \in \{1,\dots,m\}.$$

The corresponding *exponentially compensated* empirical process is $T_i(t) := S_n(s_i) + t^{-1}L_i(t)$. In addition to Z, let us define $Z_t := \sup_{i\in\{1,\dots,m\}} T_i(t)$. For notational convenience, we use from now on the shorthand notation $\sup_i/\inf_i$ when the supremum/infimum over $i \in \{1,\dots,m\}$ is taken. Redefine $f = f(t) := \exp(-tZ_t)$ and $f_{kj} = f_{kj}(t) := \mathbb{E}_n^{kj}[f]$. Here, the σ-fields $\mathscr{F}_n^{kj}$ are defined as in Section 2.1. Finally, we define $F = F(t) := \mathbb{E}[f]$ and $\mathcal{L} = \mathcal{L}(t) := \log F(t)$. The main strategy of the proof given in Section 2.3.2 is to derive a differential inequality for $\mathcal{L}$. Let $\tau = \tau(t)$ denote the minimal value of $i \in \{1,\dots,m\}$ such that $Z_t = T_i(t)$. As in the proof of concentration results for right-hand side deviations from Z from its mean, let C always denote a constant (whose value is independent of ℓ but might depend on t) which is allowed to attain different values in different contexts.

LEMMA 2.18. *Let Assumption 2.6 hold and* $\psi_\ell(t) = \frac{1}{2}(1 + e^{2(1+\Delta)t})$. *Set*

$$\ell_{kji} = \ell_{kji}(t) = \log \mathbb{E}[\exp(-tI^{kj}(s_i))].$$

Then, the following estimates hold almost surely:

a) $f_{kj}/f \leq \exp(tI^{kj}(s_\tau) + \ell_{kj\tau})$, *and*

b) $\exp(tI^{kj}(s_\tau) + \ell_{kj\tau}) \leq \psi_\ell(t) \cdot (1 + \alpha_\ell) + \beta(t) \cdot \ell^{-3/2}$ *on* Ω_{kj} *where* α_ℓ *is a monotone sequence decreasing to* 0 *as* ℓ *tends to* ∞ *and* $\beta = \beta(t) > 0$ *is monotone increasing in* t.

PROOF. For $s \in \mathcal{S}$, define $S_n^{kj}(s) := S_n(s) - I^{kj}(s)$ and

$$Z^{kj} := \sup_{s\in\mathcal{S}}(S_n^{kj}(s) + t^{-1} \log \mathbb{E}[\exp(-tS_n^{kj}(s))]).$$

Let $\tau_{kj} = \tau_{kj}(t)$ be the smallest $i \in \{1, \ldots, m\}$ such that

$$Z^{kj} = S_n^{kj}(s_i) + t^{-1} \log \mathbb{E}[\exp(-tS_n^{kj}(s_i))].$$

Then, $f \leq \exp(-tZ^{kj}) \exp(-tI^{kj}(s_{\tau_{kj}}) - \ell_{kj\tau_{kj}}(t))$, which implies $\mathbb{E}_n^{kj}[f] \leq \exp(-tZ^{kj})$. By definition of Z^{kj}, we have $\exp(-tZ^{kj}) \leq f \cdot \exp(tI^{kj}(s_\tau) + \ell_{kj\tau}(t))$, which shows statement a). In order to prove statement b), first note that $\exp(tI^{kj}(s_\tau)) \leq e^{(1+\Delta)t}$ on Ω_{kj}, and it remains to find an estimate for $\exp(\ell_{kj\tau}(t)) = \mathbb{E}[\exp(-tI^{kj}(s_\tau))]$. Consider the decomposition

$$\mathbb{E}[\exp(-tI^{kj}(s_\tau))] = \mathbb{E}[\exp(-tI^{kj}(s_\tau))\mathbb{1}_{\Omega_{kj}}] + \mathbb{E}[\exp(-tI^{kj}(s_\tau))\mathbb{1}_{\Omega_{kj}^c}]. \tag{2.17}$$

In order to bound the first term on the right-hand side of (2.17), note that $\mathbb{E}[\exp(-tI^{kj}(s_\tau))\mathbb{1}_{\Omega_{kj}}] \leq \mathbb{E}[e^{tY}]$ with $Y = -I^{kj}(s_\tau)\mathbb{1}_{\Omega_{kj}}$. By the convexity of the exponential function, we have

$$\mathbb{E}[e^{tY}] \leq \frac{1 + \Delta - \mathbb{E}Y}{2(1 + \Delta)} e^{-(1+\Delta)t} + \frac{\mathbb{E}Y + 1 + \Delta}{2(1 + \Delta)} e^{(1+\Delta)t} = \frac{1}{2}(e^{-(1+\Delta)t} + e^{(1+\Delta)t})(1 + o(1)). \tag{2.18}$$

The second term one the right-hand side of (2.17) is bounded using Hölder's inequality, Lemmata 2.5 and 2.21 as follows:

$$\mathbb{E}[\exp(-tI^{kj}(s_\tau))\mathbb{1}_{\Omega_{kj}^c}] \leq \mathbb{E}[\exp(-4tI^{kj}(s_\tau))]^{1/4} \cdot \mathbb{P}(\Omega_{kj}^c)^{3/4} \leq C \cdot \ell^{-3/2}, \tag{2.19}$$

and statement b) follows now from the combination of (2.18) and (2.19). □

LEMMA 2.19. *For* $k \in \{1, \ldots, n\}$ *and* $j \in \{1, \ldots, \ell\}$, *define positive random variables* $g_{kj} = g_{kj}(t)$ *via*

$$g_{kj} = \sum_{i=1}^{m} \mathbb{P}_n^{kj}(\tau = i) \exp(-tS_n(s_i) - L_i(t)).$$

Set $\varphi_\ell = \varphi_\ell(t) := \widetilde{\psi}_\ell \cdot \log \widetilde{\psi}_\ell$ *where* $\widetilde{\psi}_\ell = \widetilde{\psi}_\ell(t) := \psi_\ell(t) \cdot (1 + \alpha_\ell) + \beta(t) \cdot \ell^{-3/2}$ *with* ψ_ℓ, α_ℓ, *and* β *defined as in Lemma 2.18. For sufficiently large* ℓ, *let* θ_ℓ *be the unique positive solution of the equation* $\varphi_\ell(t) = 1$. *Then, for any* $t \in (0, \theta_\ell)$,

$$\sum_{k=1}^{n} \sum_{j=1}^{\ell} \mathbb{E}[(g_{kj} - f) \log(f_{kj}/f)]$$

$$\leq \frac{\varphi_\ell}{1-\varphi_\ell} \cdot \left(\sum_{k=1}^{n} \sum_{j=1}^{\ell} \mathbb{E}(g_{kj} \log(g_{kj}/\mathbb{E}_n^{kj}[g_{kj}])) - \mathbb{E}[f \log f] \right) + C \cdot \ell^{-1/2}.$$

Proof. Since S_n^{kj} is $\mathscr{F}_n^{kj}$-measurable, it is easy to verify that

$$\mathbb{E}_n^{kj}[g_{kj}] = \mathbb{E}_n^{kj}[f \cdot \exp(tI^{kj}(s_\tau) + \ell_{kj\tau})],$$

and hence,

$$\sum_{k=1}^{n} \sum_{j=1}^{\ell} \mathbb{E}[g_{kj} - f] = \sum_{k=1}^{n} \sum_{j=1}^{\ell} \mathbb{E}[f \cdot (\exp(tI^{kj}(s_\tau) + \ell_{kj\tau}) - 1)].$$

Set $\eta_{kj} = tI^{kj}(s_\tau) + \ell_{kj\tau}$. Then,

$$\begin{aligned} \sum_{k=1}^{n} \sum_{j=1}^{\ell} \mathbb{E}[g_{kj} - f] &= \sum_{k=1}^{n} \sum_{j=1}^{\ell} \mathbb{E}[f \cdot (e^{\eta_{kj}} - 1 - \widetilde{\psi}_\ell \eta_{kj})] + \widetilde{\psi}_\ell \mathbb{E}[f \sum_{k=1}^{n} \sum_{j=1}^{\ell} \eta_{kj}] \\ &= \sum_{k=1}^{n} \sum_{j=1}^{\ell} \mathbb{E}[f \cdot (e^{\eta_{kj}} - 1 - \widetilde{\psi}_\ell \eta_{kj})] - \widetilde{\psi}_\ell \mathbb{E}[f \log f], \end{aligned} \tag{2.20}$$

since $\sum_{k=1}^{n} \sum_{j=1}^{\ell} \eta_{kj} = -\log f$. Consider the first term on the right-hand side of (2.20). First, by Hölder's inequality and Lemma 2.21

$$\sum_{k=1}^{n} \sum_{j=1}^{\ell} \mathbb{E}[f \cdot (e^{\eta_{kj}} - 1 - \widetilde{\psi}_\ell \eta_{kj}) \mathbb{1}_{\Omega_{kj}^c}] \leq C \cdot \ell^{-1/2}.$$

In order to bound $\sum_{k=1}^{n} \sum_{j=1}^{\ell} \mathbb{E}[f \cdot (e^{\eta_{kj}} - 1 - \widetilde{\psi}_\ell \eta_{kj}) \mathbb{1}_{\Omega_{kj}}]$ from above, note that the function $x \mapsto e^x - 1 - x\widetilde{\psi}_\ell$ is non-increasing on the interval $(-\infty, \log \widetilde{\psi}_\ell]$. Hence, we obtain by Lemma 2.21 that

$$\begin{aligned} \sum_{k=1}^{n} \sum_{j=1}^{\ell} \mathbb{E}[f \cdot (e^{\eta_{kj}} - 1 - \widetilde{\psi}_\ell \eta_{kj}) \mathbb{1}_{\Omega_{kj}}] &\leq \widetilde{\psi}_\ell \sum_{k=1}^{n} \sum_{j=1}^{\ell} \mathbb{E}[f \log(f/f_{kj})] \\ &\quad - \sum_{k=1}^{n} \sum_{j=1}^{\ell} \mathbb{E}[(f_{kj} - f - \widetilde{\psi}_\ell f \log(f_{kj}/f)) \mathbb{1}_{\Omega_{kj}^c}] \\ &\leq \widetilde{\psi}_\ell \sum_{k=1}^{n} \sum_{j=1}^{\ell} \mathbb{E}[f \log(f/f_{kj})] + C \cdot \ell^{-1/2}. \end{aligned}$$

Putting the obtained estimates into (2.20) yields

$$\sum_{k=1}^{n} \sum_{j=1}^{\ell} \mathbb{E}[g_{kj} - f] \leq \widetilde{\psi}_\ell \cdot \left(\sum_{k=1}^{n} \sum_{j=1}^{\ell} \mathbb{E}[f \log(f/f_{kj})] - \mathbb{E}[f \log f] \right) + C \cdot \ell^{-1/2}.$$

Using the same argument as in the proof of Theorem 2.1 yields

$$\begin{aligned} &\sum_{k=1}^{n} \sum_{j=1}^{\ell} \mathbb{E}[g_{kj} - f] \\ &\leq \widetilde{\psi}_\ell \cdot \left(\sum_{k=1}^{n} \sum_{j=1}^{\ell} \mathbb{E}[g_{kj} \log(g_{kj}/\mathbb{E}_n^{kj}[g_{kj}]) + (g_{kj} - f) \log(f_{kj}/f)] - \mathbb{E}[f \log f] \right) + C \cdot \ell^{-1/2}. \end{aligned} \tag{2.21}$$

Now, in order to prove the claim assertion of the lemma, take note of the decomposition

$$\sum_{k=1}^{n}\sum_{j=1}^{\ell}\mathbb{E}[(g_{kj}-f)\log(f_{kj}/f)] = \sum_{k=1}^{n}\sum_{j=1}^{\ell}\mathbb{E}[(g_{kj}-f)\log(f_{kj}/f)\mathbb{1}_{\Omega_{kj}}] + \sum_{k=1}^{n}\sum_{j=1}^{\ell}\mathbb{E}[(g_{kj}-f)\log(f_{kj}/f)\mathbb{1}_{\Omega_{kj}^{\mathsf{c}}}]. \tag{2.22}$$

Using statement b) of Lemma 2.18, the estimate (2.21) and the definition of φ_ℓ, we can bound the first term on the right-hand side of (2.22) as follows (note that $g_{kj}-f\geq 0$):

$$\begin{aligned}\sum_{k=1}^{n}\sum_{j=1}^{\ell}\mathbb{E}[(g_{kj}-f)\log(f_{kj}/f)\mathbb{1}_{\Omega_{kj}}] &\leq \log\widetilde{\psi}_\ell\sum_{k=1}^{n}\sum_{j=1}^{\ell}\mathbb{E}[g_{kj}-f]\\ &\leq \varphi_\ell\cdot\left(\sum_{k=1}^{n}\sum_{j=1}^{\ell}[g_{kj}\log(g_{kj}/\mathbb{E}_n^{kj}[g_{kj}])+(g_{kj}-f)\log(f_{kj}/f)]-\mathbb{E}[f\log f]\right)+C\cdot\ell^{-1/2}.\end{aligned}$$

The second summand on the right-hand side of (2.22) can be bounded using Hölder's inequality, Lemma 2.5 and Lemma 2.21:

$$\sum_{k=1}^{n}\sum_{j=1}^{\ell}\mathbb{E}[(g_{kj}-f)\log(f_{kj}/f)\mathbb{1}_{\Omega_{kj}^{\mathsf{c}}}] \leq \sum_{k=1}^{n}\sum_{j=1}^{\ell}\mathbb{E}[((g_{kj}-f)\log(f_{kj}/f))^4]^{1/4}\mathbb{P}(\Omega_{kj}^{\mathsf{c}})^{3/4}\leq C\cdot\ell^{-1/2}.$$

Combining the bounds obtained for the two terms in (2.22) implies the assertion of the lemma. □

Remark 2.20. Both $\widetilde{\psi}_\ell(t)$ and $\varphi_\ell(t)$ are non-increasing in ℓ and non-decreasing in t. Hence, the solution θ_ℓ of the equation $\varphi_\ell = 1$ (which exists for sufficiently large ℓ) is non-decreasing in ℓ and the limit $\theta_\infty := \lim_{\ell\to\infty}\theta_\ell$ satisfies $\theta_\infty\in[0.46, 0.47]$ (see p. 1075 in [KR05]). The approximate value of θ_∞ is of interest for the proof of Theorem 2.16 which is done by considering different cases for the value of t (see [KR05] for details).

The simple proof of the following lemma is omitted.

Lemma 2.21. *Let Assumption 2.6 hold. Then, the estimate $\mathbb{E}[X]\leq C$ holds true, where X can be replaced by any of the following random variables:*

a) $\exp(-4tI^{kj}(s_\tau))$,

b) $(f_{kj}-f-\widetilde{\psi}_\ell f\log(f_{kj}/f))^4$,

c) $(f\cdot(e^{\eta_{kj}}-1-\widetilde{\psi}_\ell\eta_{kj}))^4$,

d) $((g_{kj}-f)\log(f_{kj}/f))^4$,

e) $(I_{kj}(s_i))^4e^{-4tI_{kj}(s_i)}$, *and*

f) $(g_{kj}\log(g_{kj}/\mathbb{E}_n^{kj}[g_{kj}]))^4$.

Here g_{kj}, η_{kj} and $\widetilde{\psi}_\ell$ are defined in Lemma 2.19 and its proof, respectively. The constant C can be chosen independent of k and j, and in statements a) and e), it can in addition be chosen independent of s_τ and s_i, respectively.

Lemma 2.22. *Let Y be a random variable with values in $(-\infty, 1+\Delta]$ and $\mathbb{E}[Y^2]<+\infty$. Then, for any positive t,*

$$\mathbb{E}[tYe^{tY}]-\mathbb{E}[e^{tY}]\log\mathbb{E}[e^{tY}]\leq\frac{\mathbb{E}[Y^2]}{(1+\Delta)^2}(1+((1+\Delta)t-1)e^{(1+\Delta)t}).$$

PROOF. The proof follows completely along the lines of the one of Lemma 4.4 in [KR05], and we thus omit it. □

REMARK 2.23. Again, there is a clear correspondence between some of the auxiliary results derived here and the ones used in [KR05]. Lemmata 2.18 and 2.19 are versions of Lemmata 4.2 and 4.3 in [KR05] tailored to our framework. As already mentioned above, Lemma 2.22 is nearly the same as Lemma 4.4 in [KR05].

2.3.2. Proof of Theorem 2.16

The key arguments of the proof follow along the proof of Theorem 1.2 in [KR05]. Since the random functions $T_i(t)$ are analytic in t, the random function $f = f(t)$ is continuous and piecewise analytic as a function in t. Its (almost everywhere existing) derivative with respect to t satisfies

$$f'(t) = -(Z_t + tZ_t')f(t)$$

where $tZ_t' = L_\tau'(t) - t^{-1}L_\tau(t)$. Thus, by the Fubini's theorem, we have

$$F(t) = 1 - \int_0^t \mathbb{E}[(Z_u + uZ_u')f(u)]du.$$

Hence, F is absolutely continuous with respect to the Lebesgue measure, with a.e. derivative in the sense of Lebesgue given by $F'(t) = -\mathbb{E}[(Z_t + tZ_t')f(t)]$. Moreover, the function $\Lambda = \log F$ has the a.e. derivative F'/F. As in the proof of Theorem 2.1, application of Proposition 4.1 from [Led96] yields

$$\mathbb{E}[f\log f] - \mathbb{E}[f]\log\mathbb{E}[f] \le \sum_{k=1}^n\sum_{j=1}^\ell \mathbb{E}[g_{kj}\log(g_{kj}/\mathbb{E}_n^{kj}[g_{kj}])] + \sum_{k=1}^n\sum_{j=1}^\ell \mathbb{E}[(g_{kj} - f)\log(f/f_{kj})] \quad (2.23)$$

for any positive integrable random variables g_{kj} such that $\mathbb{E}[g_{kj}\log g_{kj}] < \infty$. On the other hand,

$$\mathbb{E}[f(t)\log f(t)] - \mathbb{E}[f(t)]\log\mathbb{E}[f(t)] = t^2\mathbb{E}[Z_t'f(t)] + tF'(t) - F(t)\log F(t) \quad \text{a.e.} \quad (2.24)$$

Combining (2.23) and (2.24) yields

$$\begin{aligned} tF'(t) - F(t)\log F(t) \le -t^2\mathbb{E}[Z_t'f(t)] &+ \sum_{k=1}^n\sum_{j=1}^\ell \mathbb{E}[g_{kj}\log(g_{kj}/\mathbb{E}_n^{kj}[g_{kj}])] \\ &+ \sum_{k=1}^n\sum_{j=1}^\ell \mathbb{E}[(g_{kj} - f)\log(f_{kj}/f)]. \end{aligned}$$

We now specialize this estimate with the choice $g_{kj} = \sum_{i=1}^m \mathbb{P}_n^{kj}(\tau = i)\exp(-tS_n(s_i) - L_i(t))$, which coincides with the definition of g_{kj} in Lemma 2.19. Applying Lemma 2.19 and algebraic transformations yields

$$\begin{aligned} (1 - \varphi_\ell(t))(tF'(t) - F(t)\log F(t)) &\le \varphi_\ell(t)\cdot t^2\mathbb{E}[Z_t'f(t) - f(t)\log f(t)] \\ &\quad - \mathbb{E}[t^2Z_t'f(t)] + \sum_{k=1}^n\sum_{j=1}^\ell \mathbb{E}[g_{kj}\log(g_{kj}/\mathbb{E}_n^{kj}[g_{kj}])] + C\cdot\ell^{-1/2}, \end{aligned}$$

where φ_ℓ is defined in Lemma 2.19. Using the identity $\mathbb{E}[t^2 Z_t' f(t) - f(t)\log f(t)] = -tF'(t)$, we obtain

$$tF'(t)-(1-\varphi_\ell(t))F(t)\log F(t) \le -t^2\mathbb{E}[Z_t' f(t)]+\sum_{k=1}^{n}\sum_{j=1}^{\ell}\mathbb{E}[g_{kj}\log(g_{kj}/\mathbb{E}_n^{kj}[g_{kj}])]+C\cdot\ell^{-1/2}. \quad (2.25)$$

Now define $\omega_{kj} := g_{kj}/\mathbb{E}_n^{kj}[g_{kj}]$. Then, $\mathbb{E}_n^{kj}[g_{kj}\log(g_{kj}/\mathbb{E}_n^{kj}[g_{kj}])] = \mathbb{E}_n^{kj}[g_{kj}]\cdot\mathbb{E}_n^{kj}[\omega_{kj}\log\omega_{kj}]$. Using the convexity of $x\mapsto x\log x$, we conclude that

$$\mathbb{E}_n^{kj}[g_{kj}]\omega_{kj}\log\omega_{kj} \le \sum_{i=1}^{m}\mathbb{P}_n^{kj}(\tau=i)(-tI^{kj}(s_i)-\ell_{kji}(t))\exp(-tS_n(s_i)-L_i(t)),$$

and by applying the $\mathbb{E}_n^{kj}$ operator on both sides we obtain

$$\begin{aligned}\mathbb{E}_n^{kj}[g_{kj}\log(g_{kj}/\mathbb{E}_n^{kj}[g_{kj}])] &\le \sum_{i=1}^{m}\mathbb{P}_n^{kj}(\tau=i)\exp(-tS_n^{kj}(s_i)-L_i(t)+\ell_{kji}(t))(t\ell_{kji}'(t)-\ell_{kji}(t))\\ &= \mathbb{E}_n^{kj}\left[\sum_{i=1}^{m}\mathbb{1}_{\{\tau=i\}}\exp(-tS_n^{kj}-L_i(t)+\ell_{kji}(t))(t\ell_{kji}'(t)-\ell_{kji}(t))\right].\end{aligned}$$

Thus, by taking expectations,

$$\mathbb{E}[g_{kj}\log(g_{kj}/\mathbb{E}_n^{kj}[g_{kj}])] \le \mathbb{E}[f(t)\exp(tI^{kj}(s_\tau)+\ell_{kj\tau}(t))(t\ell_{kj\tau}'(t)-\ell_{kj\tau}(t))].$$

By Hölder's inequality and Lemma 2.21, we have $\mathbb{E}[g_{kj}\log(g_{kj}/\mathbb{E}_n^{kj}[g_{kj}])\mathbb{1}_{\Omega_{kj}^{\mathsf{c}}}] \le C\cdot\ell^{-3/2}$. In order to bound $\mathbb{E}[g_{kj}\log(g_{kj}/\mathbb{E}_n^{kj}[g_{kj}])\mathbb{1}_{\Omega_{kj}}]$, first note that the convexity of the functions ℓ_{kji} together with the fact that $\ell_{kji}(0)=0$ implies $t\ell_{kj\tau}'(t)-\ell_{kj\tau}(t)\ge 0$. Thus, we can use Lemma 2.18 in order to obtain

$$\mathbb{E}[g_{kj}\log(g_{kj}/\mathbb{E}_n^{kj}[g_{kj}])\mathbb{1}_{\Omega_{kj}}] \le \widetilde{\psi}_\ell(t)\cdot\mathbb{E}[(t\ell_{kj\tau}'(t)-\ell_{kj\tau}(t))f(t)].$$

By the identity $t^2 Z_t' = tL_\tau'(t)-L_\tau(t)$, we get for the first two summands on the right-hand side of (2.25) the estimate

$$-\mathbb{E}[t^2 Z_t' f(t)]+\sum_{k=1}^{n}\sum_{j=1}^{\ell}\mathbb{E}[g_{kj}\log(g_{kj}/\mathbb{E}_n^{kj}[g_{kj}])] \le (\widetilde{\psi}_\ell(t)-1)\mathbb{E}[(tL_\tau'(t)-L_\tau(t)f(t)]+C\cdot\ell^{-1/2}. \quad (2.26)$$

In order to bound the expectation on the right-hand side of the last estimate, let us first note that $tL_\tau'(t)-L_\tau(t)\le\sup_i(tL_i'(t)-L_i(t))$. In order to bound $\sup_i(tL_i'(t)-L_i(t))$, introduce (for fixed $i\in\{1,\dots,m\}$) the event $\widetilde{\Omega}_{kj}$ defined via

$$\widetilde{\Omega}_{kj} := \{I^{kj}(s_i)\ge-(1+\Delta)\}.$$

Thanks to the boundedness of the functions $s\in\mathcal{S}$, we have $\Omega_{kj}\subseteq\widetilde{\Omega}_{kj}$, hence $\widetilde{\Omega}_{kj}^{\mathsf{c}}\subseteq\Omega_{kj}^{\mathsf{c}}$. Setting $Y_{kj} := -I^{kj}(s_i)$, we obtain

$$\begin{aligned}t\ell_{kji}'(t)-\ell_{kji}(t) &\le t\mathbb{E}[\exp(tY_{kj})Y_{kj}]-\mathbb{E}[e^{tY_{kj}}]\log\mathbb{E}[e^{tY_{kj}}]\\ &\le t\mathbb{E}[\exp(tY_{kj})Y_{kj}\mathbb{1}_{\widetilde{\Omega}_{kj}^{\mathsf{c}}}]+t\mathbb{E}[\exp(tY_{kj})Y_{kj}\mathbb{1}_{\widetilde{\Omega}_{kj}}]-\mathbb{E}[e^{tY_{kj}\mathbb{1}_{\widetilde{\Omega}_{kj}}}]\log\mathbb{E}[e^{tY_{kj}\mathbb{1}_{\widetilde{\Omega}_{kj}}}]. \quad (2.27)\end{aligned}$$

The first term on the right-hand side of (2.27) is bounded using Lemma 2.21:

$$t\mathbb{E}[\exp(tY_{kj})Y_{kj}\mathbb{1}_{\widetilde{\Omega}_{kj}^{\mathsf{c}}}] \le C\cdot\mathbb{P}(\widetilde{\Omega}_{kj}^{\mathsf{c}})^{3/4} \le C\cdot\mathbb{P}(\Omega_{kj}^{\mathsf{c}})^{3/4} \le C\cdot\ell^{-3/2}.$$

The second and third term on the right-hand side of (2.27) are bounded using Lemma 2.22 which yields

$$\begin{aligned}\mathbb{E}[t\exp(tY_{kj})Y_{kj}\mathbb{1}_{\widetilde{\Omega}_{kj}}] - \mathbb{E}[e^{tY_{kj}\mathbb{1}_{\widetilde{\Omega}_{kj}}}]\log\mathbb{E}[e^{tY_{kj}\mathbb{1}_{\widetilde{\Omega}_{kj}}}] \\ = \mathbb{E}[t\exp(tY_{kj}\mathbb{1}_{\widetilde{\Omega}_{kj}})Y_{kj}\mathbb{1}_{\widetilde{\Omega}_{kj}}] - \mathbb{E}[e^{tY_{kj}\mathbb{1}_{\widetilde{\Omega}_{kj}}}]\log\mathbb{E}[e^{tY_{kj}\mathbb{1}_{\widetilde{\Omega}_{kj}}}] \\ \leq \frac{\mathbb{E}[Y_{kj}^2]}{(1+\Delta)^2}\cdot(1+((1+\Delta)t-1)e^{(1+\Delta)t}).\end{aligned}$$

Hence summing over all k and j in (2.27) yields

$$tL_i'(t) - L_i(t) \leq C\cdot\ell^{-1/2} + \frac{V_n}{(1+\Delta)^2}\cdot(1+((1+\Delta)t-1)e^{(1+\Delta)t}),$$

and this estimate holds for all $i\in\{1,\dots,m\}$. Combining the obtained estimates with (2.25) and (2.26) and letting ℓ tend to ∞, we obtain

$$tF'(t) - (1-\varphi(t))F(t)\log F(t) \leq (\psi(t)-1)F(t)V_n(1+(t-1)e^t),$$

where $\psi(t) = \frac{1}{2}(1+e^{2t})$ and $\varphi = \psi\log\psi$. Division by $F(t)$ yields

$$t\mathcal{L}'(t) - (1-\varphi(t))\mathcal{L}(t) \leq \frac{V_n}{2}(e^{2t}-1)(1+(t-1)e^t).$$

This differential inequality for $\mathcal{L}$ coincides with equation (4.21) in [KR05] and the rest of the proof follows along the lines of the one given in that paper (Lemma 4.1 in [KR05] which is used for the proof translates without changes in the proof to our framework, whereas the purely analytical Lemmata 4.5 and 4.6 in [KR05] can be borrowed unchanged). □

We conclude this chapter with the remark that in most situations of interest, it is possible to apply the concentration inequalities proved in this chapter also in setups with *non-countable* classes of measurable functions. This practice can be made rigorous by means of standard density arguments (see [Cha13] for details).

Part II.

Applications to non-parametric estimation problems

3. Non-parametric intensity estimation

In this chapter, we consider the non-parametric estimation of the intensity of a PPP on the interval $[0,1]$ from n independent observations of the process. More precisely, we assume that the intensity measure is absolutely continuous with respect to the Lebesgue measure with Radon-Nikodym derivative λ that we aim to estimate from the i.i.d. sample

$$N_1, \dots, N_n. \tag{3.1}$$

We assume that $\lambda \in \mathbb{L}^2 := \mathbb{L}^2([0,1], dx)$, the space of square-integrable real-valued functions on $[0,1]$.

3.1. Methodology: Orthonormal series estimator of the intensity

Orthonormal series estimators represent a natural approach in non-parametric statistics. In this chapter, we consider an orthonormal series estimator for the intensity λ with respect to the standard trigonometric basis $\{\varphi_j\}_{j\in\mathbb{Z}}$ where

$$\varphi_0 \equiv 1, \quad \varphi_j(x) = \sqrt{2}\cos(2\pi jx), \quad \varphi_{-j}(x) = \sqrt{2}\sin(2\pi jx), \qquad j = 1, 2, \dots$$

Setting $[\lambda]_j = \int_0^1 \varphi_j(x)\lambda(x)dx$ we have the representation

$$\lambda = \sum_{j\in\mathbb{Z}} [\lambda]_j \varphi_j \tag{3.2}$$

as a $\mathbb{L}^2$-converging series. By Campbell's theorem (see Theorem 1.14), the estimator

$$\widehat{[\lambda]}_j := \frac{1}{n}\sum_{i=1}^n \int_0^1 \varphi_j(x) dN_i(x).$$

is unbiased for all $j \in \mathbb{Z}$. Replacing the unknown Fourier coefficients in (3.2) by these estimators and truncating the series representation, we obtain the estimator

$$\widehat{\lambda}_k = \sum_{0\le|j|\le k} \widehat{[\lambda]}_j \varphi_j$$

where $k \in \mathbb{N}_0$ is a dimension parameter that has to be chosen appropriately.

3.2. Minimax theory

We will evaluate the performance of an arbitrary estimator $\widetilde{\lambda}$ of λ by means of the mean integrated squared error $\mathbb{E}[\|\widetilde{\lambda} - \lambda\|^2]$ where $\|\cdot\|$ denotes the usual $\mathbb{L}^2$-norm and expectation is taken under the true functional parameter λ (of course, the expectation operator $\mathbb{E}$ is the one associated with the distribution of the sample $N_1, \dots, N_n$ in (3.1)). Taking on the minimax point of view, we consider the *maximum risk*

$$\sup_{\lambda\in\Lambda} \mathbb{E}[\|\widetilde{\lambda} - \lambda\|^2]$$

where Λ is a class of potential intensity functions with $\Lambda \subseteq \mathbb{L}^2$. In the minimax framework, the class Λ is assumed to be known and our objective is to define a *rate optimal* estimator of λ, that is, an estimator that attains the minimax rate

$$\inf_{\widetilde{\lambda}} \sup_{\lambda \in \Lambda} \mathbb{E}[\|\widetilde{\lambda} - \lambda\|^2]$$

at least up to a multiplicative numerical constant. Here, the infimum is taken over all estimators $\widetilde{\lambda}$ that are based on the observations (3.1). In this chapter, we assume that the unknown intensity function λ belongs to the set Λ_γ^r defined via

$$\Lambda_\gamma^r := \{\lambda \in \mathbb{L}^2 : \lambda \geq 0 \quad \text{and} \quad \|\lambda\|_\gamma^2 := \sum_{j \in \mathbb{Z}} \gamma_j |[\lambda]_j|^2 \leq r\}$$

for some strictly positive symmetric sequence $\gamma = (\gamma_j)_{j \in \mathbb{Z}}$. We need the following mild assumption concerning the sequence γ.

ASSUMPTION 3.1. *$\gamma = (\gamma_j)_{j \in \mathbb{Z}}$ is a strictly positive symmetric sequence with $\gamma_0 = 1$ and the sequence $(\gamma_n)_{n \in \mathbb{N}_0}$ is non-decreasing.*

In particular, Assumption 3.1 is satisfied by the following standard choices of the sequence γ:

- $\gamma_0 = 1$, $\gamma_j = |j|^{2p}$ for $j \neq 0$ and some $p > 0$. This setting corresponds to λ belonging to some *Sobolev ellipsoid*.
- $\gamma_j = \exp(2\beta|j|)$ for all $j \in \mathbb{Z}$ and some $\beta > 0$. This setting corresponds to λ belonging to some space of *analytic functions*.
- $\gamma_j = \exp(2\beta|j|^p)$ for all $j \in \mathbb{Z}$ and some $\beta, p > 0$. This setting corresponds to λ belonging to some space of *generalized analytic functions*.

We will illustrate our abstract results by means of these three examples throughout the chapter.

3.2.1. Upper bound

The following proposition provides an upper bound for the maximum risk of the estimator $\widehat{\lambda}_k$ over the class Λ_γ^r under a suitable choice of the dimension parameter k.

PROPOSITION 3.2. *Let Assumption 3.1 hold. Consider the estimator $\widehat{\lambda}_{k_n^*}$ with dimension parameter defined as $k_n^* := \operatorname{argmin}_{k \in \mathbb{N}_0} \max\{\frac{1}{\gamma_k}, \frac{2k+1}{n}\}$. Then, for any $n \in \mathbb{N}$,*

$$\sup_{\lambda \in \Lambda_\gamma^r} \mathbb{E}[\|\widehat{\lambda}_{k_n^*} - \lambda\|^2] \lesssim \max\left\{\frac{1}{\gamma_{k_n^*}}, \frac{2k_n^* + 1}{n}\right\} =: \Psi_n.$$

PROOF. Introduce the function $\lambda_{k_n^*} := \sum_{0 \leq |j| \leq k_n^*} [\lambda]_j \varphi_j$ which is used to obtain the decomposition

$$\mathbb{E}[\|\widehat{\lambda}_{k_n^*} - \lambda\|^2] = \|\lambda - \lambda_{k_n^*}\|^2 + \mathbb{E}[\|\widehat{\lambda}_{k_n^*} - \lambda_{k_n^*}\|^2]$$

of the risk into squared bias and variance. Using the fact that $\lambda \in \Lambda_\gamma^r$ together with Assumption 3.1, it is easy to see that $\|\lambda - \lambda_{k_n^*}\|^2 \leq r\gamma_{k_n^*}^{-1}$ and $\mathbb{E}[\|\widehat{\lambda}_{k_n^*} - \lambda_{k_n^*}\|^2] \leq \sqrt{r} \cdot \frac{2k_n^*+1}{n}$ and the statement of the proposition follows. □

3.2.2. Lower bound

Under the validity of Assumption 3.1 and mild additional assumptions, the following theorem provides a minimax lower bound for the estimation of the intensity from the observations (3.1) under the mean integrated squared error.

THEOREM 3.3. *Let Assumption 3.1 hold, and further assume that*

(C1) $\Gamma := \sum_{j\in\mathbb{Z}} \gamma_j^{-1} < \infty$, *and*

(C2) $0 < \eta^{-1} := \inf_{n\in\mathbb{N}} \Psi_n^{-1} \min\{\frac{1}{\gamma_{k_n^*}}, \frac{2k_n^*+1}{n}\}$ *for some* $1 \leq \eta < \infty$

where the quantities k_n^* *and* Ψ_n *are defined in Proposition 3.2. Then, for any* $n \in \mathbb{N}$,

$$\inf_{\widetilde{\lambda}} \sup_{\lambda\in\Lambda_\gamma^r} \mathbb{E}[\|\widetilde{\lambda} - \lambda\|^2] \gtrsim \Psi_n.$$

PROOF. Define $\zeta = \min\{\frac{1}{\Gamma\eta}, \frac{16\delta}{\sqrt{r}}\}$ with $\delta = \frac{1}{2} - \frac{1}{2\sqrt{2}}$, and for each $\theta = (\theta_j)_{0\leq|j|\leq k_n^*} \in \{\pm 1\}^{2k_n^*+1}$ the function λ_θ by

$$\lambda_\theta := \left(\frac{r}{4}\right)^{1/2} + \theta_0 \left(\frac{r\zeta}{16n}\right)^{1/2} + \left(\frac{r\zeta}{16n}\right)^{1/2} \sum_{1\leq|j|\leq k_n^*} \theta_j\varphi_j = \left(\frac{r}{4}\right)^{1/2} + \left(\frac{r\zeta}{16n}\right)^{1/2} \sum_{0\leq|j|\leq k_n^*} \theta_j\varphi_j.$$

Then, the calculation

$$\begin{aligned}
\left\| \left(\frac{r\zeta}{16n}\right)^{1/2} \sum_{0\leq|j|\leq k_n^*} \theta_j\varphi_j \right\|_\infty &\leq \left(\frac{r\zeta}{16n}\right)^{1/2} \sum_{0\leq|j|\leq k_n^*} \sqrt{2} \\
&\leq \left(\frac{r\zeta}{8}\right)^{1/2} \left(\sum_{0\leq|j|\leq k_n^*} \gamma_j^{-1}\right)^{1/2} \left(\sum_{0\leq|j|\leq k_n^*} \frac{\gamma_j}{n}\right)^{1/2} \\
&\leq \left(\frac{r\zeta\Gamma}{8}\right)^{1/2} \left(\gamma_{k_n^*} \cdot \frac{2k_n^*+1}{n}\right)^{1/2} \\
&\leq \left(\frac{r\zeta\eta\Gamma}{8}\right)^{1/2} \leq \left(\frac{r}{8}\right)^{1/2}
\end{aligned}$$

shows that $\lambda_\theta \geq \sqrt{r}\cdot\delta$. In particular, λ_θ is non-negative for all $\theta \in \{\pm 1\}^{2k_n^*+1}$. Moreover $\|\lambda_\theta\|_\gamma^2 \leq r$ holds for each $\theta \in \{\pm 1\}^{2k_n^*+1}$ due to the estimate

$$\begin{aligned}
\|\lambda_\theta\|_\gamma^2 = \sum_{0\leq|j|\leq k_n^*} |[\lambda_\theta]_j|^2\, \gamma_j &= \left[\left(\frac{r}{4}\right)^{1/2} + \theta_0\left(\frac{r\zeta}{16n}\right)^{1/2}\right]^2 + \frac{r\zeta}{16} \sum_{1\leq|j|\leq k_n^*} \frac{\gamma_j}{n} \\
&\leq \frac{r}{2} + \left(\frac{r\zeta}{8n}\right) + \frac{r\zeta}{16} \cdot \gamma_{k_n^*} \sum_{1\leq|j|\leq k_n^*} \frac{1}{n} \\
&\leq \frac{r}{2} + \frac{r\zeta}{8} \cdot \gamma_{k_n^*} \cdot \frac{2k_n^*+1}{n} \leq r.
\end{aligned}$$

This estimate and the non-negativity of λ_θ together imply $\lambda_\theta \in \Lambda_\gamma^r$ for all $\theta \in \{\pm 1\}^{2k_n^*+1}$. Let $\mathbb{P}_\theta$ denote the joint distribution of the i.i.d. sample $N_1, \ldots, N_n$ when the true parameter is λ_θ. Let $\mathbb{P}_\theta^{N_i}$ denote the corresponding one-dimensional marginal distributions and $\mathbb{E}_\theta$ the expectation with respect to $\mathbb{P}_\theta$. From now on, let $\widetilde{\lambda}$ be an arbitrary estimator of λ. The key argument of the proof is the reduction scheme

$$\begin{aligned}
\sup_{\lambda\in\Lambda_\gamma^r} \mathbb{E}[\|\widetilde{\lambda} - \lambda\|^2] \geq \sup_{\theta\in\{\pm1\}^{2k_n^*+1}} \mathbb{E}_\theta[\|\widetilde{\lambda} - \lambda_\theta\|^2] &\geq \frac{1}{2^{2k_n^*+1}} \sum_{\theta\in\{\pm1\}^{2k_n^*+1}} \mathbb{E}_\theta[\|\widetilde{\lambda} - \lambda_\theta\|^2] \\
&= \frac{1}{2^{2k_n^*+1}} \sum_{\theta\in\{\pm1\}^{2k_n^*+1}} \sum_{0\leq|j|\leq k_n^*} \mathbb{E}_\theta[|[\widetilde{\lambda} - \lambda_\theta]_j|^2]
\end{aligned}$$

$$= \frac{1}{2^{2k_n^*+1}} \sum_{\theta \in \{\pm 1\}^{2k_n^*+1}} \sum_{0 \le |j| \le k_n^*} \frac{1}{2} \{\mathbb{E}_\theta[|[\widetilde{\lambda} - \lambda_\theta]_j|^2] + \mathbb{E}_{\theta^{(j)}}[|[\widetilde{\lambda} - \lambda_{\theta^{(j)}}]_j|^2]\}, \quad (3.3)$$

where for $\theta \in \{\pm 1\}^{2k_n^*+1}$ the element $\theta^{(j)} \in \{\pm 1\}^{2k_n^*+1}$ is defined by $\theta_k^{(j)} = \theta_k$ for $k \neq j$ and $\theta_j^{(j)} = -\theta_j$. Consider the Hellinger affinity $\rho(\mathbb{P}_\theta, \mathbb{P}_{\theta^{(j)}}) := \int \sqrt{d\mathbb{P}_\theta d\mathbb{P}_{\theta^{(j)}}}$. For an arbitrary estimator $\widetilde{\lambda}$ of λ we have

$$\begin{aligned}\rho(\mathbb{P}_\theta, \mathbb{P}_{\theta^{(j)}}) &\le \int \frac{|[\widetilde{\lambda} - \lambda_\theta]_j|}{|[\lambda_\theta - \lambda_{\theta^{(j)}}]_j|} \sqrt{d\mathbb{P}_\theta d\mathbb{P}_{\theta^{(j)}}} + \int \frac{|[\widetilde{\lambda} - \lambda_{\theta^{(j)}}]_j|}{|[\lambda_\theta - \lambda_{\theta^{(j)}}]_j|} \sqrt{d\mathbb{P}_\theta d\mathbb{P}_{\theta^{(j)}}} \\ &\le \left(\int \frac{|[\widetilde{\lambda} - \lambda_\theta]_j|^2}{|[\lambda_\theta - \lambda_{\theta^{(j)}}]_j|^2} d\mathbb{P}_\theta \right)^{1/2} + \left(\int \frac{|[\widetilde{\lambda} - \lambda_{\theta^{(j)}}]_j|^2}{|[\lambda_\theta - \lambda_{\theta^{(j)}}]_j|^2} d\mathbb{P}_{\theta^{(j)}} \right)^{1/2},\end{aligned}$$

from which we conclude by means of the elementary inequality $(a+b)^2 \le 2a^2 + 2b^2$ that

$$\frac{1}{2} |[\lambda_\theta - \lambda_{\theta^{(j)}}]_j|^2 \rho^2(\mathbb{P}_\theta, \mathbb{P}_{\theta^{(j)}}) \le \mathbb{E}_\theta[|[\widetilde{\lambda} - \lambda_\theta]_j|^2] + \mathbb{E}_{\theta^{(j)}}[|[\widetilde{\lambda} - \lambda_{\theta^{(j)}}]_j|^2].$$

Recall that the Hellinger distance between two probability measures $\mathbb{P}$ and $\mathbb{Q}$ is defined as $H(\mathbb{P}, \mathbb{Q}) := (\int [\sqrt{d\mathbb{P}} - \sqrt{d\mathbb{Q}}]^2)^{1/2}$. By means of Theorem A.8 *(ii)* we obtain

$$H^2(\mathbb{P}_\theta^{N_i}, \mathbb{P}_{\theta^{(j)}}^{N_i}) \le \int (\sqrt{\lambda_\theta} - \sqrt{\lambda_{\theta^{(j)}}})^2 = \int \frac{|\lambda_\theta - \lambda_{\theta^{(j)}}|^2}{(\sqrt{\lambda_\theta} + \sqrt{\lambda_{\theta^{(j)}}})^2} \le \frac{1}{4\delta\sqrt{r}} \; \|\lambda_\theta - \lambda_{\theta^{(j)}}\|_2^2 = \frac{\zeta\sqrt{r}}{16\delta n} \le \frac{1}{n}.$$

Consequently, with Lemma A.3 it follows

$$H^2(\mathbb{P}_\theta, \mathbb{P}_{\theta^{(j)}}) \le \sum_{i=1}^n H^2(\mathbb{P}_\theta^{N_i}, \mathbb{P}_{\theta^{(j)}}^{N_i}) \le 1.$$

Thus, the relation $\rho(\mathbb{P}_\theta, \mathbb{P}_{\theta^{(j)}}) = 1 - \frac{1}{2} H^2(\mathbb{P}_\theta, \mathbb{P}_{\theta^{(j)}})$ implies $\rho(\mathbb{P}_\theta, \mathbb{P}_{\theta^{(j)}}) \ge \frac{1}{2}$. Finally, putting the obtained estimates into the reduction scheme (3.3) leads to

$$\begin{aligned}\sup_{\lambda \in \Lambda_\gamma^r} \mathbb{E}[\|\widetilde{\lambda} - \lambda\|^2] &\ge \frac{1}{2^{2k_n^*+1}} \sum_{\theta \in \{\pm 1\}^{2k_n^*+1}} \sum_{0 \le |j| \le k_n^*} \frac{1}{2} \left\{ \mathbb{E}_\theta[|[\tilde{\lambda} - \lambda_\theta]_j|^2] + \mathbb{E}_{\theta^{(j)}}[|[\tilde{\lambda} - \lambda_{\theta^{(j)}}]_j|^2] \right\} \\ &\ge \sum_{0 \le |j| \le k_n^*} \frac{1}{16} |[\lambda_\theta - \lambda_{\theta^{(j)}}]_j|^2 = \frac{\zeta r}{64} \sum_{0 \le |j| \le k_n^*} \frac{1}{n} \ge \frac{\zeta r}{64\eta} \cdot \Psi_n,\end{aligned}$$

which finishes the proof of the theorem since $\widetilde{\lambda}$ was arbitrary. □

As a direct consequence of the lower bound and Proposition 3.2, we obtain that the estimator $\widehat{\lambda}_{k_n^*}$ is rate optimal under the assumptions stated in Proposition 3.2 and Theorem 3.3.

Remark 3.4. The proof of Theorem 3.3 is inspired by the proof of Theorem 2.1 in [JS13a] expanded with the essential ingredient that the Hellinger distance between two PPPs is bounded by the Hellinger distance of the corresponding intensity measures (see Theorem A.8). As in [JS13a], the mild assumption (C1) on the convergence of the series $\sum_{j\in\mathbb{Z}} \gamma_j^{-1}$ is needed only in order to guarantee the non-negativity of the candidate intensities considered in the proof.

Remark 3.5. The lower bound proof given above supplements the lower bound result in [RB03]. Note that the result in [RB03] cannot be applied for ellipsoids defined in terms of the trigonometric basis that we consider here.

3.2.3. Examples of convergence rates

Example 3.6 (Sobolev ellipsoids). *Let $\gamma_0 = 1$, $\gamma_j = |j|^{2p}$ for $j \neq 0$. Then, Assumption 3.1 is satisfied and elementary computations show that $k_n^* \asymp n^{\frac{1}{2p+1}}$ as well as $\Psi_n \asymp n^{-\frac{2p}{2p+1}}$. Furthermore, the additional conditions of Theorem 3.3 are satisfied if $p > \frac{1}{2}$.*

Example 3.7 (Analytic functions). *Let $\gamma_j = \exp(2\beta|j|)$ for $j \in \mathbb{Z}$ for some $\beta > 0$. Assumption 3.1 is also fulfilled in this case and we obtain $k_n^* \asymp \log n$ and $\Psi_n \asymp \frac{\log n}{n}$. The additional assumptions of Theorem 3.3 do not impose any additional restriction on p.*

Example 3.8 (Generalized analytic functions). *Let $\gamma_j = \exp(2\beta|j|^p)$ for $\beta, p > 0$. Assumption 3.1 is satisfied in this case and there are no additional restrictions on p due to Theorem 3.3. We have $k_n^* \asymp (\log n)^{\frac{1}{p}}$ resulting in the rate $\Psi_n \asymp \frac{(\log n)^{\frac{1}{p}}}{n}$.*

3.3. Adaptive estimation

The definition of k_n^* in Proposition 3.2 depends on the sequence γ and hence on smoothness characteristics of the functional parameter to be estimated. Thus, the estimator $\widehat{\lambda}_{k_n^*}$ is not adaptive. In the following, we propose a selection rule for the dimension parameter $k \in \mathbb{N}_0$ that is fully data-driven and does not depend on any structural pre-assumptions on λ. In order to realize this plan, we follow the model selection paradigm sketched already in the introduction and define the contrast function

$$\Upsilon_n(t) := \|t\|^2 - 2\langle \widehat{\lambda}_n, t\rangle, \quad t \in \mathbb{L}^2$$

where for $s, t \in \mathbb{L}^2$ the standard scalar product is given by $\langle s, t\rangle = \int_0^1 s(x)t(x)dx$. In addition, define the random sequence of penalties $(\textsc{Pen}_k)_{k\in\mathbb{N}_0}$ via

$$\textsc{Pen}_k := 12\eta^{-1} \cdot (\widehat{[\lambda]}_0 \vee 1) \cdot \frac{2k+1}{n}$$

for some tuning parameter $\eta \in (0,1)$. The dependence of the estimator on the parameter η will be suppressed for the sake of convenience from now on. Building on the definitions made up to now, the data-driven selection of the dimension parameter is defined as a minimizer of the penalized contrast,

$$\widehat{k}_n := \operatorname*{argmin}_{0\leq k\leq n}\{\Upsilon_n(\widehat{\lambda}_k) + \textsc{Pen}_k\}.$$

The following theorem provides a uniform upper risk bound for the adaptive estimator $\widehat{\lambda}_{\widehat{k}}$.

Theorem 3.9. *Let Assumption 3.1 hold. Then, for any $n \in \mathbb{N}$,*

$$\sup_{\lambda\in\Lambda_\gamma^r} \mathbb{E}[\|\widehat{\lambda}_{\widehat{k}_n} - \lambda\|^2] \lesssim \min_{0\leq k\leq n} \max\left\{\frac{1}{\gamma_k}, \frac{2k+1}{n}\right\} + \frac{1}{n}.$$

Proof. Let us introduce the event $\Omega := \{\eta([\lambda]_0 \vee 1) \leq \widehat{[\lambda]}_0 \vee 1 \leq \eta^{-1}([\lambda]_0 \vee 1)\}$, the definition of which is used to obtain the decomposition

$$\mathbb{E}[\|\widehat{\lambda}_{\widehat{k}_n} - \lambda\|^2] \leq \underbrace{\mathbb{E}[\|\widehat{\lambda}_{\widehat{k}_n} - \lambda\|^2 \mathbb{1}_\Omega]}_{=:\square} + \underbrace{\mathbb{E}[\|\widehat{\lambda}_{\widehat{k}_n} - \lambda\|^2 \mathbb{1}_{\Omega^c}]}_{=:\blacksquare}.$$

We establish uniform upper bounds for both terms separately.
Uniform upper bound for $\square$: Since the equation $\Upsilon_n(t) = \|\widehat{\lambda}_n - t\|^2 - \|\widehat{\lambda}_n\|^2$ holds for all $t \in \mathbb{L}^2$, we obtain that $\operatorname{argmin}_{t\in\mathcal{S}_k} \Upsilon_n(t) = \widehat{\lambda}_k$ for all $k \in \{0,\dots,n\}$ where $\mathcal{S}_k$ denotes the linear subspace of $\mathbb{L}^2$ generated by the φ_j with $j \in \{-k,\dots,k\}$. This identity combined with the definition of $\widehat{k}_n$

yields for all $k \in \{0, \dots, n\}$ the inequality chain

$$\Upsilon_n(\widehat{\lambda}_{\widehat{k}_n}) + \mathrm{PEN}_{\widehat{k}_n} \leq \Upsilon_n(\widehat{\lambda}_k) + \mathrm{PEN}_k \leq \Upsilon_n(\lambda_k) + \mathrm{PEN}_k,$$

where $\lambda_k := \sum_{0 \leq |j| \leq k} [\lambda]_j \varphi_j$ is the projection of λ on the finite-dimensional space $\mathcal{S}_k$. Hence, using the definition of the contrast, we obtain

$$\|\widehat{\lambda}_{\widehat{k}_n}\|^2 \leq \|\lambda_k\|^2 + 2\langle \widehat{\lambda}_n, \widehat{\lambda}_{\widehat{k}_n} - \lambda_k \rangle + \mathrm{PEN}_k - \mathrm{PEN}_{\widehat{k}_n}$$

for all $k \in \{0, \dots, n\}$, from which we conclude, setting $\widehat{\Theta}_n := \widehat{\lambda}_n - \lambda_n$, that

$$\|\widehat{\lambda}_{\widehat{k}_n} - \lambda\|^2 \leq \|\lambda - \lambda_k\|^2 + \mathrm{PEN}_k - \mathrm{PEN}_{\widehat{k}_n} + 2\langle \widehat{\Theta}_n, \widehat{\lambda}_{\widehat{k}_n} - \lambda_k \rangle \tag{3.4}$$

for all $k \in \{0, \dots, n\}$. Consider the set $\mathcal{B}_k := \{\lambda \in \mathcal{S}_k : \|\lambda\|^2 \leq 1\}$. By means of the inequality $2uv \leq \tau u^2 + \tau^{-1} v^2$, we obtain for every $\tau > 0$ and $\widetilde{t} \in \mathcal{S}_k, h \in \mathcal{S}_n$ that

$$2|\langle h, \widetilde{t} \rangle| \leq 2\,\|\widetilde{t}\| \sup_{t \in \mathcal{B}_k} |\langle h, t \rangle| \leq \tau \|\widetilde{t}\|^2 + \tau^{-1} \sup_{t \in \mathcal{B}_k} |\langle h, t \rangle|^2$$

Combining this estimate with (3.4), we obtain (note that $\widehat{\lambda}_{\widehat{k}_n} - \lambda_k \in \mathcal{S}_{k \vee \widehat{k}_n}$)

$$\|\widehat{\lambda}_{\widehat{k}_n} - \lambda\|^2 \leq \|\lambda - \lambda_k\|^2 + \mathrm{PEN}_k - \mathrm{PEN}_{\widehat{k}_n} + \tau \|\widehat{\lambda}_{\widehat{k}_n} - \lambda_k\|^2 + \tau^{-1} \sup_{t \in \mathcal{B}_{k \vee \widehat{k}_n}} |\langle \widehat{\Theta}_n, t \rangle|^2.$$

We have $\|\widehat{\lambda}_{\widehat{k}_n} - \lambda_k\|^2 \leq 2\|\widehat{\lambda}_{\widehat{k}_n} - \lambda\|^2 + 2\|\lambda_k - \lambda\|^2$ and $\|\lambda - \lambda_k\|^2 \leq r\gamma_k^{-1}$ for all $\lambda \in \Lambda_\gamma^r$ thanks to Assumption 3.1. Hence, specializing with $\tau = 1/4$ implies

$$\|\widehat{\lambda}_{\widehat{k}_n} - \lambda\|^2 \leq 3r\gamma_k^{-1} + 2\mathrm{PEN}_k - 2\mathrm{PEN}_{\widehat{k}_n} + 8 \sup_{t \in \mathcal{B}_{k \vee \widehat{k}_n}} |\langle \widehat{\Theta}_n, t \rangle|^2,$$

which is used to obtain

$$\begin{aligned} \|\widehat{\lambda}_{\widehat{k}_n} - \lambda\|^2 \leq 3r\gamma_k^{-1} &+ 8 \left(\sup_{t \in \mathcal{B}_{k \vee \widehat{k}_n}} |\langle \widehat{\Theta}_n, t \rangle|^2 - \frac{3([\lambda]_0 \vee 1) \cdot (2(k \vee \widehat{k}_n) + 1)}{n} \right)_+ \\ &+ \frac{24([\lambda]_0 \vee 1) \cdot (2(k \vee \widehat{k}_n) + 1)}{n} + 2\mathrm{PEN}_k - 2\mathrm{PEN}_{\widehat{k}_n}. \end{aligned}$$

Note that we have $2(k \vee \widehat{k}_n) + 1 \leq 2k + 2\widehat{k}_n + 2$. Thus, due to the definition of both the penalty and Ω we obtain

$$\begin{aligned} \|\widehat{\lambda}_{\widehat{k}_n} - \lambda\|^2 \, \mathbb{1}_\Omega \leq \Bigg\{ 3r\gamma_k^{-1} &+ 8 \left(\sup_{t \in \mathcal{B}_{k \vee \widehat{k}_n}} |\langle \widehat{\Theta}_n, t \rangle|^2 - \frac{3([\lambda]_0 \vee 1) \cdot (2(k \vee \widehat{k}_n) + 1)}{n} \right)_+ \\ &+ 24(1 + \eta^{-2})\sqrt{r} \cdot \frac{2k+1}{n} \Bigg\} \mathbb{1}_\Omega. \end{aligned}$$

Since the last estimate holds for all $k \in \{0, \dots, n\}$ and $\lambda \in \Lambda_\gamma^r$, we obtain

$$\begin{aligned} \mathbb{E}[\|\widehat{\lambda}_{\widehat{k}_n} - \lambda\|^2 \, \mathbb{1}_\Omega] \leq 24\sqrt{r}((1 + \eta^{-2}) + \sqrt{r}) \min_{0 \leq k \leq n} \max \left\{ \frac{1}{\gamma_k}, \frac{2k+1}{n} \right\} \\ + 8 \sum_{k=0}^{n} \mathbb{E} \left[\left(\sup_{t \in \mathcal{B}_k} |\langle \widehat{\Theta}_n, t \rangle|^2 - \frac{3([\lambda]_0 \vee 1)(2k+1)}{n} \right)_+ \right]. \end{aligned} \tag{3.5}$$

We now apply Lemma 3.11 from Section 3.4 which yields for $\lambda \in \Lambda_\gamma^r$ that

$$\mathbb{E}\left[\left(\sup_{t\in\mathcal{B}_k} |\langle\widehat{\Theta}_n, t\rangle|^2 - \frac{3([\lambda]_0 \vee 1)(2k+1)}{n}\right)_+\right] \\ \leq K_1\left[\frac{\sqrt{(2k+1)r}}{n}\exp\left(-K_2\sqrt{\frac{2k+1}{r}}\right) + \frac{2k+1}{n^2}\exp\left(-K_3\sqrt{n}\right)\right],$$

where K_1, K_2 and K_3 are numerical constants independent of n. The estimate $2k+1 \leq 3n$ which holds for $k \in \{0,\ldots,n\}$ implies that

$$\sum_{k=0}^{n}\left[\left(\sup_{t\in\mathcal{B}_k} |\langle\widehat{\Theta}_n, t\rangle|^2 - \frac{3(2k+1)[\lambda]_0}{n}\right)_+\right] \lesssim \sum_{k=0}^{\infty}\frac{\sqrt{2k+1}}{n}\exp\left(-K_2\sqrt{\frac{2k+1}{r}}\right) + \exp(-K_3\sqrt{n}).$$

Note that we have $\sum_{k=0}^{\infty}\sqrt{2k+1}\exp(-K_2\sqrt{(2k+1)/r}) \leq C$ for some numerical constant $C < \infty$. Thus, plugging the derived estimates into (3.5) and taking into account that all the derived estimates hold uniformly for $\lambda \in \Lambda_\gamma^r$, we obtain

$$\sup_{\lambda\in\Lambda_\gamma^r} \mathbb{E}[\|\widehat{\lambda}_{\widehat{k}_n} - \lambda\|^2 \mathbb{1}_\Omega] \lesssim \min_{0\leq k\leq n} \max\left\{\frac{1}{\gamma_k}, \frac{2k+1}{n}\right\} + \frac{1}{n} + \exp(-K_3\sqrt{n}).$$

Uniform upper bound for ■: In order to derive an upper bound for ■, first recall the definition $\lambda_k := \sum_{0\leq|j|\leq k}[\lambda]_j\varphi_j$ from above. We obtain the identity

$$\mathbb{E}[\|\widehat{\lambda}_{\widehat{k}_n} - \lambda\|^2\mathbb{1}_{\Omega^c}] = \mathbb{E}[\|\widehat{\lambda}_{\widehat{k}_n} - \lambda_{\widehat{k}_n}\|^2\mathbb{1}_{\Omega^c}] + \mathbb{E}[\|\lambda - \lambda_{\widehat{k}_n}\|^2\mathbb{1}_{\Omega^c}]. \tag{3.6}$$

Since $\|\lambda - \lambda_{\widehat{k}_n}\|^2 \leq \|\lambda\|^2 \leq r$ due to Assumption 3.1, the second term on the right-hand side of (3.6) satisfies

$$\mathbb{E}[\|\lambda - \lambda_{\widehat{k}_n}\|^2\mathbb{1}_{\Omega^c}] \leq r\mathbb{P}(\Omega^c) \lesssim \frac{1}{n}, \tag{3.7}$$

where the probability estimate for Ω^c will be obtained below. In order to bound the first term on the right-hand side of (3.6), first note that

$$\mathbb{E}[\|\widehat{\lambda}_{\widehat{k}_n} - \lambda_{\widehat{k}_n}\|^2\mathbb{1}_{\Omega^c}] \leq \sum_{0\leq|j|\leq n}\mathbb{E}[|\widehat{[\lambda]}_j - [\lambda]_j|^2\,\mathbb{1}_{\Omega^c}] \leq \mathbb{P}(\Omega^c)^{1/2}\sum_{0\leq|j|\leq n}\mathbb{E}[|\widehat{[\lambda]}_j - [\lambda]_j|^4]^{1/2}.$$

Therefrom, by applying Theorem B.1 with $p = 4$, we can conclude

$$\mathbb{E}[\|\widehat{\lambda}_{\widehat{k}_n} - \lambda_{\widehat{k}_n}\|^2\mathbb{1}_{\Omega^c}] \lesssim \mathbb{P}\left(\Omega^c\right)^{1/2},$$

and it remains to find a suitable bound for $\mathbb{P}(\Omega^c)$. We have

$$\mathbb{P}(\Omega^c) = \mathbb{P}(\widehat{[\lambda]}_0 \vee 1 < \eta([\lambda]_0 \vee 1)) + \mathbb{P}(\widehat{[\lambda]}_0 \vee 1 > \eta^{-1}([\lambda]_0 \vee 1)),$$

and the probabilities on the right-hand side can be bounded by Theorem B.2. More precisely, we have

$$\mathbb{P}(\widehat{[\lambda]}_0 \vee 1 < \eta([\lambda]_0 \vee 1)) \leq \exp(-\omega_1(\eta)n), \quad \text{and}$$
$$\mathbb{P}(\widehat{[\lambda]}_0 \vee 1 > \eta^{-1}([\lambda]_0 \vee 1)) \leq \exp(-\omega_2(\eta)n)$$

with $\omega_1(\eta) = 1 - \eta + \eta\log\eta > 0$ and $\omega_2(\eta) = 1 - \eta^{-1} - \eta^{-1}\log\eta > 0$ for all $\eta \in (0,1)$. Hence, putting together the estimates derived so far, we obtain

$$\mathbb{E}[\|\widehat{\lambda}_{\widehat{k}_n} - \lambda_{\widehat{k}_n}\|^2 \mathbb{1}_{\Omega^c}] \lesssim \frac{1}{n}. \tag{3.8}$$

Putting the estimates (3.7) and (3.8) into (3.6), and again taking into account that all the estimates hold uniformly for $\lambda \in \Lambda_\gamma^r$ yields

$$\sup_{\lambda \in \Lambda_\gamma^r} \mathbb{E}[\|\widehat{\lambda}_{\widehat{k}_n} - \lambda\|^2 \mathbb{1}_{\Omega^c}] \lesssim \frac{1}{n}.$$

Combining the derived uniform bounds for $\square$ and $\blacksquare$ implies the statement of the theorem. $\square$

Remark 3.10. The penalty term used in the definition of $\widehat{k}_n$ is non-deterministic which is in contrast to penalty terms usually used in density estimation or density deconvolution problems. The need for randomization is due to the factor $[\lambda]_0$ in the definition of H in Lemma 3.11. If r (but not γ) was known, one could proceed without randomization by choosing the penalty proportional to $\sqrt{r}(2k+1)/n$. However, the factor $\sqrt{r}$ in this definition cannot be replaced by an estimate of $\sqrt{r}$ because a reasonable estimator of $\sqrt{r}$ is not reachable from the data. Note that the penalty terms considered in [RB03] in a point process framework contain a similar random proportionality constant.

The adaptive estimator $\widehat{\lambda}_{\widehat{k}_n}$ attains the rate Ψ_n if and only if $\min_{0 \leq k \leq n} \max\{\frac{1}{\gamma_k}, \frac{2k+1}{n}\}$ has the same order as Ψ_n. Since under Assumption 3.1 it holds that $k_n^* \lesssim n$, we immediately obtain that the estimator $\widehat{\lambda}_{\widehat{k}_n}$ is rate optimal over the class Λ_γ^r. In particular, the estimator $\widehat{\lambda}_{\widehat{k}_n}$ is rate optimal in the framework of Examples 3.6, 3.7 and 3.8 where $k_n^* \asymp n^{\frac{1}{2p+1}}$, $k_n^* \asymp \log n$, and $k_n^* \asymp (\log n)^{1/p}$, respectively.

3.4. An auxiliary result

The following lemma is a version of Lemma A4 in [JS13a] adapted to our framework. In that paper, a circular deconvolution model was considered and the same way Lemma A4 in [JS13a] is obtained from a variant of Proposition 2.13 in a non-point-process framework (see Lemma B.4, Lemma A3 in [JS13a] or Lemma 1 in [CRT06]), the key ingredient for the proof of the following Lemma 3.11 is Proposition 2.13.

Lemma 3.11. *For all $k \in \{0, \dots, n\}$, we have*

$$\mathbb{E}\left[\left(\sup_{t \in \mathcal{B}_k} |\langle \widehat{\Theta}_n, t\rangle|^2 - \frac{3([\lambda]_0 \vee 1)(2k+1)}{n}\right)_+\right]$$
$$\leq K_1 \left\{ \frac{\sqrt{2k+1}\,\|\lambda\|}{n} \exp\left(-K_2 \cdot \frac{\sqrt{2k+1}}{\|\lambda\|}\right) + \frac{2k+1}{n^2} \exp\left(-K_3\sqrt{n}\right)\right\},$$

with strictly positive numerical constants K_1, K_2, and K_3.

Proof. For $t \in \mathcal{S}_k$, we define the function r_t by $r_t := \sum_{j=-k}^{k} [t]_j \varphi_j$. Then, it is readily verified that $\langle \widehat{\Theta}_n, t\rangle = \frac{1}{n}\sum_{i=1}^n \{\int_0^1 r_t(x) dN_i(x) - \int_0^1 r_t(x)\lambda(x)dx\}$. Hence, building on this definition of r_t, it remains to find constants M_1, H and υ satisfying the preconditions of Proposition 2.13.
Condition concerning M_1: We have

$$\sup_{t \in \mathcal{B}_k} \|r_t\|_\infty^2 = \sup_{t \in \mathcal{B}_k} \sup_{y \in [0,1)} |r_t(y)|^2 \leq \sup_{t \in \mathcal{B}_k} \sup_{y \in [0,1)} \left(\sum_{0 \leq |j| \leq k} |[t]_j||\varphi_j(y)|\right)^2$$

$$\leq \sup_{t\in\mathcal{B}_k} \sup_{y\in[0,1)} \left(\sum_{0\leq|j|\leq k} |[t]_j|^2 \right) \left(\sum_{0\leq|j|\leq k} \varphi_j^2(y) \right)$$
$$\leq 2k+1 =: M_1^2.$$

Condition concerning H: We have

$$\begin{aligned}\mathbb{E}[\sup_{t\in\mathcal{B}_k} |\langle\widehat{\Theta}_n, t\rangle|^2] &\leq \sup_{t\in\mathcal{B}_k} \left(\sum_{0\leq|j|\leq k} |[t]_j|^2 \right) \cdot \mathbb{E}\left[\sum_{0\leq|j|\leq k} \left| \frac{1}{n} \sum_{i=1}^n \left\{ \int_0^1 \varphi_j(x)[dN_i(x) - d\Lambda_i(x)] \right\} \right|^2 \right] \\ &\leq \frac{1}{n} \sum_{0\leq|j|\leq k} \mathrm{Var}\left(\int_0^1 \varphi_j(x) dN_1(x) \right) \\ &\leq \frac{1}{n} \sum_{0\leq|j|\leq k} \int_0^1 \varphi_j^2(x)\lambda(x)dx \\ &\leq \frac{2k+1}{n} \cdot ([\lambda]_0 \vee 1),\end{aligned}$$

and it follows from Jensen's inequality that we can choose $H := (([\lambda]_0 \vee 1) \cdot (2k+1)/n)^{1/2}$.
Condition concerning υ: We have

$$\mathrm{Var}\left(\int_0^1 r_t(x) dN_1(x) \right) = \int_0^1 |r_t(x)|^2 \lambda(x) dx. \tag{3.9}$$

Define $\mathbf{e}_j(t) = \exp(2\pi i j t)$ and set $\langle\lambda\rangle_j = \int_0^1 \lambda(t)\mathbf{e}_j(-t)dt$ using which the identity $\lambda = \sum_{j\in\mathbb{Z}} \langle\lambda\rangle_j \mathbf{e}_j$ holds. We have

$$|r_t(x)|^2 = \left\langle \sum_{0\leq|i|\leq k} \langle t\rangle_i \mathbf{e}_i(x), \sum_{0\leq|j|\leq k} \langle t\rangle_j \mathbf{e}_j(x) \right\rangle_{\mathbb{C}} = \sum_{0\leq|i|\leq k} \sum_{0\leq|j|\leq k} \langle t\rangle_i \overline{\langle t\rangle}_j \mathbf{e}_i(x)\mathbf{e}_{-j}(x),$$

and thus by means of (3.9) that $\mathrm{Var}(\int_0^1 r_t(x)dN_1(x)) = \sum_{0\leq|i|\leq k} \sum_{0\leq|j|\leq k} \langle t\rangle_i \overline{\langle t\rangle}_j \langle\lambda\rangle_{j-i}$. It follows that

$$\sup_{t\in\mathcal{B}_k} \mathrm{Var}\left(\int_0^1 r_t(x) dN_1(x) \right) = \sup_{t\in\mathcal{B}_k} \langle A\boldsymbol{t}, \boldsymbol{t}\rangle_{\mathbb{C}^{2k+1}}$$

where for $t \in \mathcal{B}_k$ we denote by $\boldsymbol{t}$ the vector $(\langle t\rangle_{-k}, \ldots, \langle t\rangle_k)$ and by A the positive semi-definite matrix $A = (\langle\lambda\rangle_{i-j})_{i,j=-k,\ldots,k}$. Hence,

$$\sup_{t\in\mathcal{B}_k} \mathrm{Var}\left(\int_0^1 r_t(x) dN_1(x) \right) = \sup_{t\in\mathcal{B}_k} \langle A^{1/2}\boldsymbol{t}, A^{1/2}\boldsymbol{t}\rangle_{\mathbb{C}^{2k+1}} = \sup_{t\in\mathcal{B}_k} \|A^{1/2}\boldsymbol{t}\|^2 = \|A\|_{\mathrm{op}}.$$

In order to bound $\|A\|_{\mathrm{op}}$, recall for an arbitrary matrix $B = (b_{ij})$ the definitions

$$\|B\|_1 := \max_j \sum_i |b_{ij}| \qquad \text{and} \qquad \|B\|_\infty := \max_i \sum_j |b_{ij}|.$$

Note that by the Cauchy-Schwarz inequality we have both $\|A\|_1 \leq \sqrt{2k+1}\|\lambda\|$ and $\|A\|_\infty \leq \sqrt{2k+1}\|\lambda\|$ and hence by the formula $\|A\|_{\mathrm{op}} \leq \sqrt{\|A\|_1 \cdot \|A\|_\infty}$ (see Corollary 2.3.2 in [GVL96]) we obtain $\|A\|_{\mathrm{op}} \leq \sqrt{2k+1} \cdot \|\lambda\|$. Thus, we can choose $\upsilon = \sqrt{2k+1} \cdot \|\lambda\| \cdot ([\lambda]_0 \vee 1)$.

The claim assertion of the lemma follows now directly from Proposition 2.13 taking $\varepsilon = \frac{1}{4}$. $\square$

4. Non-parametric inverse intensity estimation

This chapter is devoted to the problem of estimating the intensity of a PPP from *indirect observations*. This means that, in contrast to the previous chapter, we do not have direct access to realizations of the point process of interest but only to a *noisy* version. We assume that the observations take on the general form

$$N_i = \sum_j \delta_{y_{ij}}$$

where $\delta_\bullet$ denotes the Dirac measure concentrated at $\bullet$. More precisely, we assume that a generic observation N is related to the target intensity by the relation

$$y_{ij} = x_{ij} + \varepsilon_{ij} - \lfloor x_{ij} + \varepsilon_{ij} \rfloor \tag{4.1}$$

where $\widetilde{N}_i = \sum_j \delta_{x_{ij}}$ is the realization of a PPP with the target intensity function $\lambda \in \mathbb{L}^2 := \mathbb{L}^2([0,1), dx)$ (in this chapter, we consider $\mathbb{L}^2([0,1), dx)$ as the space of square-integrable *complex-valued* functions on $[0,1)$) and ε_{ij} is additive error. As already mentioned in the introduction of this thesis, concerning the relationship between the $\widetilde{N}_i$ and the N_i, we distinguish between the following two models:

1. the errors ε_{ij} in (4.1) are i.i.d. $\sim f$ for some *unknown* density function f. From now on, we refer to this model as *model 1* or the model with Poisson observations.

2. the errors ε_{ij} satisfy $\varepsilon_{ij} \equiv \varepsilon_i \sim f$, that is, all the single points from the hidden point processes $\widetilde{N}_i$ are shifted by the same amount modulo 1. We refer to this model as *model 2* or the model with Cox process observations.

Let us consider the models 1 and 2 in a more detailed way.

Model 1: Poisson observations

In the first model, we assume that the observed point processes are generated from the hidden point processes $\widetilde{N}_i$ by addition of i.i.d. errors $\varepsilon_{ij} \sim f$ to all the single points of the $\widetilde{N}_i$ and then taking the fractional part of the shifted points. This model assumption results in the following random measure representation of the observations:

$$N_i = \sum_j \delta_{x_{ij} + \varepsilon_{ij} - \lfloor x_{ij} + \varepsilon_{ij} \rfloor}.$$

Under the given assumption on the additive errors ε_{ij}, the observable point processes N_i are again Poisson. More precisely, the intensity function ℓ of the N_i is given by the circular convolution $\ell = \lambda \star f$ of the intensity λ with the error density f modulo 1:

$$\ell(t) := \int_0^1 \lambda((t - \varepsilon) - \lfloor t - \varepsilon \rfloor) f(\varepsilon) d\varepsilon, \qquad t \in [0,1). \tag{4.2}$$

From Campbell's theorem (see Theorem 1.14) it can be deduced that for all integrable functions $g : [0,1) \to \mathbb{C}$, we have

$$\mathbb{E}\left[\int_0^1 g(t) dN_i(t)\right] = \int_0^1 g(t)\ell(t)dt. \tag{4.3}$$

Let $\{\mathbf{e}_j\}_{j\in\mathbb{Z}}$ be the *complex trigonometric basis* of $\mathbb{L}^2$ where $\mathbf{e}_j(t) := \exp(2\pi i j t)$. For $j \in \mathbb{Z}$, denote with

$$[\ell]_j := \int_0^1 \ell(t)\mathbf{e}_j(-t)dt, \quad [\lambda]_j := \int_0^1 \lambda(t)\mathbf{e}_j(-t)dt, \quad [f]_j := \int_0^1 f(t)\mathbf{e}_j(-t)dt$$

the Fourier coefficients of ℓ, λ and f, respectively[1]. Setting

$$\widehat{[\ell]}_j := \frac{1}{n}\sum_{i=1}^n \int_0^1 \mathbf{e}_j(-t)dN_i(t), \tag{4.4}$$

applying the convolution theorem and exploiting (4.3), we obtain that

$$\mathbb{E}\widehat{[\ell]}_j = [\lambda]_j[f]_j \qquad \text{for all } j \in \mathbb{Z}.$$

More precisely, we have

$$\widehat{[\ell]}_j = [\lambda]_j[f]_j + \xi_j \quad \text{for all } j \in \mathbb{Z} \tag{4.5}$$

with centred random variables

$$\xi_j = \widehat{[\ell]}_j - \mathbb{E}\widehat{[\ell]}_j = \frac{1}{n}\sum_{i=1}^n \left[\int_0^1 \mathbf{e}_j(-t)dN_i(t) - \int_0^1 \mathbf{e}_j(-t)\ell(t)dt\right].$$

Model 2: Cox observations

In the second model, we assume that all the points of the hidden point process $\widetilde{N}_i$ are shifted by the same amount $\varepsilon_i \sim f$. Hence, the random measure representation of the observations reads

$$N_i = \sum_j \delta_{x_{ij}+\varepsilon_i - \lfloor x_{ij}+\varepsilon_i\rfloor}. \tag{4.6}$$

However, we assume that the errors $\varepsilon_1, \ldots, \varepsilon_n$ are mutually independent. This model has already been intensively considered in [Big+13]. Under the given assumptions, the observed point processes N_i are not Poisson in general but only Cox processes. This fact becomes evident from the following two-step procedure for the generation of observations under model 2: in the first step, random shifts $\varepsilon_i \sim f$ are generated. In the second step, conditionally on the ε_i, the N_i are drawn as independent realizations of a PPP on $[0, 1)$ whose intensity function is $\lambda(t - \varepsilon_i - \lfloor t - \varepsilon_i\rfloor)$, respectively. Thus, in this second model, the observations follow the distribution of a Cox process which is directed by the random measure with random intensity $\lambda(t - \varepsilon - \lfloor t - \varepsilon\rfloor)$ for $\varepsilon \sim f$.

We now derive a sequence space representation of the model with Cox observations similar to the Poisson case. First, notice that for $i = 1, \ldots, n$ and integrable functions g we have

$$\mathbb{E}\left[\int_0^1 g(t)dN_i(t) \mid \varepsilon_i\right] = \int_0^1 g(t)\lambda(t - \varepsilon_i - \lfloor t - \varepsilon_i\rfloor)dt$$

which implies

$$\mathbb{E}\left[\int_0^1 g(t)dN_i(t)\right] = \int_0^1 g(t)\int_0^1 \lambda(t - \varepsilon - \lfloor t - \varepsilon\rfloor)f(\varepsilon)d\varepsilon dt = \int_0^1 g(t)\ell(t)dt,$$

where $\ell = \lambda \star f$ denotes the circular convolution of the function λ and the density f defined as in (4.2). Thus, the mean measure of a generic realization N obeying model 2 has the Radon-Nikodym derivative ℓ with respect to the Lebesgue measure. Note that the mean measures of the observed point processes under models 1 and 2 coincide, but the observations in model 2 stem from

[1] Since only the effect of the errors ε_{ij} modulo $\mathbb{Z}$ is of interest, one can assume without loss of generality that f is supported on $[0, 1]$.

a Cox instead of a Poisson process. With $\widehat{[\ell]}_j$ defined as in (4.4) the relation

$$\mathbb{E}[\widehat{[\ell]}_j \mid \varepsilon_1, \ldots, \varepsilon_n] = [\lambda]_j \cdot \widetilde{[f]}_j$$

holds where $\widetilde{[f]}_j := \frac{1}{n} \sum_{i=1}^n \mathbf{e}_j(-\varepsilon_i)$. Thus, we get the following representation as a sequence space model (cf. Equation (2.4) in [Big+13]):

$$\widehat{[\ell]}_j = [\lambda]_j \cdot \widetilde{[f]}_j + \xi_j \quad \text{for all } j \in \mathbb{Z} \tag{4.7}$$

where $\xi_j := \frac{1}{n} \sum_{i=1}^n [\int_0^1 \mathbf{e}_j(-t) dN_i(t) - \int_0^1 \mathbf{e}_j(-t) \lambda(t - \varepsilon_i - \lfloor t - \varepsilon_i \rfloor) dt]$ are centred random variables for all $j \in \mathbb{Z}$. The connection between the sequence space model at hand and the standard sequence space model formulation for statistical linear inverse problems is discussed in detail in Section 2.1 of [Big+13].

Observation scheme

Estimation of the intensity λ under model 2 has been investigated in detail in [Big+13] under the assumption that the error density is known and its Fourier coefficients obey a polynomial decay. In this setup, the authors proved a minimax lower bound and proposed a wavelet-series estimator which automatically adapts to unknown smoothness. Contrary to this, we will assume that the error density f is unknown. Instead, we assume that one can observe an additional independent sample $Y_1, \ldots, Y_m$ from the error density f. This second sample only makes inference possible, and its availability ensures identifiability of the model under certain assumptions on f. Thus, our complete set of observations is given by

$$N_1, \ldots, N_n \text{ i.i.d.} \sim \mathcal{L}(N) \qquad \text{and} \qquad Y_1, \ldots, Y_m \text{ i.i.d.} \sim f \tag{4.8}$$

where N is a generic realization of the observed point process under one of the considered models.

4.1. Methodology: Orthonormal series estimator of the intensity

As in the previous chapter, we use an orthonormal series estimator as a natural device for the non-parametric estimation of λ. In contrast to Chapter 3, we consider an orthonormal series estimator in terms of the *complex* trigonometric basis $\{\mathbf{e}_j\}_{j \in \mathbb{Z}}$ where $\mathbf{e}_j(t) := \exp(2\pi i j t)$. This basis was already considered in the derivation of the sequence space representations (4.5) and (4.7) above. The considered estimators take on the form

$$\widehat{\lambda}_k = \sum_{0 \leq |j| \leq k} \widehat{[\lambda]}_j \mathbf{e}_j$$

where $\widehat{[\lambda]}_j$ is a suitable estimator of $[\lambda]_j$ and $k \in \mathbb{N}_0$ is a dimension parameter that has to be chosen appropriately. In view of equations (4.5) and (4.7), it seems natural to estimate $[\lambda]_j$ via the quotient of suitable estimators $\widehat{[\ell]}_j$ and $\widehat{[f]}_j$ of $[\ell]_j$ and $[f]_j$, respectively. Note that neither of the quantities $[\ell]_j$ and $[f]_j$ is known *a priori*. However, unbiased estimators of $[\ell]_j$ and $[f]_j$ are available by means of their empirical counterparts

$$\widehat{[\ell]}_j := \frac{1}{n} \sum_{i=1}^n \int_0^1 \mathbf{e}_j(-t) dN_i(t) \qquad \text{and} \qquad \widehat{[f]}_j := \frac{1}{m} \sum_{i=1}^m \mathbf{e}_j(-Y_i).$$

In order to account for 'too small' absolute values of $\widehat{[f]}_j$ which would result in unstable behaviour of the estimator, we insert an additional threshold by defining for $j \in \mathbb{Z}$ the event

$\Omega_j := \{|\widehat{[f]}_j|^2 \geq m^{-1}\}$ and based on the definition of Ω_j the final estimator

$$\widehat{\lambda}_k := \sum_{0\leq|j|\leq k} \frac{\widehat{[\ell]}_j}{\widehat{[f]}_j} \mathbb{1}_{\Omega_j} \mathbf{e}_j. \tag{4.9}$$

The idea of adding the indicator $\mathbb{1}_{\Omega_j}$ is taken from [NH97], and has been used for the construction of a variety of non-parametric estimators in statistical inverse problems (see, for instance, [JS13a], [JS13b]). As in the case of direct observations, the choice of the tuning parameter $k \in \mathbb{N}_0$ crucially determines the performance of the estimator.

4.2. Minimax theory

Let us first consider the estimation of $\lambda \in \mathbb{L}^2$ under models 1 and 2 from the observations in (4.8) taking on a minimax point of view. For some strictly positive sequence $\omega = (\omega_j)_{j\in\mathbb{Z}}$ of weights, introduce the weighted squared norm $\|\cdot\|_\omega^2$ defined via

$$\|g\|_\omega^2 := \sum_{j\in\mathbb{Z}} \omega_j |[g]_j|^2$$

for all $g \in \mathbb{L}^2$ such that the sum in the definition is finite. The performance of an arbitrary estimator $\widetilde{\lambda}$ of λ will be evaluated by means of the *maximum risk*

$$\sup_{\lambda\in\Lambda} \sup_{f\in\mathcal{F}} \mathbb{E}[\|\widetilde{\lambda} - \lambda\|_\omega^2]$$

for appropriately defined classes Λ of intensities and $\mathcal{F}$ of error densities. Note that in this chapter the supremum is taken both over a class of intensities and a class of error densities. Again, the benchmark for potential estimators is the *minimax risk*

$$\inf_{\widetilde{\lambda}} \sup_{\lambda\in\Lambda} \sup_{f\in\mathcal{F}} \mathbb{E}[\|\widetilde{\lambda} - \lambda\|_\omega^2]$$

where the infimum is taken over all estimators $\widetilde{\lambda}$ of λ based on the observations in (4.8). In the following, we consider abstract smoothness classes $\Lambda = \Lambda_\gamma^r$ and $\mathcal{F} = \mathcal{F}_\alpha^d$ defined in terms of strictly positive symmetric sequences $\gamma = (\gamma_j)_{j\in\mathbb{Z}}$, $\alpha = (\alpha_j)_{j\in\mathbb{Z}}$, and real numbers $r > 0$, $d \geq 1$. More precisely, we will derive minimax results under the assumption that the intensity λ is an element of the ellipsoid

$$\Lambda_\gamma^r := \{\lambda \in \mathbb{L}_2 : \lambda \geq 0 \quad \text{and} \quad \|\lambda\|_\gamma^2 := \sum_{j\in\mathbb{Z}} \gamma_j |[\lambda]_j|^2 \leq r\},$$

and the error density f belongs to the hyperrectangle

$$\mathcal{F}_\alpha^d := \{f \in \mathbb{L}^2 : f \geq 0, [f]_0 = 1 \text{ and } d^{-1} \leq |[f]_j|^2/\alpha_j \leq d \quad \forall j \in \mathbb{Z}\}.$$

The mild regularity assumptions which we impose on the sequences ω, γ, and α to obtain our results are summarized in the following assumption.

Assumption 4.1. *γ, ω and α are strictly positive symmetric sequences such that $(\omega_n\gamma_n^{-1})_{n\in\mathbb{N}_0}$ and $(\alpha_n)_{n\in\mathbb{N}_0}$ are non-increasing and $\rho := \sum_{j\in\mathbb{Z}} \alpha_j < \infty$. In addition, $\gamma_0 = \omega_0 = \alpha_0 = 1$ and $\gamma_j \geq 1$ for all $j \in \mathbb{Z}$.*

4.2.1. Upper bounds

We start our investigation with the derivation of upper bounds for the minimax risk under models 1 and 2. The bounds will turn out to be essentially the same and differ merely with respect to the numerical constants involved. They are established by considering a suitable estimator which is defined by specializing the orthonormal series estimator $\widehat{\lambda}_k$ in (4.9) with some specific choice of the dimension parameter k. This choice of the dimension parameter will be the same for both models 1 and 2. Given the sequences ω, γ and α, we put

$$k_n^* := \operatorname*{argmin}_{k\in\mathbb{N}_0} \max\left\{ \frac{\omega_k}{\gamma_k}, \sum_{0\leq|j|\leq k} \frac{\omega_j}{n\alpha_j} \right\}, \tag{4.10}$$

and, in addition,

$$\Psi_n := \max\left\{ \frac{\omega_{k_n^*}}{\gamma_{k_n^*}}, \sum_{0\leq|j|\leq k_n^*} \frac{\omega_j}{n\alpha_j} \right\}. \tag{4.11}$$

The quantity Ψ_n will turn out to be the optimal rate of convergence in terms of the sample size n under mild assumptions and k_n^* is the corresponding optimal choice of the dimension parameter which remarkably does not depend on the sample size m. Note that, formally, the definition of Ψ_n in Chapter 3 corresponds to the one in (4.11) with $\omega_j = \alpha_j = 1$ for all $j \in \mathbb{Z}$. The rate of convergence in terms of the sample size m will turn out to be given by

$$\Phi_m := \max_{j\in\mathbb{N}} \left\{ \frac{\omega_j}{\gamma_j} \cdot \min\left\{1, \frac{1}{m\alpha_j}\right\}\right\}. \tag{4.12}$$

Theorem 4.2. *Let Assumption 4.1 hold and further assume that the samples $N_1, \ldots, N_n$ and $Y_1, \ldots, Y_m$ in* (4.8) *are drawn in accordance with model 1 or 2. Then, for any $n, m \in \mathbb{N}$,*

$$\sup_{\lambda\in\Lambda_\gamma^r} \sup_{f\in\mathcal{F}_\alpha^d} \mathbb{E}[\|\widehat{\lambda}_{k_n^*} - \lambda\|_\omega^2] \lesssim \Psi_n + \Phi_m.$$

Proof. We give the proof for model 1 only. The proof for model 2 follows in complete analogy by exploiting statement ii) instead of i) in part a) of Lemma 4.15 and leads to slightly different numerical constants only.

Set $\widetilde{\lambda}_{k_n^*} := \sum_{0\leq|j|\leq k_n^*} [\lambda]_j \mathbb{1}_{\Omega_j} \mathbf{e}_j$. The proof consists in finding appropriate upper bounds for the quantities $\square$ and $\triangle$ in the estimate

$$\mathbb{E}[\|\widehat{\lambda}_{k_n^*} - \lambda\|_\omega^2] \leq 2\,\mathbb{E}[\|\widehat{\lambda}_{k_n^*} - \widetilde{\lambda}_{k_n^*}\|_\omega^2] + 2\,\mathbb{E}[\|\lambda - \widetilde{\lambda}_{k_n^*}\|_\omega^2] =: 2\square + 2\triangle. \tag{4.13}$$

Uniform upper bound for $\square$: Using the identity $\mathbb{E}\widehat{[\ell]}_j = [f]_j[\lambda]_j$ we obtain

$$\begin{aligned}
\square &= \sum_{0\leq|j|\leq k_n^*} \omega_j\, \mathbb{E}[|\widehat{[\ell]}_j/\widehat{[f]}_j - [\lambda]_j|^2 \mathbb{1}_{\Omega_j}] \\
&\leq 2 \sum_{0\leq|j|\leq k_n^*} \omega_j\, \mathbb{E}[|\widehat{[\ell]}_j/\widehat{[f]}_j - \mathbb{E}\widehat{[\ell]}_j/\widehat{[f]}_j|^2\, \mathbb{1}_{\Omega_j}] + 2 \sum_{0\leq|j|\leq k_n^*} \omega_j |[\lambda]_j|^2\, \mathbb{E}[|[f]_j/\widehat{[f]}_j - 1|^2\, \mathbb{1}_{\Omega_j}] \\
&=: 2\square_1 + 2\square_2.
\end{aligned}$$

Using the estimate $|a|^2 \leq 2\,|a-1|^2 + 2$ for $a = [f]_j/\widehat{[f]}_j$, the definition of Ω_j and the independence of $\widehat{[\ell]}_j$ and $\widehat{[f]}_j$ we get

$$\square_1 = \sum_{0\leq|j|\leq k_n^*} \omega_j\, \mathbb{E}\left[|\widehat{[\ell]}_j/\widehat{[f]}_j - \mathbb{E}\widehat{[\ell]}_j/\widehat{[f]}_j|^2 \cdot \left|\frac{\widehat{[f]}_j}{[f]_j}\right|^2 \mathbb{1}_{\Omega_j} \right]$$

$$\leq 2 \sum_{0\leq |j| \leq k_n^*} m\omega_j \frac{\operatorname{Var}(\widehat{[\ell]}_j) \operatorname{Var}(\widehat{[f]}_j)}{|[f]_j|^2} + 2 \sum_{0\leq |j| \leq k_n^*} \omega_j \frac{\operatorname{Var}(\widehat{[\ell]}_j)}{|[f]_j|^2}.$$

Applying statements a) and b) from Lemma 4.15 together with $f \in \mathcal{F}_\alpha^d$ yields

$$\square_1 \leq 4d \sum_{0\leq |j| \leq k_n^*} \omega_j \frac{[\lambda]_0}{n\alpha_j}$$

which using $\gamma_0 = 1$ (which holds due to Assumption 4.1) implies

$$\square_1 \leq 4d\sqrt{r} \sum_{0\leq |j| \leq k_n^*} \frac{\omega_j}{n\alpha_j} \leq 4d\sqrt{r} \cdot \Psi_n.$$

Now consider $\square_2$. Using the estimate $|a|^2 \leq 2\,|a-1|^2 + 2$ for $a = [f]_j/\widehat{[f]}_j$ and the definition of Ω_j yields

$$\mathbb{E}[|[f]_j/\widehat{[f]}_j - 1|^2\, \mathbb{1}_{\Omega_j}] \leq 2m \frac{\mathbb{E}[|\widehat{[f]}_j - [f]_j|^4]}{|[f]_j|^2} + 2 \frac{\operatorname{Var}(\widehat{[f]}_j)}{|[f]_j|^2}.$$

Notice that Theorem B.1 implies the existence of a constant $C > 0$ (independent of j) with $\mathbb{E}[|\widehat{[f]}_j - [f]_j|^4] \leq C/m^2$. Using this inequality in combination with assertion b) from Lemma 4.15 and $f \in \mathcal{F}_\alpha^d$ implies

$$\mathbb{E}[|[f]_j/\widehat{[f]}_j - 1|^2\, \mathbb{1}_{\Omega_j}] \leq 2d(C+1)/(m\alpha_j). \tag{4.14}$$

In addition, $\mathbb{E}[|[f]_j/\widehat{[f]}_j - 1|^2 \mathbb{1}_{\Omega_j}] \leq m \operatorname{Var}(\widehat{[f]}_j) \leq 1$ which in combination with (4.14) implies

$$\square_2 \leq 2d(C+1) \sum_{0\leq |j| \leq k_n^*} \omega_j\, |[\lambda]_j|^2 \min\left(1, \frac{1}{m\alpha_j}\right).$$

Exploiting the fact that $\lambda \in \Lambda_\gamma^r$ and the definition of Φ_m in (4.12) we obtain

$$\square_2 \leq 2dr(C+1)(1+\gamma_1/\omega_1) \cdot \Phi_m.$$

Putting together the estimates for $\square_1$ and $\square_2$ yields

$$\square \leq 8d\sqrt{r} \cdot \Psi_n + 4d(C+1)(1+\gamma_1/\omega_1) r \cdot \Phi_m.$$

Uniform upper bound for $\triangle$: $\triangle$ can be decomposed as

$$\begin{aligned} \triangle &= \sum_{j\in\mathbb{Z}} \omega_j\, |[\lambda]_j|^2\, \mathbb{E}[1 - \mathbb{1}_{\{0\leq |j| \leq k_n^*\}} \cdot \mathbb{1}_{\Omega_j}] = \sum_{|j| > k_n^*} \omega_j\, |[\lambda]_j|^2 + \sum_{0\leq |j| \leq k_n^*} \omega_j\, |[\lambda]_j|^2 \cdot \mathbb{P}(\Omega_j^c) \\ &= \triangle_1 + \triangle_2. \end{aligned}$$

$\lambda \in \Lambda_\gamma^r$ implies $\triangle_1 \leq r\omega_{k_n^*}/\gamma_{k_n^*} \leq r \cdot \Psi_n$, and Lemma 4.15 yields the estimate $\triangle_2 \leq 4dr \cdot \Phi_m$ which together imply that $\triangle \leq r \cdot \Psi_n + 4dr \cdot \Phi_m$. Putting the obtained estimates for $\square$ and $\triangle$ into (4.13) finishes the proof of the theorem. □

4.2.2. Lower bounds

In this section, we derive a lower bound for the minimax risk under model 1. For this purpose, we provide lower bounds in terms of the sample sizes n and m in (4.8), separately. The following theorem shows that the quantity Ψ_n is a lower bound for the minimax risk up to a multiplicative numerical constant.

THEOREM 4.3 (Lower bound in n for model 1). *Let Assumption 4.1 hold and further assume that*

(C1) $\Gamma := \sum_{j\in\mathbb{Z}} \gamma_j^{-1} < \infty$, *and*

(C2) $0 < \eta^{-1} := \inf_{n\in\mathbb{N}} \Psi_n^{-1} \cdot \min\left\{\frac{\omega_{k_n^*}}{\gamma_{k_n^*}}, \sum_{0\le|j|\le k_n^*} \frac{\omega_j}{n\alpha_j}\right\}$ *for some* $1 \le \eta < \infty$

where the quantities k_n^* *and* Ψ_n *are defined in* (4.10) *and* (4.11), *respectively. Then, for any* $n \in \mathbb{N}$,

$$\inf_{\widetilde{\lambda}} \sup_{\lambda\in\Lambda_\gamma^r} \sup_{f\in\mathcal{F}_\alpha^d} \mathbb{E}[\|\widetilde{\lambda} - \lambda\|_\omega^2] \ge \frac{\zeta r}{16\eta} \cdot \Psi_n$$

where $\zeta = \min\{\frac{1}{2d\Gamma\eta}, \frac{2\delta}{d\sqrt{r}}\}$ *with* $\delta = \frac{1}{2} - \frac{1}{2\sqrt{2}}$ *and the infimum is taken over all estimators* $\widetilde{\lambda}$ *of* λ *based on the observations from* (4.8) *under model 1.*

PROOF. Let us define ζ as in the statement of the theorem and for each $\theta = (\theta_j)_{0\le j\le k_n^*} \in \{\pm 1\}^{k_n^*+1}$ the function λ_θ through

$$\begin{aligned}\lambda_\theta &:= \left(\frac{r}{4}\right)^{1/2} + \theta_0 \left(\frac{r\zeta}{4n}\right)^{1/2} + \left(\frac{r\zeta}{4n}\right)^{1/2} \sum_{1\le|j|\le k_n^*} \theta_{|j|}\alpha_j^{-1/2}\mathbf{e}_j \\ &= \left(\frac{r}{4}\right)^{1/2} + \left(\frac{r\zeta}{4n}\right)^{1/2} \sum_{0\le|j|\le k_n^*} \theta_{|j|}\alpha_j^{-1/2}\mathbf{e}_j.\end{aligned}$$

Then each λ_θ is a real-valued function which is non-negative thanks to the estimate

$$\begin{aligned}\left\|\left(\frac{r\zeta}{4n}\right)^{1/2} \sum_{0\le|j|\le k_n^*} \theta_{|j|}\alpha_j^{-1/2}\mathbf{e}_j\right\|_\infty &\le \left(\frac{r\zeta}{4n}\right)^{1/2} \sum_{0\le|j|\le k_n^*} \alpha_j^{-1/2} \\ &\le \left(\frac{r\zeta}{4}\right)^{1/2} \left(\sum_{0\le|j|\le k_n^*} \gamma_j^{-1}\right)^{1/2} \left(\sum_{0\le|j|\le k_n^*} \frac{\gamma_j}{n\alpha_j}\right)^{1/2} \\ &\le \left(\frac{r\zeta\Gamma}{4}\right)^{1/2} \left(\frac{\gamma_{k_n^*}}{\omega_{k_n^*}} \sum_{0\le|j|\le k_n^*} \frac{\omega_j}{n\alpha_j}\right)^{1/2} \\ &\le \left(\frac{r\zeta\eta\Gamma}{4}\right)^{1/2} \le \left(\frac{r}{4}\right)^{1/2}.\end{aligned}$$

Moreover $\|\lambda_\theta\|_\gamma^2 \le r$ holds for each $\theta \in \{\pm 1\}^{k_n^*+1}$ due to the estimate

$$\begin{aligned}\|\lambda_\theta\|_\gamma^2 = \sum_{0\le|j|\le k_n^*} |[\lambda_\theta]_j|^2\,\gamma_j &= \left[\left(\frac{r}{4}\right)^{1/2} + \theta_0\left(\frac{r\zeta}{4n}\right)^{1/2}\right]^2 + \frac{r\zeta}{4} \sum_{1\le|j|\le k_n^*} \frac{\gamma_j}{n\alpha_j} \\ &\le \frac{r}{2} + \frac{r\zeta}{2n} + \frac{r\zeta}{4} \cdot \frac{\gamma_{k_n^*}}{\omega_{k_n^*}} \sum_{1\le|j|\le k_n^*} \frac{\omega_j}{n\alpha_j} \\ &\le \frac{r}{2} + \frac{r\zeta}{2} \cdot \frac{\gamma_{k_n^*}}{\omega_{k_n^*}} \sum_{0\le|j|\le k_n^*} \frac{\omega_j}{n\alpha_j} \le r.\end{aligned}$$

This estimate and the non-negativity of λ_θ together imply $\lambda_\theta \in \Lambda_\gamma^r$ for all $\theta \in \{\pm 1\}^{k_n^*+1}$. From now on let $f \in \mathcal{F}_\alpha^d$ be fixed and let $\mathbb{P}_\theta$ denote the joint distribution of the i.i.d. samples $N_1, \dots, N_n$ and $Y_1, \dots, Y_m$ when the true parameters are λ_θ and f, respectively. Let $\mathbb{P}_\theta^{N_i}$ denote the corresponding one-dimensional marginal distributions and $\mathbb{E}_\theta$ the expectation with respect to $\mathbb{P}_\theta$. Let $\widetilde{\lambda}$ be an arbitrary estimator of λ. The key argument of the proof is the following reduction scheme:

$$\begin{aligned}\sup_{\lambda\in\Lambda_\gamma^r}\sup_{f\in\mathcal{F}_\alpha^d}\mathbb{E}[\|\widetilde{\lambda}-\lambda\|_\omega^2] &\geq \sup_{\theta\in\{\pm1\}^{k_n^*+1}}\mathbb{E}_\theta[\|\widetilde{\lambda}-\lambda_\theta\|_\omega^2] \geq \frac{1}{2^{k_n^*+1}}\sum_{\theta\in\{\pm1\}^{k_n^*+1}}\mathbb{E}_\theta[\|\widetilde{\lambda}-\lambda_\theta\|_\omega^2]\\ &= \frac{1}{2^{k_n^*+1}}\sum_{\theta\in\{\pm1\}^{k_n^*+1}}\sum_{0\leq|j|\leq k_n^*}\omega_j\ \mathbb{E}_\theta[|[\widetilde{\lambda}-\lambda_\theta]_j|^2]\\ &= \frac{1}{2^{k_n^*+1}}\sum_{0\leq|j|\leq k_n^*}\frac{\omega_j}{2}\sum_{\theta\in\{\pm1\}^{k_n^*+1}}\{\mathbb{E}_\theta[|[\widetilde{\lambda}-\lambda_\theta]_j|^2]+\mathbb{E}_{\theta^{(j)}}[|[\widetilde{\lambda}-\lambda_{\theta^{(|j|)}}]_j|^2]\} \end{aligned}\tag{4.15}$$

where for $\theta\in\{\pm1\}^{k_n^*+1}$ and $j\in\{-k_n^*,\ldots,k_n^*\}$ the element $\theta^{(|j|)}\in\{\pm1\}^{k_n^*+1}$ is defined by $\theta_k^{(|j|)}=\theta_k$ for $k\neq|j|$ and $\theta_{|j|}^{(|j|)}=-\theta_{|j|}$. Consider the Hellinger affinity $\rho(\mathbb{P}_\theta,\mathbb{P}_{\theta^{(|j|)}}):=\int\sqrt{d\mathbb{P}_\theta d\mathbb{P}_{\theta^{(|j|)}}}$. For an arbitrary estimator $\widetilde{\lambda}$ of λ we have

$$\begin{aligned}\rho(\mathbb{P}_\theta,\mathbb{P}_{\theta^{(|j|)}}) &\leq \int\frac{|[\widetilde{\lambda}-\lambda_\theta]_j|}{|[\lambda_\theta-\lambda_{\theta^{(|j|)}}]_j|}\sqrt{d\mathbb{P}_\theta d\mathbb{P}_{\theta^{(|j|)}}}+\int\frac{|[\widetilde{\lambda}-\lambda_{\theta^{(|j|)}}]_j|}{|[\lambda_\theta-\lambda_{\theta^{(|j|)}}]_j|}\sqrt{d\mathbb{P}_\theta d\mathbb{P}_{\theta^{(|j|)}}}\\ &\leq\left(\int\frac{|[\widetilde{\lambda}-\lambda_\theta]_j|^2}{|[\lambda_\theta-\lambda_{\theta^{(|j|)}}]_j|^2}d\mathbb{P}_\theta\right)^{1/2}+\left(\int\frac{|[\widetilde{\lambda}-\lambda_{\theta^{(|j|)}}]_j|^2}{|[\lambda_\theta-\lambda_{\theta^{(|j|)}}]_j|^2}d\mathbb{P}_{\theta^{(|j|)}}\right)^{1/2}\end{aligned}$$

from which we conclude by means of the elementary inequality $(a+b)^2\leq 2a^2+2b^2$ that

$$\frac{1}{2}|[\lambda_\theta-\lambda_{\theta^{(|j|)}}]_j|^2\rho^2(\mathbb{P}_\theta,\mathbb{P}_{\theta^{(|j|)}})\leq\mathbb{E}_\theta[|[\widetilde{\lambda}-\lambda_\theta]_j|^2]+\mathbb{E}_{\theta^{(|j|)}}[|[\widetilde{\lambda}-\lambda_{\theta^{(|j|)}}]_j|^2].$$

Recall the definition of the Hellinger distance between two probability measures $\mathbb{P}$ and $\mathbb{Q}$ as $H(\mathbb{P},\mathbb{Q}):=(\int[\sqrt{d\mathbb{P}}-\sqrt{d\mathbb{Q}}]^2)^{1/2}$ and, analogously, the Hellinger distance between two finite measures ν and μ (that not necessarily have total mass equal to one) by $H(\nu,\mu):=(\int[\sqrt{d\nu}-\sqrt{d\mu}]^2)^{1/2}$ (as usual, the integral is formed with respect to any measure dominating both ν and μ). Let ν_θ denote the intensity measure of a PPP N on $[0,1)$ whose Radon-Nikodym derivative with respect to the Lebesgue measure is given by $\ell_\theta:=\lambda_\theta\star f$. Note that we have the estimate $\ell_\theta\geq\delta\sqrt{r}$ for all $\theta\in\{\pm1\}^{k_n^*+1}$ with $\delta=\frac{1}{2}-\frac{1}{2\sqrt{2}}$ due to

$$\left(\frac{r\zeta}{4n}\right)^{1/2}+\sum_{1\leq|j|\leq k_n^*}|[\lambda_\theta]_j\cdot[f]_j|\leq\left(\frac{rd\zeta}{4n}\right)^{1/2}\sum_{0\leq|j|\leq k_n^*}\alpha_j^{-1/2}\leq\frac{\sqrt{r}}{2\sqrt{2}}$$

which can be realized in analogy to the non-negativity of λ_θ shown above. We obtain

$$H^2(\nu_\theta,\nu_{\theta^{(|j|)}})=\int(\sqrt{\ell_\theta}-\sqrt{\ell_{\theta^{(|j|)}}})^2=\int\frac{|\ell_\theta-\ell_{\theta^{(|j|)}}|^2}{(\sqrt{\ell_\theta}+\sqrt{\ell_{\theta^{(|j|)}}})^2}\leq\frac{\|\ell_\theta-\ell_{\theta^{(|j|)}}\|^2}{4\delta\sqrt{r}}=\frac{\zeta d\sqrt{r}}{4\delta n}\leq\frac{1}{n}.$$

Since the distribution of the sample $Y_1,\ldots,Y_m$ does not depend on the choice of θ we obtain

$$H^2(\mathbb{P}_\theta,\mathbb{P}_{\theta^{(|j|)}})\leq\sum_{i=1}^n H^2(\mathbb{P}_\theta^{N_i},\mathbb{P}_{\theta^{(|j|)}}^{N_i})\leq\sum_{i=1}^n H^2(\nu_\theta,\nu_{\theta^{(|j|)}})\leq 1,\tag{4.16}$$

where the first estimate follows from Lemma A.3 and the second one is due to Theorem A.8 *(ii)* which can be applied since each N_i is a PPP under model 1. Thus, the relation $\rho(\mathbb{P}_\theta,\mathbb{P}_{\theta^{(|j|)}})=1-\frac{1}{2}H^2(\mathbb{P}_\theta,\mathbb{P}_{\theta^{(|j|)}})$ implies $\rho(\mathbb{P}_\theta,\mathbb{P}_{\theta^{(|j|)}})\geq\frac{1}{2}$. Finally, putting the obtained estimates into the reduction scheme (4.15) leads to

$$\sup_{\lambda\in\Lambda_\gamma^r}\sup_{f\in\mathcal{F}_\alpha^d}\mathbb{E}[\|\widetilde{\lambda}-\lambda\|_\omega^2] \geq \frac{1}{2^{k_n^*+1}}\sum_{\theta\in\{\pm 1\}^{k_n^*+1}}\sum_{0\leq|j|\leq k_n^*}\frac{\omega_j}{2}\{\mathbb{E}_\theta[|[\widetilde{\lambda}-\lambda_\theta]_j|^2]+\mathbb{E}_{\theta^{(|j|)}}[|[\widetilde{\lambda}-\lambda_{\theta^{(|j|)}}]_j|^2]\}$$
$$\geq \sum_{0\leq|j|\leq k_n^*}\frac{\omega_j}{16}|[\lambda_\theta-\lambda_{\theta^{(|j|)}}]_j|^2 = \frac{\zeta r}{16}\sum_{0\leq|j|\leq k_n^*}\frac{\omega_j}{n\alpha_j}\geq\frac{\zeta r}{16\eta}\cdot\Psi_n$$

which finishes the proof of the theorem since $\widetilde{\lambda}$ was arbitrary. □

Let us state some remarks concerning Theorem 4.3: firstly, the lower bound proportional to Ψ_n holds already in case of a known error density because only one fixed error density $f\in\mathcal{F}_\alpha^d$ is considered in the proof of Theorem 4.3. Secondly, assuming the convergence of the series $\sum_{j\in\mathbb{Z}}\gamma_j^{-1}$ through condition (C1) is necessary only in order to establish the non-negativity of the candidate intensity functions λ_θ. The same condition appeared already in the lower bound proof in the setup with direct observations (cf. Theorem 3.3 in Chapter 3). Thirdly, in the uninteresting case that r equals 0 (which we have excluded from our investigation by assuming that r is strictly positive), the lower bound equals 0 as well because in this case the only admissible intensity function is the zero function. This is in accordance with the fact that the estimator $\widehat{\lambda}_k$ in (4.9) equals the zero function almost surely if $\lambda\equiv 0$ (independent of the choice of the dimension parameter).

Remark 4.4. Unfortunately, the proof given above cannot be adopted directly to establish a lower bound for model 2. The crux of the matter here is the second estimate in (4.16) which only holds for PPPs. Thus, the establishment of such a lower bound in our framework remains an open question for future work.

We now tackle the question whether the rate Φ_m of the estimator $\lambda_{k_n^*}$ in terms of the sample size m is optimal. The following theorem provides an affirmative answer under mild assumptions.

Theorem 4.5. *Let Assumption 4.1 hold, and in addition assume that*

(C3) there exists a density f in $\mathcal{F}_\alpha^{\sqrt{d}}$ satisfying $f\geq 1/2$.

Then, for any $m\in\mathbb{N}$,

$$\inf_{\widetilde{\lambda}}\sup_{\lambda\in\Lambda_\gamma^r}\sup_{f\in\mathcal{F}_\alpha^d}\mathbb{E}[\|\widetilde{\lambda}-\lambda\|_\omega^2]\geq\frac{1}{8}\left(1-\frac{\sqrt{3}}{2}\right)\zeta^2 r d^{-1/2}\cdot\Phi_m$$

where Φ_m is defined in (4.12)*, $\zeta=\min\{\frac{1}{4\sqrt{d}},1-d^{-1/4}\}$ and the infimum is taken over all estimators $\widetilde{\lambda}$ of λ based on the observations from* (4.8) *under model 1.*

Proof. The following reduction scheme follows along a general strategy that is well-known for the establishment of lower bounds in non-parametric estimation (for a detailed account cf. [Tsy08], Chapter 2). Note that by Markov's inequality we have for an arbitrary estimator $\widetilde{\lambda}$ of λ and arbitrary $A>0$ (which will be specified below)

$$\mathbb{E}[\Phi_m^{-1}\|\widetilde{\lambda}-\lambda\|_\omega^2]\geq A\cdot\mathbb{P}(\|\widetilde{\lambda}-\lambda\|_\omega^2\geq A\Phi_m),$$

which by reduction to two hypotheses implies

$$\sup_{\lambda\in\Lambda_\gamma^r}\sup_{f\in\mathcal{F}_\alpha^d}\mathbb{E}[\Phi_m^{-1}\|\widetilde{\lambda}-\lambda\|_\omega^2]\geq A\sup_{\lambda\in\Lambda_\gamma^r}\sup_{f\in\mathcal{F}_\alpha^d}\mathbb{P}(\|\widetilde{\lambda}-\lambda\|_\omega^2\geq A\Phi_m)$$
$$\geq A\sup_{\theta\in\{\pm 1\}}\mathbb{P}_\theta(\|\widetilde{\lambda}-\lambda_\theta\|_\omega^2\geq A\Phi_m)$$

where $\mathbb{P}_\theta$ denotes the distribution when the true parameters are λ_θ and f_θ, respectively. The specific hypotheses λ_1, λ_{-1} and f_1, f_{-1} will be specified below. If λ_{-1} and λ_1 can be chosen such that $\|\lambda_1 - \lambda_{-1}\|_\omega^2 \geq 4A\Phi_m$, application of the triangle inequality yields

$$\mathbb{P}_\theta(\|\widetilde{\lambda} - \lambda_\theta\|_\omega^2 \geq A\Phi_m) \geq \mathbb{P}_\theta(\tau^* \neq \theta)$$

where τ^* denotes the *minimum distance test* defined through $\tau^* = \arg\min_{\theta\in\{\pm 1\}} \|\widetilde{\lambda} - \lambda_\theta\|_\omega^2$. Hence, we obtain

$$\begin{aligned}\inf_{\widetilde{\lambda}} \sup_{\lambda\in\Lambda_\gamma^r} \sup_{f\in\mathcal{F}_\alpha^d} \mathbb{P}(\|\widetilde{\lambda} - \lambda\|_\omega^2 \geq A\Phi_m) &\geq \inf_{\widetilde{\lambda}} \sup_{\theta\in\{\pm 1\}} \mathbb{P}_\theta(\|\widetilde{\lambda} - \lambda_\theta\|_\omega^2 \geq A\Phi_m) \\ &\geq \inf_{\tau} \sup_{\theta\in\{\pm 1\}} \mathbb{P}_\theta\left(\tau \neq \theta\right) \\ &=: p^* \end{aligned} \tag{4.17}$$

where the infimum is taken over all $\{\pm 1\}$-valued functions τ based on the observations. Thus, it remains to find hypotheses $\lambda_1, \lambda_{-1} \in \Lambda_\gamma^r$ and $f_1, f_{-1} \in \mathcal{F}_\alpha^d$ such that

$$\|\lambda_1 - \lambda_{-1}\|_\omega^2 \geq 4A\Phi_m, \tag{4.18}$$

and which allow us to bound p^* by a universal constant (independent of m) from below.

For this purpose, set $k_m^* := \arg\max_{j\in\mathbb{N}}\{\frac{\omega_j}{\gamma_j} \min(1, \frac{1}{m\alpha_j})\}$ and $a_m := \zeta \min(1, m^{-1/2}\alpha_{k_m^*}^{-1/2})$, where ζ is defined as in the statement of the theorem. Take note of the inequalities

$$1/d^{1/2} = (1 - (1 - 1/d^{1/4}))^2 \leq (1 - a_m)^2 \leq 1,$$

and

$$1 \leq (1 + a_m)^2 \leq (1 + (1 - 1/d^{1/4}))^2 = (2 - 1/d^{1/4})^2 \leq d^{1/2}$$

which in combination imply $1/d^{1/2} \leq (1 + \theta a_m)^2 \leq d^{1/2}$ for $\theta \in \{\pm 1\}$. These inequalities will be used below without further reference. For $\theta \in \{\pm 1\}$, we define

$$\lambda_\theta = \left(\frac{r}{2}\right)^{1/2} + (1 - \theta a_m)\ \left(\frac{r}{8}\right)^{1/2} d^{-1/4}\ \gamma_{k_m^*}^{-1/2}(\mathbf{e}_{k_m^*} + \mathbf{e}_{-k_m^*}).$$

Note that λ_θ is real-valued by definition. Furthermore, we have

$$\|\lambda_\theta\|_\gamma^2 = \frac{r}{2} + 2\gamma_{k_m^*}|[\lambda_\theta]_{k_m^*}|^2 \leq \frac{r}{2} + (1 + a_m)^2\frac{r}{4}d^{-1/2} \leq \frac{3r}{4},$$

and

$$|\lambda_\theta(t)| \geq \left(\frac{r}{2}\right)^{1/2} - 2\left(\frac{r}{8}\right)^{1/2} \geq 0 \qquad \forall t \in [0, 1),$$

which together imply that $\lambda_\theta \in \Lambda_\gamma^r$ for $\theta \in \{\pm 1\}$. The identity

$$\|\lambda_1 - \lambda_{-1}\|_\omega^2 = ra_m^2 d^{-1/2}\omega_{k_m^*}\gamma_{k_m^*}^{-1} = \zeta^2 r d^{-1/2} \cdot \Phi_m$$

shows that the condition in (4.18) is satisfied with $A = \zeta^2 r/(4\sqrt{d})$.

Let $f \in \mathcal{F}_\alpha^{\sqrt{d}}$ be such that $f \geq 1/2$ (the existence is guaranteed through condition (C3)) and define for $\theta \in \{\pm 1\}$

$$f_\theta = f + \theta a_m([f]_{k_m^*}\mathbf{e}_{k_m^*} + [f]_{-k_m^*}\mathbf{e}_{-k_m^*}).$$

Since $k_m^* \geq 1$ we have $\int_0^1 f_\theta(x)dx = 1$ and $f_\theta \geq 0$ holds because of the estimate

$$|f_\theta(t)| \geq 1/2 - 2a_m\alpha_{k_m^*}^{1/2}d^{1/2} \geq 0 \qquad \text{for all } t \in [0, 1).$$

For $|j| \neq k_m^*$, we have $[f]_j = [f_\theta]_j$ and thus trivially $1/d \leq |[f_\theta]_j|^2/\alpha_j \leq d$ for $|j| \neq k_m^*$ since $\mathcal{F}_\alpha^{\sqrt{d}} \subseteq \mathcal{F}_\alpha^d$. Moreover

$$1/d \leq d^{-1/2}\frac{|[f]_{\pm k_m^*}|^2}{\alpha_{\pm k_m^*}} \leq \frac{(1+\theta a_m)^2|[f]_{\pm k_m^*}|^2}{\alpha_{\pm k_m^*}} \leq d^{1/2}\frac{|[f]_{\pm k_m^*}|^2}{\alpha_{\pm k_m^*}} \leq d$$

and hence $f_\theta \in \mathcal{F}_\alpha^d$ for $\theta \in \{\pm 1\}$.

To obtain a lower bound for p^* defined in (4.17) consider the joint distribution $\mathbb{P}_\theta$ of the samples $N_1, \ldots, N_n$ and $Y_1, \ldots, Y_m$ under λ_θ and f_θ. Note that due to our construction we have $\lambda_{-1} \star f_{-1} = \lambda_1 \star f_1$. Thus $\mathbb{P}_{-1}^{N_i} = \mathbb{P}_1^{N_i}$ for all $i = 1, \ldots, n$ (due to the fact that the distribution of a *Poisson* point process is determined by its intensity) and the Hellinger distance between $\mathbb{P}_{-1}$ and $\mathbb{P}_1$ does only depend on the distribution of the sample $Y_1, \ldots, Y_m$. More precisely,

$$H^2(\mathbb{P}_{-1}, \mathbb{P}_1) = H^2(\mathbb{P}_{-1}^{Y_1,\ldots,Y_m}, \mathbb{P}_1^{Y_1,\ldots,Y_m}) \leq mH^2(\mathbb{P}_{-1}^{Y_1}, \mathbb{P}_1^{Y_1}),$$

and we proceed by bounding $H^2(\mathbb{P}_{-1}^{Y_1}, \mathbb{P}_1^{Y_1})$ from above. Recall that $f \geq 1/2$ which is used to obtain the estimate

$$H^2(\mathbb{P}_{-1}^{Y_1}, \mathbb{P}_1^{Y_1}) = \int_0^1 \frac{|f_1(x) - f_{-1}(x)|^2}{2f(x)}dx \leq \int |f_1(x) - f_{-1}(x)|^2\, dx \leq 8da_m^2\alpha_{k_m^*} \leq \frac{1}{m}.$$

Hence we have $H^2(\mathbb{P}_{-1}, \mathbb{P}_1) \leq 1$ and application of statement (ii) of Theorem 2.2 in [Tsy08] with $\alpha = 1$ implies $p^* \geq \frac{1}{2}(1 - \sqrt{3}/2)$. □

For the proof of the theorem it was sufficient to construct two hypotheses which are statistically indistinguishable but generate the lower bound Φ_m. This is in notable contrast to the proof of Theorem 4.3 where we had to construct $2^{k_n^*+1}$ hypotheses. Condition (C3) has to be imposed in order to guarantee that the considered hypotheses f_θ, $\theta \in \{\pm 1\}$ belong to $\mathcal{F}_\alpha^d$. It is easy to check that this condition is satisfied if $\sum_{j\neq 0} \alpha_j^{1/2} \leq \frac{1}{2\sqrt{d}}$.

Remark 4.6. The stated proof is only valid in model 1 and cannot be transferred directly to model 2. In the proof given above, the identity $\lambda_{-1} \star f_{-1} = \lambda_1 \star f_1$ would only imply equality of the mean measures of the two Cox process hypotheses but not equality of their distributions. We conjecture that the lower bound in Corollary 4.7 is valid for model 2 as well. Unfortunately, we do not have a proof of this conjecture in our framework up to now. The article [Big+13] provides a formidable proof of a lower bound in case of a known error density with polynomially decaying Fourier coefficients when the smoothness class of the unknown intensity is a Besov ellipsoid (see Theorem 3.1 in [Big+13]).

The following corollary merges the results of Theorems 4.3 and 4.5.

Corollary 4.7. *Under the assumptions of Theorems 4.3 and 4.5, for any $n, m \in \mathbb{N}$,*

$$\inf_{\widetilde{\lambda}} \sup_{\lambda \in \Lambda_\gamma^r} \sup_{f \in \mathcal{F}_\alpha^d} \mathbb{E}[\|\widetilde{\lambda} - \lambda\|_\omega^2] \gtrsim \max\{\Psi_n, \Phi_m\}$$

where the infimum is taken over all estimators $\widetilde{\lambda}$ of λ based on the observations from (4.8) *under model 1.*

4.2.3. Examples of convergence rates

In order to flesh out the abstract results of this chapter, we consider special choices for the sequences ω, γ and α and state the resulting rates of convergence with respect to both sample sizes n and m. For the sequence ω, we will assume throughout that $\omega_0 = 1$ and $\omega_j = |j|^{2s}$ for $j \neq 0$. As

γ	α	$\Theta(\Psi_n)$	$\Theta(\Phi_m)$	Restrictions
(pol)	(pol)	$n^{-\frac{2(p-s)}{2p+2a+1}}$	$m^{-\frac{(p-s)\wedge a}{a}}$	$p \geq s$, $p > \frac{1}{2}$, $a > \frac{1}{2}$
(exp)	(pol)	$(\log n)^{2s+2a+1} \cdot n^{-1}$	m^{-1}	$a > \frac{1}{2}$
(pol)	(exp)	$(\log n)^{-2(p-s)}$	$(\log m)^{-2(p-s)}$	$p \geq s$, $p > \frac{1}{2}$
(exp)	(exp)	$(\log n)^{2s} \cdot n^{-\frac{p}{p+a}}$	$(\log m)^{2s} \cdot m^{-p/a}$ if $a \geq p$ m^{-1} if $a < p$	

Table 4.1.: Exemplary rates of convergence for non-parametric intensity estimation from indirect observations. The rates are given in the framework of Theorems 4.2, 4.3, and 4.5 which impose the given restrictions. In all the examples $\omega_0 = 1$, $\omega_j = |j|^{2s}$ for $j \neq 0$, whereas the choices (pol) and (exp) for the sequences γ and α are explained in Section 4.2.3.

argumented in [JS13a], the resulting weighted norm corresponds to the $\mathbb{L}^2$-norm of the s^{th} weak derivative.

Choices for the sequence γ: Concerning the sequence γ we distinguish the following two scenarios:

(pol): $\gamma_0 = 0$ and $\gamma_j = |j|^{2p}$ for all $j \neq 0$ and some $p \geq 0$. This corresponds to the case when the unknown intensity function belongs to some *Sobolev space*.

(exp): $\gamma_j = \exp(2p|j|)$ for all $j \in \mathbb{Z}$ and some $p \geq 0$. In this case, λ belongs to some space of *analytic functions* (see for instance [Cav08]).

Choices for the sequence α: Concerning the sequence α we consider the following scenarios:

(pol): $\alpha_0 = 0$ and $\alpha_j = |j|^{-2a}$ for all $j \neq 0$ and some $a > \frac{1}{2}$. This corresponds to the case when the error density is *ordinary smooth*.

(exp): $\alpha_j = \exp(-2a|j|)$ for all $j \in \mathbb{Z}$ and some $a \geq 0$.

Table 4.1 summarizes the rates Ψ_n and Φ_m corresponding to the different choices of γ and α. The rates with respect to n coincide with the classical rates for non-parametric inverse problems (see, for instance, Table 1 in [Cav08] where the error variance ε^2 corresponds to n^{-1} in our setup and only the case $s = 0$ is considered).

4.3. Adaptive estimation for model 1: PPP observations

The estimator considered in Theorem 4.2 is obtained by specializing the orthonormal series estimator in (4.9) with dimension parameter k_n^* defined in (4.10). Thus, this procedure suffers from the apparent drawback that it depends on the smoothness characteristics of both λ and f, namely on the sequences γ and α. Since such characteristics are typically unavailable in advance, there is need for an adaptive selection of the dimension parameter which does not require any *a priori* knowledge on λ and f. In order to reach such an adaptive definition under model 1 we follow the procedure proposed in [JS13a] and proceed in two steps. In the first step (treated in Section 4.3.1), we assume that the class Λ_γ^r is unknown but assume the class $\mathcal{F}_\alpha^d$ of potential error densities f to be known. This assumption allows us to define a *partially adaptive* choice $\widetilde{k}$ of k. In the second step (treated in Section 4.3.2), we dispense with any knowledge on the smoothness both of λ and f and propose a *fully data-driven* choice $\widehat{k}$ of the dimension parameter.

4.3.1. Partially adaptive estimation (Λ_γ^r unknown, $\mathcal{F}_\alpha^d$ known)

First, we aim at choosing k equal to some $\widetilde{k}$ that, in contrast to k_n^* in (4.10), does no longer depend on the sequence γ but only on the sequence α. For the definition of $\widetilde{k}$ some terminology has to be introduced: for any $k \in \mathbb{N}_0$, let

$$\Delta_k^\alpha := \max_{0 \le j \le k} \omega_j \alpha_j^{-1} \quad \text{and} \quad \delta_k^\alpha := (2k+1)\Delta_k^\alpha \frac{\log(\Delta_k^\alpha \vee (k+3))}{\log(k+3)}.$$

Put $\omega_j^+ := \max_{0 \le i \le |j|} \omega_i$, and for all $n, m \in \mathbb{N}$,

$$N_n^\alpha := \inf\left\{1 \le j \le n : \frac{\alpha_j}{2j+1} < \frac{\log(n+3)\omega_j^+}{n}\right\} - 1 \wedge n,$$
$$M_m^\alpha := \inf\{1 \le j \le m : \alpha_j < 640 d m^{-1} \log(m+1)\} - 1 \wedge m,$$

and set $K_{nm}^\alpha := N_n^\alpha \wedge M_m^\alpha$. Now, denoting $\langle s, t\rangle_\omega := \sum_{j\in\mathbb{Z}} \omega_j [s]_j \overline{[t]}_j$, define the contrast function

$$\Upsilon(t) := \|t\|_\omega^2 - 2\Re\langle \widehat{\lambda}_{n\wedge m}, t\rangle_\omega, \qquad t \in \mathbb{L}^2,$$

and define the random sequence of penalties $(\widetilde{\text{PEN}}_k)_{k\in\mathbb{N}_0}$ via

$$\widetilde{\text{PEN}}_k := \frac{165}{4} d\eta^{-1} \cdot ([\widehat{\ell}]_0 \vee 1) \cdot \frac{\delta_k^\alpha}{n}$$

where $\eta \in (0,1)$ is some additional tuning parameter. The parameter η finds its way into the upper risk bound only as a numerical constant and does not have any effect on the rate of convergence. The dependence of the adaptive estimator on the specific choice of η will be suppressed for the sake of convenience in the sequel. Building on our definition of contrast and penalty, we define the partially adaptive selection of the tuning parameter k as

$$\widetilde{k} := \operatorname*{argmin}_{0 \le k \le K_{nm}^\alpha} \{\Upsilon(\widehat{\lambda}_k) + \widetilde{\text{PEN}}_k\}.$$

The following theorem provides an upper bound for the partially adaptive estimator $\widehat{\lambda}_{\widetilde{k}}$.

THEOREM 4.8. *Let Assumption 4.1 hold. Then, for any $n, m \in \mathbb{N}$,*

$$\sup_{\lambda\in\Lambda_\gamma^r} \sup_{f\in\mathcal{F}_\alpha^d} \mathbb{E}[\|\widehat{\lambda}_{\widetilde{k}} - \lambda\|_\omega^2] \lesssim \min_{0\le k\le K_{nm}^\alpha} \max\left\{\frac{\omega_k}{\gamma_k}, \frac{\delta_k^\alpha}{n}\right\} + \Phi_m + \frac{1}{m} + \frac{1}{n}$$

where the observations in (4.8) *stem from model 1.*

PROOF. Define the events $\Xi_1 := \{\eta([\ell]_0 \vee 1) \le [\widehat{\ell}]_0 \vee 1 \le \eta^{-1}([\ell]_0 \vee 1)\}$ and

$$\Xi_2 := \left\{\forall\, 0 \le |j| \le M_m^\alpha : \left|\frac{1}{\widehat{[f]}_j} - \frac{1}{[f]_j}\right| \le \frac{1}{2|[f]_j|} \quad \text{and} \quad |\widehat{[f]}_j| \ge \frac{1}{m}\right\}.$$

The identity $1 = \mathbb{1}_{\Xi_1\cap\Xi_2} + \mathbb{1}_{\Xi_2^c} + \mathbb{1}_{\Xi_1^c\cap\Xi_2}$ provides the decomposition

$$\mathbb{E}[\|\widehat{\lambda}_{\widetilde{k}} - \lambda\|_\omega^2] = \underbrace{\mathbb{E}[\|\widehat{\lambda}_{\widetilde{k}} - \lambda\|_\omega^2 \mathbb{1}_{\Xi_1\cap\Xi_2}]}_{=:\Box_1} + \underbrace{\mathbb{E}[\|\widehat{\lambda}_{\widetilde{k}} - \lambda\|_\omega^2 \mathbb{1}_{\Xi_2^c}]}_{=:\Box_2} + \underbrace{\mathbb{E}[\|\widehat{\lambda}_{\widetilde{k}} - \lambda\|_\omega^2 \mathbb{1}_{\Xi_1^c\cap\Xi_2}]}_{=:\Box_3},$$

and we will establish uniform upper bounds over the ellipsoids Λ_γ^r and $\mathcal{F}_\alpha^d$ for the three terms on the right-hand side separately.

Uniform upper bound for $\Box_1$: Denote by $\mathcal{S}_k$ the linear subspace of $\mathbb{L}^2$ spanned by the functions $\mathbf{e}_j(\cdot)$ for $j \in \{-k, \dots, k\}$. Since the identity $\Upsilon(t) = \|t - \widehat{\lambda}_k\|_\omega^2 - \|\widehat{\lambda}_k\|_\omega^2$ holds for all $t \in \mathcal{S}_k$, $k \in \{0, \dots, n \wedge m\}$, we obtain for all such k that $\operatorname{argmin}_{t\in\mathcal{S}_k} \Upsilon(t) = \widehat{\lambda}_k$. Using this identity and the definition of $\widetilde{k}$ yields for all $k \in \{0, \dots, K^{\alpha}_{nm}\}$ that

$$\Upsilon(\widehat{\lambda}_{\widetilde{k}}) + \widetilde{\text{PEN}}_{\widetilde{k}} \le \Upsilon(\widehat{\lambda}_k) + \widetilde{\text{PEN}}_k \le \Upsilon(\lambda_k) + \widetilde{\text{PEN}}_k$$

where $\lambda_k := \sum_{0 \le |j| \le k} [\lambda]_j \mathbf{e}_j$ denotes the projection of λ on the subspace $\mathcal{S}_k$. Elementary computations imply

$$\|\widehat{\lambda}_{\widetilde{k}}\|^2_\omega \le \|\lambda_k\|^2_\omega + 2\Re\langle \widehat{\lambda}_{n\wedge m}, \widehat{\lambda}_{\widetilde{k}} - \lambda_k \rangle_\omega + \widetilde{\text{PEN}}_k - \widetilde{\text{PEN}}_{\widetilde{k}} \tag{4.19}$$

for all $k \in \{0, \dots, K^{\alpha}_{nm}\}$. In addition to λ_k defined above, introduce the further abbreviations

$$\widetilde{\lambda}_k := \sum_{0 \le |j| \le k} \frac{[\widehat{\ell}]_j}{[f]_j} \mathbf{e}_j \qquad \text{and} \qquad \check{\lambda}_k := \sum_{0 \le |j| \le k} \frac{[\ell]_j}{[\widehat{f}]_j} \mathbb{1}_{\Omega_j} \mathbf{e}_j,$$

as well as

$$\Theta_k := \widehat{\lambda}_k - \check{\lambda}_k - \widetilde{\lambda}_k + \lambda_k, \quad \widetilde{\Theta}_k := \widetilde{\lambda}_k - \lambda_k, \quad \text{and} \quad \check{\Theta}_k := \check{\lambda}_k - \lambda_k.$$

Using these abbrevations and the identity $\widehat{\lambda}_{n\wedge m} - \lambda_{n\wedge m} = \Theta_{n\wedge m} + \widetilde{\Theta}_{n\wedge m} + \check{\Theta}_{n\wedge m}$, we deduce from (4.19) that

$$\begin{aligned} \|\widehat{\lambda}_{\widetilde{k}} - \lambda\|^2_\omega \le \|\lambda - \lambda_k\|^2_\omega + \widetilde{\text{PEN}}_k - \widetilde{\text{PEN}}_{\widetilde{k}} + 2\Re\langle \widetilde{\Theta}_{n\wedge m}, \widehat{\lambda}_{\widetilde{k}} - \lambda_k \rangle_\omega \\ + 2\Re\langle \Theta_{n\wedge m}, \widehat{\lambda}_{\widetilde{k}} - \lambda_k \rangle_\omega + 2\Re\langle \check{\Theta}_{n\wedge m}, \widehat{\lambda}_{\widetilde{k}} - \lambda_k \rangle_\omega \end{aligned} \tag{4.20}$$

for all $k \in \{0, \dots, K^{\alpha}_{nm}\}$. Define $\mathcal{B}_k := \{\lambda \in \mathcal{S}_k : \|\lambda\|_\omega \le 1\}$. For every $\tau > 0$ and $h \in \mathcal{S}_{n\wedge m}, t \in \mathcal{S}_k$, the estimate $2uv \le \tau u^2 + \tau^{-1} v^2$ implies

$$2\left|\langle h, \widetilde{t} \rangle_\omega\right| \le 2\|\widetilde{t}\|_\omega \sup_{t \in \mathcal{B}_k} |\langle h, t \rangle_\omega| \le \tau \|\widetilde{t}\|^2_\omega + \tau^{-1} \sup_{t \in \mathcal{B}_k} |\langle h, t \rangle_\omega|^2.$$

Because $\widehat{\lambda}_{\widetilde{k}} - \lambda_k \in \mathcal{S}_{\widetilde{k} \vee k}$, combining the last estimate with (4.20) we get

$$\begin{aligned} \|\widehat{\lambda}_{\widetilde{k}} - \lambda\|^2_\omega \le \|\lambda - \lambda_k\|^2_\omega + 3\tau \|\widehat{\lambda}_{\widetilde{k}} - \lambda_k\|^2_\omega + \widetilde{\text{PEN}}_k - \widetilde{\text{PEN}}_{\widetilde{k}} + \\ + \tau^{-1} \sup_{t \in \mathcal{B}_{k \vee \widetilde{k}}} |\langle \widetilde{\Theta}_{n\wedge m}, t \rangle_\omega|^2 + \tau^{-1} \sup_{t \in \mathcal{B}_{k \vee \widetilde{k}}} |\langle \Theta_{n\wedge m}, t \rangle_\omega|^2 + \tau^{-1} \sup_{t \in \mathcal{B}_{k \vee \widetilde{k}}} |\langle \check{\Theta}_{n\wedge m}, t \rangle_\omega|^2. \end{aligned}$$

Note that $\|\widehat{\lambda}_{\widetilde{k}} - \lambda_k\|^2_\omega \le 2\|\widehat{\lambda}_{\widetilde{k}} - \lambda\|^2_\omega + 2\|\lambda_k - \lambda\|^2_\omega$ and $\|\lambda - \lambda_k\|^2_\omega \le r\omega_k \gamma_k^{-1}$ for all $\lambda \in \Lambda^r_\gamma$ since $(\omega_n \gamma_n^{-1})_{n \in \mathbb{N}_0}$ is non-increasing due to Assumption 4.1. Specializing with $\tau = 1/8$ we obtain

$$\begin{aligned} \|\widehat{\lambda}_{\widetilde{k}} - \lambda\|^2_\omega \le 7r\omega_k \gamma_k^{-1} + 4\widetilde{\text{PEN}}_k - 4\widetilde{\text{PEN}}_{\widetilde{k}} + 32 \sup_{t \in \mathcal{B}_{k \vee \widetilde{k}}} |\langle \widetilde{\Theta}_{n\wedge m}, t \rangle_\omega|^2 \\ + 32 \sup_{t \in \mathcal{B}_{k \vee \widetilde{k}}} |\langle \Theta_{n\wedge m}, t \rangle_\omega|^2 + 32 \sup_{t \in \mathcal{B}_{k \vee \widetilde{k}}} |\langle \check{\Theta}_{n\wedge m}, t \rangle_\omega|^2. \end{aligned} \tag{4.21}$$

Combining the facts that $\mathbb{1}_{\Omega_j} \mathbb{1}_{\Xi_2} = \mathbb{1}_{\Xi_2}$ for $0 \le |j| \le M^{\alpha}_m$ and $K^{\alpha}_{nm} \le M^{\alpha}_m$ by definition, we obtain for all $0 \le |j| \le K^{\alpha}_{nm}$ the estimate

$$|[f]_j/[\widehat{f}]_j \mathbb{1}_{\Omega_j} - 1|^2 \mathbb{1}_{\Xi_2} = |[f]_j|^2 \cdot |1/[\widehat{f}]_j - 1/[f]_j|^2 \mathbb{1}_{\Xi_2} \le 1/4.$$

Hence, $\sup_{t \in \mathcal{B}_k} |\langle \Theta_{n\wedge m}, t \rangle_\omega|^2 \mathbb{1}_{\Xi_2} \le \frac{1}{4} \sup_{t \in \mathcal{B}_k} |\langle \widetilde{\Theta}_{n\wedge m}, t \rangle_\omega|^2$ for all $0 \le k \le K^{\alpha}_{nm}$. Thus, from (4.21) we obtain

$$\|\widehat{\lambda}_{\widetilde{k}} - \lambda\|_\omega^2 \mathbb{1}_{\Xi_1\cap\Xi_2} \leq 7r\omega_k\gamma_{\widetilde{k}}^{-1} + 40\left(\sup_{t\in\mathcal{B}_{k\vee\widetilde{k}}} |\langle\widetilde{\Theta}_{n\wedge m}, t\rangle_\omega|^2 - \frac{33d([\ell]_0\vee 1)\delta^\alpha_{k\vee\widetilde{k}}}{8n}\right)_+ \\ + (165d([\ell]_0\vee 1)\delta^\alpha_{k\vee\widetilde{k}}/n + 4\widetilde{\mathrm{PEN}}_k - 4\widetilde{\mathrm{PEN}}_{\widetilde{k}})\mathbb{1}_{\Xi_1\cap\Xi_2} + 32\sup_{t\in\mathcal{B}_{K^\alpha_{nm}}} |\langle\check{\Theta}_{n\wedge m}, t\rangle_\omega|^2.$$

Exploiting the definition of both the penalty $\widetilde{\mathrm{PEN}}$ and the event Ξ_1, we obtain

$$\mathbb{E}[\|\widehat{\lambda}_{\widetilde{k}} - \lambda\|_\omega^2 \mathbb{1}_{\Xi_1\cap\Xi_2}] \leq C(d,r) \min_{0\leq k\leq K^\alpha_{nm}} \max\left\{\frac{\omega_k}{\gamma_k}, \frac{\delta^\alpha_k}{n}\right\} \\ + 40\sum_{k=0}^{K^\alpha_{nm}} \mathbb{E}\left[\left(\sup_{t\in\mathcal{B}_k} |\langle\widetilde{\Theta}_{n\wedge m}, t\rangle_\omega|^2 - \frac{33([\ell]_0\vee 1)d\delta^\alpha_k}{8n}\right)_+\right] \\ + 32\mathbb{E}\left[\sup_{t\in\mathcal{B}_{K^\alpha_{nm}}} |\langle\check{\Theta}_{n\wedge m}, t\rangle_\omega|^2\right]. \tag{4.22}$$

Applying Lemma 4.17 with $\delta^*_k = d\delta^\alpha_k$ and $\Delta^*_k = d\Delta^\alpha_k$ yields

$$\mathbb{E}\left[\left(\sup_{t\in\mathcal{B}_k} |\langle\widetilde{\Theta}_{n\wedge m}, t\rangle_\omega|^2 - \frac{33d([\ell]_0\vee 1)\delta^\alpha_k}{8n}\right)_+\right] \leq K_1\left[\frac{d\|f\|\|\lambda\|\Delta^\alpha_k}{n}\exp\left(-K_2\frac{\delta^\alpha_k}{\|f\|^2\|\lambda\|^2\Delta^\alpha_k}\right) \right. \\ \left. + \quad \frac{d\delta^\alpha_k}{n^2}\exp(-K_3\sqrt{n})\right].$$

Using statement a) of Lemma 4.16 and the fact that $K^\alpha_{nm} \leq n$ by definition, we obtain that

$$\sum_{k=0}^{K^\alpha_{nm}} \mathbb{E}\left[\left(\sup_{t\in\mathcal{B}_k} |\langle\widetilde{\Theta}_{n\wedge m}, t\rangle_\omega| - \frac{33d([\ell]_0\vee 1)\delta^\alpha_k}{8n}\right)_+\right] \\ \lesssim \frac{d^{3/2}\sqrt{r\rho}}{n}\sum_{k=0}^{\infty}\Delta^\alpha_k \exp\left(-\frac{2K_2k}{\sqrt{dr\rho}}\cdot\frac{\log(\Delta^\alpha_k\vee(k+3))}{\log(k+3)}\right) + \exp(-K_3\sqrt{n})$$

where the last estimate is due to the fact that $\|f\|^2 \leq d\rho$ for all $f \in \mathcal{F}^d_\alpha$ and $\|\lambda\|^2 \leq r$ for all $\lambda \in \Lambda^r_\gamma$. Note that we have

$$\sum_{k=0}^{\infty}\Delta^\alpha_k \exp\left(-\frac{2K_2k}{\sqrt{dr\rho}}\cdot\frac{\log(\Delta^\alpha_k\vee(k+3))}{\log(k+3)}\right) \leq C < \infty$$

with a numerical constant C which implies

$$\sum_{k=0}^{K^\alpha_{nm}} \mathbb{E}\left[\left(\sup_{t\in\mathcal{B}_k} |\langle\widetilde{\Theta}_{n\wedge m}, t\rangle_\omega| - \frac{33d([\ell]_0\vee 1)\delta^\alpha_k}{8n}\right)_+\right] \lesssim \frac{1}{n}.$$

The last term on the right-hand side of (4.22) is bounded by means of Lemma 4.18 which immediately yields

$$\mathbb{E}\left[\sup_{t\in\mathcal{B}_{K^\alpha_{nm}}} |\langle\check{\Theta}_{n\wedge m}, t\rangle_\omega|^2\right] \lesssim \Phi_m.$$

Combining the preceeding estimates, which hold uniformly for all $\lambda \in \Lambda^r_\gamma$ and $f \in \mathcal{F}^d_\alpha$, we conclude from Equation (4.22) that

$$\sup_{\lambda\in\Lambda_\gamma^r}\sup_{f\in\mathcal{F}_\alpha^d}\mathbb{E}[\|\widehat{\lambda}_{\widetilde{k}}-\lambda\|_\omega^2\,\mathbb{1}_{\Xi_1\cap\Xi_2}]\lesssim\min_{0\le k\le K_{nm}^\alpha}\max\left\{\frac{\omega_k}{\gamma_k},\frac{\delta_k^\alpha}{n}\right\}+\Phi_m+\frac{1}{n}.$$

Uniform upper bound for $\square_2$: Define $\breve{\lambda}_k:=\sum_{0\le|j|\le k}[\lambda]_j\mathbb{1}_{\Omega_j}\mathbf{e}_j$. Note that $\|\widehat{\lambda}_k-\breve{\lambda}_k\|_\omega^2\le\|\widehat{\lambda}_{k'}-\breve{\lambda}_{k'}\|_\omega^2$ for $k\le k'$ and $\|\breve{\lambda}_k-\lambda\|_\omega^2\le\|\lambda\|_\omega^2$ for all $k\in\mathbb{N}_0$. Consequently, since $0\le\widetilde{k}\le K_{nm}^\alpha$, we obtain the estimate

$$\begin{aligned}\mathbb{E}[\|\widehat{\lambda}_{\widetilde{k}}-\lambda\|_\omega^2\,\mathbb{1}_{\Xi_2^c}]&\le 2\mathbb{E}[\|\widehat{\lambda}_{\widetilde{k}}-\breve{\lambda}_{\widetilde{k}}\|_\omega^2\,\mathbb{1}_{\Xi_2^c}]+2\,\mathbb{E}[\|\breve{\lambda}_{\widetilde{k}}-\lambda\|_\omega^2\mathbb{1}_{\Xi_2^c}]\\&\le 2\mathbb{E}[\|\widehat{\lambda}_{K_{nm}^\alpha}-\breve{\lambda}_{K_{nm}^\alpha}\|_\omega^2\,\mathbb{1}_{\Xi_2^c}]+2\,\|\lambda\|_\omega^2\,\mathbb{P}(\Xi_2^c),\end{aligned}$$

and due to Assumption 4.1 and Lemma 4.20 it is easily seen that $\|\lambda\|_\omega^2\cdot\mathbb{P}(\Xi_2^c)\lesssim m^{-4}$. Using the definition of Ω_j, we further obtain

$$\begin{aligned}\mathbb{E}[\|\widehat{\lambda}_{K_{nm}^\alpha}-\breve{\lambda}_{K_{nm}^\alpha}\|_\omega^2\,\mathbb{1}_{\Xi_2^c}]&\le 2m\sum_{0\le|j|\le K_{nm}^\alpha}\omega_j\{\mathbb{E}[|[\widehat{\ell}]_j-[\ell]_j|^2\,\mathbb{1}_{\Xi_2^c}]+\mathbb{E}[|[f]_j[\lambda]_j-[\widehat{f}]_j[\lambda]_j|^2\,\mathbb{1}_{\Xi_2^c}]\}\\&\le 2m\sum_{0\le|j|\le K_{nm}^\alpha}\omega_j(\mathbb{E}[|[\widehat{\ell}]_j-[\ell]_j|^4])^{1/2}\mathbb{P}(\Xi_2^c)^{1/2}\\&\quad+2m\sum_{0\le|j|\le K_{nm}^\alpha}\omega_j\,|[\lambda]_j|^2\,(\mathbb{E}[|[\widehat{f}]_j-[f]_j|^4])^{1/2}\mathbb{P}\,(\Xi_2^c)^{1/2}\\&\lesssim m\mathbb{P}(\Xi_2^c)^{1/2}\sum_{0\le|j|\le K_{nm}^\alpha}\frac{\omega_j}{n}+\mathbb{P}(\Xi_2^c)^{1/2}\sum_{0\le|j|\le K_{nm}^\alpha}\omega_j|[\lambda]_j|^2\end{aligned}\tag{4.23}$$

where the last estimate follows by applying Theorem B.1 with $p=4$ two times. If $K_{nm}^\alpha=0$, Lemma 4.20 implies

$$\mathbb{E}[\|\widehat{\lambda}_{K_{nm}^\alpha}-\breve{\lambda}_{K_{nm}^\alpha}\|_\omega^2\,\mathbb{1}_{\Xi_2^c}]\lesssim\frac{1}{nm}+\frac{1}{m^2}.$$

Otherwise, if $K_{nm}^\alpha>0$, we exploit $\omega_j\le\omega_j^+\alpha_j^{-1}$, $K_{nm}^\alpha\le N_n^\alpha$ and the definition of N_n^α to bound the first term on the right-hand side of (4.23). The second term on the right-hand side of (4.23) can be bounded from above by noting that $\omega_j\le\gamma_j$ thanks to Assumption 4.1. We obtain

$$\mathbb{E}[\|\widehat{\lambda}_{K_{nm}^\alpha}-\breve{\lambda}_{K_{nm}^\alpha}\|_\omega^2\,\mathbb{1}_{\Xi_2^c}]\lesssim m\mathbb{P}(\Xi_2^c)^{1/2}\left(\sum_{0\le|j|\le N_n^\alpha}\frac{1}{2|j|+1}\right)\frac{1}{\log(n+3)}+\mathbb{P}(\Xi_2^c)^{1/2}.$$

Thanks to the logarithmic increase of the harmonic series, $N_n^\alpha\le n$ and Lemma 4.20, the last estimate implies

$$\mathbb{E}[\|\widehat{\lambda}_{K_{nm}^\alpha}-\breve{\lambda}_{K_{nm}^\alpha}\|_\omega^2\,\mathbb{1}_{\Xi_2^c}]\lesssim\frac{1}{m}+\frac{1}{m^2},$$

if $K_{nm}^\alpha>0$, and thus

$$\mathbb{E}[\|\widehat{\lambda}_{K_{nm}^\alpha}-\breve{\lambda}_{K_{nm}^\alpha}\|_\omega^2\,\mathbb{1}_{\Xi_2^c}]\lesssim\frac{1}{m}+\frac{1}{m^2},$$

independent of the actual value of K_{nm}^α. Using the obtained estimates, which hold uniformly for $\lambda\in\Lambda_\gamma^r$ and $f\in\mathcal{F}_\alpha^d$, we conclude

$$\sup_{\lambda\in\Lambda_\gamma^r}\sup_{f\in\mathcal{F}_\alpha^d}\mathbb{E}[\|\widehat{\lambda}_{\widetilde{k}}-\lambda\|_\omega^2\,\mathbb{1}_{\Xi_2^c}]\lesssim\frac{1}{m}.$$

Uniform upper bound for $\square_3$: In order to find a uniform upper bound for $\square_3$, first recall the definition $\check{\lambda}_k := \sum_{0\le|j|\le k}[\lambda]_j \mathbb{1}_{\Omega_j}\mathbf{e}_j$, and consider the estimate

$$\mathbb{E}[\|\widehat{\lambda}_{\widetilde{k}} - \lambda\|_\omega^2 \mathbb{1}_{\Xi_1^c\cap\Xi_2}] \le 2\mathbb{E}[\|\widehat{\lambda}_{\widetilde{k}} - \check{\lambda}_{\widetilde{k}}\|_\omega^2 \mathbb{1}_{\Xi_1^c\cap\Xi_2}] + 2\mathbb{E}[\|\check{\lambda}_{\widetilde{k}} - \lambda\|_\omega^2 \mathbb{1}_{\Xi_1^c\cap\Xi_2}]. \tag{4.24}$$

Using the estimate $\|\check{\lambda}_{\widetilde{k}} - \lambda\|_\omega^2 \le \|\lambda\|_\omega^2$, we obtain for $\lambda\in\Lambda_\gamma^r$ by means of Lemma 4.19 that

$$\mathbb{E}[\|\check{\lambda}_{\widetilde{k}} - \lambda\|_\omega^2 \mathbb{1}_{\Xi_1^c\cap\Xi_2}] \le r\mathbb{P}(\Xi_1^c) \lesssim \frac{1}{n}$$

which controls the second term on the right-hand side of (4.24). We now bound the first term on the right-hand side of (4.24). If $K_{nm}^\alpha = 0$, we have $\widetilde{k}=0$, and by means of the Cauchy-Schwarz inequality and Theorem B.1 it is easily seen that

$$\mathbb{E}[\|\widehat{\lambda}_{\widetilde{k}} - \check{\lambda}_{\widetilde{k}}\|_\omega^2 \mathbb{1}_{\Xi_1^c\cap\Xi_2}] \lesssim \frac{1}{n}.$$

Otherwise, $K_{nm}^\alpha > 0$, and we need the following further estimate which is easily verified:

$$\begin{aligned}\mathbb{E}[\|\widehat{\lambda}_{\widetilde{k}} - \check{\lambda}_{\widetilde{k}}\|_\omega^2 \mathbb{1}_{\Xi_1^c\cap\Xi_2}] \le 3 &\sum_{0\le|j|\le K_{nm}^\alpha} \omega_j \mathbb{E}[|[\ell]_j/\widehat{[f]}_j - [\ell]_j/[f]_j|^2 \mathbb{1}_{\Xi_1^c\cap\Xi_2}]\\ &+ 3\sum_{0\le|j|\le K_{nm}^\alpha} \omega_j \mathbb{E}[|\widehat{[\ell]}_j - [\ell]_j|^2/|[f]_j|^2 \mathbb{1}_{\Xi_1^c\cap\Xi_2}]\\ &+ 3\sum_{0\le|j|\le K_{nm}^\alpha} \omega_j \mathbb{E}[|\widehat{[\ell]}_j - [\ell]_j|^2\cdot|1/\widehat{[f]}_j - 1/[f]_j|^2 \mathbb{1}_{\Xi_1^c\cap\Xi_2}].\end{aligned} \tag{4.25}$$

We start by bounding the first term on the right-hand side of (4.25). Using the definition of Ξ_2 and $\omega_j \le \gamma_j$, we obtain for all $\lambda\in\Lambda_\gamma^r$ that

$$\sum_{0\le|j|\le K_{nm}^\alpha} \omega_j \mathbb{E}[|[\ell]_j/\widehat{[f]}_j - [\ell]_j/[f]_j|^2 \mathbb{1}_{\Xi_1^c\cap\Xi_2}] \le \frac{r}{4}\cdot\mathbb{P}(\Xi_1^c) \lesssim \frac{1}{n}.$$

Since $|[f]_j|^{-2} \le d\alpha_j$ for $f\in\mathcal{F}_\alpha^d$, the Cauchy-Schwarz inequality in combination with Theorem B.1 implies for the second term on the right-hand side of (4.25) that

$$\sum_{0\le|j|\le K_{nm}^\alpha} \omega_j \mathbb{E}[|\widehat{[\ell]}_j - [\ell]_j|^2/|[f]_j|^2 \mathbb{1}_{\Xi_1^c\cap\Xi_2}] \lesssim \mathbb{P}(\Xi_1^c)^{1/2} \sum_{0\le|j|\le K_{nm}^\alpha} \frac{\omega_j^+}{n\alpha_j}.$$

We exploit the definition of N_n^α together with $K_{nm}^\alpha \le N_n^\alpha$ in order to obtain

$$\sum_{0\le|j|\le K_{nm}^\alpha} \omega_j \mathbb{E}[|\widehat{[\ell]}_j - [\ell]_j|^2/|[f]_j|^2 \mathbb{1}_{\Xi_1^c\cap\Xi_2}] \lesssim \frac{\mathbb{P}(\Xi_1^c)^{1/2}}{\log(n+3)} \sum_{0\le|j|\le N_n^\alpha} \frac{1}{2|j|+1}$$

from which by the logarithmic growth of the harmonic series and Lemma 4.19 we can conclude that

$$\sum_{0\le|j|\le K_{nm}^\alpha} \omega_j \mathbb{E}[|\widehat{[\ell]}_j - [\ell]_j|^2/|[f]_j|^2 \mathbb{1}_{\Xi_1^c\cap\Xi_2}] \lesssim \frac{1}{n},$$

independent of the actual value of K_{nm}^α. Finally, the third and last term on the right-hand side of (4.25) can be bounded from above the same way after exploiting the definition of Ξ_2, and we obtain

$$\sum_{0\le|j|\le K_{nm}^\alpha} \omega_j \mathbb{E}[|\widehat{[\ell]}_j - [\ell]_j|^2\cdot|1/\widehat{[f]}_j - 1/[f]_j|^2 \mathbb{1}_{\Xi_1^c\cap\Xi_2}] \lesssim \frac{1}{n}.$$

Putting together the derived estimates, that again hold uniformly for all $\lambda\in\Lambda_\gamma^r$ and $f\in\mathcal{F}_\alpha^d$, we obtain

$$\sup_{\lambda \in \Lambda_\gamma^r} \sup_{f \in \mathcal{F}_\alpha^d} \mathbb{E}[\|\widehat{\lambda}_{\widetilde{k}} - \lambda\|_\omega^2 \mathbb{1}_{\Xi_1^c \cap \Xi_2}] \lesssim \frac{1}{n}.$$

Finally, the statement of the theorem follows by combining the obtained uniform upper bounds for $\square_1$, $\square_2$, and $\square_3$. $\square$

4.3.2. Fully adaptive estimation (Λ_γ^r and $\mathcal{F}_\alpha^d$ unknown)

We now also dispense with the knowledge of the smoothness of the error density f and propose an adaptive choice $\widehat{k}$ of the dimension parameter such that the resulting estimator $\widehat{\lambda}_{\widehat{k}}$ adapts to the unknown smoothness of both λ and f and attains the optimal rate of convergence in a variety of scenarios. As in the case of partially adaptive estimation, we have to introduce some notation first. For $k \in \mathbb{N}_0$, let

$$\widehat{\Delta}_k := \max_{0 \le j \le k} \frac{\omega_j}{|\widehat{[f]}_j|^2} \mathbb{1}_{\Omega_j} \qquad \text{and} \qquad \widehat{\delta}_k := (2k+1)\widehat{\Delta}_k \frac{\log(\widehat{\Delta}_k \vee (k+4))}{\log(k+4)}.$$

For $n, m \in \mathbb{N}$, set

$$\widehat{N}_n := \inf\{1 \le j \le n : |\widehat{[f]}_j|^2/(2j+1) < \log(n+4)\omega_j^+/n\} - 1 \wedge n,$$
$$\widehat{M}_m := \inf\{1 \le j \le m : |\widehat{[f]}_j|^2 < m^{-1}\log m\} - 1 \wedge m,$$

and $\widehat{K}_{nm} := \widehat{N}_n \wedge \widehat{M}_m$. We consider the same contrast function as in the partially adaptive case but define the random sequence $(\widehat{\mathrm{PEN}}_k)_{k \in \mathbb{N}_0}$ of penalities now by

$$\widehat{\mathrm{PEN}}_k := 1375\eta^{-1} \cdot ([\widehat{\ell}]_0 \vee 1) \cdot \frac{\widehat{\delta}_k}{n}.$$

Note that this definition does not depend on the knowledge of the sequence α. Using this definition of a completely data-driven penalty, we define the fully adaptive selection $\widehat{k}$ of the dimension parameter k by means of

$$\widehat{k} := \operatorname*{argmin}_{0 \le k \le \widehat{K}_{nm}} \{\Upsilon(\widehat{\lambda}_k) + \widehat{\mathrm{PEN}}_k\}.$$

In order to state and prove the upper risk bound of the estimator $\widehat{\lambda}_{\widehat{k}}$, we have to introduce some further notation. We keep the definition of Δ_k^α from Section 4.3.1 but slightly redefine δ_k^α as

$$\delta_k^\alpha := (2k+1)\Delta_k^\alpha \frac{\log(\Delta_k^\alpha \vee (k+4))}{\log(k+4)}.$$

For $k \in \mathbb{N}_0$, we also define

$$\Delta_k := \max_{0 \le j \le k} \omega_j / |[f]_j|^2 \qquad \text{and} \qquad \delta_k := (2k+1)\Delta_k \frac{\log(\Delta_k \vee (k+4))}{\log(k+4)},$$

which can be regarded as analogues of Δ_k^α and δ_k^α in Section 4.3.1 in the case of a known error density f. Finally, for $n, m \in \mathbb{N}$, define

$$N_n^{\alpha-} := \inf\{1 \le j \le n : \alpha_j/(2j+1) < 4d\log(n+4)\omega_j^+/n\} - 1 \wedge n,$$
$$N_n^{\alpha+} := \inf\{1 \le j \le n : \alpha_j/(2j+1) < \log(n+4)\omega_j^+/(4dn)\} - 1 \wedge n,$$
$$M_m^{\alpha-} := \inf\{1 \le j \le m : \alpha_j < 4dm^{-1}\log m\} - 1 \wedge m,$$
$$M_m^{\alpha+} := \inf\{1 \le j \le m : 4d\alpha_j < m^{-1}\log m\} - 1 \wedge m,$$

and set $K_{nm}^{\alpha-} := N_n^{\alpha-} \wedge M_m^{\alpha-}$, $K_{nm}^{\alpha+} := N_n^{\alpha+} \wedge M_m^{\alpha+}$. In contrast to the proof of Theorem 4.8 we have to impose an additional assumption for the proof of an upper risk bound of $\widehat{\lambda}_{\widehat{k}}$:

Assumption 4.9. $\exp(-m\alpha_{M_m^{\alpha+}+1}/(128d)) \leq C(\alpha, d)m^{-5}$ *for all* $m \in \mathbb{N}$.

Theorem 4.10. *Let Assumptions 4.1 and 4.9 hold. Then, for any* $n, m \in \mathbb{N}$,

$$\sup_{\lambda \in \Lambda_\gamma^r} \sup_{f \in \mathcal{F}_\alpha^d} \mathbb{E}[\|\widehat{\lambda}_{\widehat{k}} - \lambda\|_\omega^2] \lesssim \min_{0 \leq k \leq K_{nm}^{\alpha-}} \max\left\{\frac{\omega_k}{\gamma_k}, \frac{\delta_k^\alpha}{n}\right\} + \Phi_m + \frac{1}{m} + \frac{1}{n}$$

where the observations in (4.8) *stem from model 1.*

Proof. Consider the event

$$\Xi_3 := \{N_n^{\alpha-} \wedge M_m^{\alpha-} \leq \widehat{K}_{nm} \leq N_n^{\alpha+} \wedge M_m^{\alpha+}\} \tag{4.26}$$

in addition to the event Ξ_1 introduced in the proof of Theorem 4.8 and the slightly redefined event Ξ_2 defined as

$$\Xi_2 := \{\forall 0 \leq |j| \leq M_m^{\alpha+} : |1/\widehat{[f]}_j - 1/[f]_j| \leq 1/(2|[f]_j|) \text{ and } |\widehat{[f]}_j| \geq 1/m\}.$$

Defining $\Xi := \Xi_1 \cap \Xi_2 \cap \Xi_3$, the identity $1 = \mathbb{1}_\Xi + \mathbb{1}_{\Xi_2^c} + \mathbb{1}_{\Xi_1^c \cap \Xi_2} + \mathbb{1}_{\Xi_1 \cap \Xi_2 \cap \Xi_3^c}$ motivates the decomposition

$$\begin{aligned}\mathbb{E}[\|\widehat{\lambda}_{\widehat{k}} - \lambda\|_\omega^2] = \mathbb{E}[\|\widehat{\lambda}_{\widehat{k}} - \lambda\|_\omega^2 \mathbb{1}_\Xi] + \mathbb{E}[\|\widehat{\lambda}_{\widehat{k}} - \lambda\|_\omega^2 \mathbb{1}_{\Xi_2^c}]& \\ + \mathbb{E}[\|\widehat{\lambda}_{\widehat{k}} - \lambda\|_\omega^2 \mathbb{1}_{\Xi_1^c \cap \Xi_2}] + \mathbb{E}[\|\widehat{\lambda}_{\widehat{k}} - \lambda\|_\omega^2 \mathbb{1}_{\Xi_1 \cap \Xi_2 \cap \Xi_3^c}]& \\ =: \square_1 + \square_2 + \square_3 + \square_4,&\end{aligned}$$

and we establish uniform upper risk bounds for the four terms on the right-hand side separately.

Uniform upper bound for $\square_1$: On Ξ we have the estimate $\frac{1}{4}\Delta_k \leq \widehat{\Delta}_k \leq \frac{9}{4}\Delta_k$, and thus

$$\frac{1}{4}[\Delta_k \vee (k+4)] \leq \widehat{\Delta}_k \vee (k+4) \leq \frac{9}{4}[\Delta_k \vee (k+4)]$$

for all $k \in \{0, \dots, M_m^{\alpha+}\}$. This last estimate implies

$$\begin{aligned}\frac{2k+1}{4}\Delta_k \frac{\log(\Delta_k \vee (k+4))}{\log(k+4)}\left(1 - \frac{\log 4}{\log(k+4)} \frac{\log(k+4)}{\log(\Delta_k \vee (k+4))}\right) \leq \widehat{\delta}_k& \\ \leq \frac{9(2k+1)}{4}\Delta_k \frac{\log(\Delta_k \vee (k+4))}{\log(k+4)}\left(1 + \frac{\log(9/4)}{\log(k+4)} \frac{\log(k+4)}{\log(\Delta_k \vee (k+4))}\right),&\end{aligned}$$

from which we conclude $\frac{3}{100} \cdot \delta_k \leq \widehat{\delta}_k \leq \frac{17}{5} \cdot \delta_k$. Putting $\textsc{Pen}_k := \frac{165}{4}\eta^{-1}([\widehat{\ell}]_0 \vee 1) \cdot \frac{\delta_k}{n}$, we observe that on Ξ_2 the estimate

$$\textsc{Pen}_k \leq \widehat{\textsc{Pen}}_k \leq \frac{340}{3}\textsc{Pen}_k$$

holds for all $k \in \{0, \dots, M_m^{\alpha+}\}$. Note that on Ξ we have $\widehat{k} \leq M_m^{\alpha+}$ which implies

$$(\textsc{Pen}_{k \vee \widehat{k}} + \widehat{\textsc{Pen}}_k - \widehat{\textsc{Pen}}_{\widehat{k}})\mathbb{1}_\Xi \leq (\textsc{Pen}_k + \textsc{Pen}_{\widehat{k}} + \widehat{\textsc{Pen}}_k - \widehat{\textsc{Pen}}_{\widehat{k}})\mathbb{1}_\Xi \leq \frac{343}{3}\textsc{Pen}_k \mathbb{1}_\Xi. \tag{4.27}$$

Now, we can proceed by mimicking the derivation of (4.22) in the proof of Theorem 4.8. More precisely, replacing the penalty term $\widetilde{\textsc{pen}}_k$ used in that proof by $\widehat{\textsc{Pen}}_k$, using the definition of $\textsc{Pen}_k$

above and (4.27), we obtain

$$\begin{aligned}\mathbb{E}[\|\widehat{\lambda}_{\widehat{k}} - \lambda\|_\omega^2 \mathbb{1}_\Xi] &\leq 7r\omega_k\gamma_k^{-1} + 40\sum_{k=0}^{N_n^{\alpha+}} \mathbb{E}\left[\left(\sup_{t\in\mathcal{B}_k} |\langle\widetilde{\Theta}_{n\wedge m}, t\rangle_\omega|^2 - \frac{33([\ell]_0 \vee 1)\delta_k}{8n}\right)_+\right] \\ &\quad + 32\mathbb{E}[\sup_{t\in\mathcal{B}_{K_{nm}^{\alpha+}}} |\langle\check{\Theta}_{n\wedge m}, t\rangle_\omega|^2] + 4\mathbb{E}[(\mathrm{PEN}_{k\vee\widehat{k}} + \widehat{\mathrm{PEN}}_k - \widehat{\mathrm{PEN}}_{\widehat{k}})\mathbb{1}_\Xi] \\ &\leq 7r\omega_k\gamma_k^{-1} + 40\sum_{k=0}^{N_n^{\alpha+}} \mathbb{E}\left[\left(\sup_{t\in\mathcal{B}_k} |\langle\widetilde{\Theta}_{n\wedge m}, t\rangle_\omega|^2 - \frac{33([\ell]_0 \vee 1)\delta_k}{8n}\right)_+\right] \\ &\quad + 32\mathbb{E}[\sup_{t\in\mathcal{B}_{K_{nm}^{\alpha+}}} |\langle\check{\Theta}_{n\wedge m}, t\rangle_\omega|^2] + \frac{1372}{3}\mathrm{PEN}_k.\end{aligned}$$

As in the proof of Theorem 4.8, the second and the third term are bounded applying Lemmata 4.17 (with $\delta_k^* = \delta_k$ and $\Delta_k^* = \Delta_k$) and 4.18, respectively. Hence, by means of an obvious adaption of statement a) in Lemma 4.16 (with N_n^α replaced by $N_n^{\alpha+}$) and the estimates

$$\Delta_k \leq d\Delta_k^\alpha, \quad \delta_k \leq d\zeta_d\delta_k^\alpha, \quad \frac{\delta_k}{\Delta_k} \geq 2k\zeta_d^{-1}\frac{\log(\Delta_k^\alpha \vee (k+4))}{\log(k+4)}$$

with $\zeta_d = \log(4d)/\log(4)$, we obtain in analogy to the way of proceeding in the proof of Theorem 4.8 that

$$\sup_{\lambda\in\Lambda_\gamma^r}\sup_{f\in\mathcal{F}_\alpha^d} \mathbb{E}[\|\widehat{\lambda}_{\widehat{k}} - \lambda\|_\omega^2 \mathbb{1}_\Xi] \lesssim \min_{0\leq k\leq K_{nm}^{\alpha-}} \max\left\{\frac{\omega_k}{\gamma_k}, \frac{\delta_k^\alpha}{n}\right\} + \Phi_m + \frac{1}{n}. \tag{4.28}$$

Upper bound for $\square_2$: The uniform upper bound for $\square_2$ can be derived in analogy to the bound for $\square_2$ in the proof of Theorem 4.8 using Assumption 4.9 instead of statement b) from Lemma 4.16 in the proof of Lemma 4.20. Hence, we obtain

$$\sup_{\lambda\in\Lambda_\gamma^r}\sup_{f\in\mathcal{F}_\alpha^d} \mathbb{E}[\|\widehat{\lambda}_{\widetilde{k}} - \lambda\|_\omega^2 \mathbb{1}_{\Xi_2^c}] \lesssim \frac{1}{m}. \tag{4.29}$$

Upper bound for $\square_3$: The term $\square_3$ can also be bounded analogously to the bound established for $\square_3$ in the proof of Theorem 4.8 (here, we do not have to exploit the additional Assumption 4.9), and we get

$$\sup_{\lambda\in\Lambda_\gamma^r}\sup_{f\in\mathcal{F}_\alpha^d} \mathbb{E}[\|\widehat{\lambda}_{\widetilde{k}} - \lambda\|_\omega^2 \mathbb{1}_{\Xi_1^c\cap\Xi_2}] \lesssim \frac{1}{n}. \tag{4.30}$$

Upper bound for $\square_4$: To find a uniform upper bound for the term $\square_4$, one can use exactly the same decompositions as in the proof of the uniform upper bound for $\square_3$ in Theorem 4.8 by replacing the probability of Ξ_1^c with the one of Ξ_3^c. Doing this, we obtain by means of Lemma 4.21 that

$$\sup_{\lambda\in\Lambda_\gamma^r}\sup_{f\in\mathcal{F}_\alpha^d} \mathbb{E}[\|\widehat{\lambda}_{\widetilde{k}} - \lambda\|_\omega^2 \mathbb{1}_{\Xi_1\cap\Xi_2\cap\Xi_3^c}] \lesssim \frac{1}{m}. \tag{4.31}$$

The result of the theorem now follows by combining (4.28), (4.29), (4.30) and (4.31). □

Note that the only additional prerequisite of Theorem 4.10 in contrast to Theorem 4.8 is the validity of Assumption 4.9.

4.3.3. Examples of convergence rates

We consider the same configurations for the sequences ω, γ and α as in Section 4.2.3. In particular, we assume that $\omega_0 = 1$ and $\omega_j = |j|^{2s}$ for all $j \neq 0$. The different configurations for γ and α will be investigated in the following (compare also with the minimax rates of convergence given in Table 4.1). Note that the additional Assumption 4.9 is satisfied in all the considered cases. Let us define $k_n^\diamond := \operatorname{argmin}_{k \in \mathbb{N}_0} \max \left\{ \frac{\omega_k}{\gamma_k}, \frac{\delta_k^\alpha}{n} \right\}$, that is, $k_n^\diamond$ realizes the best compromise between squared bias and penalty.

Scenario (pol)-(pol): In this scenario, $k_n^\diamond \asymp n^{\frac{1}{2p+2a+1}}$ and $N_n^{\alpha-} \asymp (n/\log n)^{\frac{1}{2s+2a+1}}$. First assume that $N_n^{\alpha-} \leq M_m^{\alpha-}$. In case that $s < p$, the rate with respect to n is $n^{-\frac{2(p-s)}{2p+2a+1}}$ which is the minimax optimal rate. In case that $s = p$, it holds $N_n^{\alpha-} \lesssim k_n^\diamond$ and the rate is $(n/\log(n))^{-\frac{2(p-s)}{2p+2a+1}}$ which is minimax optimal up to a logarithmic factor. Assume now that $M_m^{\alpha-} \leq N_n^{\alpha-}$. If $k_n^\diamond \lesssim M_m^{\alpha-}$, then the estimator obtains the optimal rate with respect to n and m. Otherwise, $M_m^{\alpha-} \asymp (m/\log m)^{1/(2a)}$ yields the contribution $(m/\log m)^{-\frac{p-s}{a}}$ to the rate.

Scenario (exp)-(pol): $N_n^{\alpha-} \asymp (n/\log n)^{1/(2a+2s+1)}$ as in scenario (pol)-(pol). Since $k_n^\diamond \asymp \log n$, it holds $k_n^\diamond \lesssim N_n^{\alpha-}$ and the optimal rate with respect to n holds in case that $k_n^\diamond \lesssim M_m^{\alpha-}$. Otherwise, the bias-penalty tradeoff generates the contribution $(M_m^{\alpha-})^{2s} \cdot \exp(-2p \cdot M_m^{\alpha-})$ to the rate.

Scenario (pol)-(exp): It holds that $k_n^\diamond \asymp N_n^{\alpha-}$ and again the sample size n is no obstacle for attaining the optimal rate of convergence. If $k_n^\diamond \lesssim M_m^{\alpha-}$, the minimax optimal rate is also attained. If $M_m^{\alpha-} \lesssim k_n^\diamond$, we get the rate $(\log m)^{-2(p-s)}$ which coincides with the optimal rate with respect to the sample size m.

Scenario (exp)-(exp): We have $N_n^{\alpha-} \asymp \log n$ and $k_1 \leq k_n^\diamond \leq k_2$ where k_1 is the solution of $k_1^2 \exp((2a+2p)k_1) \asymp n$ and k_2 the solution of $\exp((2a+2p)k_2) \asymp n$. Thus, we have $k_n^\diamond \asymp N_n^{\alpha-}$ and computation of $\frac{\omega_{k_1}}{\gamma_{k_1}}$ and $\frac{\delta_{k_2}^\alpha}{n}$ shows that only a loss by a logarithmic factor can occur as far as $k_n^\diamond \leq N_n^{\alpha-} \wedge M_m^{\alpha-}$. If $M_m^{\alpha-} \leq k_n^\diamond$, the contribution to the rate from the trade-off between squared bias and penalty is determined by $(M_m^{\alpha-})^{2s} \cdot \exp(-2pM_m^{\alpha-})$ which deteriorates the optimal rate with respect to m at most by a logarithmic factor.

We have not considered the case that the Fourier coefficients of the error density obey a power-exponential decay, that is $\alpha_j = \exp(-2\kappa |j|^a)$ for some $\kappa > 0$ and arbitrary $a > 0$. Indeed, for our definition of the quantity $M_m^{\alpha+}$, Assumption 4.9 is in general not satisfied in this case. This shortage can be removed by considering a more elaborate choice of the quantities $M_m^{\alpha-}$, $M_m^{\alpha+}$, and $\widehat{M}_m$ as was considered in [JS13a] but we do not include this here.

4.4. Adaptive estimation for model 2: Cox process observations

Unfortunately, the approach from Section 4.3 cannot be transferred in order to obtain an upper risk bound for an adaptive estimator in the case of Cox observations. Thus, in this section, we follow another approach. The price we have to pay is that we can only obtain rates which are optimal up to some additional logarithmic factors. Again we split our investigation into the partially adaptive and the fully adaptive case.

4.4.1. Partially adaptive estimation

We define $\mathbb{D}_k^\alpha := \sum_{0 \leq |j| \leq k} \frac{\omega_j}{\alpha_j}$ which might be interpreted as the dimension of the model associated with the linear subspace spanned by the $\mathbf{e}_j$ for $j \in \{-k, \ldots, k\}$ for the inverse problem at hand. In addition, we define the quantities N_n^α, M_m^α, and K_{nm}^α as well as the contrast function Υ exactly as in Section 4.3. However, we replace the definition of the penalty given in the case of Poisson observations with

$$\widetilde{\mathrm{PEN}}_k := 2000\eta^{-1} \cdot ([\widehat{\ell}]_0 \vee 1) \cdot \frac{d\mathbb{D}_k^\alpha \log(n+2)}{n} + 2000\eta^{-2} \cdot ([\widehat{\ell}]_0^2 \vee 1) \cdot \frac{d\mathbb{D}_k^\alpha \log(n+2)}{n}$$

where $\eta \in (0,1)$ is an additional tuning parameter that effects the rate of convergence only by a numerical constant. Based on this updated definition of the penalty we define the adaptive selection of the dimension parameter in the case of Cox observations by means of

$$\widetilde{k} := \operatorname*{argmin}_{0 \leq k \leq K_{nm}^\alpha} \{\Upsilon(\widehat{\lambda}_k) + \widetilde{\text{PEN}}_k\}.$$

Theorem 4.11. *Let Assumption 4.1 hold. Then, for any $n, m \in \mathbb{N}$,*

$$\sup_{\lambda \in \Lambda_\gamma^r} \sup_{f \in \mathcal{F}_\alpha^d} \mathbb{E}[\|\widehat{\lambda}_{\widetilde{k}} - \lambda\|_\omega^2] \lesssim \min_{0 \leq k \leq K_{nm}^\alpha} \max\left\{\frac{\omega_k}{\gamma_k}, \frac{\mathbb{D}_k^\alpha \log(n+2)}{n}\right\} + \Phi_m + \frac{1}{m} + \frac{1}{n}$$

where the observations in (4.8) *stem from model 2.*

The following proof of Theorem 4.11 turns out to be more intricate than the one of Theorem 4.8 due to the fact that we need to smuggle in an additional term. In order to deal with this term we have to apply consequences of Talagrand type concentration inequalities both for Poisson processes (see Proposition 2.13) and the analogue result for 'ordinary' random variables (see Lemma B.4 in the appendix).

Proof. We define all the sets Ξ_1, Ξ_2 and (based on the updated definition of the penalty) the terms $\square_1$, $\square_2$ and $\square_3$ as in the proof of Theorem 4.8. We use the decomposition

$$\mathbb{E}[\|\widehat{\lambda}_{\widetilde{k}} - \lambda\|_\omega^2] = \square_1 + \square_2 + \square_3$$

established in the proof of Theorem 4.8 and use exactly the same arguments as in that proof to bound the terms $\square_2$ and $\square_3$. Thus, it remains to find an appropriate uniform bound for $\square_1$. In order to get such a bound, we first proceed as in the proof of Theorem 4.8 in order to obtain on $\Xi_1 \cap \Xi_2$ the estimate

$$\|\widehat{\lambda}_{\widetilde{k}} - \lambda\|_\omega^2 \leq 7r\omega_k\gamma_k^{-1} + 4\widetilde{\text{PEN}}_k - 4\widetilde{\text{PEN}}_{\widetilde{k}} + 40 \sup_{t \in \mathcal{B}_{k \vee \widetilde{k}}} |\langle \widetilde{\Theta}_{n\wedge m}, t\rangle_\omega|^2 + 32 \sup_{t \in \mathcal{B}_{k \vee \widetilde{k}}} |\langle \check{\Theta}_{n\wedge m}, t\rangle_\omega|^2 \quad (4.32)$$

(here, $\widetilde{\Theta}$ and $\check{\Theta}$ are defined as in the proof of Theorem 4.8). Let us now introduce the function

$$\ddot{\lambda}_k := \sum_{0 \leq |j| \leq k} \frac{\mathbb{E}[[\widehat{\ell}]_j | \boldsymbol{\varepsilon}]}{[f]_j} \mathbf{e}_j$$

where $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)$ is the vector containing the *unobservable* shifts ε_i in (4.6). Using the decomposition $\widetilde{\Theta}_{n\wedge m} = \widetilde{\lambda}_{n\wedge m} - \lambda_{n\wedge m} = \widetilde{\lambda}_{n\wedge m} - \ddot{\lambda}_{n\wedge m} + \ddot{\lambda}_{n\wedge m} - \lambda_{n\wedge m}$ and setting

$$\Theta_{n\wedge m}^{(1)} = \widetilde{\lambda}_{n\wedge m} - \ddot{\lambda}_{n\wedge m} \qquad \text{and} \qquad \Theta_{n\wedge m}^{(2)} = \ddot{\lambda}_{n\wedge m} - \lambda_{n\wedge m}$$

we obtain from (4.32) that on $\Xi_1 \cap \Xi_2$

$$\|\widehat{\lambda}_{\widetilde{k}} - \lambda\|_\omega^2 \leq 7r\omega_k\gamma_k^{-1} + 4\widetilde{\text{PEN}}_k - 4\widetilde{\text{PEN}}_{\widetilde{k}} + 80 \sup_{t \in \mathcal{B}_{k \vee \widetilde{k}}} |\langle \Theta_{n\wedge m}^{(1)}, t\rangle_\omega|^2$$
$$+ 80 \sup_{t \in \mathcal{B}_{k \vee \widetilde{k}}} |\langle \Theta_{n\wedge m}^{(2)}, t\rangle_\omega|^2 + 32 \sup_{t \in \mathcal{B}_{k \vee \widetilde{k}}} |\langle \check{\Theta}_{n\wedge m}, t\rangle_\omega|^2.$$

Following along the lines of the proof of Theorem 4.8 we obtain that

$$\mathbb{E}[\|\widehat{\lambda}_{\widetilde{k}} - \lambda\|_\omega^2 \mathbb{1}_{\Xi_1 \cap \Xi_2}] \leq C(d,r) \min_{0 \leq k \leq K_{nm}^\alpha} \max\left\{\frac{\omega_k}{\gamma_k}, \frac{\mathbb{D}_k^\alpha \log(n+2)}{n}\right\}$$

$$+ 80 \sum_{k=0}^{K^\alpha_{nm}} \mathbb{E}\left[\left(\sup_{t\in\mathcal{B}_k} |\langle \Theta^{(1)}_{n\wedge m}, t\rangle_\omega|^2 - \frac{100\log(n+2) d\mathbb{D}^\alpha_k([\ell]_0 \vee 1)}{n}\right)_+\right]$$
$$+ 80 \sum_{k=0}^{K^\alpha_{nm}} \mathbb{E}\left[\left(\sup_{t\in\mathcal{B}_k} |\langle \Theta^{(2)}_{n\wedge m}, t\rangle_\omega|^2 - \frac{100\log(n+2) d\mathbb{D}^\alpha_k([\ell]_0^2 \vee 1)}{n}\right)_+\right]$$
$$+ 32\mathbb{E}[\sup_{t\in\mathcal{B}_{k\vee\widetilde{k}}} |\langle \check{\Theta}_{n\wedge m}, t\rangle_\omega|^2]. \tag{4.33}$$

We have

$$\mathbb{E}\left[\left(\sup_{t\in\mathcal{B}_k} |\langle \Theta^{(1)}_{n\wedge m}, t\rangle_\omega|^2 - \frac{100\log(n+2) d\mathbb{D}^\alpha_k([\ell]_0 \vee 1)}{n}\right)_+\right]$$
$$= \mathbb{E}\left[\mathbb{E}\left[\left(\sup_{t\in\mathcal{B}_k} |\langle \Theta^{(1)}_{n\wedge m}, t\rangle_\omega|^2 - \frac{100\log(n+2) d\mathbb{D}^\alpha_k([\ell]_0 \vee 1)}{n}\right)_+ |\boldsymbol{\varepsilon}\right]\right].$$

We apply Lemma 4.23 with $\delta^*_k = d\mathbb{D}^\alpha_k$ in order to obtain

$$\mathbb{E}\left[\left(\sup_{t\in\mathcal{B}_k} |\langle \Theta^{(1)}_{n\wedge m}, t\rangle_\omega|^2 - \frac{100\log(n+2) d\mathbb{D}^\alpha_k([\ell]_0 \vee 1)}{n}\right)_+ |\boldsymbol{\varepsilon}\right]$$
$$\lesssim \frac{\mathbb{D}^\alpha_k}{n^3} + \frac{\mathbb{D}^\alpha_k}{n^2}\exp(-K_2\sqrt{n\log(n+2)}).$$

Hence

$$\mathbb{E}\left[\left(\sup_{t\in\mathcal{B}_k} |\langle \Theta^{(1)}_{n\wedge m}, t\rangle_\omega|^2 - \frac{100\log(n+2) d\mathbb{D}^\alpha_k([\ell]_0 \vee 1)}{n}\right)_+\right] \lesssim \frac{\mathbb{D}^\alpha_k}{n^3} + \frac{\mathbb{D}^\alpha_k}{n^2}\exp(-K_2\sqrt{n\log(n+2)}).$$

We have $K^\alpha_{nm} \leq N^\alpha_n$ and hence by the definition of N^α_n that for $k \in \{0,\dots,K^\alpha_{nm}\}$

$$\mathbb{D}^\alpha_k \leq \mathbb{D}^\alpha_{N^\alpha_n} = \sum_{0\leq|j|\leq N^\alpha_n} \frac{\omega_j}{\alpha_j} \leq \frac{n}{\log(n+3)} \sum_{0\leq|j|\leq N^\alpha_n} \frac{1}{2|j|+1} \lesssim n$$

where we obtain the last estimate thanks to the logarithmic increase of the harmonic series. Due to $K^\alpha_{nm} \leq n$ we get

$$\sum_{k=0}^{K^\alpha_{nm}} \mathbb{E}\left[\left(\sup_{t\in\mathcal{B}_k} |\langle \Theta^{(1)}_{n\wedge m}, t\rangle_\omega|^2 - \frac{100\log(n+2) d\mathbb{D}^\alpha_k([\ell]_0 \vee 1)}{n}\right)_+\right] \lesssim \frac{1}{n}.$$

Applying Lemma 4.24 with $\delta^*_k = d\mathbb{D}^\alpha_k$ we obtain that

$$\mathbb{E}\left[\left(\sup_{t\in\mathcal{B}_k} |\langle \Theta^{(2)}_{n\wedge m}, t\rangle_\omega|^2 - \frac{100\log(n+2) d\mathbb{D}^\alpha_k([\ell]_0^2 \vee 1)}{n}\right)_+\right] \lesssim \frac{\mathbb{D}^\alpha_k}{n}\exp\left(-2\log(n+2)\right)$$
$$+ \frac{\mathbb{D}^\alpha_k}{n^2}\exp(-K_2\sqrt{n\log(n+2)}).$$

Using the relation $\mathbb{D}^\alpha_k \lesssim n$ established above we obtain

$$\sum_{k=0}^{K^\alpha_{nm}} \mathbb{E}\left[\left(\sup_{t\in\mathcal{B}_k} |\langle \Theta^{(2)}_{n\wedge m}, t\rangle_\omega|^2 - \frac{100\log(n+2) d\mathbb{D}^\alpha_k([\ell]_0^2 \vee 1)}{n}\right)_+\right] \lesssim \frac{1}{n}.$$

Finally, bounding the last term on the right-hand side of (4.33) by means of Lemma 4.18 we obtain from (4.33) using the obtained estimates that

$$\sup_{\lambda \in \Lambda_\gamma^r} \sup_{f \in \mathcal{F}_\alpha^d} \mathbb{E}[\|\widehat{\lambda}_{\widetilde{k}} - \lambda\|_\omega^2 \mathbb{1}_{\Xi_1 \cap \Xi_2}] \lesssim \min_{0 \le k \le K_{nm}^\alpha} \max\left\{\frac{\omega_k}{\gamma_k}, \frac{\mathbb{D}_k^\alpha \log(n+2)}{n}\right\} + \Phi_m + \frac{1}{n}.$$

This shows the desired uniform upper bound for $\square_1$ and combining it with the bounds for $\square_2$ and $\square_3$ yields the result. $\square$

4.4.2. Fully adaptive estimation

In the fully adaptive case, we replace the 'model dimension' $\mathbb{D}_k^\alpha$ from Section 4.4.1 by its natural estimate

$$\widehat{\mathbb{D}}_k := \sum_{0 \le |j| \le k} \frac{\omega_j}{|\widehat{[f]}_j|^2} \mathbb{1}_{\Omega_j}.$$

Based on the definition of $\widehat{\mathbb{D}}_k$ we define

$$\widehat{\text{PEN}}_k := 8000\eta^{-1} \cdot ([\widehat{\ell}]_0 \vee 1) \cdot \frac{\widehat{\mathbb{D}}_k \log(n+2)}{n} + 8000\eta^{-2} \cdot ([\widehat{\ell}]_0^2 \vee 1) \cdot \frac{\widehat{\mathbb{D}}_k \log(n+2)}{n}.$$

Note that this definition of the penalty is fully data-driven. We define the contrast function Υ exactly as in Section 4.3.1. For $n, m \in \mathbb{N}$, set

$$\widehat{N}_n := \inf\{1 \le j \le n : |\widehat{[f]}_j|^2/(2j+1) < \log(n+3)\omega_j^+/n\} - 1 \wedge n$$
$$\widehat{M}_m := \inf\{1 \le j \le m : |\widehat{[f]}_j|^2 < m^{-1} \log m\} - 1 \wedge m,$$

and $\widehat{K}_{nm} := \widehat{N}_n \wedge \widehat{M}_m$. We define the fully data-driven choice $\widehat{k}$ of k in analogy to the approach for model 1 via

$$\widehat{k} := \operatorname*{argmin}_{0 \le k \le \widehat{K}_{nm}} \{\Upsilon(\widehat{\lambda}_k) + \widehat{\text{PEN}}_k\}.$$

For the statement and the proof of the following theorem, define for $n, m \in \mathbb{N}$ the quantities

$$N_n^{\alpha-} := \inf\{1 \le j \le n : \alpha_j/(2j+1) < 4d \log(n+3)\omega_j^+/n\} - 1 \wedge n,$$
$$N_n^{\alpha+} := \inf\{1 \le j \le n : \alpha_j/(2j+1) < \log(n+3)\omega_j^+/(4dn)\} - 1 \wedge n,$$
$$M_m^{\alpha-} := \inf\{1 \le j \le m : \alpha_j < 4dm^{-1} \log m\} - 1 \wedge m,$$
$$M_m^{\alpha+} := \inf\{1 \le j \le m : 4d\alpha_j < m^{-1} \log m\} - 1 \wedge m,$$

$K_{nm}^{\alpha-} := N_n^{\alpha-} \wedge M_m^{\alpha-}$, and $K_{nm}^{\alpha+} := N_n^{\alpha+} \wedge M_m^{\alpha+}$. Note that the proof of the following theorem requires the validity of Assumption 4.9 again.

THEOREM 4.12. *Let Assumptions 4.1 and 4.9 hold. Then, for any $n, m \in \mathbb{N}$,*

$$\sup_{\lambda \in \Lambda_\gamma^r} \sup_{f \in \mathcal{F}_\alpha^d} \mathbb{E}[\|\widehat{\lambda}_{\widehat{k}} - \lambda\|_\omega^2] \lesssim \min_{0 \le k \le K_{nm}^{\alpha-}} \max\left\{\frac{\omega_k}{\gamma_k}, \frac{\mathbb{D}_k^\alpha \log(n+2)}{n}\right\} + \Phi_m + \frac{1}{m} + \frac{1}{n}$$

where $\mathbb{D}_k^\alpha := \sum_{0 \le |j| \le k} \frac{\omega_j}{\alpha_j}$.

PROOF. We define the sets Ξ_i for $i = 1, 2, 3$ and Ξ as in the proof of Theorem 4.10 and consider the decomposition

$$\mathbb{E}[\|\widehat{\lambda}_{\widehat{k}} - \lambda\|_\omega^2] = \square_1 + \square_2 + \square_3 + \square_4$$

where $\square_i$, $i = 1, 2, 3, 4$ are also defined as in the proof of Theorem 4.10. The terms $\square_2$, $\square_3$, and $\square_4$ are bounded exactly as in the proof of Theorem 4.10 and it remains to find an appropriate bound for $\square_1$. Set $\mathbb{D}_k := \sum_{0\leq|j|\leq k} \frac{\omega_j}{|[f]_j|^2}$ and

$$\textsc{Pen}_k = 2000\eta^{-1}\cdot([\widehat{\ell}]_0 \vee 1)\cdot\frac{\mathbb{D}_k\log(n+2)}{n} + 2000\eta^{-2}\cdot([\widehat{\ell}]_0^2 \vee 1)\cdot\frac{\mathbb{D}_k\log(n+2)}{n}.$$

From the definition of $\textsc{Pen}_k$ and $\widehat{\textsc{Pen}}_k$ one immediately obtains that on Ξ

$$\textsc{Pen}_k \leq \widehat{\textsc{Pen}}_k \leq 9\textsc{Pen}_k$$

from which one follows that

$$(\textsc{Pen}_{k\vee\widehat{k}} + \widehat{\textsc{Pen}}_k - \widehat{\textsc{Pen}}_{\widehat{k}})\mathbb{1}_\Xi \leq (\textsc{Pen}_k + \textsc{Pen}_{\widehat{k}} + \widehat{\textsc{Pen}}_k - \widehat{\textsc{Pen}}_{\widehat{k}})\mathbb{1}_\Xi \leq 10\textsc{Pen}_k.$$

Now, combining the argumentation from the proofs of Theorems 4.10 and 4.11 one can show that

$$\sup_{\lambda\in\Lambda_\gamma^r}\sup_{f\in\mathcal{F}_\alpha^d}\mathbb{E}[\|\widehat{\lambda}_{\widehat{k}} - \lambda\|_\omega^2\mathbb{1}_\Xi] \leq \min_{0\leq k\leq K_{nm}^{\alpha-}}\max\left\{\frac{\omega_k}{\gamma_k}, \frac{\mathbb{D}_k^\alpha\log(n+2)}{n}\right\} + \Phi_m + \frac{1}{n}.$$

The claim assertion of the theorem follows now by combining the bounds established for $\square_1$, $\square_2$, $\square_3$, and $\square_4$. $\square$

Remark 4.13. Of course, the approach presented in this section can also be applied to the case of Poisson observations but since the logarithmic factor in the rates is unavoidable we would obtain worse rates than using the approach from Section 4.3. Using the approach presented in this section we are not able to dispense with the additional logarithmic factor in the rates neither in case of model 1 nor model 2. Note that in case that the error density f is known (which is, vaguely spoken, equivalent to $m = \infty$) we regain the adaptive rate established in [Big+13] for the case that the unknown intensity is ordinary smooth and the Fourier coefficients of f obey a polynomial decay. However, our results are more general since we do not exclusively consider the case of polynomially decreasing Fourier coefficients.

Remark 4.14. Needless to say, the numerical constants in the definition of the penalty are ridiculously large which makes our rate optimal estimator nearly useless for small sample sizes. Hence there is still research necessary to establish an estimator which performs well both from a theoretical point of view and also yields good results for simulations with relatively small sample sizes. Another approach would be to calibrate numerical constants in the penalty by means of a simulation study as was done, for instance, in [CRT06].

4.4.3. Examples of convergence rates

Note that in all the scenarios considered in Table 4.1 we have $k_n^\diamond \lesssim N_n^{\alpha-}$ where $k_n^\diamond$ denotes the optimal trade-off between the squared bias ω_k/γ_k and the term $\mathbb{D}_k^\alpha\log(n+2)/n$. Computations similar to the ones leading to the rates in Table 4.1 show that the rates with respect to the sample size n are those from the minimax framework in Table 4.1 with n replaced with $n/\log(n+2)$ as long as $k_n^\diamond \leq N_n^{\alpha-} \wedge M_m^{\alpha-}$. If $M_m^{\alpha-} \leq k_n^\diamond$, $M_m^{\alpha-}$ contributes to the rate exactly with the same contribution as in Section 4.3.3.

4.5. Auxiliary results

4.5.1. Auxiliary results for Section 4.2

Lemma 4.15. *With the notations introduced in the main part of the present chapter, the following assertions hold true:*

a) *i)* $\mathrm{Var}(\widehat{[\ell]}_j) \leq [\lambda]_0/n$ *under model 1 and*

ii) $\mathrm{Var}(\widehat{[\ell]}_j) \leq 2(|[\lambda]|_j^2 + [\lambda]_0)/n$ *under model 2.*

b) $\mathrm{Var}(\widehat{[f]}_j) \leq 1/m$,

c) $\mathbb{P}(\Omega_j^c) = \mathbb{P}(|\widehat{[f]}_j|^2 < 1/m) \leq \min\{1, 4d/(m\alpha_j)\} \quad \forall f \in \mathcal{F}_\alpha^d$.

Proof. The proof of statement i) in a) is given by the identity

$$\mathrm{Var}(\widehat{[\ell]}_j) = \frac{1}{n}\,\mathrm{Var}\left(\int_0^1 \mathbf{e}_j(t)dN_1(t)\right) = \frac{1}{n}\int_0^1 |\mathbf{e}_j(t)|^2(\lambda \star f)(t)dt = \frac{1}{n}\cdot[\lambda]_0.$$

To prove ii), the identity $\mathbb{E}\widehat{[\ell]}_j = [\lambda]_j[f]_j$ implies

$$\mathrm{Var}(\widehat{[\ell]}_j) := \mathbb{E}[|\widehat{[\ell]}_j - \mathbb{E}\widehat{[\ell]}_j|^2] \leq 2\mathbb{E}[|\widetilde{[f]}_j[\lambda]_j - [f]_j[\lambda]_j|^2] + 2\,\mathbb{E}[|\xi_j|^2] =: 2V_1 + 2V_2$$

where $V_1 \leq |[\lambda]_j|^2 \cdot \mathrm{Var}(\widetilde{[f]}_j) \leq |[\lambda]_j|^2/n$. Here, the estimate $\mathrm{Var}(\widetilde{[f]}_j) \leq 1/n$ is easily derived in analogy to the proof of part b). In order to bound V_2 from above, notice

$$\begin{aligned}
\mathbb{E}[|\xi_j|^2] &= \frac{1}{n}\,\mathbb{E}\left[\mathbb{E}\left[\left|\int_0^1 \mathbf{e}_j(-t)\left\{dN_1(t) - \lambda(t-\varepsilon_1-\lfloor t-\varepsilon_1\rfloor)dt\right\}\right|^2 \mid \varepsilon_1\right]\right]\\
&= \frac{1}{n}\,\mathbb{E}\left[\int_0^1 |\mathbf{e}_j(-t)|^2\,\lambda(t-\varepsilon_1-\lfloor t-\varepsilon_1\rfloor)dt\right]\\
&= \frac{1}{n}\,\mathbb{E}\left[\int_0^1 \lambda(t-\varepsilon_1-\lfloor t-\varepsilon_1\rfloor)dt\right]\\
&= [\lambda]_0/n.
\end{aligned}$$

The assertion follows now by combining the obtained bounds for V_1 and V_2.

For the proof of b), note that we have $\mathrm{Var}(\widehat{[f]}_j) = \frac{1}{m}\,\mathrm{Var}\left(\mathbf{e}_j(-Y_1)\right)$ and the assertion follows from the estimate

$$\mathrm{Var}(\mathbf{e}_j(-Y_1)) = \mathbb{E}[|\mathbf{e}_j(-Y_1)|^2] - |\mathbb{E}\left[\mathbf{e}_j(-Y_1)\right]|^2 \leq \mathbb{E}[|\mathbf{e}_j(-Y_1)|^2] = 1.$$

For the proof of c), we consider two cases: if $|[f]_j|^2 < 4/m$ we have $1 < \frac{4d}{m\alpha_j}$ because $f \in \mathcal{F}_\alpha^d$ and the statement is evident. Otherwise, $|[f]_j|^2 \geq 4/m$ which implies

$$\mathbb{P}(|\widehat{[f]}_j|^2 < 1/m) \leq \mathbb{P}(|\widehat{[f]}_j|/\,|[f]_j| < 1/2) \leq \mathbb{P}(|\widehat{[f]}_j/[f]_j - 1| > 1/2).$$

Applying Chebyshev's inequality and exploiting the definition of $\mathcal{F}_\alpha^d$ yields

$$\mathbb{P}(|\widehat{[f]}_j|^2 < 1/m) \leq 4/\,|[f]_j|^2 \cdot \mathrm{Var}(\widehat{[f]}_j) \leq 4d/(m\alpha_j)$$

and statement c) follows. □

4.5.2. Auxiliary results for Section 4.3

Lemma 4.16. *Let Assumption 4.1 hold. Then the following assertions hold true:*

a) $\delta_j^\alpha/n \leq 1$ *for all* $n \in \mathbb{N}$ *and* $0 \leq j \leq N_n^\alpha$,

b) $\exp\left(-m\alpha_{M_m^\alpha}/(128d)\right) \leq C(d)m^{-5}$ *for all* $m \in \mathbb{N}$, *and*

c) $\min_{1\leq j\leq M_m^\alpha} |[f]_j|^2 \geq 2m^{-1}$ *for all* $m \in \mathbb{N}$.

PROOF. a) In case $N_n^\alpha = 0$, we have $\delta^\alpha_{N_n^\alpha} = 1$ and there is nothing to show. Otherwise $0 < N_n^\alpha \leq n$, and by definition of N_n^α we have $(2j+1)\Delta_j^\alpha \leq n/\log(n+3)$ for $0 \leq j \leq N_n^\alpha$ which by definition of δ_j^α implies that

$$\delta_j^\alpha \leq \frac{n}{\log(n+3)} \cdot \frac{\log(n/((2j+1)\log(n+3)) \vee (j+3))}{\log(j+3)}.$$

We consider two cases: In the first case, $n/((2j+1)\log(n+3)) \vee (j+3) = j+3$. Then $n \geq 1$ directly implies the estimate $\delta_j^\alpha \leq n$. In the second case, we have $n/((2j+1)\log(n+3)) \vee (j+3) = n/((2j+1)\log(n+3))$, and therefrom

$$\delta_j^\alpha \leq n\log(n)/(\log(n+3)\log(j+3)) \leq n,$$

and thus $\delta_j^\alpha \leq n$ in both cases. Division by n yields the claim assertion.

b) Note that, due to Assumption 4.1, we have $M_m^\alpha > 0$ for all sufficiently large m and that it is sufficient to show the desired inequality for such values of m. By the definition of M_m^α, we have $\alpha_{M_m^\alpha} \geq 640dm^{-1} \cdot \log(m+1)$ which implies

$$\exp(-m\alpha_{M_m^\alpha}/(128d)) \leq \exp(-5\log m) = m^{-5},$$

and the assertion follows.

c) The assertion follows from the observation that

$$\min_{1 \leq j \leq M_m^\alpha} |[f]_j|^2 \geq \min_{1 \leq j \leq M_m^\alpha} \frac{\alpha_j}{d} = \frac{\alpha_{M_m^\alpha}}{d} \geq 640m^{-1} \cdot \log(m+1)$$

combined with the fact that $640m^{-1} \cdot \log(m+1) \geq 2m^{-1}$ for all $m \in \mathbb{N}$.

□

LEMMA 4.17. *Let $(\delta_k^*)_{k\in\mathbb{N}_0}$ and $(\Delta_k^*)_{k\in\mathbb{N}_0}$ be sequences such that for all $k \in \mathbb{N}_0$,*

$$\delta_k^* \geq \sum_{0 \leq |j| \leq k} \frac{\omega_j}{|[f]_j|^2} \qquad \text{and} \qquad \Delta_k^* \geq \max_{0 \leq |j| \leq k} \frac{\omega_j}{|[f]_j|^2}.$$

Then, for all $k \in \{1, \ldots, n \wedge m\}$, we have

$$\mathbb{E}\left[\left(\sup_{t\in\mathcal{B}_k} |\langle\widetilde{\Theta}_{n\wedge m}, t\rangle|^2 - \frac{33\delta_k^*([\ell]_0 \vee 1)}{8n}\right)_+\right] \\ \leq K_1\left\{\frac{\|f\|\,\|\lambda\|\,\Delta_k^*}{n}\exp\left(-K_2 \cdot \frac{\delta_k^*}{\|f\|\,\|\lambda\|\,\Delta_k^*}\right) + \frac{\delta_k^*}{n^2}\exp\left(-K_3\sqrt{n}\right)\right\}$$

with positive numerical constants K_1, K_2, and K_3.

PROOF. We start the proof with the observation that, putting $r_t = \sum_{0\leq|j|\leq k} \omega_j \overline{[t]}_{-j}[f]_{-j}^{-1}\mathbf{e}_j$, we have

$$\frac{1}{n}\sum_{i=1}^{n}\int_0^1 r_t(x)[dN_i(x) - \ell(x)dx] = \langle\widetilde{\Theta}_{n\wedge m}, t\rangle_\omega.$$

Thus, we are in the framework of Proposition 2.13 and it remains to find suitable quantities M_1, H and v that satisfy the pre-conditions of that proposition.

Condition concerning M_1:

$$\sup_{t\in\mathcal{B}_k}\|r_t\|_\infty^2 = \sup_{t\in\mathcal{B}_k}\sup_{y\in[0,1)}|r_t(y)|^2 \le \sup_{t\in\mathcal{B}_k}\sup_{y\in[0,1)}\left(\sum_{0\le|j|\le k}\omega_j\overline{[t]}_{-j}[f]_{-j}^{-1}|\mathbf{e}_j(y)|\right)^2$$
$$\le \sup_{t\in\mathcal{B}_k}\sup_{y\in[0,1)}\left(\sum_{0\le|j|\le k}\omega_j|[t]_j|^2\right)\left(\sum_{0\le|j|\le k}\frac{\omega_j}{|[f]_j|^2}\right) \le \delta_k^* =: M_1^2.$$

Condition concerning H:

$$\mathbb{E}[\sup_{t\in\mathcal{B}_k}|\langle\Xi_k,t\rangle_\omega|^2] \le \sup_{t\in\mathcal{B}_k}\left(\sum_{0\le|j|\le k}\omega_j|[t]_j|^2\right)$$
$$\cdot\,\mathbb{E}\left[\sum_{0\le|j|\le k}\frac{\omega_j}{|[f]_j|^2}\left|\frac{1}{n}\sum_{i=1}^n\int_0^1\mathbf{e}_j(x)[dN_i(x)-\ell(x)dx]\right|^2\right]$$
$$\le \frac{1}{n}\sum_{0\le|j|\le k}\frac{\omega_j}{|[f]_j|^2}\cdot\operatorname{Var}\left(\int_0^1\mathbf{e}_j(x)dN_1(x)\right)$$
$$= \frac{1}{n}\sum_{0\le|j|\le k}\frac{\omega_j}{|[f]_j|^2}\cdot\int_0^1\ell(x)dx = \frac{[\ell]_0}{n}\sum_{0\le|j|\le k}\frac{\omega_j}{|[f]_j|^2}.$$

Hence, by Jensen's inequality it follows that we can choose $H^2 = \frac{\delta_k^*}{n}\cdot([\ell]_0\vee 1)$.

Condition concerning υ: First, note that

$$\operatorname{Var}\left(\int_0^1 r_t(x)dN_i(x)\right) = [\ell]_0\cdot\mathbb{E}[|r_t(X)|^2]$$

where X is a random variable with density proportional to ℓ. It remains to find an appropriate bound for $\mathbb{E}[|r_t(X)|^2] = \mathbb{E}[\langle r_t(X), r_t(X)\rangle]$ (here $\langle\cdot,\cdot\rangle$ denotes the standard scalar product in $\mathbb{C}$). By some calculations it follows that

$$\mathbb{E}[|r_t(X)|^2] = \frac{1}{[\ell]_0}\langle AD_\omega\boldsymbol{t}, D_\omega\boldsymbol{t}\rangle_{\mathbb{C}^{2k+1}}$$

where $\boldsymbol{t} = ([t]_i)_{i=-k,\dots,k}$, $D_\omega\in\mathbb{R}^{(2k+1)\times(2k+1)}$ is the diagonal matrix with diagonal $(\omega_i)_{i=-k,\dots,k}$ and the matrix $A = ([A]_{ij})_{i,j=-k,\dots,k}\in\mathbb{R}^{(2k+1)\times(2k+1)}$ is given by means of

$$[A]_{ij} = [f]_i^{-1}\overline{[f]}_j^{-1}[f]_{i-j}[\lambda]_{i-j}.$$

The matrix A is positive semi-definite and we obtain for any $t\in\mathcal{B}_k$

$$\operatorname{Var}\left(\int_0^1 r_t(x)dN_i(x)\right) = \langle AD_\omega\boldsymbol{t}, D_\omega\boldsymbol{t}\rangle_{\mathbb{C}^{2k+1}} = \langle\sqrt{A}D_\omega\boldsymbol{t},\sqrt{A}D_\omega\boldsymbol{t}\rangle_{\mathbb{C}^{2k+1}}$$
$$= \|\sqrt{A}D_\omega\boldsymbol{t}\|^2 = \|\sqrt{A D_\omega}\|_{\mathrm{op}}^2 \le \|\sqrt{D_\omega}A\sqrt{D_\omega}\|_{\mathrm{op}}$$

where the last identity holds since $\|S\|_{\mathrm{op}}^2\le\|S^*S\|$ for a linear operator S between Hilbert spaces. One has the decomposition $A = D_{[f]^{-1}}BD_{\overline{[f]}^{-1}}$ with $B = ([f]_{i-j}[\lambda]_{i-j})_{i,j=-k,\dots,k}$ from which we conclude

$$\operatorname{Var}\left(\int_0^1 r_t(x)dN_i(x)\right) \le \Delta_k^*\cdot\|B\|_{\mathrm{op}},$$

and it remains to find a suitable bound for $\|B\|_{\mathrm{op}}$. Note that $\|B\|_{\mathrm{op}} \leq \sqrt{\|B\|_1 \cdot \|B\|_\infty}$ where

$$\|B\|_1 = \max_j \sum_i |b_{ij}|, \quad \text{and}$$
$$\|B\|_\infty = \max_i \sum_j |b_{ij}|$$

(see Corollary 2.3.2 in [GVL96]). The Cauchy-Schwarz inequality shows that $\|B\|_1, \|B\|_\infty \leq \|\lambda\| \cdot \|f\|$ and we can finally conclude that

$$\operatorname{Var}\left(\int_0^1 r_t(x) dN_i(x)\right) \leq \Delta_k^* \cdot \|f\| \cdot \|\lambda\| \cdot ([\ell]_0 \vee 1) =: \upsilon.$$

The statement of the lemma follows now by applying Proposition 2.13 with $\varepsilon = \frac{1}{64}$. □

Lemma 4.18. *Let $m \in \mathbb{N}$ and $k \in \mathbb{N}_0$. Then*

$$\sup_{\lambda \in \Lambda_\gamma^r} \mathbb{E}\left[\sup_{t \in \mathcal{B}_k} |\langle \check{\Theta}_{n\wedge m}, t\rangle_\omega|^2\right] \leq C(d,r) \cdot \Phi_m.$$

Proof. Note that $\lambda \in \Lambda_\gamma^r$ implies

$$\mathbb{E}[\sup_{t \in \mathcal{B}_k} |\langle \check{\Theta}_{n\wedge m}, t\rangle_\omega|^2] \leq r \sup_{-k \leq j \leq k} \omega_j \gamma_j^{-1}\, \mathbb{E}[|[f]_j/\widehat{[f]}_j \cdot \mathbb{1}_{\Omega_j} - 1|^2]$$

Thus, recalling the definition of Φ_m in (4.12), it suffices to show that

$$\mathbb{E}[|[f]_j/\widehat{[f]}_j \cdot \mathbb{1}_{\Omega_j} - 1|^2] \leq C(d,r) \min\{1, 1/(m\alpha_j)\},$$

which can be realized by means of the identity

$$\mathbb{E}[|[f]_j/\widehat{[f]}_j \cdot \mathbb{1}_{\Omega_j} - 1|^2] = \mathbb{E}[|[f]_j/\widehat{[f]}_j \mathbb{1}_{\Omega_j} - 1|^2 \cdot \mathbb{1}_{\Omega_j}] + \mathbb{P}(\Omega_j^{\mathsf{c}}) =: \Box + \triangle.$$

The bound $\Box \leq C(d,r)\min\{1, 1/(m\alpha_j)\}$ was already derived in the proof of Theorem 4.2. For $\triangle$, the corresponding upper bound can be obtained from statement c) of Lemma 4.15. □

Lemma 4.19. *Let Assumption 4.1 hold and consider the event Ξ_1 defined in Theorem 4.8. Then, for any $n \in \mathbb{N}$, $\mathbb{P}(\Xi_1^{\mathsf{c}}) \leq 2\exp(-Cn)$ with a numerical constant $C = C(\eta) > 0$.*

Proof. Note that

$$\mathbb{P}(\Xi_1^{\mathsf{c}}) = \mathbb{P}(\widehat{[\ell]}_0 \vee 1 < \eta([\ell]_0 \vee 1)) + \mathbb{P}(\widehat{[\ell]}_0 \vee 1 > \eta^{-1}([\ell]_0 \vee 1)),$$

and the two terms on the right-hand side can be bounded by Chernoff bounds for Poisson distributed random variables (see Theorem B.2). More precisely, we have

$$\mathbb{P}(\widehat{[\ell]}_0 \vee 1 < \eta([\ell]_0 \vee 1)) \leq \exp(-\omega_1(\eta)n) \qquad \text{and}$$
$$\mathbb{P}(\widehat{[\ell]}_0 \vee 1 > \eta^{-1}([\ell]_0 \vee 1)) \leq \exp(-\omega_2(\eta)n)$$

with $\omega_1(\eta) = 1 - \eta + \eta\log\eta > 0$ and $\omega_2(\eta) = 1 - \eta^{-1} - \eta^{-1}\log\eta > 0$ for all $\eta \in (0,1)$. □

Lemma 4.20. *Let Assumption 4.1 hold and consider the event Ξ_2 defined in the proof of Theorem 4.8. Then, for any $m \in \mathbb{N}$, $\mathbb{P}(\Xi_2^{\mathsf{c}}) \leq C(d)m^{-4}$.*

Proof. The complement Ξ_2^c of Ξ_2 is

$$\Xi_2^c = \{\exists 1 \leq |j| \leq M_m^\alpha : |[f]_j/\widehat{[f]}_j - 1| > 1/2 \text{ or } |\widehat{[f]}_j|^2 < 1/m\}.$$

Owing to statement c) from Lemma 4.16 we have $|[f]_j|^2 \geq 2/m$ for all $1 \leq |j| \leq M_m^\alpha$. In case that $|\widehat{[f]}_j|^2 < 1/m$ a direct calculation using the reverse triangle inequality shows that $|\widehat{[f]}_j/[f]_j - 1| \geq 1/\sqrt{2} - 1 > 1/4$. In case that $|[f]_j/\widehat{[f]}_j - 1| > \frac{1}{2}$, one obtains $|\widehat{[f]}_j/[f]_j - 1| > 1/3$, and thus together we have

$$\Xi_2^c \subseteq \{\exists 1 \leq |j| \leq M_m^\alpha : |\widehat{[f]}_j/[f]_j - 1| > 1/4\}.$$

Now, Hoeffding's inequality implies for $|j| \leq M_m^\alpha$ that

$$\mathbb{P}(|\widehat{[f]}_j/[f]_j - 1| > 1/4) \leq 4\exp\left(-\frac{m|[f]_j|^2}{128}\right) \leq 4\exp\left(-\frac{m\alpha_{M_m^\alpha}}{128d}\right),$$

and the statement of the lemma follows from statement b) of Lemma 4.16 and the estimate $M_m^\alpha \leq m$ which holds by definition of M_m^α. □

Lemma 4.21. *Let Assumptions 4.1 and 4.9 hold. The event Ξ_3 defined in (4.26) satisfies $\mathbb{P}(\Xi_3^c) \leq C(\alpha, d)m^{-4}$ for all $m \in \mathbb{N}$.*

Proof. Let us consider the random sets

$$\Xi_{31} := \{N_n^{\alpha-} \wedge M_m^{\alpha-} > \widehat{K}_{nm}\} \quad \text{and} \quad \Xi_{32} := \{\widehat{K}_{nm} > N_n^{\alpha+} \wedge M_m^{\alpha+}\}.$$

Then, $\Xi_3^c = \Xi_{31} \cup \Xi_{32}$ and we establish bounds for $\mathbb{P}(\Xi_{31})$ and $\mathbb{P}(\Xi_{32})$, separately.

Upper bound for $\mathbb{P}(\Xi_{31})$: We use the identity $\Xi_{31} = \{\widehat{N}_n < K_{nm}^{\alpha-}\} \cup \{\widehat{M}_m < K_{nm}^{\alpha-}\}$. Owing to the definition of $N_n^{\alpha-}$, we have $|[f]_j|^2/((2j+1)\omega_j^+) \geq 4\log(n+4)/n$ for all $1 \leq j \leq N_n^{\alpha-}$, which yields

$$\begin{aligned}\{\widehat{N}_n < K_{nm}^{\alpha-}\} &\subseteq \{\exists 1 \leq j \leq K_{nm}^{\alpha-} : |\widehat{[f]}_j|^2/((2j+1)\omega_j^+) < \log(n+4)/n\} \\ &\subseteq \bigcup_{1 \leq j \leq K_{nm}^{\alpha-}} \{|\widehat{[f]}_j|/|[f]_j| \leq 1/2\} \\ &\subseteq \bigcup_{1 \leq j \leq K_{nm}^{\alpha-}} \{|\widehat{[f]}_j/[f]_j - 1| \geq 1/2\}.\end{aligned}$$

In a similar way, we obtain $\{\widehat{M}_m < K_{nm}^{\alpha-}\} \subseteq \bigcup_{1 \leq j \leq K_{nm}^{\alpha-}} \{|\widehat{[f]}_j/[f]_j - 1| \geq 1/2\}$. Thus, since $M_m^{\alpha-} \leq M_m^{\alpha+}$ by definition, we have

$$\Xi_{31} \subseteq \bigcup_{1 \leq j \leq M_m^{\alpha+}} \{|\widehat{[f]}_j/[f]_j - 1| \geq 1/2\}.$$

Applying Hoeffding's inequality as in the proof of Lemma 4.20 and exploiting Assumption 4.9 yields

$$\mathbb{P}(\Xi_{31}) \leq 4 \sum_{1 \leq j \leq M_m^{\alpha+}} \exp\left(-\frac{m|[f]_j|^2}{128}\right) \leq C(\alpha, d) \cdot m^{-4}. \tag{4.34}$$

Upper bound for $\mathbb{P}(\Xi_{32})$: First, note that $\Xi_{32} = \{\widehat{N}_n > K_{nm}^{\alpha+}\} \cap \{\widehat{M}_m > K_{nm}^{\alpha+}\}$. In particular, $K_{nm}^{\alpha+} < n \wedge m$. If $K_{nm}^{\alpha+} = N_n^{\alpha+} < n$, we obtain

$$\begin{aligned}\Xi_{32} &\subseteq \{\widehat{N}_n > N_n^{\alpha+}\} \subseteq \{\forall 1 \leq j \leq N_n^{\alpha+} + 1 : |\widehat{[f]}_j|^2/((2j+1)\omega_j^+) \geq \log(n+4)/n\} \\ &\subseteq \{|\widehat{[f]}_{N_n^{\alpha+}+1}|/|[f]_{N_n^{\alpha+}+1}| \geq 2\} \subseteq \{|\widehat{[f]}_{N_n^{\alpha+}+1}/[f]_{N_n^{\alpha+}+1} - 1| \geq 1\}.\end{aligned}$$

Analogously, if $K_{nm}^{\alpha+} = M_m^{\alpha+} < m$, using $m^{-1}\log m \geq 4|[f]_{M_m^{\alpha+}+1}|^2$ yields

$$\Xi_{32} \subseteq \{\widehat{M}_m > M_m^{\alpha+}\} \subseteq \{|\widehat{[f]}_{M_m^{\alpha+}+1}/[f]_{M_m^{\alpha+}+1} - 1| \geq 1\}$$

and thus $\Xi_{32} \subseteq \{|\widehat{[f]}_{K_{nm}^{\alpha+}+1}/[f]_{K_{nm}^{\alpha+}+1} - 1| \geq 1\}$. Application of Hoeffding's inequality and exploiting Assumption 4.9 yields

$$\mathbb{P}(\Xi_{32}) \leq 4\exp\left(-\frac{m|[f]_{K_{nm}^{\alpha+}+1}|^2}{128}\right) \leq 4\exp\left(-\frac{m\alpha_{M_m^{\alpha+}+1}}{128d}\right) \leq C(\alpha,d)m^{-5}. \tag{4.35}$$

The statement of the lemma follows by combining Equations (4.34) and (4.35). □

4.5.3. Auxiliary results for Section 4.4

The following result is a condtional version of Proposition 2.13. Since the proof is exactly the same as the one in the unconditional case we omit its proof.

Proposition 4.22. *Let $N_1,\dots,N_n$ be independent Cox processes driven by* finite *random measures $\eta_1,\dots,\eta_n$ (that is, given η_i, N_i is a PPP with intensity measure η_i) that are conditionally independent given $\eta_1,\dots,\eta_n$. Set $\nu_n(r) = \frac{1}{n}\sum_{k=1}^n\{\int_{\mathbb{X}} r(x)dN_k(x) - \int_{\mathbb{X}} r(x)d\eta_k(x)\}$ for r contained in a countable class of complex-valued measurable functions. Then, for any $\varepsilon > 0$, there exist constants c_1, $c_2 = \frac{1}{6}$, and c_3 such that*

$$\mathbb{E}\left[\left(\sup_{r\in\mathcal{R}}|\nu_n(r)|^2 - c(\varepsilon)H^2\right)_+ |\boldsymbol{\eta}\right] \leq c_1\left\{\frac{\upsilon}{n}\exp\left(-c_2\varepsilon\frac{nH^2}{\upsilon}\right) + \frac{M_1^2}{C^2(\varepsilon)n^2}\exp\left(-c_3C(\varepsilon)\sqrt{\varepsilon}\frac{nH}{M_1}\right)\right\}$$

where $C(\varepsilon) = (\sqrt{1+\varepsilon}-1)\wedge 1$, $c(\varepsilon) = 4(1+2\varepsilon)$ and M_1, H and υ are such that (denoting $\boldsymbol{\eta} = (\eta_1,\dots,\eta_n)$)

$$\sup_{r\in\mathcal{R}}\|r\|_\infty \leq M_1, \quad \mathbb{E}[\sup_{r\in\mathcal{R}}|\nu_n(r)||\boldsymbol{\eta}] \leq H, \quad \sup_{r\in\mathcal{R}}\operatorname{Var}\left(\int_{\mathbb{X}} r(x)dN_k(x)|\boldsymbol{\eta}\right) \leq \upsilon \quad \forall k.$$

Lemma 4.23. *Let $(\delta_k^*)_{k\in\mathbb{N}_0}$ be a sequence such that $\delta_k^* \geq \sum_{0\leq|j|\leq k}\frac{\omega_j}{|[f]_j|^2}$ for all $k\in\mathbb{N}_0$. Then,*

$$\mathbb{E}\left[\left(\sup_{t\in\mathcal{B}_k}|\langle\Theta_{n\wedge m}^{(1)},t\rangle|^2 - \frac{100\log(n+2)\delta_k^*([\ell]_0\vee 1)}{n}\right)_+ |\,\boldsymbol{\varepsilon}\right] \leq K_1\left\{\frac{2\delta_k^*([\ell]_0\vee 1)}{n}\exp\left(-2\log(n+2)\right) + \frac{\delta_k^*}{n^2}\exp\left(-K_2\sqrt{n\log(n+2)}\right)\right\}$$

with positive numerical constants K_1 and K_2.

Proof. Putting $r_t = \sum_{0\leq|j|\leq k}\omega_j[f]_{-j}^{-1}\overline{[t]}_{-j}\mathbf{e}_j$, it is easy to check that given $\boldsymbol{\varepsilon}$

$$\langle\Theta_{n\wedge m}^{(1)},t\rangle_\omega = \frac{1}{n}\sum_{i=1}^n\int_0^1 r_t(x)(dN_i(x) - \lambda_{\varepsilon_i}(x)dx)$$

where $\lambda_{\varepsilon_i}(x) = \lambda(x-\varepsilon_i-\lfloor x-\varepsilon_i\rfloor)$. Thus, we are in the framework of Proposition 4.22 and it remains to find suitable constants M_1, H, and υ satisfying its preconditions.

Condition concerning M_1: We have

$$\sup_{t \in \mathcal{B}_k} \|r_t\|_\infty^2 = \sup_{t \in \mathcal{B}_k} \sup_{y \in [0,1)} |r_t(y)|^2 \leq \sum_{0 \leq |j| \leq k} \frac{\omega_j}{|[f]_j|^2} \leq \delta_k^*,$$

and one can choose $M_1 = (\delta_k^*)^{\frac{1}{2}}$.
Condition concerning H: We have

$$\mathbb{E}[\sup_{t \in \mathcal{B}_k} |\langle \Theta_{n\wedge m}^{(1)}, t \rangle_\omega|^2 \mid \boldsymbol{\varepsilon}] = \frac{[\ell]_0}{n} \cdot \sum_{0 \leq |j| \leq k} \frac{\omega_j}{|[f]_j|^2} \leq \frac{([\ell]_0 \vee 1)\delta_k^* \log(n+2)}{n},$$

and one can choose $H = \left(\frac{([\ell]_0 \vee 1)\delta_k^* \log(n+2)}{n}\right)^{1/2}$.
Condition concerning υ: It holds that

$$\operatorname{Var}\left(\int_0^1 r_t(x) N_k(x) \mid \boldsymbol{\varepsilon}\right) = \int_0^1 |r_t(x)|^2 \lambda_{\varepsilon_k}(x) dx \leq \left(\sum_{0 \leq |j| \leq k} \frac{\omega_j}{|[f]_j|^2}\right) \cdot [\ell]_0 \leq \delta_k^* \cdot ([\ell]_0 \vee 1),$$

and one can choose $\upsilon = \delta_k^* \cdot ([\ell]_0 \vee 1)$. The statement of the lemma follows now by applying Proposition 4.22 with $\varepsilon = 12$. □

Lemma 4.24. *Let $(\delta_k^*)_{k \in \mathbb{N}_0}$ be a sequence such that $\delta_k^* \geq \sum_{0 \leq |j| \leq k} \frac{\omega_j}{|[f]_j|^2}$ for all $k \in \mathbb{N}_0$. Then*

$$\mathbb{E}\left[\left(\sup_{t \in \mathcal{B}_k} |\langle \Theta_{n\wedge m}^{(2)}, t \rangle|^2 - \frac{100 \log(n+2)\delta_k^*([\ell]_0^2 \vee 1)}{n}\right)_+\right] \\ \leq K_1 \left\{\frac{\delta_k^*([\ell]_0^2 \vee 1)}{n} \exp(-2\log(n+2)) + \frac{([\ell]_0^2 \vee 1)\delta_k^*}{n^2} \cdot \exp(-K_2\sqrt{n \log(n+2)})\right\}$$

with strictly positive numerical constants K_1 and K_2.

Proof. We define $r_t' = \sum_{0 \leq |j| \leq k} \omega_j [f]_{-j}^{-1} \overline{[t]}_{-j} \mathbf{e}_j$ which coincides with the definition of r_t in the proof of Lemma 4.23. Then, we have

$$\langle \Theta_{n\wedge m}^{(2)}, t \rangle_\omega = \frac{1}{n} \sum_{i=1}^n \int_0^1 r_t'(x) \lambda_{\varepsilon_i}(x) dx - \int_0^1 r_t'(x) \ell(x) dx$$

where λ_ε is defined as in the proof of Lemma 4.23. Setting $r_t(\varepsilon_i) := \int_0^1 r_t'(x) \lambda_{\varepsilon_i}(x) dx$, we are in the framework of Proposition B.4 and it remains to find suitable constants M_1, H and υ satisfying the preconditions of that proposition.
Condition concerning M_1: Note that the definition of r_t' is the same as the definition of r_t in the proof of Lemma 4.23. Thus we obtain

$$\sup_{t \in \mathcal{B}_k} \|r_t\|_\infty = \sup_{\varepsilon \in [0,1)} \sup_{t \in \mathcal{B}_k} |\int_0^1 r_t'(x) \lambda_\varepsilon(x) dx| \leq (\delta_k^*)^{1/2} \cdot \sup_{\varepsilon \in [0,1)} \int_0^1 \lambda_\varepsilon(x) dx = (\delta_k^*)^{1/2} \cdot ([\ell]_0 \vee 1),$$

and we can take $M_1 = (\delta_k^*)^{1/2} \cdot ([\ell]_0 \vee 1)$.
Condition concerning H: We have

$$\mathbb{E}[\sup_{t \in \mathcal{B}_k} |\langle \Theta_{n\wedge m}^{(2)}, t \rangle_\omega|^2] \leq \left(\sum_{0 \leq |j| \leq k} \frac{\omega_j}{|[f]_j|^2}\right) \frac{1}{n} \mathbb{E}[|\int_0^1 \mathbf{e}_j(x)(\lambda_{\varepsilon_1}(x)dx - \ell(x)dx)|^2] \\ \leq \frac{\delta_k^*[\ell]_0^2}{n} \leq \frac{\delta_k^*[\ell]_0^2 \log(n+2)}{n},$$

and we can set $H = \left(\frac{\delta_k^* \log(n+2)}{n}\right)^{1/2} \cdot ([\ell]_0 \vee 1)$.
Condition concerning υ: It holds

$$\operatorname{Var}(r_t(\varepsilon_k)) \leq \mathbb{E}\left[\left|\int_0^1 r_t'(x)\lambda_{\varepsilon_k}(x)dx\right|^2\right] \leq [\ell]_0^2 \cdot \mathbb{E}\left[\int_0^1 |r_t'(x)|^2 \frac{\lambda_{\varepsilon_k}(x)}{[\lambda]_0} dx\right] \leq ([\ell]_0^2 \vee 1) \cdot \delta_k^*,$$

and we define $\upsilon = ([\ell]_0^2 \vee 1) \cdot \delta_k^*$. Now that statement of the lemma follows from Proposition B.4 (together with Remark B.5) with $\varepsilon = 12$. □

5. Non-parametric Poisson regression

In this chapter, we consider a non-parametric Poisson regression model. We assume that the observations are given by an i.i.d. sample

$$(X_1, Y_1), \dots, (X_n, Y_n) \in [0,1] \times \mathbb{N}_0 \tag{5.1}$$

satisfying the relationship

$$Y_i | X_i \sim \mathcal{P}(T\lambda(X_i)).$$

Here, $\mathcal{P}(\alpha)$ denotes the Poisson distribution with parameter $\alpha \geq 0$[1], $T > 0$ and $\lambda : [0,1] \to [0,\infty)$ is a non-negative function. The aim of this chapter is to derive an adaptive estimator of the unknown function λ from the observations (5.1).

The classical distinction in non-parametric regression is made between *random* and *deterministic* design: in the deterministic design framework, one assumes that the X_i are predetermined and fixed sampling points. Most frequently, the so-called *equidistant* deterministic design where $X_i = \frac{i}{n}$ for $i = 1, \dots, n$ is considered. In the random design framework, one assumes that $X_1, \dots, X_n$ is an i.i.d. sample drawn according to some known probability density function $f : [0,1] \to \mathbb{R}$.

In this thesis, we restrict ourselves to the random design case. It is intuitively appealing that the explanatory variables X_i should be scattered over the interval $[0,1]$ in a sufficiently uniform way to make a reasonable estimate of λ over the whole interval possible. In order to ensure this, we will assume that the density f is bounded away from zero (see Assumption 5.2 below) which is a standard assumption in non-parametric regression (see, for instance, [Sto82] but also [Che07] for a study that does not use such an assumption).

5.1. Methodology: Orthonormal series estimator of the regression function

As in Chapter 3, we assume that the unknown functional parameter belongs to the space $\mathbb{L}^2 := \mathbb{L}^2([0,1], dx)$ of square-integrable and real-valued functions. In addition, we again use an orthonormal series estimator in terms of the trigonometric orthonormal basis $\{\varphi_j\}_{j\in\mathbb{Z}}$ given by

$$\varphi_0 \equiv 1, \quad \varphi_j(x) = \sqrt{2}\cos(2\pi j x), \quad \varphi_{-j}(x) = \sqrt{2}\sin(2\pi j x), \quad j = 1, 2, \dots$$

The Fourier coefficients of a function $\lambda \in \mathbb{L}^2$ are denoted with

$$[\lambda]_j := \int_0^1 \lambda(x)\varphi_j(x)dx$$

leading to the $\mathbb{L}^2$-convergent representation $\lambda = \sum_{j\in\mathbb{Z}} [\lambda]_j \varphi_j$. As in the previous chapters, we consider projection estimators of the form

$$\widehat{\lambda}_k := \sum_{0 \leq |j| \leq k} \widehat{[\lambda]}_j \varphi_j$$

[1] By convention, we define the Poisson distribution with parameter $\alpha = 0$ to be the probability distribution degenerated at 0.

where $\widehat{[\lambda]}_j$ is an appropriate estimator of $[\lambda]_j$ for all $j \in \mathbb{Z}$ and $k \in \mathbb{N}_0$ a dimension parameter. Under the assumption that $f(x) > 0$ for all $x \in [0,1]$, we have for the bivariate random variable (X,Y) with $X \sim f$ and $Y|X \sim \mathcal{P}(T\lambda(X))$ for all $j \in \mathbb{Z}$ the identity

$$\mathbb{E}\left[\frac{Y}{f(X)}\varphi_j(X)\right] = \mathbb{E}\left[\mathbb{E}\left[\frac{Y}{f(X)}\varphi_j(X)|X\right]\right] = \mathbb{E}\left[\frac{\varphi_j(X)}{f(X)}\cdot \mathbb{E}[Y|X]\right] = T\ \mathbb{E}\left[\frac{\lambda(X)}{f(X)}\varphi_j(X)\right]$$
$$= T\int_0^1 \frac{\lambda(x)}{f(x)}\varphi_j(x)f(x)dx = T\int_0^1 \lambda(x)\varphi_j(x)dx = T[\lambda]_j,$$

and thus

$$\widehat{[\lambda]}_j := \frac{1}{nT}\sum_{i=1}^n \frac{Y_i}{f(X_i)}\varphi_j(X_i) \tag{5.2}$$

is an unbiased estimator of $[\lambda]_j$ for all $j \in \mathbb{Z}$.

5.2. Minimax theory

As in Chapter 3, we evaluate the performance of an arbitrary estimator $\widetilde{\lambda}$ by means of its maximum risk defined through

$$\sup_{\lambda\in\Lambda} \mathbb{E}[\|\widetilde{\lambda}-\lambda\|^2],$$

and aim at finding an estimator that attains the minimax risk defined through

$$\inf_{\widetilde{\lambda}}\sup_{\lambda\in\Lambda} \mathbb{E}[\|\widetilde{\lambda}-\lambda\|^2]$$

at least up to a multiplicative numerical constant. We work with the same abstract smoothness assumptions as in the previous chapters, that is, we will assume that λ belongs to some ellipsoid

$$\Lambda_\gamma^r = \{\lambda \in \mathbb{L}^2 : \lambda \geq 0 \text{ and } \|\lambda\|_\gamma^2 := \sum_{j\in\mathbb{Z}} \gamma_j |[\lambda]_j|^2 \leq r\}$$

for some $r > 0$ and a strictly positive symmetric sequence $\gamma = (\gamma_j)_{j\in\mathbb{Z}}$. We will impose the following assumption on γ which coincides with Assumption 3.1 in Chapter 3.

Assumption 5.1. *$\gamma = (\gamma_j)_{j\in\mathbb{Z}}$ is a strictly positive symmetric sequence with $\gamma_0 = 1$ and the sequence $(\gamma_n)_{n\in\mathbb{N}_0}$ is non-decreasing.*

In addition, we need the following assumption on the density f.

Assumption 5.2. *$f(x) \geq f_0 > 0$ for all $x \in [0,1]$.*

5.2.1. Upper bound

We start our investigation with the derivation of an upper bound for the estimator $\widehat{\lambda}_k$ with $\widehat{[\lambda]}_j$ defined in (5.2) and suitably chosen dimension parameter $k \in \mathbb{N}_0$.

Theorem 5.3. *Let Assumptions 5.1 and 5.2 hold. Then, for any $n \in \mathbb{N}$,*

$$\sup_{\lambda\in\Lambda_\gamma^r} \mathbb{E}[\|\widehat{\lambda}_{k_n^*}-\lambda\|^2] \lesssim \min_{k\in\mathbb{N}_0}\max\left\{\frac{1}{\gamma_k}, \frac{2k+1}{n}\right\} =: \Psi_n$$

for k_n^* chosen as $k_n^* := \operatorname{argmin}_{k\in\mathbb{N}_0}\max\left\{\frac{1}{\gamma_k},\frac{2k+1}{n}\right\}$.

Proof. We have the bias-variance decomposition

$$\mathbb{E}[\|\widehat{\lambda}_{k_n^*} - \lambda\|^2] \leq \sum_{|j|>k_n^*} |[\lambda]_j|^2 + \sum_{0\leq|j|\leq k_n^*} \mathbb{E}[|\widehat{[\lambda]}_j - [\lambda]_j|^2] =: \mathfrak{b}^2 + \mathfrak{v}.$$

From the definition of Λ_γ^r it can be deduced under the validity of Assumption 5.1 that

$$\mathfrak{b}^2 \leq r \cdot \gamma_{k_n^*}^{-1}.$$

For the variance term, we obtain for arbitrary $\lambda \in \Lambda_\gamma^r$ that

$$\begin{aligned}
\mathfrak{v} &= \sum_{0\leq|j|\leq k_n^*} \frac{1}{n^2T^2}\mathbb{E}\left[\left|\sum_{i=1}^n \frac{Y_i\varphi_j(X_i)}{f(X_i)} - \sum_{i=1}^n T[\lambda]_j\right|^2\right]\\
&= \sum_{0\leq|j|\leq k_n^*} \frac{1}{nT^2}\mathbb{E}\left[\left|\frac{Y_1\varphi_j(X_1)}{f(X_1)} - T[\lambda]_j\right|^2\right]\\
&\leq \sum_{0\leq|j|\leq k_n^*} \frac{1}{nT^2}\left\{T^2\int_0^1 \frac{\varphi_j^2(x)}{f(x)}\lambda^2(x)dx + T\int_0^1 \frac{\varphi_j^2(x)}{f(x)}\lambda(x)dx\right\}\\
&\leq 2\cdot\frac{2k_n^*+1}{n}\left(\frac{r}{f_0} + \frac{[\lambda]_0}{Tf_0}\right) \lesssim \frac{2k_n^*+1}{n}.
\end{aligned}$$

The statement of the theorem follows now by combining the obtained bounds for $\mathfrak{b}^2$ and $\mathfrak{v}$. □

5.2.2. Lower bound

Theorem 5.4. *Let Assumption 5.1 hold, and further assume that*

(C1) $\Gamma := \sum_{j\in\mathbb{Z}} \gamma_j^{-1} < \infty$, *and*

(C2) $0 < \eta^{-1} := \inf_{n\in\mathbb{N}} \Psi_n^{-1} \min\{\frac{1}{\gamma_{k_n^*}}, \frac{2k_n^*+1}{n}\}$ *for some* $1 \leq \eta < \infty$

where the quantities k_n^* *and* Ψ_n *are defined in Theorem 5.3. Then, for any* $n \in \mathbb{N}$,

$$\inf_{\widetilde{\lambda}} \sup_{\lambda\in\Lambda_\gamma^r} \mathbb{E}[\|\widetilde{\lambda} - \lambda\|^2] \gtrsim \Psi_n$$

where the infimum is taken over all estimators $\widetilde{\lambda}$ *of* λ.

Proof. For each $\theta = (\theta_j)_{0\leq|j|\leq k_n^*} \in \{\pm1\}^{2k_n^*+1}$ we define the function λ_θ exactly as in the proof of Theorem 3.3 with ζ from this proof replaced with $\zeta = \min\{\frac{1}{\Gamma\eta}, \frac{8\delta}{T\sqrt{r}}\}$ where $\delta = \frac{1}{2} - \frac{1}{2\sqrt{2}}$. Then one can proceed exactly as in the proof of Theorem 3.3 in order to show that $\lambda_\theta \in \Lambda_\gamma^r$ for all $\theta \in \{\pm1\}^{2k_n^*+1}$.

Consider the following reduction argument which holds for an arbitrary estimator $\widetilde{\lambda}$ of λ. In contrast to the argument in the proof of Theorem 3.3, it contains conditional instead of unconditional expectations. More precisely, denote $\boldsymbol{X} = (X_1,\ldots,X_n)$, $\boldsymbol{Y} = (Y_1,\ldots,Y_n)$. Then

$$\begin{aligned}
\sup_{\lambda\in\Lambda_\gamma^r} \mathbb{E}[\|\widetilde{\lambda} - \lambda\|^2] &\geq \frac{1}{2^{2k_n^*+1}} \sum_{\theta\in\{\pm1\}^{2k_n^*+1}} \sum_{0\leq|j|\leq k_n^*} \mathbb{E}[\mathbb{E}_\theta[|[\widetilde{\lambda} - \lambda_\theta]_j|^2|\mathbf{X}]]\\
&= \frac{1}{2^{2k_n^*+1}} \sum_{0\leq|j|\leq k_n^*} \sum_{\theta\in\{\pm1\}^{2k_n^*+1}} \frac{1}{2}\{\mathbb{E}[\mathbb{E}_\theta[|[\widetilde{\lambda} - \lambda_\theta]_j|^2|\mathbf{X}]] + \mathbb{E}[\mathbb{E}_{\theta^{(j)}}[|[\widetilde{\lambda} - \lambda_{\theta^{(j)}}]_j|^2|\mathbf{X}]]\} \qquad (5.3)
\end{aligned}$$

where for $\theta \in \{\pm1\}^{2k_n^*+1}$ the element $\theta^{(j)} \in \{\pm1\}^{2k_n^*+1}$ is defined by $\theta_k^{(j)} = \theta_k$ for $k \neq j$ and

$\theta_j^{(j)} = -\theta_j$. Consider the Hellinger affinity $\rho(\mathbb{P}_\theta^{\mathbf{Y}|\mathbf{X}}, \mathbb{P}_{\theta^{(j)}}^{\mathbf{Y}|\mathbf{X}}) := \int \sqrt{d\mathbb{P}_\theta^{\mathbf{Y}|\mathbf{X}} d\mathbb{P}_{\theta^{(j)}}^{\mathbf{Y}|\mathbf{X}}}$. We have

$$\begin{aligned}\rho(\mathbb{P}_\theta^{\mathbf{Y}|\mathbf{X}}, \mathbb{P}_{\theta^{(j)}}^{\mathbf{Y}|\mathbf{X}}) &\leq \int \frac{|[\widetilde{\lambda} - \lambda_\theta]_j|}{|[\lambda_\theta - \lambda_{\theta^{(j)}}]_j|} \sqrt{d\mathbb{P}_\theta^{\mathbf{Y}|\mathbf{X}} d\mathbb{P}_{\theta^{(j)}}^{\mathbf{Y}|\mathbf{X}}} + \int \frac{|[\widetilde{\lambda} - \lambda_{\theta^{(j)}}]_j|}{|[\lambda_\theta - \lambda_{\theta^{(j)}}]_j|} \sqrt{d\mathbb{P}_\theta^{\mathbf{Y}|\mathbf{X}} d\mathbb{P}_{\theta^{(j)}}^{\mathbf{Y}|\mathbf{X}}} \\ &\leq \left(\int \frac{|[\widetilde{\lambda} - \lambda_\theta]_j|^2}{|[\lambda_\theta - \lambda_{\theta^{(j)}}]_j|^2} d\mathbb{P}_\theta^{\mathbf{Y}|\mathbf{X}} \right)^{1/2} + \left(\int \frac{|[\widetilde{\lambda} - \lambda_{\theta^{(j)}}]_j|^2}{|[\lambda_\theta - \lambda_{\theta^{(j)}}]_j|^2} d\mathbb{P}_{\theta^{(j)}}^{\mathbf{Y}|\mathbf{X}} \right)^{1/2}.\end{aligned}$$

By means of the estimate $(a+b)^2 \leq 2a^2 + 2b^2$ we obtain

$$\frac{1}{2} |[\lambda_\theta - \lambda_{\theta^{(j)}}]_j|^2 \rho^2(\mathbb{P}_\theta^{\mathbf{Y}|\mathbf{X}}, \mathbb{P}_{\theta^{(j)}}^{\mathbf{Y}|\mathbf{X}}) \leq \mathbb{E}_\theta[|[\widetilde{\lambda} - \lambda_\theta]_j|^2 | \mathbf{X}] + \mathbb{E}_{\theta^{(j)}}[|[\widetilde{\lambda} - \lambda_{\theta^{(j)}}]_j|^2 | \mathbf{X}].$$

Recall the definition of the Hellinger distance,

$$H(\mathbb{P}_\theta^{\mathbf{Y}|\mathbf{X}}, \mathbb{P}_{\theta^{(j)}}^{\mathbf{Y}|\mathbf{X}}) := \left(\int \left[\sqrt{\mathbb{P}_\theta^{\mathbf{Y}|\mathbf{X}}} - \sqrt{\mathbb{P}_{\theta^{(j)}}^{\mathbf{Y}|\mathbf{X}}} \right]^2 \right)^{1/2}.$$

Let N_i be a PPP on $[0, T]$ with constant intensity equal to $\lambda(X_i)$. Consider the transformation which maps the point process N_i to $Y_i = N_i([0, T])$. Using Lemma A.4 we can conclude

$$\begin{aligned}H^2(\mathbb{P}_\theta^{Y_i|\mathbf{X}}, \mathbb{P}_{\theta^{(j)}}^{Y_i|\mathbf{X}}) \leq H^2(\mathbb{P}_\theta^{N_i|\mathbf{X}}, \mathbb{P}_{\theta^{(j)}}^{N_i|\mathbf{X}}) &\leq \int_0^T (\sqrt{\lambda_\theta(X_i)} - \sqrt{\lambda_{\theta^{(j)}}(X_i)})^2 \\ &= \int_0^T \frac{|\lambda_\theta(X_i) - \lambda_{\theta^{(j)}}(X_i)|^2}{(\sqrt{\lambda_\theta(X_i)} + \sqrt{\lambda_{\theta^{(j)}}(X_i)})^2} \\ &\leq \frac{T\zeta\sqrt{r}}{8n\delta} \leq \frac{1}{n}.\end{aligned}$$

Since $Y_1, \ldots, Y_n$ are independent conditionally on $X_1, \ldots, X_n$ we obtain by Lemma A.3 that

$$H^2(\mathbb{P}_\theta^{\mathbf{Y}|\mathbf{X}}, \mathbb{P}_{\theta^{(j)}}^{\mathbf{Y}|\mathbf{X}}) \leq \sum_{i=1}^n H^2(\mathbb{P}_\theta^{Y_i|\mathbf{X}}, \mathbb{P}_{\theta^{(j)}}^{Y_i|\mathbf{X}}) \leq 1.$$

Hence the relation $\rho(\mathbb{P}_\theta^{\mathbf{Y}|\mathbf{X}}, \mathbb{P}_{\theta^{(j)}}^{\mathbf{Y}|\mathbf{X}}) = 1 - \frac{1}{2} H^2(\mathbb{P}_\theta^{\mathbf{Y}|\mathbf{X}}, \mathbb{P}_{\theta^{(j)}}^{\mathbf{Y}|\mathbf{X}})$ implies $\rho(\mathbb{P}_\theta^{\mathbf{Y}|\mathbf{X}}, \mathbb{P}_{\theta^{(j)}}^{\mathbf{Y}|\mathbf{X}}) \geq \frac{1}{2}$. Putting this estimate into the reduction scheme (5.3) yields

$$\begin{aligned}\sup_{\lambda \in \Lambda_\gamma^r} \mathbb{E}[\|\widetilde{\lambda} - \lambda\|^2] &\geq \frac{1}{2^{2k_n^*+1}} \sum_{\theta \in \{\pm 1\}^{2k_n^*+1}} \sum_{0 \leq |j| \leq k_n^*} \frac{1}{2} \mathbb{E}[\mathbb{E}_\theta[|[\widetilde{\lambda} - \lambda_\theta]_j|^2 | \mathbf{X}] + \mathbb{E}_{\theta^{(j)}}[|[\widetilde{\lambda} - \lambda_{\theta^{(j)}}]_j|^2 | \mathbf{X}]] \\ &\geq \frac{1}{16} \sum_{0 \leq |j| \leq k_n^*} |[\lambda_\theta - \lambda_{\theta^{(j)}}]_j|^2 \geq \frac{\zeta r}{64} \sum_{0 \leq |j| \leq k_n^*} \frac{1}{n} = \frac{\zeta r}{64} \cdot \frac{2k_n^* + 1}{n}.\end{aligned}$$

Since the last estimate holds for arbitrary $\widetilde{\lambda}$, we obtain the claim assertion thanks to Assumption (C2). □

Theorems 5.3 and 5.4 show that under the stated assumptions the minimax rate of convergence is given by Ψ_n, and that this rate is attained by the estimator $\widehat{\lambda}_{k_n^*}$. For the examples of sequences γ considered in Chapter 3, we obtain exactly the same rates for the Poisson regression model as for intensity estimation. Note that, as in the previous chapters, the estimator $\widehat{\lambda}_{k_n^*}$ is not fully data-driven but depends on *a priori* knowledge concerning the class of potential regression functions, namely the knowledge of the sequence γ.

5.3. Adaptive estimation

In order to construct an adaptive estimator we make again use of the model selection paradigm already applied in Chapters 3 and 4. Again, our derivation of the adaptive estimator is split into two parts. In the first part, we will construct a preliminary estimator whose definition is based on the knowledge of an upper bound of the regression function. In the second part, we replace this known upper bound by an appropriate estimator of $\|\lambda\|_\infty$ in order to obtain a fully data-driven estimator.

5.3.1. Known upper bound of the regression function

Denoting $\langle s,t\rangle := \int_0^1 s(x)t(x)dx$ for $s,t\in\mathbb{L}^2$, we define the contrast function

$$\Upsilon_n(t) := \|t\|^2 - 2\langle\widehat{\lambda}_n, t\rangle, \qquad t\in\mathbb{L}^2.$$

Our definition of the penalty term in this section is based on the validity of the following assumption.

Assumption 5.5. *We know some $\xi>0$ such that $\|\lambda\|_\infty\le\xi$.*

Based on the knowledge of ξ, we define the penalty via

$$\textsc{Pen}_k := 24\mu\cdot\frac{2k+1}{n} + 400\mu\cdot(2k+1)\cdot\frac{\log(n+2)}{nT} \tag{5.4}$$

where $\mu = \frac{1\vee\xi^2}{f_0^2}\ge 1$. The resulting data-driven choice of the dimension parameter $k\in\mathbb{N}_0$ is as in the previous chapters defined as the minimizer of the penalized contrast, that is

$$\widetilde{k}_n := \operatorname*{argmin}_{0\le k\le n}\{\Upsilon_n(\widehat{\lambda}_k) + \textsc{Pen}_k\}.$$

Theorem 5.6. *Let Assumptions 5.1, 5.2, and 5.5 hold. Then, for any $n\in\mathbb{N}$,*

$$\sup_{\substack{\lambda\in\Lambda_\gamma^r\\ \|\lambda\|_\infty\le\xi}} \mathbb{E}[\|\widehat{\lambda}_{\widetilde{k}_n}-\lambda\|^2] \lesssim \min_{0\le k\le n}\max\left\{\gamma_k^{-1},\textsc{Pen}_k\right\} + \frac{1}{n}.$$

Proof. Using the same arguments as in the proof of Theorem 3.9 we can derive the inequality chain

$$\|\widehat{\lambda}_{\widetilde{k}_n}-\lambda\|^2 \le \|\lambda_k-\lambda\|^2 + 2\langle\widehat{\lambda}_n-\lambda_n, \widehat{\lambda}_{\widetilde{k}_n}-\lambda_k\rangle + \textsc{Pen}_k - \textsc{Pen}_{\widetilde{k}_n}$$

with $\lambda_k := \sum_{0\le|j|\le k}[\lambda]_j\varphi_j$ for $k\in\{0,\ldots,n\}$. Putting

$$\widetilde{[\lambda]}_j := \frac{1}{n}\sum_{i=1}^n\frac{\lambda(X_i)}{f(X_i)}\varphi_j(X_i) \qquad\text{and}\qquad \widetilde{\lambda}_n := \sum_{0\le|j|\le n}\widetilde{[\lambda]}_j\varphi_j$$

we obtain

$$\|\widehat{\lambda}_{\widetilde{k}_n}-\lambda\|^2 \le \|\lambda_k-\lambda\|^2 + 2\langle\widehat{\Theta}_n, \widehat{\lambda}_{\widetilde{k}_n}-\lambda_k\rangle + 2\langle\widetilde{\Theta}_n, \widehat{\lambda}_{\widetilde{k}_n}-\lambda_k\rangle + \textsc{Pen}_k - \textsc{Pen}_{\widetilde{k}_n}$$

where $\widehat{\Theta}_n := \widehat{\lambda}_n - \widetilde{\lambda}_n$ and $\widetilde{\Theta}_n := \widetilde{\lambda}_n - \lambda_n$. Set $\mathcal{B}_k := \{\lambda\in\mathcal{S}_k : \|\lambda\|^2\le 1\}$. Using the estimate $2uv\le\tau u^2+\tau^{-1}v^2$ for positive τ we can conclude

$$\|\widehat{\lambda}_{\widetilde{k}_n}-\lambda\|^2 \le \|\lambda_k-\lambda\|^2 + 2\tau\|\widehat{\lambda}_{\widetilde{k}_n}-\lambda_k\|^2 + \tau^{-1}\sup_{t\in\mathcal{B}_{k\vee\widetilde{k}_n}}|\langle\widehat{\Theta}_n,t\rangle|^2$$

$$+\tau^{-1}\sup_{t\in\mathcal{B}_{k\vee\widetilde{k}_n}}|\langle\widetilde{\Theta}_n,t\rangle|^2+\textsc{Pen}_k-\textsc{Pen}_{\widetilde{k}_n}.$$

Note that $\|\widehat{\lambda}_{\widetilde{k}_n}-\lambda_k\|^2\le 2\|\widehat{\lambda}_{\widetilde{k}_n}-\lambda\|^2+2\|\lambda_k-\lambda\|^2$ and $\|\lambda-\lambda_k\|^2\le r\gamma_k^{-1}$ for all $\lambda\in\Lambda_\gamma^r$ thanks to Assumption 5.1. Taking $\tau=1/8$ we obtain

$$\begin{aligned}
\|\widehat{\lambda}_{\widetilde{k}_n}-\lambda\|^2 &\le 3r\gamma_k^{-1}+16\sup_{t\in\mathcal{B}_{k\vee\widetilde{k}_n}}|\langle\widehat{\Theta}_n,t\rangle|^2+16\sup_{t\in\mathcal{B}_{k\vee\widetilde{k}_n}}|\langle\widetilde{\Theta}_n,t\rangle|^2+2\textsc{Pen}_k-2\textsc{Pen}_{\widetilde{k}_n}\\
&\le 3r\gamma_k^{-1}+16\left(\sup_{t\in\mathcal{B}_{k\vee\widetilde{k}_n}}|\langle\widehat{\Theta}_n,t\rangle|^2-50\mu\cdot\frac{(2(k\vee\widetilde{k}_n)+1)\log(n+2)}{nT}\right)_+\\
&\quad+16\left(\sup_{t\in\mathcal{B}_{k\vee\widetilde{k}_n}}|\langle\widetilde{\Theta}_n,t\rangle|^2-3\mu\cdot\frac{2(k\vee\widetilde{k}_n)+1}{n}\right)_+\\
&\quad+800\mu\cdot\frac{(2(k\vee\widetilde{k}_n)+1)\log(n+2)}{nT}+48\mu\cdot\frac{2(k\vee\widetilde{k}_n)+1}{n}+2\textsc{Pen}_k-2\textsc{Pen}_{\widetilde{k}_n}.
\end{aligned}$$

By definition of the penalty and roughly bounding the brackets $(\ldots)_+$ by summing over all potential values of k, this implies

$$\begin{aligned}
\|\widehat{\lambda}_{\widetilde{k}_n}-\lambda\|^2 \le 3r\gamma_k^{-1}&+16\sum_{k=0}^{n}\left(\sup_{t\in\mathcal{B}_k}|\langle\widehat{\Theta}_n,t\rangle|^2-50\mu\cdot\frac{(2k+1)\log(n+2)}{nT}\right)_+\\
&+16\sum_{k=0}^{n}\left(\sup_{t\in\mathcal{B}_k}|\langle\widetilde{\Theta}_n,t\rangle|^2-3\mu\cdot\frac{2k+1}{n}\right)_+ +4\textsc{Pen}_k.
\end{aligned}$$

Consequently, taking expectations and into account that the last estimate holds for arbitrary k, we obtain

$$\begin{aligned}
\sup_{\substack{\lambda\in\Lambda_\gamma^r\\ \|\lambda\|_\infty\le\xi}}\mathbb{E}[\|\widehat{\lambda}_{\widetilde{k}_n}-\lambda\|^2] &\le \min_{0\le k\le n}\{3r\gamma_k^{-1}+4\textsc{Pen}_k\}\\
&\quad+16\sup_{\substack{\lambda\in\Lambda_\gamma^r\\ \|\lambda\|_\infty\le\xi}}\sum_{k=0}^{n}\underbrace{\mathbb{E}\left[\left(\sup_{t\in\mathcal{B}_k}|\langle\widehat{\Theta}_n,t\rangle|^2-50\mu\cdot\frac{(2k+1)\log(n+2)}{nT}\right)_+\right]}_{=:\square_k}\\
&\quad+16\sup_{\substack{\lambda\in\Lambda_\gamma^r\\ \|\lambda\|_\infty\le\xi}}\sum_{k=0}^{n}\underbrace{\mathbb{E}\left[\left(\sup_{t\in\mathcal{B}_k}|\langle\widetilde{\Theta}_n,t\rangle|^2-3\mu\cdot\frac{2k+1}{n}\right)_+\right]}_{=:\blacksquare_k}. \qquad (5.5)
\end{aligned}$$

We now use Lemmata 5.9 and 5.11 in order to bound the terms $\square_k$ and $\blacksquare_k$ which yields for $k\in\{0,\ldots,n\}$ that

$$\begin{aligned}
\square_k &\le K_1'\left\{\frac{(2k+1)\mu}{nT}\exp(-2\log(n+2))+\frac{(2k+1)\mu}{n^2T^2}\exp(-K_2'\sqrt{nT})\right\}\quad\text{and}\\
\blacksquare_k &\le K_1\left\{\frac{\mu}{n}\exp(-K_2(2k+1))+\frac{2k+1}{n^2}\exp(-K_3\sqrt{n})\right\}.
\end{aligned}$$

Putting these estimates into (5.5), using the estimate $2k+1\le 3n$ for $k\le n$, and the convergence of $\sum_{k=0}^{\infty}\exp(-K_2(2k+1))$, we obtain that

$$\sup_{\substack{\lambda\in\Lambda_\gamma^r\\ \|\lambda\|_\infty\le\xi}}\mathbb{E}[\|\widehat{\lambda}_{\widetilde{k}_n}-\lambda\|^2]\lesssim\min_{0\le k\le n}\max\{\gamma_k^{-1},\textsc{Pen}_k\}+\frac{1}{n}+\exp(-\kappa\sqrt{n})$$

with $\kappa = K_2'\sqrt{T} \wedge K_3$. $\square$

Since the penalty term $\textsc{Pen}_k$ differs from the variance term in Theorem 5.3 by addition of an extra logarithmic factor and $k_n^* \leq n$, the estimator $\widehat{\lambda}_{\widetilde{k}_n}$ is rate optimal only up to a logarithmic factor.

5.3.2. Unknown upper bound of the regression function

We now propose an adaptive estimator of the regression function λ that does not depend on *a priori* knowledge of an upper bound for $\|\lambda\|_\infty$, and is thus fully data-driven. Not surprisingly, the key idea is to replace the quantity ξ in the definition of the penalty in (5.4) by an appropriate estimator of $\|\lambda\|_\infty$. For the construction of the estimator of $\|\lambda\|_\infty$, we follow an approach that was used in [Com01] in the context of adaptive estimation of the spectral density from a stationary Gaussian sequence. More precisely, the estimator of $\|\lambda\|_\infty$ is obtained as the plug-in estimator $\|\widehat{\lambda}_{\mathfrak{m}}\|_\infty$ where $\widehat{\lambda}_{\mathfrak{m}}$ is a suitable projection estimator of λ in some space of piecewise polynomials. The following brief disgression provides a short overview of piecewise polynomials.

Piecewise polynomials

The presentation in this section is based on [BM97] and provides in a nutshell the basic properties of piecewise polynomials that we will use in the following. As in the whole chapter, we restrict ourselves to piecewise polynomials defined on $[0, 1]$. The linear space $\mathcal{P}_{\mathfrak{m}}$ of piecewise polynomials is characterized by the 'model' $\mathfrak{m} = (q, \{b_0, \ldots, b_D : 0 = b_0 < b_1 < \ldots < b_D = 1\})$. Here, $q \in \mathbb{N}_0$ is the maximal degree of the admissible polynomials and the knots $b_0, b_1, \ldots, b_D$ define a partition of $[0, 1]$ into D intervals. The dimension of $\mathcal{P}_{\mathfrak{m}}$ is $\mathbb{D}_{\mathfrak{m}} = D \cdot (q + 1)$.

The point of origin in order to find a convenient basis are the *Legendre polynomials*. We recall that the set of Legendre polynomials $\{Q_j\}_{j\in\mathbb{N}_0}$ is a family of orthogonal polynomials in $\mathbb{L}^2([-1, 1], dx)$ where each Q_j is a polynomial of degree j with

$$|Q_j(x)| \leq 1 \quad \text{for all } x \in [-1, 1], \quad Q_j(1) = 1, \quad \int_{-1}^{1} Q_j^2(t)dt = \frac{2}{2j + 1}.$$

Hence, $\{R_j\}_{j\in\mathbb{N}_0}$ with

$$R_j(x) = \sqrt{\frac{2j + 1}{b - a}} Q_j \left(\frac{2}{b - a} x - \left(1 - \frac{2a}{b - a} \right) \right)$$

is an orthonormal basis for the space of polynomials on $[a, b]$ (cf. [DL93], p. 328 for an explicit representation of the polynomials R_j). If P is a polynomial of degree $\leq q$ with representation $P(x) = \sum_{j=0}^q a_j R_j(x)$, then

$$|P(x)|^2 \leq \left(\sum_{j=0}^{q} a_j^2 \right) \left(\sum_{j=0}^{q} \frac{2j + 1}{b - a} \right) = \frac{(q + 1)^2}{b - a} \left(\sum_{j=0}^{q} a_j^2 \right),$$

and thus $\|P\|_\infty \leq \frac{q+1}{\sqrt{b-a}} \cdot \|P\|$. For our purposes, it is sufficient to consider *regular piecewise polynomials* where $b_i = i/M$ for some $M \in \mathbb{N}$ and $i = 0, \ldots, M$. In this case, one can write $\mathfrak{m} = (q, M)$ instead of $\mathfrak{m} = (q, \{0, 1/M, \ldots, 1\})$. For a space $\mathcal{P}_{\mathfrak{m}}$ of piecewise polynomials we denote with $\{\varphi_\eta\}_{\eta\in\mathcal{I}_{\mathfrak{m}}}$ the orthonormal basis obtained from transformed Legendre polynomials as above (then, $|\mathcal{I}_{\mathfrak{m}}| = \mathbb{D}_{\mathfrak{m}} = M \cdot (q + 1)$ if $\mathfrak{m} = (q, M)$).

Definition of the fully data-driven estimator

Let $\widehat{\lambda}_{\mathfrak{m}}$ be the projection estimator of λ on the space of regular piecewise polynomials $\mathcal{P}_{\mathfrak{m}}$ with $\mathfrak{m} = (q, M)$ as introduced above. We substitute the quantity ξ in the definition of the penalty term in the previous section with $\|\widehat{\lambda}_{\mathfrak{m}}\|_\infty$. Precise assumptions on the 'model' $\mathfrak{m}$, that is, on q and M will be stated in Theorem 5.7 below. We replace the deterministic penalty $\textsc{Pen}_k$ by the random penalty

$$\widehat{\textsc{Pen}}_k = 384\widehat{\mu} \cdot \frac{2k+1}{n} + 6400\widehat{\mu} \cdot (2k+1) \cdot \frac{\log(n+2)}{nT} \tag{5.6}$$

where $\widehat{\mu} = \frac{1 \vee \|\widehat{\lambda}_{\mathfrak{m}}\|_\infty^2}{f_0^2} \geq 1$. Keeping the contrast function Υ_n from Section 5.3.1, we define

$$\widehat{k}_n := \operatorname*{argmin}_{0 \leq k \leq n} \{\Upsilon_n(\widehat{\lambda}_k) + \widehat{\textsc{Pen}}_k\}.$$

The following theorem provides a risk bound for the fully data-driven estimator $\widehat{\lambda}_{\widehat{k}_n}$.

Theorem 5.7. *Let Assumptions 5.1 and 5.2 hold, and further assume that*

($\mathfrak{m}$*1)* $\|\lambda - \lambda_{\mathfrak{m}}\|_\infty \leq \frac{1}{4}\|\lambda\|_\infty$ *where* $\lambda_{\mathfrak{m}}$ *denotes the projection of* λ *on* $\mathcal{P}_{\mathfrak{m}}$*, and*

($\mathfrak{m}$*2) the model* $\mathfrak{m} = (q, M)$ *in the definition of the auxiliary estimator* $\widehat{\lambda}_{\mathfrak{m}}$ *satisfies*

$$\mathbb{D}_{\mathfrak{m}} \leq \frac{1}{4\sqrt{10}} \cdot \frac{\sqrt{f_0 \wedge f_0^2 T}}{(q+1)^{3/2}} \cdot \frac{\sqrt{n}}{\log(n+2)}.$$

Then, for any $n \in \mathbb{N}$*,*

$$\mathbb{E}[\|\widehat{\lambda}_{\widehat{k}_n} - \lambda\|^2] \lesssim \min_{0 \leq k \leq n} \max\left\{\frac{1}{\gamma_k}, \textsc{Pen}_k\right\} + \frac{1}{n}$$

where $\textsc{Pen}_k = 24\mu \cdot \frac{2k+1}{n} + 400\mu \cdot (2k+1) \cdot \frac{\log(n+2)}{nT}$ *and* $\mu = \frac{1 \vee \xi^2}{f_0^2} \geq 1$.

Remark 5.8. The additonal Assumptions ($\mathfrak{m}$1) and ($\mathfrak{m}$2) are inspired by the assumptions made in Theorem 2 of [Com01].

Proof. Introduce the event $\Xi := \left\{\left|\frac{\|\widehat{\lambda}_{\mathfrak{m}}\|_\infty \vee 1}{\|\lambda\|_\infty \vee 1} - 1\right| < \frac{3}{4}\right\}$. It is readily verified that on Ξ it holds that

$$\|\lambda\|_\infty \vee 1 \leq 4(\|\widehat{\lambda}_{\mathfrak{m}}\|_\infty \vee 1) \qquad \text{and} \qquad \|\widehat{\lambda}_{\mathfrak{m}}\|_\infty \vee 1 \leq \frac{7}{4}(\|\lambda\|_\infty \vee 1).$$

These estimates will be used below without further reference. We consider the decomposition

$$\mathbb{E}[\|\widehat{\lambda}_{\widehat{k}_n} - \lambda\|^2] \leq \mathbb{E}[\|\widehat{\lambda}_{\widehat{k}_n} - \lambda\|^2 \mathbb{1}_\Xi] + \mathbb{E}[\|\widehat{\lambda}_{\widehat{k}_n} - \lambda\|^2 \mathbb{1}_{\Xi^c}] =: \Box_1 + \Box_2.$$

In the sequel, we will derive uniform upper bounds for the terms $\Box_1$ and $\Box_2$, respectively.
Uniform upper bound for $\Box_1$*:* In analogy to the proof of Theorem 5.6 one can derive

$$\|\widehat{\lambda}_{\widehat{k}_n} - \lambda\|^2 \leq \|\lambda_k - \lambda\|^2 + 2\tau\|\widehat{\lambda}_{\widehat{k}_n} - \lambda_k\|^2 + \tau^{-1} \sup_{t \in \mathcal{B}_{k \vee \widehat{k}_n}} |\langle \widehat{\Theta}_n, t\rangle|^2$$
$$+ \tau^{-1} \sup_{t \in \mathcal{B}_{k \vee \widehat{k}_n}} |\langle \widetilde{\Theta}_n, t\rangle|^2 + \widehat{\textsc{Pen}}_k - \widehat{\textsc{Pen}}_{\widehat{k}_n}$$

for all $k \in \{0, \ldots, n\}$ and all the appearing quantities are defined exactly as in the proof of Theorem 5.6. Using the same arguments as in that proof, one obtains by specializing with $\tau = 1/8$

and setting $\mu = \frac{1 \vee \xi^2}{f_0^2}$ (recall that ξ satisfies $\|\lambda\|_\infty \leq \xi$) that

$$\begin{aligned}
\|\widehat{\lambda}_{\widehat{k}_n} - \lambda\|^2 &\leq 3r\gamma_k^{-1} + 16 \sup_{t \in \mathcal{B}_{k \vee \widehat{k}_n}} |\langle \widehat{\Theta}_n, t\rangle|^2 + 16 \sup_{t \in \mathcal{B}_{k \vee \widehat{k}_n}} |\langle \widetilde{\Theta}_n, t\rangle|^2 + 2\widehat{\text{PEN}}_k - 2\widehat{\text{PEN}}_{\widehat{k}_n} \\
&\leq 3r\gamma_k^{-1} + 16 \left(\sup_{t \in \mathcal{B}_{k \vee \widehat{k}_n}} |\langle \widehat{\Theta}_n, t\rangle|^2 - 50\mu \cdot \frac{(2(k \vee \widehat{k}_n) + 1)\log(n+2)}{nT} \right)_+ \\
&\quad + 16 \left(\sup_{t \in \mathcal{B}_{k \vee \widehat{k}_n}} |\langle \widetilde{\Theta}_n, t\rangle|^2 - 3\mu \cdot \frac{2(k \vee \widehat{k}_n) + 1}{n} \right)_+ \\
&\quad + 800\mu \cdot \frac{(2(k \vee \widehat{k}_n) + 1)\log(n+2)}{nT} + 48\mu \cdot \frac{2(k \vee \widehat{k}_n) + 1}{n} + 2\widehat{\text{PEN}}_k - 2\widehat{\text{PEN}}_{\widehat{k}_n}.
\end{aligned}$$

By definition of Ξ and the random penalty function, we obtain using the estimate $2(k \vee \widehat{k}_n) + 1 \leq 2k + 2\widehat{k}_n + 2$ that

$$\begin{aligned}
\|\widehat{\lambda}_{\widehat{k}_n} - \lambda\|^2 \mathbb{1}_\Xi &\leq 3r\gamma_k^{-1} + 16 \left(\sup_{t \in \mathcal{B}_{k \vee \widehat{k}_n}} |\langle \widehat{\Theta}_n, t\rangle|^2 - 50\mu \cdot \frac{(2(k \vee \widehat{k}_n) + 1)\log(n+2)}{nT} \right)_+ \\
&\quad + 16 \left(\sup_{t \in \mathcal{B}_{k \vee \widehat{k}_n}} |\langle \widetilde{\Theta}_n, t\rangle|^2 - 3\mu \cdot \frac{2(k \vee \widehat{k}_n) + 1}{n} \right)_+ + 100\text{PEN}_k.
\end{aligned}$$

Bounding the terms in the brackets $(\ldots)_+$ by summing over all admissible values of k and taking expectations on both sides yield

$$\begin{aligned}
\sup_{\substack{\lambda \in \Lambda_\gamma^r \\ \|\lambda\|_\infty \leq \xi}} \mathbb{E}[\|\widehat{\lambda}_{\widehat{k}_n} - \lambda_k\|^2 \mathbb{1}_\Xi] &\leq 3r\gamma_k^{-1} + 100\text{PEN}_k \\
&\quad + 16 \sup_{\substack{\lambda \in \Lambda_\gamma^r \\ \|\lambda\|_\infty \leq \xi}} \sum_{k=0}^{n} \mathbb{E}\left[\left(\sup_{t \in \mathcal{B}_{k \vee \widehat{k}_n}} |\langle \widehat{\Theta}_n, t\rangle|^2 - 50\mu \cdot \frac{(2k+1)\log(n+2)}{nT} \right)_+ \right] \\
&\quad + 16 \sup_{\substack{\lambda \in \Lambda_\gamma^r \\ \|\lambda\|_\infty \leq \xi}} \sum_{k=0}^{n} \mathbb{E}\left[\left(\sup_{t \in \mathcal{B}_{k \vee \widehat{k}_n}} |\langle \widetilde{\Theta}_n, t\rangle|^2 - 3\mu \cdot \frac{2k+1}{n} \right)_+ \right].
\end{aligned}$$

Applying Lemmata 5.9 and 5.11 as in the proof of Theorem 5.6 finally implies

$$\sup_{\substack{\lambda \in \Lambda_\gamma^r \\ \|\lambda\|_\infty \leq \xi}} \mathbb{E}[\|\widehat{\lambda}_{\widehat{k}_n} - \lambda\|^2 \mathbb{1}_\Xi] \lesssim \min_{0 \leq k \leq n} \max\{\gamma_k^{-1}, \text{PEN}_k\} + \frac{1}{n} + \exp(-\kappa\sqrt{n})$$

for some numerical constant $\kappa > 0$.

Uniform upper bound for $\square_2$*:* For $\lambda \in \Lambda_\gamma^r$, take note of the estimate

$$\begin{aligned}
\mathbb{E}[\|\widehat{\lambda}_{\widehat{k}_n} - \lambda\|^2 \mathbb{1}_{\Xi^c}] &\leq \mathbb{E}[\|\widehat{\lambda}_{\widehat{k}_n} - \lambda_{\widehat{k}_n}\|^2 \mathbb{1}_{\Xi^c}] + \mathbb{E}[\|\lambda_{\widehat{k}_n} - \lambda\|^2 \mathbb{1}_{\Xi^c}] \\
&\leq \mathbb{P}(\Xi^c)^{1/2} \sum_{0 \leq |j| \leq \widehat{k}_n} \mathbb{E}[|\widehat{[\lambda]}_j - [\lambda]_j|^4]^{1/2} + r\mathbb{P}(\Xi^c) \\
&\lesssim \frac{2\widehat{k}_n + 1}{n} \mathbb{P}(\Xi^c)^{1/2} + r\mathbb{P}(\Xi^c)
\end{aligned}$$

where $\lambda_{\widehat{k}_n} = \sum_{0 \leq |j| \leq \widehat{k}_n} [\lambda]_j \mathbf{e}_j$ and we used Theorem B.1 with $p = 4$. Because $2k + 1 \leq 3n$ for all

$k \in \{0, \dots, n\}$ it suffices to show that $\mathbb{P}(\Xi^c) \lesssim n^{-2}$. Note that we have

$$|\|\widehat{\lambda}_{\mathfrak{m}}\|_\infty - \|\lambda\|_\infty| \le \|\widehat{\lambda}_{\mathfrak{m}} - \lambda_{\mathfrak{m}}\|_\infty + \|\lambda_{\mathfrak{m}} - \lambda\|_\infty \le \|\widehat{\lambda}_{\mathfrak{m}} - \lambda_{\mathfrak{m}}\|_\infty + \frac{1}{4}\|\lambda\|_\infty \tag{5.7}$$

where the last estimate holds due to Assumption ($\mathfrak{m}1$). Put $I_j = [\frac{j-1}{M}, \frac{j}{M}]$ for $j = 1, \dots, M$ and let $\{\varphi_\eta\}_{\eta \in \mathcal{I}_{\mathfrak{m}}}$ be the basis of $\mathcal{P}_{\mathfrak{m}}$ consisting of transformed Legendre polynomials (see the disgression on piecewise polynomials above). We have

$$\begin{aligned}
\|\widehat{\lambda}_{\mathfrak{m}} - \lambda_{\mathfrak{m}}\|_\infty &= \sup_{1 \le j \le M} \|(\widehat{\lambda}_{\mathfrak{m}} - \lambda_{\mathfrak{m}})\mathbb{1}_{I_j}\|_\infty \\
&\le \sup_{1 \le j \le M} (q+1)\sqrt{M}\|(\widehat{\lambda}_{\mathfrak{m}} - \lambda_{\mathfrak{m}})\mathbb{1}_{I_j}\| \\
&\le (q+1)^{3/2}\sqrt{M} \sup_{\eta \in \mathcal{I}_{\mathfrak{m}}} |\langle \widehat{\lambda}_{\mathfrak{m}} - \lambda_{\mathfrak{m}}, \varphi_\eta \rangle| \\
&\le (q+1)^{3/2}\sqrt{M}\{ \sup_{\eta \in \mathcal{I}_{\mathfrak{m}}} |\langle \widehat{\lambda}_{\mathfrak{m}} - \mathbb{E}[\widehat{\lambda}_{\mathfrak{m}}|\mathbf{X}], \varphi_\eta \rangle| + \sup_{\eta \in \mathcal{I}_{\mathfrak{m}}} |\langle \mathbb{E}[\widehat{\lambda}_{\mathfrak{m}}|\mathbf{X}] - \lambda_{\mathfrak{m}}, \varphi_\eta \rangle| \} \\
&= (q+1)\sqrt{\mathbb{D}_{\mathfrak{m}}}\{ \sup_{\eta \in \mathcal{I}_{\mathfrak{m}}} |\nu(\varphi_\eta)| + \sup_{\eta \in \mathcal{I}_{\mathfrak{m}}} |\widetilde{\nu}(\varphi_\eta)| \}
\end{aligned}$$

where $\nu(\varphi_\eta) := \langle \widehat{\lambda}_{\mathfrak{m}} - \mathbb{E}[\widehat{\lambda}_{\mathfrak{m}}|\mathbf{X}], \varphi_\eta \rangle$, and $\widetilde{\nu}(\varphi_\eta) := \langle \mathbb{E}[\widehat{\lambda}_{\mathfrak{m}}|\mathbf{X}] - \lambda_{\mathfrak{m}}, \varphi_\eta \rangle$. Using (5.7) and the estimate $|a \vee 1 - b \vee 1| \le |a - b|$, we obtain

$$\begin{aligned}
\mathbb{P}(\Xi^c) &= \mathbb{P}(|\|\widehat{\lambda}_{\mathfrak{m}}\|_\infty \vee 1 - \|\lambda\|_\infty \vee 1| \ge 3/4 \cdot (\|\lambda\|_\infty \vee 1)) \\
&\le \mathbb{P}(\|\widehat{\lambda}_{\mathfrak{m}} - \lambda_{\mathfrak{m}}\|_\infty \ge 1/2 \cdot (\|\lambda\|_\infty \vee 1)) \\
&\le \mathbb{P}((q+1)\sqrt{\mathbb{D}_{\mathfrak{m}}} \sup_{\eta \in \mathcal{I}_{\mathfrak{m}}} |\nu_n(\varphi_\eta)| \ge 1/4 \cdot (\|\lambda\|_\infty \vee 1)) \\
&\quad + \mathbb{P}((q+1)\sqrt{\mathbb{D}_{\mathfrak{m}}} \sup_{\eta \in \mathcal{I}_{\mathfrak{m}}} |\widetilde{\nu}_n(\varphi_\eta)| \ge 1/4 \cdot (\|\lambda\|_\infty \vee 1)) \\
&\le \sum_{\eta \in \mathcal{I}_{\mathfrak{m}}} \left[\mathbb{P}\left(\nu_n(\varphi_\eta) \ge \xi\right) + \mathbb{P}\left(-\nu_n(\varphi_\eta) \ge \xi\right) + \mathbb{P}\left(\widetilde{\nu}_n(\varphi_\eta) \ge \xi\right) + \mathbb{P}\left(-\widetilde{\nu}_n(\varphi_\eta) \ge \xi\right) \right],
\end{aligned}$$

where $\xi = \frac{\|\lambda\|_\infty \vee 1}{4 \cdot (q+1) \cdot \sqrt{\mathbb{D}_{\mathfrak{m}}}}$. We will now obtain upper bounds for the probabilities on the right-hand side via Bernstein type inequalities. Note that $\|\varphi_\eta\| = 1$ and $\|\varphi_\eta\|_\infty \le \sqrt{(q+1)\mathbb{D}_{\mathfrak{m}}}$. By application of Proposition B.7 we obtain

$$\begin{aligned}
\mathbb{P}\left(\pm\nu_n(\varphi_\eta) \ge \xi\right) &\le \exp\left(-\frac{nT\xi^2}{2\|\varphi_\eta\|_\infty^2 \|\lambda\|_\infty / f_0^2 + 2/3 \cdot \xi \cdot \|\varphi_\eta\|_\infty / f_0} \right) \\
&\le \exp\left(-\frac{1}{4}\left(\frac{nT\xi^2}{\|\varphi_\eta\|_\infty^2 (\|\lambda\|_\infty \vee 1)/f_0^2} \wedge \frac{3nT\xi}{\|\varphi_\eta\|_\infty / f_0} \right) \right) \\
&\le \exp\left(-\frac{nTf_0^2(\|\lambda\|_\infty \vee 1)}{64(q+1)^3\mathbb{D}_{\mathfrak{m}}^2} \right) \le \exp\left(-\frac{nTf_0^2}{64(q+1)^3\mathbb{D}_{\mathfrak{m}}^2} \right)
\end{aligned}$$

Analogously, exploiting Proposition B.6, we get

$$\mathbb{P}\left(\pm\widetilde{\nu}_n(\varphi_\eta) \ge \xi\right) \le \exp\left(-\frac{nf_0}{64(q+1)^3\mathbb{D}_{\mathfrak{m}}^2} \right),$$

and hence

$$\mathbb{P}(\Xi^c) \le 4\mathbb{D}_{\mathfrak{m}} \exp\left(-\frac{n(f_0 \wedge Tf_0^2)}{64(q+1)^3\mathbb{D}_{\mathfrak{m}}^2} \right).$$

Assumption ($\mathfrak{m}2$) finally implies $\mathbb{P}(\Xi^c) \le \frac{1}{\sqrt{10}} \frac{\sqrt{f_0 \wedge f_0^2 T}}{(q+1)^{3/2}} \frac{\sqrt{n}}{\log(n+2)} \cdot n^{-5/2} \lesssim \frac{1}{n^2}$. □

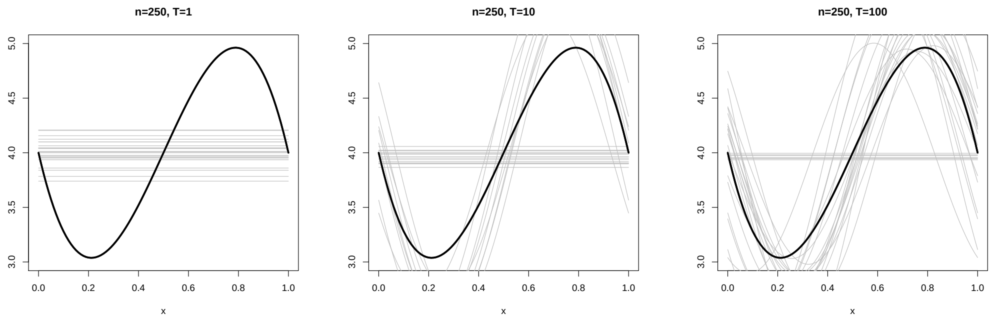


Figure 3.: Outcome of 25 replications (in grey) of the fully data-driven estimator in the non-parametric Poisson regression model for $n = 250$ and different values of T. The true regression function (in black) is given through $\lambda(x) = 20x(1-x)(x-0.5)+4$.

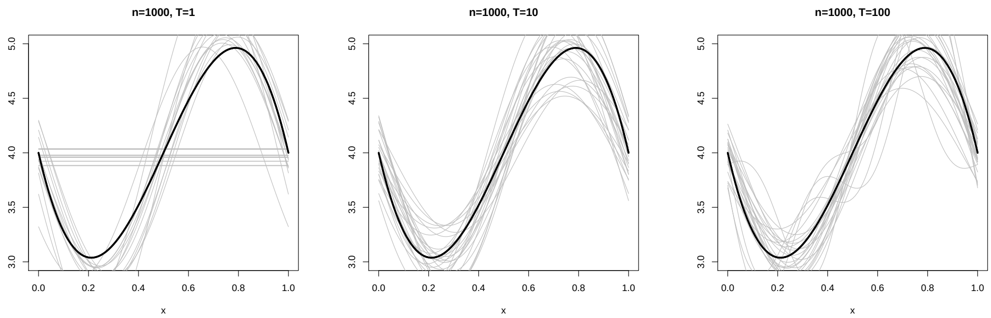


Figure 4.: Outcome of 25 replications (in grey) of the fully data-driven estimator in the non-parametric Poisson regression model for $n = 1000$ and different values of T. The true regression function (in black) is given through $\lambda(x) = 20x(1-x)(x-0.5)+4$.

Note that we the have considered the parameter $T > 0$ as a fixed constant in our setup. However, our analysis has *en passant* shown that our estimator performs better for larger values of T which is intuitively clear. It might be worth to have a closer look on the effect of the parameter T and, more precisely, the interplay of n and T in future work.

As in the previous chapter, the numerical constants in the definition of the penalty in (5.6) are too large in order to obtain a practicable estimator for small sample sizes. As usual in model selection frameworks, a reasonable constant for the definition of the penalty might be found by means of some calibration experiments. Figures 3 and 4 provide, for the sake of illustration, outcomes of some simulations for the fully data-driven estimator for different values of n and T in the case that the constant is set equal to 2 (in the Gaussian regression framework, this choice of the constant is known as Mallow's C_p, cf. [BBM99], p. 313). The unknown regression function in this illustrative simulation is $\lambda(x) = 20x(1-x)(x-0.5)+4$ and the auxiliary estimator was $\widehat{\lambda}_{\mathfrak{m}}$ with $\mathfrak{m} = (0, 10)$, that is, $\widehat{\lambda}_{\mathfrak{m}}$ is chosen as a histogram estimator. As one would expect from the definition of the penalty, smaller values of n and T favour the selection of less complex models with few basis functions.

5.3.3. Auxiliary results

Lemma 5.9. *For all $k \in \{0, \dots, n\}$, we have*

$$\mathbb{E}\left[\left(\sup_{t \in \mathcal{B}_k} |\langle \widetilde{\Theta}_n, t\rangle|^2 - 3\mu \cdot \frac{2k+1}{n}\right)_+\right] \leq K_1 \left\{\frac{\mu}{n}\exp(-K_2(2k+1)) + \frac{2k+1}{n^2}\exp(-K_3\sqrt{n})\right\}$$

with strictly positive numerical constants K_1, K_2, and K_3.

PROOF. With $t \in \mathcal{B}_k$, we associate the function

$$r_t(x) := \sum_{0\le|j|\le k} [t]_j \lambda(x) \frac{\varphi_j(x)}{f(x)}.$$

Evidently, for $X \sim f$ we have $\mathbb{E}[r_t(X)] = \sum_{0\le|j|\le k}[t]_j[\lambda]_j$. Consequently, one has the identity

$$\langle \widetilde{\Theta}_n, t\rangle = \frac{1}{n}\sum_{i=1}^n r_t(X_i) - \mathbb{E}[r_t(X_i)],$$

and $\langle \widetilde{\Theta}_n, t\rangle$ will take the role of $\nu_n(\cdot)$ in Lemma B.4. We now check the preconditions concerning the existence of suitable constants M_1, H and υ in the framework of Lemma B.4.
Condition concerning M_1: We have

$$\begin{aligned}\sup_{t\in\mathcal{B}_k}\|r_t\|_\infty^2 = \sup_{t\in\mathcal{B}_k}\sup_{y\in[0,1]}|r_t(y)|^2 &\le \sup_{t\in\mathcal{B}_k}\sup_{y\in[0,1]}\left(\sum_{0\le|j|\le k}|[t]_j|^2\right)\left(\sum_{0\le|j|\le k}\lambda^2(y)\cdot\frac{\varphi_j^2(y)}{f^2(y)}\right)\\ &\le \frac{\|\lambda\|_\infty^2}{f_0^2}\cdot(2k+1) \le \mu\cdot(2k+1),\end{aligned}$$

and we can put $M_1 := (\mu\cdot(2k+1))^{1/2}$.
Condition concerning H: We have

$$\begin{aligned}\mathbb{E}[\sup_{t\in\mathcal{B}_k}|\langle\widetilde{\Theta}_n,t\rangle|^2] &\le \frac{1}{n^2}\mathbb{E}\left[\sup_{t\in\mathcal{B}_k}\left(\sum_{0\le|j|\le k}|[t]_j|^2\right)\left(\sum_{0\le|j|\le k}\left|\sum_{i=1}^n\left\{\frac{\varphi_j(X_i)}{f(X_i)}\lambda(X_i)-[\lambda]_j\right\}\right|^2\right)\right]\\ &\le \frac{1}{n}\sum_{0\le|j|\le k}\operatorname{Var}\left(\frac{\varphi_j(X_1)}{f(X_1)}\lambda(X_1)\right) \le \frac{1}{n}\sum_{0\le|j|\le k}\mathbb{E}\left[\left(\frac{\varphi_j(X_1)}{f(X_1)}\lambda(X_1)\right)^2\right]\\ &\le \frac{2k+1}{n}\cdot\frac{\|\lambda\|_\infty^2}{f_0} \le \mu\cdot\frac{2k+1}{n},\end{aligned}$$

and thus by Jensen's inequality we can put $H := \left(\frac{\mu\cdot(2k+1)}{n}\right)^{1/2}$.
Condition concerning υ: For arbitrary $t \in \mathcal{B}_k$, it holds

$$\operatorname{Var}(r_t(X)) = \operatorname{Var}\left(\sum_{0\le|j|\le k}[t]_j\frac{\varphi_j(X)}{f(X)}\lambda(X)\right) \le \mathbb{E}\left[\left(\sum_{0\le|j|\le k}[t]_j\frac{\varphi_j(X)}{f(X)}\lambda(X)\right)^2\right] \le \mu.$$

Thus, we can take $\upsilon := \mu$ and the statement of the lemma follows now by applying Lemma B.4 with $\varepsilon = \frac{1}{4}$. □

In order to deal with the terms $\square_k$ in the proof of Theorem 5.6 we need to the following conditional version of Proposition 2.13. Since the proof is exactly the same as in the unconditional case (replacing all probabilities and expectations by their conditional counterparts), we omit the proof.

LEMMA 5.10. *Let $N_1, \ldots, N_n$ be independent Cox processes driven by finite random measures $\eta_1, \ldots, \eta_n$ (that is, given η_i, N_i is a PPP with intensity measure η_i) that are conditionally independent given $\eta_1, \ldots, \eta_n$. Set $\nu_n(r) = \frac{1}{n}\sum_{k=1}^n\{\int_{\mathbb{X}} r(x)dN_k(x) - \int_{\mathbb{X}} r(x)d\eta_k(x)\}$ for r contained in a countable class of real-valued measurable functions. Then, for any $\varepsilon > 0$, there exist constants*

c_1, $c_2 = \frac{1}{6}$, and c_3 such that

$$\mathbb{E}\left[\left(\sup_{r\in\mathcal{R}}|\nu_n(r)|^2 - c(\varepsilon)H^2\right)_+ |\boldsymbol{\eta}\right] \leq c_1\left\{\frac{v}{n}\exp\left(-c_2\varepsilon\frac{nH^2}{v}\right) + \frac{M_1^2}{C^2(\varepsilon)n^2}\exp\left(-c_3C(\varepsilon)\sqrt{\varepsilon}\frac{nH}{M_1}\right)\right\}$$

where $C(\varepsilon) = (\sqrt{1+\varepsilon}-1)\wedge 1$, $c(\varepsilon) = 2(1+2\varepsilon)$ and M_1, H and v are such that (denoting $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)$)

$$\sup_{r\in\mathcal{R}}\|r\|_\infty \leq M_1, \quad \mathbb{E}[\sup_{r\in\mathcal{R}}|\nu_n(r)||\boldsymbol{\eta}] \leq H, \quad \sup_{r\in\mathcal{R}}\operatorname{Var}\left(\int_{\mathbb{X}} r(x)dN_k(x)|\boldsymbol{\eta}\right) \leq v \quad \forall k.$$

We need Lemma 5.10 to prove the following Lemma 5.11. The crucial fact that we will exploit here is that the constants M_1, H and v in the statement of Lemma 5.10 can be chosen independently from the underlying directing measure in our specific setup. Thus, we obtain the identical bound also for the unconditional case.

LEMMA 5.11. *With the notation from the proof of Theorem 5.6 it holds for all $k \in \{0,\ldots,n\}$*

$$\mathbb{E}\left[\left(\sup_{t\in\mathcal{B}_k}|\langle\widehat{\Theta}_n, t\rangle|^2 - 50\mu\cdot\frac{(2k+1)\log(n+2)}{nT}\right)_+\right] \leq K_1'\left\{\frac{(2k+1)\mu}{nT}\exp(-2\log(n+2)) + \frac{(2k+1)\mu}{n^2T^2}\exp(-K_2'\sqrt{nT})\right\}$$

with strictly positive numerical constants K_1' and K_2'.

PROOF. Given $\boldsymbol{X} = (X_1,\ldots,X_n)$, we can write Y_i as $\int_0^T dN_i(s)$ where N_i is a Poisson process with homogeneous intensity equal to $\lambda(X_i)$. Thus, conditional on $\boldsymbol{X}$, it holds

$$\begin{aligned}\langle\widehat{\Theta}_n, t\rangle &= \frac{1}{nT}\sum_{0\leq|j|\leq k}[t]_j\sum_{i=1}^n\left\{\int_0^T\frac{\varphi_j(X_i)}{f(X_i)}dN_i(s) - \frac{\varphi_j(X_i)}{f(X_i)}\cdot T\lambda(X_i)\right\}\\ &= \frac{1}{n}\sum_{i=1}^n\left\{\int_0^T r_t(s)dN_i(s) - \int_0^T r_t(s)\lambda(X_i)ds\right\}\end{aligned}$$

where r_t is the function given by $r_t(s) := \frac{1}{T}\sum_{0\leq|j|\leq k}[t]_j\frac{\varphi_j(X_i)}{f(X_i)}$ (note that this is a constant function given $\boldsymbol{X}$). We now check the preconditions concerning the existence of suitable constants M_1, H and v from Lemma 5.10.

Condition concerning M_1: We have

$$\begin{aligned}\sup_{t\in\mathcal{B}_k}\|r_t\|_\infty^2 = \sup_{t\in\mathcal{B}_k}\frac{1}{T^2}\left(\sum_{0\leq|j|\leq k}[t]_j\frac{\varphi_j(X_i)}{f(X_i)}\right)^2 &\leq \sup_{t\in\mathcal{B}_k}\frac{1}{T^2}\left(\sum_{0\leq|j|\leq k}|[t]_j|^2\right)\cdot\left(\sum_{0\leq|j|\leq k}\frac{\varphi_j^2(X_i)}{f^2(X_i)}\right)\\ &\leq \frac{2k+1}{T^2f_0^2},\end{aligned}$$

and we can take $M_1 := \frac{1}{T}\sqrt{\mu\cdot(2k+1)}$.

Condition concerning H: It holds

$$\mathbb{E}[\sup_{t\in\mathcal{B}_k}|\langle\widehat{\Theta}_n, t\rangle|^2|\boldsymbol{X}] \leq \sup_{t\in\mathcal{B}_k}\left(\sum_{0\leq|j|\leq k}|[t]_j|^2\right)$$

$$\begin{aligned}
&\cdot \mathbb{E}\left[\sum_{0\le|j|\le k} |\frac{1}{nT}\sum_{i=1}^{n}\left\{\int_0^T \frac{\varphi_j(X_i)}{f(X_i)}[dN_i(s)-\lambda(X_i)ds]\right\}|^2|\boldsymbol{X}\right]\\
&\le \frac{1}{nT^2}\sum_{0\le|j|\le k}\mathrm{Var}\left(\int_0^T \frac{\varphi_j(X_1)}{f(X_1)}dN_1(s)|X_1\right)\\
&= \frac{1}{nT^2}\sum_{0\le|j|\le k}\int_0^T \frac{\varphi_j^2(X_i)}{f^2(X_i)}\lambda(X_i)ds\\
&\le \frac{2k+1}{n}\cdot\frac{1}{T}\cdot\frac{\|\lambda\|_\infty}{f_0^2}\le\frac{2k+1}{n}\cdot\frac{\mu}{T}.
\end{aligned}$$

Thus, we can put $H := \left(\frac{(2k+1)\mu\log(n+2)}{nT}\right)^{1/2}$.
Condition concerning υ: For arbitrary $k\in\{0,\dots,n\}$ and $t\in\mathcal{B}_k$ it holds

$$\mathrm{Var}\left(\int_0^T r_t(s)dN_k(s)|X_k\right)=\int_0^T |r_t(s)|^2\lambda(X_k)ds\le T\cdot\|\lambda\|_\infty\cdot\|r_t\|_\infty^2\le\frac{\xi}{Tf_0^2}\cdot(2k+1),$$

and we can put $\upsilon := \frac{\mu}{T}\cdot(2k+1)$.

We can apply Lemma 5.10 with $\varepsilon = 12$ which yields

$$\begin{aligned}
&\mathbb{E}\left[\left(\sup_{t\in\mathcal{B}_k}|\langle\widehat{\Theta}_n,t\rangle|^2-50\mu\cdot\frac{(2k+1)\log(n+2)}{nT}\right)_+|\boldsymbol{X}\right]\le\\
&\qquad K_1'\left\{\frac{(2k+1)\mu}{nT}\exp(-2\log(n+2))+\frac{(2k+1)\mu}{n^2T^2}\exp(-K_2'\sqrt{nT\log(n+2)})\right\}.
\end{aligned}$$

Because the right-hand side of the last estimate does not depend on $\boldsymbol{X}$, taking expectations on both sides implies the assertion of the lemma. □

6. Conclusion and perspectives

In the first part of this thesis, we have derived concentration inequalities for maxima of empirical processes associated with Poisson point processes. In the second part, we have considered different non-parametric models related to point processes and demonstrated that the concentration results from the first part turn out to be useful for the theoretical study of adaptive non-parametric estimators.

During the research which led to the results of this thesis, some questions have arisen that might be worth being dealt with in future research projects:

Concerning the first part of the thesis, it might be of interest whether the concentration results derived in Chapter 2 in a setup with Poisson processes can be transferred to more general point process setups, for instance setups with Cox processes. Moreover, our method of proof might also be appropriate in order to derive concentration inequalities for general stochastic integrals: our results from Chapter 2 might then be seen as special cases where the integrator is just a Poisson process.

In the second part of the thesis, we have assumed throughout that the observations in the considered non-parametric estimation problems are i.i.d. samples. It should be possible to transfer methodology recently derived in [AJ16a] and [AJ16b] (these papers dispense with the independence assumption and replace it with suitable mixing-conditions), to at least some of our problems.

In the context of Chapter 4, it might be of interest to study adaptive estimation procedures not only for the two models considered in this thesis, but under the more general assumption that the errors ε_{ij} are only stationary (note that some of the arguments used in the proofs of Chapter 4 fail to hold in this general framework). In addition, the question of lower bounds with respect to both sample sizes n and m remains open in the setup of model 2.

For the analysis of the Poisson regression problem in Chapter 5, we have restricted ourselves to an orthonormal series estimator in terms of the standard trigonometric basis. Since the properties of this basis (for instance, the boundedness of the basis functions) were exploited only at some places, it would be of interest to investigate whether our analysis can be performed also under weaker assumptions that are often used in papers using model selection techniques (cf., for instance, Assumption N in [BM97] or Assumption 4 in [Com01]).

Besides non-parametric estimation, non-parametric testing following along the guidelines of the general theory developed in [IS03] might be considered, for instance, in the setup of Chapter 4. There is already some work on hypothesis testing for Poisson point processes in case of direct observations, that is, the framework of Chapter 3, see [IK07], for instance. Furthermore, there exist already papers on non-parametric testing for inverse problems, for example [ISS12] where a Gaussian sequence space model is considered. A starting point for a research project here might be to combine ideas both from [IK07] and [ISS12] in order to develop non-parametric testing procedures for the setup of Chapter 4.

Part III.

Appendix

A. Hellinger distance between (probability) measures

Let $(\Omega, \mathscr{A})$ be a measurable space and $\mathbb{P}_1, \mathbb{P}_2$ be probability measures on $(\Omega, \mathscr{A})$. Let us assume that there exists a σ-finite measure ν on $(\Omega, \mathscr{A})$ such that $\mathbb{P}_1 \ll \nu$ and $\mathbb{P}_2 \ll \nu$. By the Radon-Nikodym theorem, $\mathbb{P}_1$, $\mathbb{P}_2$ have densities with respect to ν which we denote with $p_1 = \frac{d\mathbb{P}_1}{d\nu}$ and $p_2 = \frac{d\mathbb{P}_2}{d\nu}$, respectively.

Definition A.1. The *Hellinger distance* between $\mathbb{P}_1$ and $\mathbb{P}_2$ is defined via

$$H(\mathbb{P}_1, \mathbb{P}_2) = \left(\int \left[\sqrt{d\mathbb{P}_1} - \sqrt{d\mathbb{P}_2} \right]^2 \right)^{1/2} = \left(\int (\sqrt{p_1} - \sqrt{p_2})^2 d\nu \right)^{1/2}.$$

Remark A.2. The Hellinger distance $H(\mathbb{P}_1, \mathbb{P}_2)$ does not depend on the choice of the dominating measure ν.

Lemma A.3 ([Rei89], Lemma 3.3.10 (i)). *Let $\mathbb{P} = \bigotimes_{i=1}^n \mathbb{P}_i$, $\mathbb{Q} = \bigotimes_{i=1}^n \mathbb{Q}_i$ be product probability measures. Then*

$$H^2(\mathbb{P}, \mathbb{Q}) \leq \sum_{i=1}^n H^2(\mathbb{P}_i, \mathbb{Q}_i).$$

Let $\mathbb{P}_1$ and $\mathbb{P}_2$ be probability measures on the same measurable space and T a measurable map into another measurable space. Denote by $\mathbb{P}_i^T$ the probability measure induced by $\mathbb{P}_i$ and T, that is $\mathbb{P}_i^T(B) = \mathbb{P}_i(T \in B)$.

Lemma A.4 ([Rei89], Lemma 3.3.13).

$$H(\mathbb{P}_1^T, \mathbb{P}_2^T) \leq H(\mathbb{P}_1, \mathbb{P}_2).$$

Definition A.5. The *Hellinger affinity* between $\mathbb{P}_1$ and $\mathbb{P}_2$ is defined via

$$\rho(\mathbb{P}_1, \mathbb{P}_2) = \int \sqrt{d\mathbb{P}_1 d\mathbb{P}_2} d\nu.$$

Lemma A.6 ([Tsy08], Section 2.4).

$$\rho(\mathbb{P}_1, \mathbb{P}_2) = 1 - \frac{H^2(\mathbb{P}_1, \mathbb{P}_2)}{2}.$$

In analogy to the definition for probability measures, one can also define the Hellinger distance between measures μ_1 and μ_2.

Definition A.7. Let μ_1 and μ_2 be measures on the same measure space. Then, the Hellinger distance between μ_1 and μ_2 is defined via

$$H(\mu_1, \mu_2) = \left(\int [\sqrt{h_1} - \sqrt{h_2}]^2 d\mu_0 \right)^{1/2}$$

where h_i is a density of μ_i with respect to the measure μ_0.

Theorem A.8 ([Rei93], Theorem 3.2.1). *For $i = 1, 2$, let N_i be Poisson processes with finite intensity measures μ_i, respectively. Then* *(i)* $H^2(\mathbb{P}^{N_1}, \mathbb{P}^{N_2}) = 2\left\{1 - \exp(-\frac{1}{2}H^2(\mu_1, \mu_2))\right\}$,

(ii) $H(\mathbb{P}^{N_1}, \mathbb{P}^{N_2}) \leq H(\mu_1, \mu_2)$.

B. Auxiliary results

Theorem B.1 ([Pet95], Theorem 2.10). *Let $X_1, \dots, X_n$ be independent random variables with zero means, and let $p \geq 2$. Then*

$$\mathbb{E}\left[|\sum_{k=1}^{n} X_k|^p\right] \leq C(p) n^{p/2-1} \sum_{k=1}^{n} \mathbb{E}[|X_k|^p]$$

where $C(p)$ is a positive constant depending only on p.

Theorem B.2 (Chernoff bound for Poisson distributed random variables, [MU05], Theorem 5.4). *Let X be Poisson random variable with parameter μ.*

(i) If $x > \mu$, then

$$\mathbb{P}(X \geq x) \leq \frac{e^{-\mu}(e\mu)^x}{x^x};$$

(ii) If $x < \mu$, then

$$\mathbb{P}(X \leq x) \leq \frac{e^{-\mu}(e\mu)^x}{x^x}.$$

Theorem B.3 (Hoeffding's inequality, [BLM16], Theorem 2.8). *Let $X_1, \dots, X_n$ be independent random variables such that X_i takes its values in $[a_i, b_i]$ almost surely for all $i \leq n$. Let*

$$S = \sum_{i=1}^{n} (X_i - \mathbb{E}X_i).$$

Then for every $t > 0$,

$$\mathbb{P}(S \geq t) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right).$$

A consequence from the classical Talagrand inequality

The following lemma is a consequence from Talagrand's inequality and is taken from [CL15]. For a detailed proof, we refer to [Cha13].

Lemma B.4. *Let $X_1, \dots, X_n$ be i.i.d. random variables with values in some Polish space and define $\nu_n(s) = \frac{1}{n}\sum_{i=1}^{n} s(X_i) - \mathbb{E}[s(X_i)]$, for s belonging to a countable class $\mathcal{S}$ of measurable real-valued functions. Then, for any $\varepsilon > 0$, there exist positive constants c_1, $c_2 = \frac{1}{6}$, and c_3 such that*

$$\mathbb{E}\left[\left(\sup_{s \in \mathcal{S}} |\nu_n(s)|^2 - c(\varepsilon)H^2\right)_+\right] \leq c_1 \left\{\frac{\upsilon}{n} \exp\left(-c_2 \varepsilon \frac{nH^2}{\upsilon}\right) + \frac{M_1^2}{C^2(\varepsilon) n^2} \exp\left(-c_3 C(\varepsilon)\sqrt{\varepsilon}\frac{nH}{M_1}\right)\right\},$$

with $C(\varepsilon) = (\sqrt{1+\varepsilon} - 1) \wedge 1$, $c(\varepsilon) = 2(1 + 2\varepsilon)$ and

$$\sup_{s \in \mathcal{S}} \|s\|_\infty \leq M_1, \quad \mathbb{E}[\sup_{s \in \mathcal{S}} |\nu_n(s)|] \leq H, \quad \textit{and} \quad \sup_{s \in \mathcal{S}} \mathrm{Var}(s(X_1)) \leq \upsilon.$$

Remark B.5. In the case that one wants to consider complex-valued functions s, the statement of Lemma B.4 holds true with the quantity $c(\varepsilon)$ replaced with $c(\varepsilon) = 4(1 + 2\varepsilon)$.

Bernstein type inequalities

Proposition B.6 (Bernstein's inequality, [BLM16], Corollary 2.11). *Let $X_1, \dots, X_n$ be independent real-valued random variables with $|X_i| \le b$ for some $b > 0$ almost surely for all $i \le n$. Let $S = \sum_{i=1}^n (X_i - \mathbb{E}X_i)$ and $\upsilon = \sum_{i=1}^n \mathbb{E}[X_i^2]$. Then*

$$\mathbb{P}(S \ge t) \le \exp\left(-\frac{t^2}{2(\upsilon + bt/3)}\right).$$

Proposition B.7 ([RB03], Proposition 7). *Let N be a PPP on some measurable space $(\mathbb{X}, \mathscr{X})$ with finite intensity measure μ. Let g be a measurable function on $(\mathbb{X}, \mathscr{X})$, essentially bounded, such that $\int_{\mathbb{X}} g^2(x)\mu(dx) > 0$. Then*

$$\mathbb{P}\left(\int_{\mathbb{X}} g(x)(dN(x) - \mu(dx)) \ge t\right) \le \exp\left(-\frac{t^2}{2(\int_{\mathbb{X}} g^2(x)\mu(dx) + \|g\|_\infty t/3)}\right), \qquad t > 0.$$

Notation

$\mathbb{N}$	$\{1, 2, \ldots\}$
$\mathbb{N}_0$	$\{0, 1, 2, \ldots\}$
$\mathbb{Z}$	$\{\ldots, -2, -1, 0, 1, 2, \ldots\}$
$\mathbb{R}$	Set of real numbers
$\mathbb{C}$	Set of complex numbers
$\Re z$	Real part of a complex number z
$\Im z$	Imaginary part of a complex number z
$\mathbb{1}_A$	Indicator function of the event A
$a_n \lesssim b_n$	$\exists C > 0$ such that $a_n \leq C b_n \; \forall n \in \mathbb{N}$
$a_n \asymp b_n$	$a_n \lesssim b_n$ and $b_n \lesssim a_n$ hold simultaneously
$\mathrm{argmin}_{t \in T} f(t)$ ($T \subset \mathbb{N}_0$ finite)	By convention the minimal $t^\star \in T$ such that $f(t^\star) = \min_{t \in T} f(t)$
$H(\mu, \nu)$	Hellinger distance between (probability) measures $\mathbb{P}_1$ and $\mathbb{P}_2$
$\rho(\mathbb{P}_1, \mathbb{P}_2)$	Hellinger affinity between probability measures $\mathbb{P}_1$ and $\mathbb{P}_2$
$\\| \cdot \\|$	$\mathbb{L}^2$ norm
$\\| \cdot \\|_\infty$	Sup norm
$\\| \cdot \\|_{\mathrm{op}}$	Operator norm

Acronyms

LCCB space	Locally compact second countable Hausdorff space
PPP	Poisson point process

Bibliography

[AB06] Antoniadis, A. and Bigot, J. Poisson inverse problems. *Ann. Statist.* 34 (2006), 2132–2158.

[AC10] Arlot, S. and Celisse, A. A survey of cross-validation procedures for model selection. *Statistics Surveys* 4 (2010), 40–79.

[AJ16a] Asin, N. and Johannes, J. Adaptive non-parametric estimation in the presence of dependence. *ArXiv preprint* (2016). arXiv: 1602.00531.

[AJ16b] Asin, N. and Johannes, J. Adaptive non-parametric instrumental regression in the presence of dependence. *ArXiv preprint* (2016). arXiv: 1604.01992.

[Ans48] Anscombe, F. J. The transformation of Poisson, binomial and negative-binomial data. *Biometrika* 35 (1948), 246–254.

[Bar00] Baraud, Y. Model selection for regression on a fixed design. *Probab. Theory Related Fields* 117 (2000), 467–493.

[BB09] Baraud, Y. and Birgé, L. Estimating the intensity of a random measure by histogram type estimators. *Probab. Theory Related Fields* 143 (2009), 239–284.

[BBM99] Barron, A., Birgé, L., and Massart, P. Risk bounds for model selection via penalization. *Probab. Theory Related Fields* 113 (1999), 301–413.

[BH09] Bauwens, L. and Hautsch, N. *Modelling financial high frequency data using point processes.* In: *Handbook of financial time series.* Ed. by Mikosch, T. et al. Springer, Berlin, 2009, 953–979.

[Big+10] Bigot, J. et al. Nonparametric estimation of covariance functions by model selection. *Electron. J. Stat.* 4 (2010), 822–855.

[Big+13] Bigot, J. et al. Intensity estimation of non-homogeneous Poisson processes from shifted trajectories. *Electron. J. Stat.* 7 (2013), 881–931.

[Bir07] Birgé, L. *Model selection for Poisson processes.* In: *IMS Lecture Notes - Monograph Series.* Institute of Mathematical Statistics, 2007, 32–64.

[BLM16] Boucheron, S., Lugosi, G., and Massart, P. *Concentration inequalities: a nonasymptotic theory of independence.* Oxford University Press, Oxford, 2016.

[BM97] Birgé, L. and Massart, P. *From model selection to adaptive estimation.* In: *Festschrift for Lucien Le Cam.* Ed. by Pollard, D., Torgersen, E., and Yang, G. L. Springer, New York, 1997, 55–87.

[BM98] Birgé, L. and Massart, P. Minimum contrast estimators on sieves: exponential bounds and rates of convergence. *Bernoulli* 4 (1998), 329–375.

[Bré81] Brémaud, P. *Point processes and queues, martingale dynamics.* Springer, Berlin, 1981.

[BTW07] Bunea, F., Tsybakov, A. B., and Wegkamp, M. H. Aggregation for Gaussian regression. *Ann. Statist.* 35 (2007), 1674–1697.

[Cav08] Cavalier, L. Nonparametric statistical inverse problems. *Inverse Probl.* 24 (2008).

[CGC15] Comte, F. and Genon-Catalot, V. Adaptive Laguerre density estimation for mixed Poisson models. *Electron. J. Stat.* 9 (2015), 1113–1149.

[Cha13] Chagny, G. Estimation adaptative avec des données transformées ou incomplètes. Application à des modèles de survie. PhD thesis. 2013. url: `https://tel.archives-ouvertes.fr/tel-00863141/`.

[Che07] Chesneau, C. Regression with random design: A minimax study. *Statist. Probab. Lett.* 77 (2007), 40–53.

[CIS76] Cirel'son, B. S., Ibragimov, I. A., and Sudakov, V. N. *Norms of Gaussian sample functions*. In: *Proceedings of the Third Japan-U.S.S.R. Symposium on Probability Theory*. Springer, Berlin, 1976, 20–41.

[CJ02] Cavalier, L. and Ja-Yong Koo. Poisson intensity estimation for tomographic data using a wavelet shrinkage approach. *IEEE T. Inform. Theory* 48 (2002), 2794–2802.

[CJ12] Comte, F. and Johannes, J. Adaptive functional linear regression. *Ann. Statist.* 40 (2012), 2765–2797.

[CL10] Comte, F. and Lacour, C. Pointwise deconvolution with unknown error distribution. *C. R. Math. Acad. Sci. Paris* 348 (2010), 323–326.

[CL11] Comte, F. and Lacour, C. Data-driven density estimation in the presence of additive noise with unknown distribution: Data-driven density estimation. *J. Roy. Statist. Soc. Ser. B* 73 (2011), 601–627.

[CL15] Chagny, G. and Lacour, C. Optimal adaptive estimation of the relative density. *TEST* 24 (3) (2015), 605–631.

[Com+15] Comte, F. et al. Estimation of the jump size density in a mixed compound Poisson process: jump density in a mixed compound Poisson. *Scand. J. Stat.* 42 (2015), 1023–1044.

[Com01] Comte, F. Adaptive estimation of the spectrum of a stationary Gaussian sequence. *Bernoulli* 7 (2001), 267–298.

[CP02] Carota, C. and Parmigiani, G. Semiparametric regression for count data. *Biometrika* 89 (2002), 265–281.

[CRT06] Comte, F., Rozenholc, Y., and Taupin, M.-L. Penalized contrast estimator for adaptive density deconvolution. *Canad. J. Statist.* 34 (2006), 431–452.

[CT98] Cameron, A. C. and Trivedi, P. K. *Regression analysis of count data*. Cambridge University Press, Cambridge, 1998.

[DH93] Diggle, P. J. and Hall, P. A Fourier approach to nonparametric deconvolution of a density estimate. *J. Roy. Stat. Soc. Ser. B* 55 (1993), 523–531.

[DL93] DeVore, R. A. and Lorentz, G. G. *Constructive approximation*. Springer, Berlin, 1993.

[DTM98] Diggle, P. J., Tawn, J. A., and Moyeed, R. A. Model-based geostatistics. *J. Roy. Stat. Soc. Ser. C* 47 (1998), 299–350.

[DVJ03] Daley, D. J. and Vere-Jones, D. *An introduction to the theory of point processes*. Second edition. Vol. I. Springer, New York, 2003.

[GN00] Grègoire, G. and Nembè, J. Convergence rates for the minimum complexity estimator of counting process intensities. *Nonparametric Statistics* 12 (2000), 611–643.

[GVL96] Golub, G. H. and Van Loan, C. F. *Matrix computations*. Third edition. John Hopkins University Press, Baltimore, 1996.

[HWMZ03] Helmers, R., Wayan Mangku, I., and Zitikis, R. Consistent estimation of the intensity function of a cyclic Poisson process. *J. Multivariate Anal.* 84 (2003), 19–39.

[IK07] Ingster, Y. I. and Kutoyants, Y. A. Nonparametric hypothesis testing for intensity of the Poisson process. *Math. Methods Statist.* 16 (2007), 217–245.

[IPR16] Ivanoff, S., Picard, F., and Rivoirard, V. Adaptive Lasso and group-Lasso for functional Poisson regression. *J. Mach. Learn. Res.* 17 (2016), 1–46.

[IS03] Ingster, Y. I. and Suslina, I. A. *Nonparametric goodness-of-fit testing under Gaussian models.* Springer, New York, 2003.

[ISS12] Ingster, Y. I., Sapatinas, T., and Suslina, I. A. Minimax signal detection in ill-posed inverse problems. *Ann. Statist.* 40 (2012), 1524–1549.

[Joh09] Johannes, J. Deconvolution with unknown error distribution. *Ann. Statist.* 37 (2009), 2301–2323.

[JS13a] Johannes, J. and Schwarz, M. Adaptive circular deconvolution by model selection under unknown error distribution. *Bernoulli* 19 (2013), 1576–1611.

[JS13b] Johannes, J. and Schwarz, M. Adaptive Gaussian inverse regression with partially unknown operator. *Comm. Statist. Theory Methods* 42 (2013), 1343–1362.

[Kal76] Kallenberg, O. *Random measures.* Akademie-Verlag, Berlin, 1976.

[Kar91] Karr, A. F. *Point processes and their statistical inference.* Marcel Dekker, New York, 1991.

[Kle03] Klein, T. Inégalités de concentration, martingales et arbres aléatoires. PhD thesis. 2003. url: `http://www.math.univ-toulouse.fr/~klein/theseweb.pdf`.

[Kol99] Kolaczyk, E. D. Wavelet shrinkage estimation of certain Poisson intensity signals using corrected thresholds. *Statist. Sinica* 9 (1999), 119–135.

[KR05] Klein, T. and Rio, E. Concentration around the mean for maxima of empirical processes. *Ann. Probab.* 33 (2005), 1060–1077.

[Kro16] Kroll, M. Concentration inequalities for Poisson point processes with application to adaptive intensity estimation. *arXiv preprint* (2016). arXiv: `1612.07901`.

[Kro17] Kroll, M. Nonparametric intensity estimation from indirect point process observations under unknown error distribution. *arXiv preprint* (2017). arXiv: `1703.05619`.

[Kut98] Kutoyants, Y. A. *Statistical inference for spatial Poisson processes.* Springer, New York, 1998.

[KYS13] Kamo, K.-I., Yanagihara, H., and Satoh, K. Bias-corrected AIC for selecting variables in Poisson regression models. *Comm. Statist. Theory Methods* 42 (2013), 1911–1921.

[Lac08] Lacour, C. Adaptive estimation of the transition density of a particular hidden Markov chain. *J. Multivariate Anal.* 99 (2008), 787–814.

[Led96] Ledoux, M. On Talagrand's deviation inequalities for product measures. *ESAIM Probab. Stat.* 1 (1996), 63–87.

[Lep91] Lepski, O. V. On a problem of adaptive estimation in Gaussian white noise. *Theor. Probab. Appl.* 35 (1991), 454–466.

[LM09] Lecué, G. and Mendelson, S. Aggregation via empirical risk minimization. *Probab. Theory Related Fields* 145 (2009), 591–613.

[Mas00] Massart, P. About the constants in Talagrand's concentration inequalities for empirical processes. *Ann. Probab.* 28 (2000), 863–884.

[Mas07] Massart, P. *Concentration inequalities and model selection.* Springer, Berlin, 2007.

[McD89] McDiarmid, C. *On the method of bounded differences.* In: *Surveys in Combinatorics.* Cambridge University Press, Cambridge, 1989, 148–188.

[Mik+07] Mikkelsen, T. S. et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448 (2007), 553–560.

[MU05] MITZENMACHER, M. and UPFAL, E. *Probability and computing: an introduction to randomized algorithms and probabilistic analysis.* Cambridge University Press, Cambridge, 2005.

[MW04] MØLLER, J. and WAAGEPETERSEN, R. P. *Statistical inference and simulation for spatial point processes.* Chapman & Hall/CRC, Boca Raton, 2004.

[NH97] NEUMANN, M. and HÖSSJER, O. On the effect of estimating the error density in nonparametric deconvolution. *J. Nonparametr. Stat.* 7 (1997), 307–330.

[Pet95] PETROV, V. V. *Limit theorems of probability theory.* Oxford Science Publications, Oxford, 1995.

[PS00] PENTTINEN, A. and STOYAN, D. Recent applications of point process methods in forestry statistics. *Statist. Sci.* 15 (1) (2000), 61–78.

[PW04] PATIL, P. N. and WOOD, A. T. A. Counting process intensity estimation by orthogonal wavelet methods. *Bernoulli* 10 (2004), 1–24.

[RB03] REYNAUD-BOURET, P. Adaptive estimation of the intensity of inhomogeneous Poisson processes via concentration inequalities. *Probab. Theory Related Fields* 126 (2003), 103–153.

[RBR10] REYNAUD-BOURET, P. and RIVOIRARD, V. Near optimal thresholding estimation of a Poisson intensity on the real line. *Electron. J. Stat.* 4 (2010), 172–238.

[Rei89] REISS, R.-D. *Approximate distributions of order statistics.* Springer, New York, 1989.

[Rei93] REISS, R.-D. *A course on point processes.* Springer, New York, 1993.

[Res87] RESNICK, S. I. *Extreme values, regular variation and point processes.* Springer, New York, 1987.

[RT12] RIGOLLET, P. and TSYBAKOV, A. B. Sparse estimation by exponential weighting. *Statist. Sci.* 27 (2012), 558–575.

[Rud82] RUDEMO, M. Empirical choice of histograms and kernel density estimators. *Scand. J. Statist.* 9 (1982), 65–78.

[San14] SANSONNET, L. Wavelet thresholding estimation in a Poissonian interactions model with application to genomic data. *Scand. J. Statist.* 41 (2014), 200–226.

[Sau13] SAUMARD, A. Optimal model selection in heteroscedastic regression using piecewise polynomial functions. *Electron. J. Stat.* 7 (2013), 1184–1223.

[Sch14] SCHMISSER, E. Non-parametric adaptive estimation of the drift for a jump diffusion process. *Stochastic Process. Appl.* 124 (1) (2014), 883–914.

[Ser09] SERFOZO, R. *Basics of applied stochastic processes.* Springer, Berlin, 2009.

[SS87] SHAMIR, E. and SPENCER, J. Sharp concentration of the chromatic number on random graphs $G_{n,p}$. *Combinatorica* 7 (1987), 121–129.

[Sto+13] STOYAN, D. et al. *Stochastic geometry and its applications.* Third edition. John Wiley & Sons, Chichester, 2013.

[Sto82] STONE, C. J. Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* 10 (1982), 1040–1053.

[Str13] STROUP, W. W. *Generalized linear mixed models.* Chapman & Hall/CRC, Boca Raton, 2013.

[SVB10] SCHWARZ, M. and VAN BELLEGEM, S. Consistent density deconvolution under partially known error distribution. *Statist. Probab. Lett.* 80 (2010), 236–241.

[Tal95] TALAGRAND, M. Concentration of measure and isoperimetric inequalities in product spaces. *Publ. Math. Inst. Hautes Études Sci.* 81 (1995), 73–205.

[Tal96] TALAGRAND, M. New concentration inequalities in product spaces. *Invent. Math.* 126 (1996), 505–563.

[Tsy08] TSYBAKOV, A. B. *Introduction to nonparametric estimation.* Springer, New York, 2008.

[Win08] WINKELMANN, R. *Econometric analysis of count data.* Springer, Berlin, 2008.

[WN07] WILLETT, R. M. and NOWAK, R. D. Multiscale Poisson intensity and density estimation. *IEEE T. Inform. Theory* 53 (2007), 3171–3187.

[Yur76] YURINSKI, V. Exponential inequalities for sums of random vectors. *J. Multivariate Anal.* 6 (1976), 473–499.

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig angefertigt und keine anderen als die angegebenen Hilfsmittel verwendet habe.

Mannheim, den

Martin Kroll