

Efficient estimation of functionals in nonparametric boundary models

MARKUS REIß¹ and LEONIE SELK²

¹*Institute of Mathematics, Humboldt-Universität zu Berlin, Germany.*

E-mail: mreiss@mathematik.hu-berlin.de

²*Institute of Mathematics, Universität Hamburg, Germany. E-mail: leonie.selk@math.uni-hamburg.de*

For nonparametric regression with one-sided errors and a boundary curve model for Poisson point processes, we consider the problem of efficient estimation for linear functionals. The minimax optimal rate is obtained by an unbiased estimation method which nevertheless depends on a Hölder condition or monotonicity assumption for the underlying regression or boundary function.

We first construct a simple blockwise estimator and then build up a nonparametric maximum-likelihood approach for exponential noise variables and the point process model. In that approach also non-asymptotic efficiency is obtained (UMVU: uniformly minimum variance among all unbiased estimators). The proofs rely essentially on martingale stopping arguments for counting processes and the point process geometry. The estimators are easily computable and a small simulation study confirms their applicability.

Keywords: completeness; frontier estimation; monotone boundary; nonparametric MLE; optional stopping; Poisson point process; shape constraint; sufficiency; support estimation; UMVU

1. Introduction

For regression models

$$\mathcal{Y}_i = g(i/n) + \varepsilon_i, \quad i = 1, \dots, n, \quad (1.1)$$

the estimation of linear functionals of the regression function g is well understood if (ε_i) are uncorrelated with mean zero and variance $\sigma^2 > 0$. Then the discrete functionals

$$\vartheta^{(n)} = \frac{1}{n} \sum_{i=1}^n g(i/n)w(i/n) \quad \text{for some function } w : [0, 1] \rightarrow \mathbb{R} \quad (1.2)$$

can be estimated by the plug-in version $\hat{\vartheta}_n = \frac{1}{n} \sum_{i=1}^n \mathcal{Y}_i w(i/n)$ without bias and with variance $\frac{\sigma^2}{n^2} \sum_{i=1}^n w(i/n)^2$. By the Gauß–Markov theorem, $\hat{\vartheta}_n$ has minimal variance among all linear and unbiased estimators. In the Gaussian case, $\hat{\vartheta}_n$ is even UMVU (uniformly of minimum variance among all unbiased estimators). In the corresponding continuous-time signal-in-white-noise model $dY(t) = g(t)dt + \sigma n^{-1/2} dW_t$, $t \in [0, 1]$, with a Brownian motion W and some $g \in L^2([0, 1])$ the plug-in estimator $\hat{\vartheta} = \int_0^1 w(t) dY(t)$ is equally an unbiased estimator of

$$\vartheta = \int_0^1 g(t)w(t) dt \quad \text{for some } w \in L^2([0, 1]) \quad (1.3)$$

of variance $\frac{\sigma^2}{n} \int_0^1 w(t)^2 dt$. By the Riesz representation theorem, we can thus estimate any linear L^2 -continuous functional of g with parametric rate $n^{-1/2}$.

In certain applications, however, the function g is determined as the boundary or frontier function of the observations, which can be modeled equivalently by one-sided errors (ε_i) . The prototypical case is that (ε_i) are i.i.d. with $\varepsilon_i \geq 0$ and for some $\lambda > 0$

$$P(\varepsilon_i \leq x) = \lambda x + O(x^2) \quad \text{as } x \downarrow 0, \quad (1.4)$$

for example, $\varepsilon_i \sim \text{Exp}(\lambda)$. In that case, the parametric rate for the location model (i.e., assuming g to be constant) is with n^{-1} much faster than in the regular case. These irregular statistical models have also found considerable theoretical interest, for example, in the recent work by Baraud and Birgé [2]. A rate-optimal estimator is given by the extreme value statistics $\min_i \mathcal{Y}_i$. For the nonparametric problem of estimating the function g in L^2 -loss, the optimal rate is $n^{-\beta/(\beta+1)}$ for g in a Hölder ball of regularity $\beta \in (0, 1]$ and radius $R > 0$:

$$g \in \mathcal{C}^\beta(R) = \{f : [0, 1] \rightarrow \mathbb{R} \mid \forall x, y \in [0, 1] : |f(y) - f(x)| \leq R|y - x|^\beta\}. \quad (1.5)$$

This is achieved by a local polynomial estimator $\hat{g}_{n,h}$ as in the regular case, see, for example, [9] for a construction and a survey of the large literature on that topic. A plug-in estimator $\hat{\vartheta}_n := \int_0^1 \hat{g}_{n,h_n}(x)w(x) dx$ with optimal bandwidth h_n to estimate ϑ in (1.3) can achieve at best the rate $n^{-\beta/(\beta+1/2)}$. This is due to a pointwise bias of order h^β and a pointwise variance of order $(nh)^{-2}$, which after integration and by using independence results in a total mean squared error of order $h^{2\beta} + n^{-2}h^{-1}$ for the plug-in estimator. The standardised rate $n^{-\beta/(\beta+1/2)}$ is not optimal and for $\beta < 1/2$ even slower than $n^{-1/2}$ in the regular case. At the heart of the problem is the usual nonparametric bias bound, which cannot be improved by averaging.

Here we show that the optimal estimation rate for $\vartheta^{(n)}$ under one-sided errors is $n^{-(\beta+1/2)/(\beta+1)}$ for $g \in \mathcal{C}^\beta(R)$. The improvement over the plug-in estimator is achieved by an unbiased estimation procedure. The bias is exactly zero for the case of exponentially distributed errors and it is asymptotically negligible under (1.4) for $\beta > 1/2$. Compared with standard nonparametric results, it is remarkable that an unbiased estimator can be constructed whose rate is nevertheless worse than the parametric rate (n^{-1} in this case). The risk bound comes from a trade-off between two terms in the variance instead of the usual bias-variance trade-off.

As for mean regression with the signal-in-white-noise model, also for one-sided errors an analogous continuous-time model is most useful in exhibiting the main statistical structure. It is given by observing a Poisson point process (PPP) on $[0, 1] \times \mathbb{R}$ of intensity

$$\lambda_g(x, y) = n\mathbf{1}(y \geq g(x)), \quad x \in [0, 1], y \in \mathbb{R}, \quad (1.6)$$

see, for example, [11] or [4] for point process properties and Figure 1 below for an illustration. For sufficiently regular g , this model can be shown to be asymptotically equivalent to the regression-type model (1.1) with $\lambda = 1$ in (1.4), cf. [18]. At the same time, this serves as a canonical model for support boundary estimation from i.i.d. observations. For instance, Girard and Jacob [6] propose projection based estimators for g in this model class and derive convergence rates as well as limit distributions, already relying on bias reduction techniques. Also Bibinger, Jirak and Reiß [3] use it as an agnostic model for limit order books in financial markets.

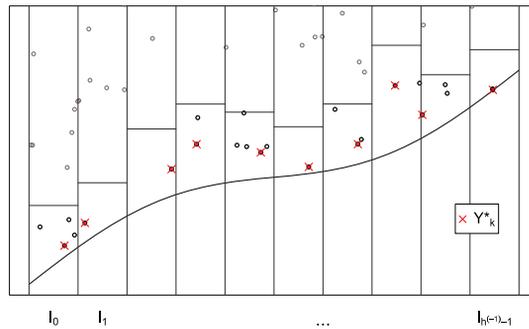


Figure 1. Construction of $\hat{\vartheta}^{\text{block}}$. Circles indicate PPP observations (X_i, Y_i) , crosses blockwise minima Y_k^* and horizontal lines the upper boundaries $Y_k^* + Rh^\beta$.

We first develop the methods in the fundamental PPP model for ϑ from (1.3) and then transfer them explicitly to the discrete model (1.1). By a blocking technique, ϑ can be estimated without bias and at the minimax optimal rate. The method is then extended to the one-sided regression setting. Using Lepski's method, we are then able to provide also an adaptive estimator, that is an estimator which does not rely on the smoothness parameters β, R and still attains the minimax rate up to a logarithmic factor. In a second step, we can even construct an estimator of ϑ which is UMVU. This non-asymptotic efficiency result is based on a nonparametric maximum-likelihood approach, where the maximum-likelihood estimator (MLE) \hat{g}^{MLE} is not only explicit, but also forms a sufficient and complete statistics. In parallel with Gaussian mean regression, we thus have the UMVU-property of the estimator, but its asymptotic rate is worse than for parametric location estimation. Still, we are able to prove its asymptotic normality and to provide a self-normalising CLT such that asymptotic inference is feasible. The MLE approach equally works for the class of monotone functions g .

The regression-type model (1.1) with one-sided errors and the PPP model (1.6) have a similar structure as models considered for density support estimation or image boundary recovery problems. Let us review briefly the literature on functional estimation for these statistical models. Many asymptotic results for the expected area of the convex hull for i.i.d. observations are based on the classical results by Rényi and Sulanke [19]. Based on these results, the ideas of the present paper have been used by Baldin and Reiß [1] to construct an UMVU estimator for the volume of a convex body. For image recovery problems Korostel'ev and Tsybakov [13] describe already the rate $n^{-(\beta+1/2)/(\beta+1)}$ obtained for the functional $\int_0^1 g(x) dx$. The upper bound is based on a localisation step and loses a logarithmic factor. By threefold sample splitting, Gayraud [5] has constructed an estimator achieving this rate exactly for the related density support area estimation. An interesting linear programming approach is proposed by Zhirar, Yuditskiĭ and Nazin [23]. Yet, these estimators are analysed asymptotically and lack the non-asymptotic unbiasedness and UMVU property we have found here. Many other estimators are concerned with the estimation of the density support set or the regression-type function itself, not of the area or other functionals, let us mention the work by Mammen and Tsybakov [17] for connections to classification problems. Specifically, a nonparametric MLE approach under monotonicity has been developed by Korostel'ev, Simar and Tsybakov [12] for the asymptotically exact risk in estimating the den-

sity support set in Hausdorff distance. In Gaussian mean regression a nonparametric MLE over regular function classes is equivalent to a least-squares approach with roughness penalty, leading, for example, to smoothing splines. Under shape constraints the MLE is a well studied object, see, for example, [7], but usually results are derived asymptotically.

In the next section, we shall develop a simple block-wise estimator. Based on optional stopping for an intrinsic martingale, we prove that it is unbiased under the PPP model and under exponential noise in the regression-type model. For more general regression noise, the required compensation cannot be achieved exactly, but it comes close to the model with corresponding exponential noise. The third part of that section presents the adaptive estimator, while the final part presents the lower bound implying that the rate is indeed optimal. The nonparametric MLE approach is presented in Section 3, first for the class of Hölder functions, then for monotone functions. The derivation of the completeness of the nonparametric MLE and the stopping arguments for the intrinsic martingale are intriguing. For the MLE under Lipschitz conditions, we obtain central limit theorems which allow for feasible confidence sets.

In Section 4, we discuss some implications of the results, in particular concerning estimating coefficients in a projection estimator approach. Extensions and limitations are mentioned and a small simulation study shows that the estimators are numerically feasible and have satisfying finite-sample properties. Most proofs are instructive and reveal some beautiful interplay between statistics, probability and geometry such that in the [Appendix](#) we only provide some technical lemmata (some of independent interest) and the more involved proofs of the adaptive rate and the CLT. The notation follows the usual conventions. We write $a_n \lesssim b_n$ or $a_n = O(b_n)$ to say that a_n is bounded by a constant multiple of b_n and $a_n \sim b_n$ for $a_n \lesssim b_n$ as well as $b_n \lesssim a_n$. Moreover, $a_n = o(b_n)$ means $a_n/b_n \rightarrow 0$ and $a_n \asymp b_n$ stands for $a_n/b_n \rightarrow 1$.

2. Simple rate-optimal estimation

2.1. Block-wise estimation in the PPP model

Let $(X_j, Y_j)_{j \geq 1}$ denote the observations of the Poisson point process (PPP) with intensity (1.6). We shall estimate ϑ from (1.3) without any bias. To grasp the main idea, suppose that $w(x) = 1$ holds and that we know a deterministic function $\bar{g} : [0, 1] \rightarrow \mathbb{R}$ with the property $\bar{g} \geq g$ (point-wise). Then the number of PPP observations below the graph of \bar{g} is Poisson-distributed with intensity equal to n times the area between g and \bar{g} :

$$\sum_{j \geq 1} \mathbf{1}(Y_j \leq \bar{g}(X_j)) \sim \text{Pois} \left(n \int_0^1 (\bar{g} - g)(x) dx \right).$$

This yields an unbiased pseudo-estimator $\bar{\vartheta}$:

$$\bar{\vartheta} := \int_0^1 \bar{g}(x) dx - \frac{1}{n} \sum_{i \geq 1} \mathbf{1}(Y_i \leq \bar{g}(X_i)) \quad \Rightarrow \quad \mathbb{E}[\bar{\vartheta}] = \int_0^1 (\bar{g} - (\bar{g} - g))(x) dx = \vartheta.$$

The larger the area between the graphs, the larger is the Poisson parameter and thus the variance of $\bar{\vartheta}$.

Now, we shall define an empirical substitute for \bar{g} , which by stopping time arguments keeps the unbiasedness, but is nevertheless sufficiently close to g . We partition $[0, 1]$ in subintervals $I_k = [kh, (k + 1)h]$ of length h with $h^{-1} \in \mathbb{N}$ and note that the block-wise minimum $Y_k^* := \min_{j: X_j \in I_k} Y_j$ satisfies $Y_k^* \geq \min_{x \in I_k} g(x)$. By the Hölder property of g , we conclude that $g(x) \leq Y_k^* + Rh^\beta$ holds for all $x \in I_k$ and thus $Y_k^* + Rh^\beta$ is a local upper bound for g , see also Figure 1. We thus estimate the functional locally on these blocks by

$$\hat{\vartheta}_k := (Y_k^* + Rh^\beta)\bar{w}_k - \frac{1}{nh} \sum_{i \geq 1} \mathbf{1}(X_i \in I_k, Y_i \leq Y_k^* + Rh^\beta)w(X_i),$$

where $\bar{w}_k = \frac{1}{h} \int_{I_k} w(x) dx$ and the true local parameter is $\vartheta_k := \frac{1}{h} \int_{I_k} g(x)w(x) dx$.

Theorem 2.1. *The estimator $\hat{\vartheta}^{\text{block}} = \sum_{k=0}^{h^{-1}-1} \hat{\vartheta}_k h$ satisfies with $\|w\|_{L^2}^2 = \int_0^1 w(x)^2 dx$*

$$\mathbb{E}[\hat{\vartheta}^{\text{block}}] = \vartheta, \quad \text{Var}(\hat{\vartheta}^{\text{block}}) \leq \frac{2Rh^\beta + (nh)^{-1}}{n} \|w\|_{L^2}^2.$$

In particular, the asymptotically optimal block size $h \asymp (2\beta Rn)^{-1/(\beta+1)}$ yields

$$\limsup_{n \rightarrow \infty} \sup_{g \in C^\beta(R)} n^{(2\beta+1)/(\beta+1)} \text{Var}(\hat{\vartheta}^{\text{block}}) \leq \frac{\beta + 1}{\beta} (2\beta R)^{1/(\beta+1)} \|w\|_{L^2}^2.$$

Proof. Let us study the weighted counting process

$$N(t) := \sum_{i \geq 1} \mathbf{1}(X_i \in I_k, Y_i \leq t)w(X_i), \quad t \in \mathbb{R}.$$

The pure counting process $\sum_i \mathbf{1}(X_i \in I_k, Y_i \leq t)$ is a point process in t with deterministic intensity $\lambda_t = n \int_{I_k} (t - g(x))_+ dx$. Hence, $(N(t), t \in \mathbb{R})$ is a process with independent increments satisfying (e.g., via Proposition 2.32 in [11])

$$\mathbb{E}[N(t)] = \int_{I_k} n(t - g(x))_+ w(x) dx, \quad \text{Var}(N(t)) = \int_{I_k} n(t - g(x))_+ w(x)^2 dx.$$

In particular, $M(t) = N(t) - \mathbb{E}[N(t)]$ is a càdlàg martingale with respect to the filtration

$$\mathcal{F}_t = \sigma((X_i, Y_i)\mathbf{1}(Y_i \leq t), i \geq 1), \quad t \in \mathbb{R}, \tag{2.1}$$

with mean zero and predictable quadratic variation $\langle M \rangle_t = \text{Var}(N(t))$.

Now note that $\tau := Y_k^* + Rh^\beta$ is an (\mathcal{F}_t) -stopping time with

$$\begin{aligned} P(\tau \geq t) &= \exp\left(-n \int_{I_k} (t - Rh^\beta - g(x))_+ dx\right) \\ &\leq \exp\left(-nh\left(t - \max_{x \in I_k} g(x) - Rh^\beta\right)\right) \end{aligned} \tag{2.2}$$

for $t \geq \max_{x \in I_k} g(x) + Rh^\beta$. In particular, τ has finite expectation and Lemma A.1 on optional stopping yields

$$\mathbb{E}[M(\tau)] = 0 \quad \Rightarrow \quad \mathbb{E}[N(\tau)] = n \int_{I_k} \mathbb{E}[(\tau - g(x))_+] w(x) dx$$

and

$$\text{Var}(M(\tau)) = \mathbb{E}[\langle M \rangle_\tau] = n \int_{I_k} \mathbb{E}[(\tau - g(x))_+] w(x)^2 dx.$$

Noting $\tau \geq g(x)$ we have

$$\mathbb{E}[(\tau - g(x))_+] = \mathbb{E}[Y_k^*] + Rh^\beta - g(x) \quad \text{for all } x \in I_k.$$

The identity

$$\hat{\vartheta}_k = \tau \bar{w}_k - \frac{1}{nh} N(\tau) = \vartheta_k - \frac{1}{nh} M(\tau)$$

implies $\mathbb{E}[\hat{\vartheta}_k] = \vartheta_k$ and

$$\text{Var}(\hat{\vartheta}_k) = \frac{1}{n^2 h^2} \text{Var}(M(\tau)) = \frac{1}{n h^2} \int_{I_k} \mathbb{E}[Y_k^* + Rh^\beta - g(x)] w(x)^2 dx.$$

A rough universal bound, using that $Y_k^* - \max_{x \in I_k} g(x)$ is stochastically smaller than the minimum in y of a PPP with intensity $n\mathbf{1}(x \in I_k, y \geq 0)$, yields with a random variable $E \sim \text{Exp}(nh)$

$$\mathbb{E}[Y_k^*] \leq \mathbb{E}\left[\max_{x \in I_k} g(x) + E\right] \leq g(x) + Rh^\beta + (nh)^{-1}.$$

This implies

$$\text{Var}(\hat{\vartheta}_k) \leq \frac{2Rh^\beta + (nh)^{-1}}{nh^2} \int_{I_k} w(x)^2 dx.$$

We conclude for the final estimator $\hat{\vartheta}^{\text{block}} = \sum_{k=0}^{h^{-1}-1} \hat{\vartheta}_k h$ by the independence of $(\hat{\vartheta}_k)_k$ that

$$\mathbb{E}[\hat{\vartheta}^{\text{block}}] = \vartheta, \quad \text{Var}(\hat{\vartheta}^{\text{block}}) \leq \frac{2Rh^\beta + (nh)^{-1}}{n} \int_0^1 w(x)^2 dx.$$

Finally, insertion of the asymptotically optimal h yields the variance bound. □

2.2. Blockwise estimation in the regression-type model

We consider the equi-distant regression model (1.1) where (ε_i) are i.i.d. satisfying (1.4). The primary example will be $\varepsilon_i \sim \text{Exp}(\lambda)$, but any distribution on \mathbb{R}^+ with a Lipschitz continuous density f_ε at zero and $f_\varepsilon(0) = \lambda$ will be covered, as soon as some loose tail bound at infinity holds.

Since the observation design is discrete, our parameter of interest becomes $\vartheta^{(n)}$ from (1.2). In analogy with the PPP case we build an estimator for $\vartheta_k = \frac{1}{nh} \sum_{i \in \tilde{I}_k} g(i/n)w(i/n)$ on each block of indices $\tilde{I}_k := \{i : kh < \frac{i}{n} \leq (k+1)h\}$, where $h^{-1}, nh \in \mathbb{N}$:

$$\tilde{\vartheta}_k := \frac{1}{nh} \sum_{i \in \tilde{I}_k} (\mathcal{Y}_i \wedge (\mathcal{Y}_k^* + Rh^\beta) - \lambda^{-1} \mathbf{1}(\mathcal{Y}_i \leq \mathcal{Y}_k^* + Rh^\beta))w(i/n). \tag{2.3}$$

Here, $\mathcal{Y}_k^* = \min_{i \in \tilde{I}_k} \mathcal{Y}_i$ is again the minimal observation on each block. In contrast to the PPP-estimator, the empirical upper bound for g on I_k is given by the minimum of $\mathcal{Y}_k^* + Rh^\beta$ and \mathcal{Y}_i , which for the rate-optimal choice of h , however, has negligible impact. We obtain the following result where $\|w\|_p = (\frac{1}{n} \sum_{i=1}^n |w(i/n)|^p)^{1/p}$ denotes the standardised ℓ^p -norm.

Theorem 2.2. *Let the i.i.d. error variables ε_i satisfy (1.4) as well as $\bar{F}_\varepsilon(y) = P(\varepsilon_i > y) \lesssim (1+y)^{-\rho}$ for some $\rho > 0$. For $g \in C^\beta(\mathbb{R})$, the estimator $\tilde{\vartheta}_n^{\text{block}} = \sum_{k=0}^{h^{-1}-1} \tilde{\vartheta}_k h$ satisfies for $h \rightarrow 0$ with $nh \rightarrow \infty$ uniformly in n, h, R, β*

$$\begin{aligned} |\mathbb{E}[\tilde{\vartheta}_n^{\text{block}} - \vartheta^{(n)}]| &\lesssim (Rh^\beta + (nh)^{-1})^2 \|w\|_1, \\ \text{Var}(\tilde{\vartheta}_n^{\text{block}}) &\lesssim h(Rh^\beta + (nh)^{-1})^2 \|w\|_2^2. \end{aligned}$$

In particular, uniformly over $\beta \geq \beta_0 > 1/2, R \leq R_0 < \infty$ we obtain with the rate-optimal block size $h \sim (Rn)^{-1/(\beta+1)}$

$$\begin{aligned} (\mathbb{E}[\tilde{\vartheta}_n^{\text{block}} - \vartheta^{(n)}])^2 &= o(\text{Var}(\tilde{\vartheta}_n^{\text{block}})), \\ \text{Var}(\tilde{\vartheta}_n^{\text{block}}) &\lesssim R^{1/(\beta+1)} n^{-(2\beta+1)/(\beta+1)} \|w\|_2^2. \end{aligned}$$

In the case $\varepsilon_i \sim \text{Exp}(\lambda)$ we have for any $\beta \in (0, 1], R, \lambda > 0$ the more precise result

$$\mathbb{E}[\tilde{\vartheta}_n^{\text{block}}] = \vartheta, \quad \text{Var}(\tilde{\vartheta}_n^{\text{block}}) \leq \frac{2Rh^\beta + (n\lambda h)^{-1}}{n\lambda} \|w\|_2^2.$$

Remark 2.3. *The result and proof for $\varepsilon_i \sim \text{Exp}(\lambda)$ are exactly as in the PPP model. For other distributions of ε_i the estimator is only asymptotically unbiased, but for Hölder regularity $\beta > 1/2$ the bias is negligible with respect to the stochastic error. A side remark is that for strong asymptotic equivalence with the PPP model in Le Cam’s sense the necessary minimal regularity $\beta > 1$ is higher, see [18].*

Proof. Fix a block with index k and consider for $t \in \mathbb{R}$

$$M(t) := \sum_{i \in \tilde{I}_k} (\mathbf{1}(\mathcal{Y}_i \leq t) + \log \bar{F}_\varepsilon(\mathcal{Y}_i \wedge t - g(i/n)))w(i/n). \tag{2.4}$$

With respect to the filtration $\mathcal{F}_t = \sigma(\mathcal{Y}_i \mathbf{1}(\mathcal{Y}_i \leq t), i \in \tilde{I}_k)$, $(M(t), t \in \mathbb{R})$ defines a martingale with $\mathbb{E}[M(t)] = 0$ and quadratic variation $\langle M \rangle_t = \sum_{i \in \tilde{I}_k} (-\log \bar{F}_\varepsilon(\mathcal{Y}_i \wedge t - g(i/n)))w(i/n)^2$: just note that the compensator of $\sum_{i \in \tilde{I}_k} \mathbf{1}(\mathcal{Y}_i \leq t)$ equals the integrated hazard function $\sum_{i \in \tilde{I}_k} \int_0^{\mathcal{Y}_i \wedge t - g(i/n)} \bar{F}_\varepsilon(s)^{-1} dF_\varepsilon(s)$. Moreover, $\tau := \mathcal{Y}_k^* + Rh^\beta$ is a stopping time with respect to (\mathcal{F}_t) . From the representation

$$\tilde{\vartheta}_k - \vartheta_k = \frac{1}{n\lambda h} \left(\sum_{i \in \tilde{I}_k} G_\varepsilon(\mathcal{Y}_i \wedge \tau - g(i/n))w(i/n) - M(\tau) \right) \tag{2.5}$$

with $G_\varepsilon(z) := \lambda z + \log \bar{F}_\varepsilon(z)$

and the stopping Lemma A.1 in combination with $\mathbb{E}[\tau] < \infty$ due to the moment bound from Lemma A.2 below we conclude

$$\mathbb{E}[\tilde{\vartheta}_k - \vartheta_k] = \frac{1}{n\lambda h} \sum_{i \in \tilde{I}_k} \mathbb{E}[G_\varepsilon(\mathcal{Y}_i \wedge \tau - g(i/n))]w(i/n).$$

In the case $\varepsilon_i \sim \text{Exp}(\lambda)$, we have $\log(\bar{F}_\varepsilon(z)) = -\lambda z$, that is $G_\varepsilon(z) = 0$, and the estimator is unbiased.

By Assumption (1.4), there is some $\delta > 0$ such that $G_\varepsilon(z) = O(z^2)$ holds for $z \in [0, \delta]$. For $z \geq \delta$, we have also $G_\varepsilon(z) \leq \lambda z = O(z^2)$ and $G_\varepsilon(z) \geq -|\log(\bar{F}_\varepsilon(z))|$. By Lemma A.2, Cauchy–Schwarz inequality, monotonicity of \bar{F}_ε and $\bar{F}_\varepsilon(\varepsilon_i) \sim U[0, 1]$ this leads to

$$\begin{aligned} & |\mathbb{E}[\tilde{\vartheta}_k - \vartheta_k]| \\ & \lesssim \sum_{i \in \tilde{I}_k} \frac{\mathbb{E}[(\tau - g(i/n))^2] + \mathbb{E}[\log(\bar{F}_\varepsilon(\mathcal{Y}_i - g(i/n)))^2]^{1/2} P(|\log(\bar{F}_\varepsilon(\tau - g(i/n)))| > \delta)^{1/2}}{nh} \\ & \quad \times |w(i/n)| \\ & \lesssim \sum_{i \in \tilde{I}_k} \frac{\mathbb{E}[(\tau - g(i/n))^2] + \mathbb{E}[\log(\bar{F}_\varepsilon(\varepsilon_i))^2]^{1/2} P(\bar{F}_\varepsilon(\min_{i \in \tilde{I}_k} \varepsilon_i + 2Rh^\beta) < e^{-\delta})^{1/2}}{nh} \\ & \quad \times |w(i/n)| \\ & \lesssim \frac{(Rh^\beta + (nh)^{-1})^2 + \bar{F}_\varepsilon(\bar{F}_\varepsilon^{-1}(e^{-\delta}) - 2Rh^\beta)^{nh/2}}{nh} \sum_{i \in \tilde{I}_k} |w(i/n)|. \end{aligned}$$

With $h \rightarrow 0$ and $nh \rightarrow \infty$ the second term in the numerator converges geometrically fast to zero and the assertion for the bias of $\tilde{\vartheta}_n^{\text{block}}$ follows.

For the variance bound we use $\text{Var}(A + B) \leq 2\text{Var}(A) + 2\text{Var}(B)$, $\text{Var}(\sum_{i \in \tilde{I}_k} A_i) \leq nh \sum_{i \in \tilde{I}_k} \mathbb{E}[A_i^2]$, $|G_\varepsilon(z)| \lesssim z + |\log \bar{F}_\varepsilon(z)|\mathbf{1}(z > \delta)$ and the stopping Lemma A.1 to obtain in

analogy with the bias part:

$$\begin{aligned}
 \text{Var}(\tilde{\vartheta}_k) &= \frac{1}{(n\lambda h)^2} \text{Var}\left(\sum_{i \in \tilde{I}_k} G_\varepsilon(\mathcal{Y}_i \wedge \tau - g(i/n))w(i/n) - M(\tau)\right) \\
 &\leq \frac{2}{(n\lambda h)^2} \sum_{i \in \tilde{I}_k} \mathbb{E}[(nhG_\varepsilon(\mathcal{Y}_i \wedge \tau - g(i/n)))^2 + |\log \bar{F}_\varepsilon(\mathcal{Y}_i \wedge \tau - g(i/n))|]w(i/n)^2 \\
 &\lesssim \sum_{i \in \tilde{I}_k} \left(\frac{\mathbb{E}[(\tau - g(i/n))^2] + \mathbb{E}[|\log \bar{F}_\varepsilon(\varepsilon_i)|^4]^{1/2} P(\bar{F}_\varepsilon(\min_{i \in \tilde{I}_k} \varepsilon_i + 2Rh^\beta) < e^{-\delta})^{1/2}}{nh} \right. \\
 &\quad \left. + \frac{\mathbb{E}[\tau - g(i/n)] + \mathbb{E}[|\log \bar{F}_\varepsilon(\varepsilon_i)|^2]^{1/2} P(\bar{F}_\varepsilon(\min_{i \in \tilde{I}_k} \varepsilon_i + 2Rh^\beta) < e^{-\delta})^{1/2}}{(nh)^2} \right) \\
 &\quad \times w(i/n)^2 \\
 &\lesssim \frac{(Rh^\beta + (nh)^{-1})^2 + (Rh^\beta + (nh)^{-1})(nh)^{-1} + \bar{F}_\varepsilon(\bar{F}_\varepsilon^{-1}(e^{-\delta}) - 2Rh^\beta)^{nh/2}}{nh} \\
 &\quad \times \sum_{i \in \tilde{I}_k} w(i/n)^2 \\
 &\lesssim \frac{(Rh^\beta + (nh)^{-1})^2}{nh} \sum_{i \in \tilde{I}_k} w(i/n)^2.
 \end{aligned}$$

For the global estimator, we infer $\text{Var}(\tilde{\vartheta}_n^{\text{block}}) \lesssim h(Rh^\beta + (nh)^{-1})^2 \|w\|_2^2$ by independence of $(\tilde{\vartheta}_k)$. It remains to insert the rate-optimal choice of h and to note that $n^{-4\beta/(\beta+1)} = o(n^{-(2\beta+1)/(\beta+1)})$ holds for $\beta > 1/2$.

Finally, in the case $\varepsilon_i \sim \text{Exp}(\lambda)$ we have $\mathbb{E}[\mathcal{Y}_k^* - \max_i g(i/n)] \leq (n\lambda h)^{-1}$ and $\text{Var}(\tilde{\vartheta}_k) = \frac{\mathbb{E}[(M)_\tau]}{(n\lambda h)^2}$. Consequently,

$$\text{Var}(\tilde{\vartheta}_k) \leq \sum_{i \in \tilde{I}_k} \frac{\mathbb{E}[\mathcal{Y}_k^* + Rh^\beta - g(i/n)]}{\lambda(nh)^2} w(i/n)^2 \leq \frac{2Rh^\beta + (n\lambda h)^{-1}}{\lambda(nh)^2} \sum_{i \in \tilde{I}_k} w(i/n)^2,$$

which by independence gives the asserted bound for $\text{Var}(\tilde{\vartheta}_n^{\text{block}})$. □

2.3. Adaptive estimation

We now address the question of choosing the block size h in a data-driven way, not assuming the regularity parameters R and β to be known. We apply Lepski’s method [16] and treat the general regression-type model (1.1). The main technical work is devoted to obtaining explicit critical values in Proposition A.4 of the Appendix. To this end, the critical values are defined via

the compensator of an exponential counting process and are thus itself again stochastic. While an explicit non-asymptotic risk analysis is clearly possible, we focus here on the asymptotic risk, showing that by the versatility of Lepski's method rate-optimal adaptive estimation up to logarithmic factors is indeed possible in our non-regular situation.

For a choice $n^{-1}(\log n)^2 \leq h_1 < \dots < h_M \leq 1$ of block sizes h_m with $h_m^{-1}, nh_m \in \mathbb{N}$ consider the corresponding blockwise estimators

$$\tilde{\vartheta}_{n,h_m}^{\text{block}} = \frac{1}{n} \sum_{k=0}^{h_m^{-1}-1} \sum_{i \in \tilde{I}_{k,h_m}} (\mathcal{Y}_i \wedge (\mathcal{Y}_{k,h_m}^* + (nh_m)^{-1}) - \lambda^{-1} \mathbf{1}(\mathcal{Y}_i \leq \mathcal{Y}_{k,h_m}^* + (nh_m)^{-1})) w(i/n),$$

where the subscript h_m marks all quantities depending on the block size. Remark that the intercept Rh_m^β in (2.3) has been replaced by the asymptotically balanced size $(nh_m)^{-1}$, which does not depend on the unknown R and β . Among $(h_m)_{1 \leq m \leq M}$ we select the block size adaptively via

$$\hat{h} := \inf \{ h_m \mid \exists m' \leq m : |\tilde{\vartheta}_{n,h_{m'}}^{\text{block}} - \tilde{\vartheta}_{n,h_{m+1}}^{\text{block}}| > \kappa_{m+1} + \kappa_{m'} \} \wedge h_M$$

with critical values (k_i denotes the block k with $i \in \tilde{I}_k$)

$$\kappa_m = \sum_{i=1}^n \left(\mathbf{1}(\mathcal{Y}_i \leq \mathcal{Y}_{k_i,h_m}^* + (nh_m)^{-1}) \frac{H_{\sqrt{c \log n}}(h_m^{1/2} w(i/n))}{n \lambda h_m^{1/2}} \right) + \frac{(C c \log n)^2 \|w\|_1}{(nh_m)^2 \lambda} + \frac{\sqrt{c \log n}}{2n \lambda h_m^{1/2}}.$$

Here, the function $H_x(y) = \frac{\log(1-2x|y|)}{-2x} - |y|$ and the constant $C > 0$ with property $|\lambda z + \log(\bar{F}_\varepsilon(z))| \leq C^2 z^2$ for $z \in [0, \delta]$ are used and $c > 0$ is specified below. Asymptotically, we have $H_x(y) \approx xy^2$ as $x \rightarrow 0$ and $C \approx -(f'_\varepsilon(0) + f_\varepsilon(0)^2)$ in the case of a differentiable density f_ε of ε_i around zero (note $C = 0$ for $\varepsilon_i \sim \text{Exp}(\lambda)$). Then the proof in the [Appendix](#) yields the following risk bounds.

Theorem 2.4. *Assume $g \in \mathcal{C}^\beta(R)$, $\sup_x |w(x)| < \infty$ and that $f_\varepsilon/\bar{F}_\varepsilon$ is bounded. The adaptive estimator $\tilde{\vartheta}_n^{\text{block}} = \tilde{\vartheta}_{n,\hat{h}}^{\text{block}}$ satisfies with $h^* := \sup \{ h_m \mid Rh_m^\beta \leq (nh_m)^{-1} \} \vee h_1$ and for $\underline{\lambda} \in (0, \lambda)$, n sufficiently large*

$$\begin{aligned} \mathbb{E}[(\tilde{\vartheta}_n^{\text{block}} - \vartheta^{(n)})^2 \mathbf{1}(\hat{h} < h^*)] &\lesssim M(n^{-c} + n^{(1-\underline{\lambda}c)/2}), \\ \mathbb{E}[(\tilde{\vartheta}_n^{\text{block}} - \tilde{\vartheta}_{n,h^*}^{\text{block}})^2 \mathbf{1}(\hat{h} \geq h^*)] &\lesssim \frac{(\log n)^4}{(nh^*)^4} + \frac{M \log n}{n^2 h^*}. \end{aligned}$$

Choosing $c > 5\lambda^{-1} \vee 2$ and asymptotically $h_0 \sim (\log n)^2 n^{-1}$, $h_m \sim h_0 q^m$ for $m = 1, \dots, M$, $q > 1$, $M \sim \log n$, the adaptive estimator exhibits the asymptotic rate

$$\mathbb{E}[(\tilde{\vartheta}_n^{\text{block}} - \vartheta^{(n)})^2] \lesssim (\log n)^2 n^{-(2\beta+1)/(\beta+1)} + (\log n)^4 n^{-4\beta/(\beta+1)}.$$

In particular, for $\beta \geq 1/2$ the estimator achieves the minimax optimal rate up to a logarithmic factor.

Remark 2.5. The oracle-type block size h^* is the largest block size among (h_m) such that $\tilde{\vartheta}_{n,h_m}^{\text{block}}$ remains unbiased, except for the distribution bias induced in the case $\varepsilon_i \approx \text{Exp}(\lambda)$. As the proof shows, in the case $\varepsilon_i \sim \text{Exp}(\lambda)$ not only the critical values ($C = 0$), but also the bounds simplify. We obtain $\mathbb{E}[(\tilde{\vartheta}_n^{\text{block}} - \tilde{\vartheta}_{n,h^*}^{\text{block}})^2 \mathbf{1}(\hat{h} > h^*)] \lesssim \frac{M \log n}{n^2 h^*}$ and thus the minimax optimal rate up to a log-factor for any $\beta > 0$.

More elaborate arguments in the proof could give a smaller exponent for the logarithmic factor, but it is quite likely that some logarithmic factor has to be paid necessarily for adaptation, cf. [9] for a related result. Similarly, the hypotheses that w and $f_\varepsilon/\bar{F}_\varepsilon$ are uniformly bounded are certainly not necessary, but permit more concise and transparent bounds.

2.4. Rate optimality

We prove that the rate $R^{1/(2\beta+2)} n^{-(\beta+1/2)/(\beta+1)}$ is optimal in a minimax sense over $\mathcal{C}^\beta(R)$. The proof is conducted for the PPP model, the regression case with $\varepsilon_i \sim \text{Exp}(\lambda)$ can be treated in the same way.

Theorem 2.6. For estimating $\vartheta = \int_0^1 g(x)w(x) dx$, $w \in L^2([0, 1])$, over the parameter class $\mathcal{C}^\beta(R)$, $\beta \in (0, 1]$, $R > 0$, the following asymptotic lower bound holds:

$$\liminf_{n \rightarrow \infty} \inf_{\hat{\vartheta}_n} \sup_{g \in \mathcal{C}^\beta(R)} R^{-1/(\beta+1)} n^{(2\beta+1)/(\beta+1)} \|w\|_{L^2}^{-2} \mathbb{E}_g [(\hat{\vartheta}_n - \vartheta)^2] > 0.$$

The infimum extends over all estimators $\hat{\vartheta}_n$ from the PPP model with intensity (1.6).

Proof. The proof is based on a Bayesian risk bound, which clearly provides a lower bound for the minimax risk, see [13] for similar approaches. Take an independent Bernoulli sequence (ε_k) , that is, $P(\varepsilon_k = 1) = p$, $P(\varepsilon_k = 0) = 1 - p$ with $p \in (0, 1)$, and set for a triangular kernel $K(y) = 2 \min(y, 1 - y) \mathbf{1}_{[0,1]}(y)$

$$g(x) = \sum_{k=0}^{h^{-1}-1} \varepsilon_k g_k(x) \quad \text{with } g_k(x) = c R h^\beta K((x - kh)/h),$$

where $h \in (0, 1)$ with $h^{-1} \in \mathbb{N}$ will be chosen later. Then for $c > 0$ sufficiently small, we have $g \in \mathcal{C}^\beta(R)$ for all h and all realisations of (ε_k) . We interpret this specification of g as a prior on $\mathcal{C}^\beta(R)$ and we shall make use of the independence of prior as well as the observation laws on different blocks $I_k = [kh, (k + 1)h)$. For each k , we obtain from the Bayes formula the posterior probability given the observations of the PPP in interval I_k (cf. the likelihood derivation in (3.1) below)

$$\hat{\varepsilon}_k := P(\varepsilon_k = 1 | (X_i, Y_i)_{i \geq 1}) = \frac{p e^{n \int g_k} \mathbf{1}(\forall X_i \in I_k : Y_i \geq g_k(X_i))}{1 - p + p e^{n \int g_k}}.$$

Using that ε_k are 0–1-valued, we have $\hat{\varepsilon}_k = \mathbb{E}[\varepsilon_k | (X_i, Y_i)_{i \geq 1}]$ and $\text{Var}(\varepsilon_k | (X_i, Y_i)_{i \geq 1}) = \hat{\varepsilon}_k(1 - \hat{\varepsilon}_k)$. Therefore, the Bayes-optimal estimator of ϑ under squared loss is given by the posterior

mean

$$\hat{\vartheta} = \sum_{k=0}^{h^{-1}-1} \hat{\varepsilon}_k \int_{I_k} g_k(x) w(x) dx.$$

Using independence and $\mathbb{E}[\hat{\varepsilon}_k - \varepsilon_k] = 0$, its Bayes risk is calculated as

$$\begin{aligned} \mathbb{E}[(\hat{\vartheta} - \vartheta)^2] &= \sum_{k=0}^{h^{-1}-1} \text{Var}(\hat{\varepsilon}_k - \varepsilon_k) \left(\int_{I_k} g_k(x) w(x) dx \right)^2 \\ &= \sum_{k=0}^{h^{-1}-1} \mathbb{E}[\text{Var}(\varepsilon_k | (X_i, Y_i)_{i \geq 1})] \left(\int_{I_k} g_k(x) w(x) dx \right)^2 \\ &= \sum_{k=0}^{h^{-1}-1} \frac{p(1-p)}{(1-p + p e^{n \int g_k})^2} \left(\int_{I_k} g_k(x) w(x) dx \right)^2. \end{aligned}$$

We now choose $h = \lceil (cRn)^{1/(\beta+1)} \rceil^{-1}$ such that $n \int_{I_k} g_k(x) dx = cRh^{\beta+1}n \leq 1$ holds. Then the Bayes risk is bounded in order by

$$\mathbb{E}[(\hat{\vartheta} - \vartheta)^2] \gtrsim R^2 h^{2\beta+1} \sum_{k=0}^{h^{-1}-1} \left(\int_{I_k} \frac{K((x - kh)/h)}{\|K((\bullet - kh)/h)\|_{L^2}} w(x) dx \right)^2.$$

The same argument over the shifted blocks $I'_k = [(k + 1/2)h, (k + 3/2)h)$ implies that the minimax risk is bounded by the maximum (and thus the average) over the respective Bayes risks:

$$\inf_{\hat{\vartheta}_n} \sup_{g \in \mathcal{C}^\beta(R)} \mathbb{E}_g[(\hat{\vartheta}_n - \vartheta)^2] \gtrsim R^2 h^{2\beta+1} \sum_{k=0}^{2h^{-1}-2} \left(\int_{I_k} \frac{K((x - kh/2)/h)}{\|K((\bullet - kh/2)/h)\|_{L^2}} w(x) dx \right)^2.$$

The tent functions $K((x - kh/2)/h)$, $k = 0, 1, \dots, 2h^{-1} - 2$ form a Riesz basis for their linear span V_h (see, e.g., Example 2.1 in [22]), which is the space of all linear splines with knots at $kh/2$, vanishing at the boundary. This means that the sum in the last display is larger than a constant times the $L^2([0, 1])$ -norm of the orthogonal projection of w onto V_h . As $\bigcup_{h>0} V_h$ is dense in $L^2([0, 1])$, the L^2 -norm of the projections of w onto V_h converges for $h \rightarrow 0$ to the L^2 -norm of w . Insertion of $h \sim (Rn)^{-1/(\beta+1)}$ yields the desired lower bound rate $R^{1/(\beta+1)} n^{-(2\beta+1)/(\beta+1)} \|w\|_{L^2}^2$. \square

3. Nonparametric maximum-likelihood

3.1. The MLE over $\mathcal{C}^\beta(R)$

Let us study the nonparametric maximum-likelihood estimator (MLE) in the class $\mathcal{C}^\beta(R)$. Denote by P_g the law of the observations in the PPP model with intensity $\lambda_g(x, y) = n\mathbf{1}(y \geq g(x))$.

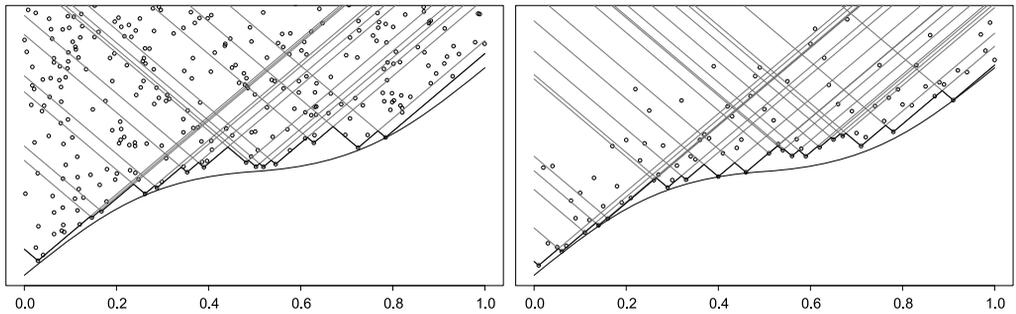


Figure 2. Construction of \hat{g}^{MLE} in the PPP model ($n = 100$, left) and in the regression-type model ($n = 100$, $\varepsilon_i \sim \text{Exp}(1)$, right) for $\beta = 1$. Thin lines indicate $x \mapsto Y_j + R|x - X_j|$ at observations on the graph of \hat{g}^{MLE} .

Then for $g \geq g_0$ by Theorem 1.3 in [14] and the fact that the PPP intensities coincide outside the compact set $[0, 1] \times [\min g_0, \max g]$ we obtain the Radon–Nikodym-derivative

$$\frac{dP_g}{dP_{g_0}} = \exp\left(n \int_0^1 (g - g_0)(x) dx\right) \mathbf{1}(\forall i : Y_i \geq g(X_i)).$$

A simple probability measure P_0 dominating all P_g , $g \in C^\beta(\mathbb{R})$ (where g need not be bounded from below), is given by the PPP model with intensity $\lambda_0(x, y) = n(e^y \wedge 1)$ and yields again via Theorem 1.3 in [14] the likelihood

$$\begin{aligned} \mathcal{L}(g) &= \frac{dP_g}{dP_0} \\ &= \left(\prod_{j \geq 1} \frac{n \mathbf{1}(Y_j \geq g(X_j))}{n(e^{Y_j} \wedge 1)} \right) \exp\left(-n \int_0^1 \int_{-\infty}^{\infty} (\mathbf{1}(y \geq g(x)) - e^y \wedge 1) dy dx\right) \\ &= \left(\prod_{j \geq 1} e^{(-Y_j)_+} \mathbf{1}(Y_j \geq g(X_j)) \right) \exp\left(-n \int_0^1 (-1 - g(x)) dx\right) \\ &= \exp\left(n + \sum_{j \geq 1} (-Y_j)_+\right) \exp\left(n \int_0^1 g(x) dx\right) \mathbf{1}(\forall j \geq 1 : Y_j \geq g(X_j)). \end{aligned} \tag{3.1}$$

The first factor is independent of g and we obtain thus the same structure as under P_{g_0} above. The MLE over $C^\beta(\mathbb{R})$ is the function \hat{g} that maximizes $\int_0^1 g$ over all $g \in C^\beta(\mathbb{R})$ with $g(X_j) \leq Y_j$ for all j . We can write explicitly

$$\hat{g}^{\text{MLE}}(x) = \min_{j \geq 1} (Y_j + R|x - X_j|^\beta),$$

since the right-hand side even maximises $g(x)$ pointwise over the considered class of g , see also Figure 2. The corresponding likelihood (with respect to n -dimensional Lebesgue measure) in the

regression-type model (1.1) with $\varepsilon_i \sim \text{Exp}(\lambda)$ i.i.d. is given by

$$\mathcal{L}^{\text{regr}}(g) = \lambda^n \exp\left(-\lambda \sum_{i=1}^n \mathcal{Y}_i\right) \exp\left(\lambda \sum_{i=1}^n g(i/n)\right) \mathbf{1}(\forall i = 1, \dots, n : \mathcal{Y}_i \geq g(i/n)).$$

The maximum-likelihood estimator over $\mathcal{C}^\beta(R)$ is then similarly given by

$$\hat{g}^{\text{MLE-regr}}(x) = \min_{i=1, \dots, n} (\mathcal{Y}_i + R|x - i/n|^\beta), \quad x \in [0, 1], \tag{3.2}$$

see Figure 2 for an illustration of the two constructions of the MLE. They are both quickly determined numerically. In the sequel, we shall focus on the MLE in the PPP model and only hint at the parallel theory for the regression-type model under exponential noise. We abstain from analysing the exponential MLE under non-exponential noise in the regression-type model because the results must be asymptotic in nature and will be comparable to Theorem 2.2.

Proposition 3.1. *The nonparametric MLE $(\hat{g}^{\text{MLE}}(x), x \in [0, 1])$ is a sufficient and complete statistics for $\mathcal{C}^\beta(R)$.*

Proof. By definition of \hat{g}^{MLE} the likelihood (3.1) can be written as

$$\mathcal{L}(g) = \exp\left(n + \sum_{j \geq 1} (-Y_j)_+\right) \exp\left(n \int_0^1 g(x) dx\right) \mathbf{1}(g \leq \hat{g}^{\text{MLE}})$$

such that by Neyman’s factorisation criterion (e.g., [15]) \hat{g}^{MLE} is a sufficient statistics for this parameter class.

Let us remark that by definition \hat{g}^{MLE} is an element of $\mathcal{C}^\beta(R)$. Since $\mathcal{C}^\beta(R)$, equipped with its \mathcal{C}^β -norm, is not separable, we equip it with the Borel σ -algebra generated by the uniform (supremum) norm, which is generated by all point evaluations. Measurability of the estimator \hat{g}^{MLE} is then easily established since all point evaluations $\hat{g}^{\text{MLE}}(x), x \in [0, 1]$, are measurable as minima of countably many random variables.

For completeness, we now consider any statistic $T : \mathcal{C}^\beta(R) \rightarrow \mathbb{R}$ satisfying $\mathbb{E}_g[T(\hat{g}^{\text{MLE}})] = 0$ for all $g \in \mathcal{C}^\beta(R)$, which is Borel measurable with respect to the uniform norm. For $g \in \mathcal{C}^\beta(R)$ denote by $[g, \infty) := \{h \in \mathcal{C}^\beta(R) | h \geq g\}$ the “bracket” between g and ∞ , that is all functions whose graphs lie above g . Noting $[g, \infty) \cap [h, \infty) = [g \vee h, \infty)$ where the maximum $g \vee h$ is again in $\mathcal{C}^\beta(R)$, the family $\{[g, \infty) | g \in \mathcal{C}^\beta(R)\}$ is an \cap -stable generator of the uniform Borel σ -algebra in $\mathcal{C}^\beta(R)$: for any $x_0 \in [0, 1], y_0 \in \mathbb{R}$ we have $\{h \in \mathcal{C}^\beta(R) | h(x_0) \geq y_0\} = [y_0 - R|\bullet - x_0|^\beta, \infty)$ by the Hölder condition and $\{[y_0, \infty) | y_0 \in \mathbb{R}\}$ generates the Borel σ -algebra on \mathbb{R} .

From $\mathbb{E}_g[T(\hat{g}^{\text{MLE}})] = 0$, we obtain by using the likelihood under P_0

$$e^{n \int (g+1)} \mathbb{E}_0[T(\hat{g}^{\text{MLE}}) e^{\sum_{j \geq 1} (-Y_j)_+} \mathbf{1}(\hat{g}^{\text{MLE}} \in [g, \infty))] = 0.$$

Splitting $T = T^+ - T^-$ with non-negative T^+, T^- , we infer that the measures $B \mapsto \mathbb{E}_0[T^\pm(\hat{g}^{\text{MLE}}) e^{\sum_{j \geq 1} (-Y_j)_+} \mathbf{1}(\hat{g}^{\text{MLE}} \in B)]$ agree on $\{[g, \infty) | g \in \mathcal{C}^\beta(R)\}$ and thus by the unique-

ness theorem for all uniform Borel sets B in $\mathcal{C}^\beta(R)$, in particular for $B = \{T > 0\}$ and $B = \{T < 0\}$. This implies $T^+(\hat{g}^{\text{MLE}})e^{\sum_{j \geq 1}(-Y_j)_+} = T^-(\hat{g}^{\text{MLE}})e^{\sum_{j \geq 1}(-Y_j)_+}$ P_0 -a.s. and thus $T(\hat{g}^{\text{MLE}}) = 0$ P_g -a.s. for all $g \in \mathcal{C}^\beta(R)$. \square

In analogy with the block-wise estimator $\hat{\vartheta}^{\text{block}}$ we set

$$\hat{\vartheta}^{\text{MLE}} := \int_0^1 \hat{g}^{\text{MLE}}(x)w(x) dx - \frac{1}{n} \sum_{j \geq 1} \mathbf{1}(\hat{g}^{\text{MLE}}(X_j) = Y_j)w(X_j).$$

This means that $\hat{\vartheta}^{\text{MLE}}$ is obtained by a plug-in of the nonparametric MLE \hat{g}^{MLE} into the functional minus a bias correction which counts the relative number of observations on the graph of \hat{g}^{MLE} . The striking result is that this estimator is not only unbiased, but even uniformly of minimum variance among all unbiased estimators for the class $\mathcal{C}^\beta(R)$ (UMVU).

Theorem 3.2. *The estimator $\hat{\vartheta}^{\text{MLE}}$ is for each finite sample size n UMVU over the class $\mathcal{C}^\beta(R)$ with*

$$\begin{aligned} \text{Var}(\hat{\vartheta}^{\text{MLE}}) &= \frac{1}{n} \int_0^1 \mathbb{E}[\hat{g}^{\text{MLE}}(x) - g(x)]w(x)^2 dx \\ &\leq \left(\Gamma(\beta/(\beta + 1))\beta(2R/(\beta + 1))^{1/(\beta+1)} n^{-(2\beta+1)/(\beta+1)} + \frac{1}{n^2} e^{-2\beta Rn/(\beta+1)} \right) \|w\|_{L^2}^2. \end{aligned}$$

For $n \rightarrow \infty$, we obtain

$$\text{Var}(\hat{\vartheta}^{\text{MLE}}) \leq (2 + o(1))R^{1/(\beta+1)}n^{-(2\beta+1)/(\beta+1)} \|w\|_{L^2}^2.$$

Proof. Let us define another weighted counting process

$$\bar{N}(t) = \sum_{j \geq 1} \mathbf{1}\left(Y_j \leq t \wedge \min_{i \geq 1} (Y_i + R|X_j - X_i|^\beta)\right)w(X_j), \quad t \in \mathbb{R}.$$

Note $\{\min_{i \geq 1} (Y_i + R|X_j - X_i|^\beta) < t\} = \{\min_{i: Y_i < t} (Y_i + R|X_j - X_i|^\beta) < t\}$ for each t and that the pure ($w = 1$) counting process has stochastic intensity

$$\bar{\lambda}_t = n \int_0^1 \int_{[g(x), t]} \mathbf{1}\left(\min_{i: Y_i < s} (Y_i + R|x - X_i|^\beta) \geq s\right) ds dx.$$

Consequently, \bar{N} is adapted to (\mathcal{F}_t) from (2.1) and we obtain by compensation (cf. Proposition 2.32 in [11]) the (\mathcal{F}_t) -martingale

$$\bar{M}(t) = \bar{N}(t) - n \int_0^1 \int_{[g(x), t]} \mathbf{1}\left(\min_{i: Y_i \leq s} (Y_i + R|x - X_i|^\beta) \geq s\right) ds w(x) dx.$$

The main observation is the identity

$$\begin{aligned} \lim_{t \rightarrow \infty} (\bar{N}(t) - \bar{M}(t)) &= n \int_0^1 \int_{g(x)}^{\infty} \mathbf{1}(\min_{i \geq 1} (Y_i + R|x - X_i|^\beta) \geq s) ds w(x) dx \\ &= n \int_0^1 \int_{g(x)}^{\infty} \mathbf{1}(\hat{g}^{\text{MLE}}(x) \geq s) ds w(x) dx \\ &= n \int_0^1 (\hat{g}^{\text{MLE}}(x) - g(x)) w(x) dx, \end{aligned}$$

which tells us that

$$\bar{N}(\infty) := \lim_{t \rightarrow \infty} \bar{N}(t) = \sum_{j \geq 1} \mathbf{1}(\hat{g}^{\text{MLE}}(X_j) \geq Y_j) w(X_j) = \sum_{j \geq 1} \mathbf{1}(\hat{g}^{\text{MLE}}(X_j) = Y_j) w(X_j)$$

simultaneously counts the weighted number of points (X_j, Y_j) on the graph of \hat{g}^{MLE} and equals the scaled bias $n(\int_0^1 \hat{g}^{\text{MLE}}(x) w(x) dx - \vartheta)$ up to a martingale term. We thus have $\hat{\vartheta}^{\text{MLE}} = \int_0^1 \hat{g}^{\text{MLE}}(x) w(x) dx - \frac{1}{n} \bar{N}(\infty) = \vartheta - \frac{1}{n} \bar{M}(\infty)$ where

$$\bar{M}(\infty) = \sum_j \mathbf{1}(Y_j \leq \hat{g}^{\text{MLE}}(x)) w(X_j) - \int_0^1 (\hat{g}^{\text{MLE}}(x) - g(x)) w(x) dx$$

is the a.s. and L^2 -limit of the L^2 -bounded martingale \bar{M} with

$$\begin{aligned} \langle \bar{M} \rangle_t &= n \int_0^1 \int_{[g(x), t]} \mathbf{1}(\min_{i: Y_i < s} (Y_i + R|x - X_i|^\beta) \geq s) ds w(x)^2 dx \\ &\uparrow n \int_0^1 (\hat{g}^{\text{MLE}}(x) - g(x)) w(x)^2 dx =: \langle \bar{M} \rangle_\infty \quad \text{as } t \uparrow \infty. \end{aligned}$$

We obtain from $\mathbb{E}[\hat{\vartheta}^{\text{MLE}} - \vartheta] = \frac{1}{n} \mathbb{E}[-\bar{M}(\infty)]$, $\text{Var}(\hat{\vartheta}^{\text{MLE}}) = \frac{1}{n^2} \text{Var}(\bar{M}(\infty))$ the result (use Lemma A.1 with $\tau = \infty$)

$$\mathbb{E}[\hat{\vartheta}^{\text{MLE}}] = \vartheta$$

and

$$\text{Var}(\hat{\vartheta}^{\text{MLE}}) = \frac{1}{n^2} E[\langle \bar{M} \rangle_\infty] = \frac{1}{n} \int_0^1 \mathbb{E}[\hat{g}^{\text{MLE}}(x) - g(x)] w(x)^2 dx.$$

Hence, $\hat{\vartheta}^{\text{MLE}}$ is an unbiased estimator and by the Lehmann–Scheffé theorem $\hat{\vartheta}^{\text{MLE}}$, derived from a sufficient and complete statistics, is uniformly of minimum variance among all unbiased estimators (e.g., [15]).

To bound the variance we use a universal, but somewhat rough deviation bound for $s \geq 0$ and $x \in [0, 1]$:

$$\begin{aligned}
 P(\hat{g}^{\text{MLE}}(x) - g(x) \geq s) &= \exp\left(-n \int_0^1 (s - R|\xi - x|^\beta + g(x) - g(\xi))_+ d\xi\right) \\
 &\leq \exp\left(-n \int_0^1 (s - 2R|\xi - x|^\beta)_+ d\xi\right) \tag{3.3} \\
 &\leq \begin{cases} \exp\left(-n \frac{2R}{\beta + 1} (s/2R)^{(\beta+1)/\beta}\right), & s \in [0, 2R], \\ \exp(-n(s - 2R/(\beta + 1))), & s > 2R. \end{cases}
 \end{aligned}$$

In the first step, we have evaluated the probability that no observation lies in $\{(\xi, \eta) | \eta + R|x - \xi|^\beta < g(x) + s\}$ using the PPP property. Integrating this survival function bound, we obtain directly

$$\begin{aligned}
 \mathbb{E}[\hat{g}^{\text{MLE}}(x) - g(x)] &= \int_0^\infty P(\hat{g}^{\text{MLE}}(x) - g(x) \geq s) ds \\
 &\leq \int_0^{2R} \exp\left(-n \frac{2R}{\beta + 1} (s/2R)^{(\beta+1)/\beta}\right) ds + \int_{2R}^\infty e^{-n(s-2R/(\beta+1))} ds \tag{3.4} \\
 &= \Gamma(\beta/(\beta + 1))\beta(2R/(\beta + 1))^{1/(\beta+1)} n^{-\beta/(\beta+1)} + \frac{1}{n} e^{-2\beta Rn/(\beta+1)}.
 \end{aligned}$$

Insertion and a numerical evaluation then yield (the maximal constant being attained for $\beta \rightarrow 0$) $\text{Var}(\hat{\vartheta}^{\text{MLE}}) \leq (2 + o(1))R^{1/(\beta+1)} \|w\|_{L^2}^2 n^{-(2\beta+1)/(\beta+1)}$. □

Remark 3.3. The MLE $\hat{g}^{\text{MLE-regr}}$ from (3.2) for the regression-type model is by the same (or simpler) arguments a sufficient and complete statistics over $\mathcal{C}^\beta(R)$. It gives rise to the estimator

$$\hat{\vartheta}^{\text{MLE-regr}} = \frac{1}{n} \sum_{i=1}^n (\hat{g}^{\text{MLE-regr}}(i/n) - \lambda^{-1} \mathbf{1}(\hat{g}^{\text{MLE-regr}}(i/n) = \mathcal{Y}_i)) w(i/n).$$

Then for $\text{Exp}(\lambda)$ -distributed errors $\hat{\vartheta}^{\text{MLE-regr}}$ is an unbiased estimator of $\vartheta^{(n)}$ with $\text{Var}(\hat{\vartheta}^{\text{MLE-regr}}) = \frac{1}{n^2 \lambda} \sum_{i=1}^n \mathbb{E}[\hat{g}^{\text{MLE-regr}}(i/n) - g(i/n)] w(i/n)^2$. This follows analogously from the corresponding counting process $\tilde{N}(t)$, replacing X_j in the PPP case by j/n . The asymptotic upper bound for the regression model as $n \rightarrow \infty$ is the same as for the PPP model, but with the noise level $1/n$ replaced by $1/(n\lambda)$, provided w^2 is Riemann-integrable.

While $\hat{\vartheta}^{\text{MLE}}$ as an UMVU estimator enjoys very desirable finite sample properties of its risk, for inference questions we are also in need of distributional properties, at least asymptotically. A

priori, in our Poisson-type boundary models it might not be clear whether the limiting distribution is Gaussian, but in fact this is the case since we average over the interval $[0, 1]$. The proof of the following central limit theorems for the Lipschitz case is slightly more technical and therefore given in the [Appendix](#). Note that a central limit theorem for the blockwise estimator $\hat{\vartheta}_n^{\text{block}}$ follows far more easily due to Lindeberg’s theorem, profiting from the independence between blocks.

Theorem 3.4. For $g \in C^1(R)$ (Lipschitz case), $\sup_{x \in [0,1]} |w(x)| < \infty$ and $\text{Var}(\hat{\vartheta}_n^{\text{MLE}}) \sim n^{-3/2}$, indicating the dependence of $\hat{\vartheta}^{\text{MLE}}$ on n , the following central limit theorems hold as $n \rightarrow \infty$:

$$\frac{n^{1/2}(\hat{\vartheta}_n^{\text{MLE}} - \vartheta)}{(\int_0^1 \mathbb{E}[\hat{g}^{\text{MLE}}(x) - g(x)]w(x)^2 dx)^{1/2}} \Rightarrow N(0, 1),$$

$$\frac{n^{1/2}(\hat{\vartheta}_n^{\text{MLE}} - \vartheta)}{(\int_0^1 (\hat{g}^{\text{MLE}}(x) - g(x))w(x)^2 dx)^{1/2}} \Rightarrow N(0, 1).$$

Furthermore, the following self-normalising version is valid:

$$\frac{n^{1/2}(\hat{\vartheta}_n^{\text{MLE}} - \vartheta)}{(\frac{1}{n} \sum_{j \geq 1} \mathbf{1}(\hat{g}^{\text{MLE}}(X_j) = Y_j)w(X_j)^2)^{1/2}} \Rightarrow N(0, 1).$$

Remark 3.5. The “super-efficient” case $\text{Var}(\hat{\vartheta}_n^{\text{MLE}}) = o(n^{-3/2})$ is to some extent degenerate and might possibly result in non-Gaussian limit laws. A lower estimate in (3.3) above shows that $\mathbb{E}[\hat{g}^{\text{MLE}}(x) - g(x)] \gtrsim n^{-3/2}$ holds as soon as the function g satisfies $|g(y) - g(x)| \leq R'|y - x|$ for $R' < R$ and y in a neighbourhood of x . This means that $\text{Var}(\hat{\vartheta}^{\text{MLE}}) \sim n^{-3/2}$ and the CLTs above are applicable whenever g has a local Lipschitz constant smaller than R , at least on some subinterval. Note that in this case we also get “for free” the nice geometric result that the number of observations on the graph of \hat{g}^{MLE} is of order $n^{1/2}$ (in mean) because of

$$\frac{1}{n} \sum_{j \geq 1} \mathbf{1}(\hat{g}^{\text{MLE}}(X_j) = Y_j) \sim \int_0^1 \mathbb{E}[\hat{g}^{\text{MLE}}(x) - g(x)] dx \sim n^{-1/2}.$$

The standard deviation is of smaller order as the proof of Theorem 3.4 shows.

An immediate consequence of the selfnormalising CLT is the following inference statement.

Corollary 3.6. Under the assumptions of Theorem 3.4

$$\mathcal{I}_n := [\hat{\vartheta}_n^{\text{MLE}} - \hat{\sigma}_n q_{1-\alpha/2}, \hat{\vartheta}_n^{\text{MLE}} + \hat{\sigma}_n q_{1-\alpha/2}], \quad \hat{\sigma}_n^2 := \frac{1}{n^2} \sum_{j \geq 1} \mathbf{1}(\hat{g}^{\text{MLE}}(X_j) = Y_j)w(X_j)^2,$$

with $q_{1-\alpha/2}$ the $(1 - \alpha/2)$ -quantile of $N(0, 1)$, is a confidence interval for ϑ with asymptotic coverage $1 - \alpha$.

Also the asymptotic variance can be determined explicitly.

Corollary 3.7. *Under the assumptions of Theorem 3.4, we obtain*

$$\text{Var}(\hat{\vartheta}_n^{\text{MLE}}) = \left(\frac{\sqrt{\pi}}{2} + o(1)\right)n^{-3/2} \int_0^1 \sqrt{(R^2 - g'(x)^2)/Rw(x)^2} dx,$$

where g' denotes the weak derivative of the Lipschitz function g , and thus

$$n^{3/4}(\hat{\vartheta}_n^{\text{MLE}} - \vartheta) \Rightarrow N\left(0, \int_0^1 \sqrt{(R^2 - g'(x)^2)/Rw(x)^2} dx\right).$$

Proof. A Lipschitz function g is absolutely continuous, hence a.e. differentiable and necessarily $|g'(x)| \leq R$ holds a.e. For $x \in (0, 1)$ where $g'(x)$ exists we obtain, arguing by dominated convergence using (3.3),

$$\begin{aligned} &P(n^{1/2}(\hat{g}_n^{\text{MLE}}(x) - g(x)) \geq z) \\ &= \exp\left(-n \int_0^1 (g(x) + zn^{-1/2} - R|\xi - x| - g(\xi))_+ d\xi\right) \\ &= \exp\left(-\int_{-n^{1/2}x}^{n^{1/2}(1-x)} (n^{1/2}(g(x) - g(x + n^{-1/2}u)) + z - R|u|)_+ du\right) \\ &\rightarrow \exp\left(-\int_{-\infty}^{\infty} (z - R|u| - g'(x)u)_+ du\right) = \exp\left(-\frac{R}{R^2 - g'(x)^2} z^2\right). \end{aligned}$$

By integrating this survival function over $z \in \mathbb{R}^+$ and applying dominated convergence due to the uniform bound (3.3), we conclude

$$n^{1/2} \mathbb{E}[\hat{g}_n^{\text{MLE}}(x) - g(x)] \rightarrow \sqrt{(R^2 - g'(x)^2)/R} \frac{\sqrt{\pi}}{2}.$$

Integration over x yields by another application of dominated convergence in view of (3.4) the asymptotic expression for $\text{Var}(\hat{\vartheta}_n^{\text{MLE}})$. □

The last corollary shows that for constant g the asymptotic variance (rescaled by $n^{3/2}$) equals $\sqrt{R\pi} \|w\|_{L^2}^2/2$ and is largest among all admissible g while for linear g with slope $\pm R$ the rescaled asymptotic variance vanishes, that is, the convergence rate is faster than $n^{-3/2}$. In Figure 2, we see indeed that \hat{g}^{MLE} is closest to g where g has largest slope. Notice that the bias correction via point counts gives a precise meaning for this observation. So far, our methods of proof do not extend to the β -Hölder case with $\beta < 1$ or to $w \in L^2$ because we need to control the difference to a block-wise partitioned MLE. The strategy of proof does neither apply to the monotone MLE, as introduced next.

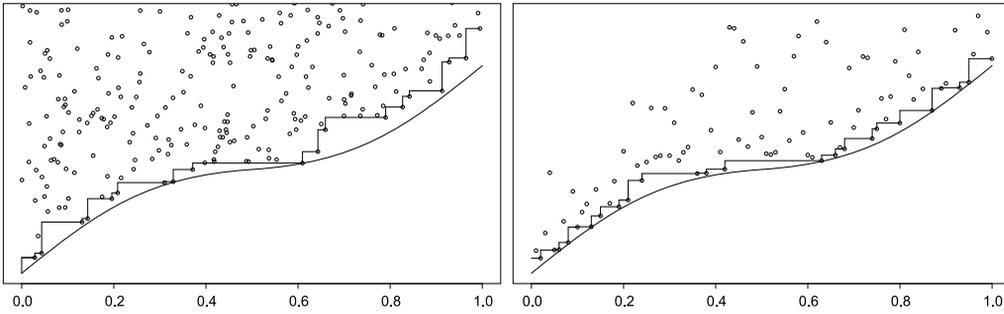


Figure 3. Construction of the estimator \hat{g}^{Mon} in the PPP model ($n = 100$, left) and in the regression-type model ($n = 100$, $\varepsilon_i \sim \text{Exp}(1)$, right).

3.2. MLE under monotonicity

Let us consider the general nonparametric class

$$\mathcal{M} := \{g : [0, 1) \rightarrow \mathbb{R} | g \text{ is increasing and left-continuous}\}$$

of monotone, that is (not necessarily strictly) increasing functions. Since monotone g have at most countably many jumps, the observations for left- and right-continuous versions of g are a.s. identical. Then the nonparametric MLE for the PPP model over this class is given by

$$\hat{g}^{\text{Mon}}(x) = \min_{i: X_i \geq x} Y_i, \quad x \in [0, 1),$$

which is obvious from the fact that any $g \in \mathcal{M}$ with $g(X_i) \leq Y_i$ for all i necessarily satisfies $g \leq \hat{g}^{\text{Mon}}$, see also Figure 3. Note that a.s. $\hat{g}^{\text{Mon}}(x) < \infty$ holds for $x \in [0, 1)$, but $\lim_{x \uparrow 1} \hat{g}^{\text{Mon}}(x) = \infty$.

Proposition 3.8. *The nonparametric MLE $(\hat{g}^{\text{Mon}}(x), x \in [0, 1))$ is a sufficient and complete statistics for \mathcal{M} .*

Proof. Sufficiency follows again from the likelihood representation

$$\mathcal{L}(g) = \exp\left(n + \sum_{j \geq 1} (-Y_j)_+\right) \exp\left(n \int_0^1 g(x) dx\right) \mathbf{1}(g \leq \hat{g}^{\text{Mon}}),$$

using $g \in L^1$ because of $g(x) \in [g(0), g(1)]$ by monotonicity, and the factorisation criterion. For completeness we equip \mathcal{M} with the ball σ -algebra for the uniform norm, cf. Examples 1.7.3, 1.7.4 in [21], which is generated by the point evaluations $f \mapsto f(x)$, $x \in [0, 1)$. In particular, this implies that \hat{g}^{Mon} is measurable because its point evaluations are measurable. For fixed $x_0 \in [0, 1)$, $y_0 \in \mathbb{R}$ we have the bracket representation

$$\{g \in \mathcal{M} | g(x_0) \geq y_0\} = \bigcup_{n \in \mathbb{N}} [y_0 - n \mathbf{1}_{[0, x_0)}(x), \infty).$$

Noting that maxima of monotone functions are again monotone, the brackets form an \cap -stable generator of the ball σ -algebra. The proof now follows exactly that of Proposition 3.1. \square

In analogy with the $C^\beta(R)$ -case, we build the estimator

$$\hat{\vartheta}^{\text{Mon}} := \int_0^1 \hat{g}^{\text{Mon}}(x)w(x) dx - \frac{1}{n} \sum_{j \geq 1} \mathbf{1}(\hat{g}^{\text{Mon}}(X_j) = Y_j)w(X_j)$$

that will enjoy similar nice properties. We have to consider, however, weight functions w whose support stays away from $x = 1$ in order to avoid problems arising from $\hat{g}^{\text{Mon}}(x) \uparrow \infty$ as $x \uparrow 1$.

Theorem 3.9. *Assume $\text{supp}(w) \subseteq [0, 1)$. Then the estimator $\hat{\vartheta}^{\text{Mon}}$ is for each finite sample size n UMVU over the class \mathcal{M} with*

$$\text{Var}(\hat{\vartheta}^{\text{Mon}}) = \frac{1}{n} \int_0^1 \mathbb{E}[\hat{g}^{\text{Mon}}(x) - g(x)]w(x)^2 dx.$$

For bounded w it satisfies

$$\text{Var}(\hat{\vartheta}^{\text{Mon}}) \leq \left(\frac{3\pi(g(1) - g(0))}{2} \right)^{1/2} \|w\|_\infty^2 n^{-3/2} + O(n^{-2}).$$

Remark 3.10.

1. For $w \in L^p$ with $p > 4$ the proof below still yields the rate $n^{-1/2}$ using Hölder’s inequality instead of a supremum norm bound. For monotonous g with bounded weak derivative g' , that is, Lipschitz-continuous g , the asymptotic constant turns out to be exactly $\sqrt{\pi/2} \int_0^1 w(x)^2 \sqrt{g'(x)} dx$ by a dominated convergence argument.
2. Concerning the support of w , the proof shows that the remainder $O(n^{-2})$ is in fact $2n^{-2}(1 - \sup\{x : w(x) \neq 0\})^{-1}$ and for varying weight functions w_n we may allow a shrinking distance ε_n of $\text{supp}(w_n)$ from 1 such that $\varepsilon_n n^{1/2} \rightarrow \infty$, implying a negligible order compared to $n^{-3/2}$.
3. The rate $n^{-3/2}$ is minimax optimal for the mean squared error over the class \mathcal{M} . This follows by adapting the proof of Theorem 2.6 for the Lipschitz case $\beta = 1$. We may just add to $g(x)$ a linear slope Ax with $A > 2cR$ such that $g \in \mathcal{M}$ holds for any realisation of (ε_k) .

Proof. The proof follows along the lines of the proof for Theorem 3.2. Here the weighted counting process is

$$\bar{N}(t) = \sum_{j \geq 1} \mathbf{1}\left(Y_j \leq t \wedge \min_{i: X_i \geq X_j} Y_i\right)w(X_j), \quad t \in \mathbb{R}.$$

Its intensity is $\bar{\lambda}_t = n \int_0^1 \int_{[g(x), t]} \mathbf{1}(\min_{i: X_i \geq x} Y_i \geq s) ds dx$ and compensation yields the corresponding martingale $\bar{M}(t)$. The same limiting arguments, by restriction to the support of w , then

yield again $\mathbb{E}[\hat{\vartheta}^{\text{Mon}}] = \vartheta$ and

$$\text{Var}(\hat{\vartheta}^{\text{Mon}}) = \frac{1}{n} \int_0^1 \mathbb{E}[\hat{g}^{\text{Mon}}(x) - g(x)] w(x)^2 dx.$$

It remains to estimate the last expectation. Suppose $w(x) = 0$ for $x > 1 - \varepsilon$ and some $\varepsilon > 0$ and let

$$g'_\infty(x) := \sup_{0 < h \leq 1-x} \frac{g(x+h) - g(x)}{h} \in [0, \infty], \quad g(x+) = \lim_{y \downarrow x} g(y), x \in [0, 1 - \varepsilon],$$

be the maximal function for the measure-valued derivative of g . Then for $x \in [0, 1 - \varepsilon]$ with $g'_\infty(x) \in (0, \infty)$

$$\begin{aligned} \sqrt{n} \mathbb{E}[\hat{g}^{\text{Mon}}(x) - g(x)] &= \int_0^\infty P(\hat{g}^{\text{Mon}}(x) - g(x) \geq sn^{-1/2}) ds \\ &= \int_0^\infty \exp\left(-n \int_x^1 (sn^{-1/2} + g(x) - g(\xi))_+ d\xi\right) ds \\ &= \int_0^\infty \exp\left(-\int_0^{n^{1/2}(1-x)} (s + n^{1/2}(g(x) - g(x + n^{-1/2}u)))_+ du\right) ds \\ &\leq \int_0^\infty \exp\left(-\int_0^{n^{1/2}(1-x) \wedge s/g'_\infty(x)} (s - g'_\infty(x)u) du\right) ds \\ &\leq \int_0^\infty \exp(-s(n^{1/2}(1-x) \wedge s/g'_\infty(x))/2) ds \\ &\leq (\pi g'_\infty(x)/2)^{1/2} + 2n^{-1/2} \varepsilon^{-1} e^{-n\varepsilon^2/(2g'_\infty(x))}, \end{aligned}$$

which trivially continues to hold with obvious extension if $g'_\infty(x) \in \{0, \infty\}$. We shall now establish a weak- L^1 -estimate for g'_∞ by adapting and improving (in the constant) classical results [20], Theorem 7.4. For $\zeta > 0$ we define

$$B_\zeta = \{x \in [0, 1 - \varepsilon] | g'_\infty(x) \geq \zeta\}$$

and we shall prove $|B_\zeta| \leq \zeta^{-1} \|g\|_{\text{BV}}$ with $|B_\zeta|$ denoting the Lebesgue measure of B_ζ , $\|g\|_{\text{BV}} = g(1) - g(0)$.

To this end we construct a family $(x_i, h_i)_{i \in J}$, J some countable set, in $B_\zeta \times \mathbb{R}^+$ with $g(x_i + h_i) - g(x_i) \geq \zeta h_i$ and $[x_i, x_i + h_i] \cap [x_j, x_j + h_j] = \emptyset$ for all $i \neq j$ such that $\bigcup_{i \in J} [x_i, x_i + h_i] \supseteq B_\zeta$ holds. We proceed by (transfinite) recursion: since $x \mapsto g(x+)$ is right-continuous, so is g'_∞ and thus $x_0 := \inf B_\zeta$ lies in B_ζ such that there is some $h_0 > 0$ with $g(x_0 + h_0) - g(x_0) \geq \zeta h_0$. Then define $x_1 := \min(B_\zeta \setminus [x_0, x_0 + h_0])$, which is again in B_ζ by right-continuity such that some $h_1 > 0$ exists with $g(x_1 + h_1) - g(x_1) \geq \zeta h_1$ and so on. Having thus defined $(x_i, h_i)_{i \in I'}$ for all (possibly infinite) ordinal numbers I' smaller than some given ordinal I we add $x_I =$

$\inf(B_\zeta \cap \bigcap_{i \in \cup I'} [x_i + h_i, 1 - \varepsilon]) \in B_\zeta$ with some corresponding h_I until B_ζ is exhausted by $\bigcup_{i \in I} [x_i, x_i + h_i)$. Then we just estimate

$$g(1) - g(0) \geq \sum_{i \in J} g(x_i + h_i) - g(x_i) \geq \zeta \sum_{i \in J} h_i \geq \zeta |B_\zeta|.$$

Using $\mathbb{E}[\hat{g}^{\text{Mon}}(x) - g(x)] \leq (\pi g'_\infty(x)/(2n))^{1/2} + 2n^{-1}\varepsilon^{-1}$ and for $a > 0$ the integral bound

$$\int_0^{1-\varepsilon} g'_\infty(x)^{1/2} dx \leq \int_0^{1/a} z^{-1/2} d|B_{z^{-1}}| + a^{1/2}(1 - \varepsilon) \leq \|g\|_{\text{BV}} \int_0^{1/a} z^{-1/2} dz + a^{1/2},$$

derived from $|B_{z^{-1}}| \leq z \|g\|_{\text{BV}}$, we obtain with $a = \|g\|_{\text{BV}}$

$$\int_0^{1-\varepsilon} \mathbb{E}[\hat{g}^{\text{Mon}}(x) - g(x)] dx \leq (3\pi \|g\|_{\text{BV}})^{1/2} (2n)^{-1/2} + O(n^{-1})$$

for $g \in \mathcal{M}$. The assertion follows by pulling $\|w\|_\infty$ out of the integral. □

In the regression-type model the nonparametric MLE over \mathcal{M} is likewise $\hat{g}^{\text{Mon-regr}}(x) = \min_{i: x \leq i/n} \mathcal{Y}_i$. Then by the same arguments

$$\hat{\vartheta}_n^{\text{Mon-regr}} := \frac{1}{n} \sum_{i=1}^n (\hat{g}^{\text{Mon-regr}}(i/n) - \lambda^{-1} \mathbf{1}(\hat{g}^{\text{Mon-regr}}(i/n) = \mathcal{Y}_i)) w(i/n)$$

is an unbiased estimator of $\frac{1}{n} \sum_{i=1}^n g(i/n) w(i/n)$ under $\text{Exp}(\lambda)$ -noise with $\text{Var}(\hat{\vartheta}_n^{\text{Mon-regr}}) = \frac{1}{n^2 \lambda} \sum_{i=1}^n \mathbb{E}[\hat{g}^{\text{Mon-regr}}(i/n) - g(i/n)] w(i/n)^2$. Note that at the right end-point $\mathbb{E}[\hat{g}^{\text{Mon-regr}}(1) - g(1)] = \lambda^{-1}$ holds, but that summand only contributes $(n\lambda)^{-2} w(1)^2$ to the total variance which is usually negligible.

4. Discussion

An important application for the estimation of functionals are orthogonal series estimators, also called projection estimators. Let $(\varphi_m)_{m \geq 1}$ be an orthonormal basis of $L^2([0, 1])$. Then we can form the estimator $\hat{g}_M = \sum_{m=1}^M \hat{\vartheta}_m \varphi_m$ of g where $\hat{\vartheta}_m$ estimates the coefficient $\langle g, \varphi_m \rangle_{L^2}$, that is, $w = \varphi_m$ in our notation. Using our estimators for $g \in \mathcal{C}^\beta(\mathcal{R})$, we thus obtain as stochastic error in the L^2 -risk:

$$\mathbb{E}[\|\hat{g}_M - \mathbb{E}[\hat{g}_M]\|_{L^2}^2] = \sum_{m=1}^M \text{Var}(\hat{\vartheta}_m) \lesssim Mn^{-(2\beta+1)/(\beta+1)}.$$

For L^2 -Sobolev spaces H^s of regularity s and standard bases like (trigonometric) polynomials, splines or wavelets we have the bias bound $\sum_{m > M} \langle g, \varphi_m \rangle^2 \lesssim M^{-2s}$. We always have

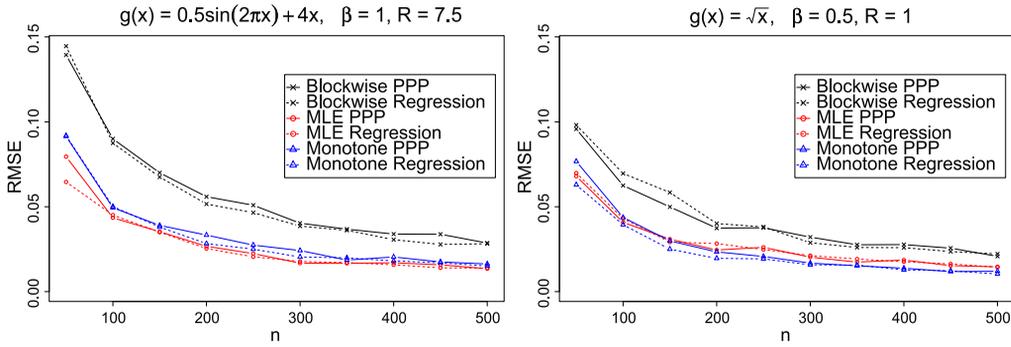


Figure 4. Monte-Carlo errors for the different estimators and two functions g .

$g \in \mathcal{C}^\beta(R) \Rightarrow g \in H^\beta$ such that

$$\mathbb{E}[\|\hat{g}_M - g\|_{L^2}^2] \lesssim M^{-2\beta} + Mn^{-(2\beta+1)/(\beta+1)} \sim n^{-2\beta/(\beta+1)} \quad \text{for } M \sim n^{1/(\beta+1)}$$

follows. This seems to be the first rate-optimal estimation result for series estimators in one-sided regression, cf. [6,9] for (optimal) rates and other approaches in the literature. We may, of course, also have $g \in H^s$ for some $s > \beta$, but then the derived rate is slower than the optimal $n^{-2s/(2s+1)}$. The unbiased estimation method essentially relies on a uniform control of the variation of g and we do not know whether similar results can be obtained for Sobolev (or Besov) instead of Hölder balls.

Concerning the function class \mathcal{G} over which the nonparametric MLE is feasible and for which the derived estimator of ϑ exhibits nice non-asymptotic properties, it was only essential for the stopping arguments as well as the completeness property that constants lie in \mathcal{G} and that for $g_1, g_2 \in \mathcal{G}$ also $g_1 \wedge g_2$ and $g_1 \vee g_2$ are in \mathcal{G} . Thus also $\mathcal{G} = \mathcal{M} \cap \mathcal{C}^\beta(R)$ or extensions to the multivariate case $\mathcal{G} = \mathcal{C}_d^\beta(R) = \{g : [0, 1]^d \rightarrow \mathbb{R} \mid |g(x) - g(y)| \leq R|x - y|^\beta\}$ are possible. For smoothness degrees $\beta > 1$ or other shape constraints like convexity our method does not transfer directly, but may possibly be adapted, see, for example, [1], where also the intensity λ is estimated.

Finally, in a small simulation example we investigate the behaviour of the blockwise estimator, the MLE and the monotone MLE for $\vartheta = \int_0^1 g(x) dx$ on finite samples. We simulate the PPP model as well as the regression-type model with $\text{Exp}(1)$ -distributed noise. For two different monotone regression functions g the RMSE (root mean squared error) is estimated in $M = 200$ Monte Carlo repetitions. On the left-hand side of Figure 4 the RMSE results for $g(x) = 0.5 \sin(2\pi x) + 4x$ are shown and on the right-hand side those for $g(x) = \sqrt{x}$. It can be seen that all three estimators work well even for the small sample size $n = 50$ and that their performances in the PPP and the regression model are comparable.

The blockwise estimator does not perform so much worse than the ML estimators. From our theoretical results this is to be expected: the ratio of the upper bounds for the nonasymptotic

variance of $\hat{\vartheta}^{\text{MLE}}$ and of $\hat{\vartheta}^{\text{block}}$ is given by

$$\frac{\Gamma(\beta/(\beta + 1))\beta(\beta + 1)^{-1/(\beta+1)}}{\beta^{-\beta/(\beta+1)}(\beta + 1)} = \Gamma(\beta/(\beta + 1))\beta^{(2\beta+1)/(\beta+1)}(\beta + 1)^{-(\beta+2)/(\beta+1)},$$

which approaches one for $\beta \downarrow 0$, has a minimum 0.54 at $\beta \approx 0.47$ and then increases to about 0.63 for $\beta \uparrow 1$. Note, however, that both upper bounds are not tight. Since the simple blockwise approach is faster to compute, which is particularly relevant for any adaptive estimator, and is theoretically easier to analyse than the MLE (especially for the CLT, but also for an adaptive procedure), we conclude that both approaches are attractive and should be considered in their own right.

Appendix

A.1. Technical results

We formulate a stopping theorem for continuous-time martingales, which does not seem readily available in the literature.

Lemma A.1. *Let $(M(t), t \geq t_0)$ be a càdlàg martingale with $M(t_0) = 0$ and let τ be a stopping time with values in $[t_0, \infty]$, both on some filtered probability space. If $\mathbb{E}[\langle M \rangle_\tau]$ is finite, then $\mathbb{E}[M(\tau)] = 0$ and $\mathbb{E}[M(\tau)^2] = \mathbb{E}[\langle M \rangle_\tau]$ hold.*

Proof. From the Burkholder–Davis–Gundy inequality (Theorem 26.12 in [10]) and the identity $\mathbb{E}[M_\tau] = \mathbb{E}[\langle M \rangle_\tau]$ (e.g., by Proposition 4.50(c) in [8] for M^τ), we conclude $\mathbb{E}[\sup_{t \geq t_0} M_{t \wedge \tau}^2] \lesssim \mathbb{E}[\langle M \rangle_\tau]$. Hence, $(|M_{t \wedge \tau}|^p)_{t \geq t_0}$, $p \in \{1, 2\}$, is uniformly integrable and by optional stopping $\mathbb{E}[M_\tau] = \lim_{t \rightarrow \infty} \mathbb{E}[M_{t \wedge \tau}] = 0$ follows as well as $\mathbb{E}[M_\tau^2] = \mathbb{E}[\langle M \rangle_\tau] = \mathbb{E}[\langle M \rangle_\tau]$. \square

A moment bound for the stopping time in the proof of Theorem 2.2 is provided.

Lemma A.2. *Under the assumptions of Theorem 2.2, we have for $\tau = \mathcal{Y}_k^* + Rh^\beta$*

$$\mathbb{E}[(\tau - g(i/n))^p]^{1/p} \lesssim Rh^\beta + (n\lambda h)^{-1}$$

as $nh \rightarrow \infty$ for any $p > 0$.

Proof. The property $\mathcal{Y}_k^* \leq \max_{i \in \tilde{I}_k} g(i/n) + \min_{i \in \tilde{I}_k} \varepsilon_i$ implies for $nh \rightarrow \infty$

$$P\left(n\lambda h \left(\mathcal{Y}_k^* - \max_{i \in \tilde{I}_k} g(i/n)\right) \geq z\right) \leq \bar{F}_\varepsilon(z/n\lambda h)^{nh} = e^{nh \log \bar{F}_\varepsilon(z/n\lambda h)} \rightarrow e^{-z}.$$

Using $\bar{F}_\varepsilon(z/nh)^{nh} \lesssim (1 + z/nh)^{-nh\rho}$, we establish

$$\lim_{R \rightarrow \infty} \sup_{n,h} \int_R^\infty z^{p-1} P\left(n\lambda h \left(\mathcal{Y}_k^* - \max_{i \in \tilde{I}_k} g(i/n)\right) \geq z\right) dz = 0$$

for any $p \geq 1$ such that by uniform integrability

$$\limsup_{nh \rightarrow \infty} \mathbb{E} \left[\left(n\lambda h \left| \mathcal{Y}_k^* - \max_{i \in \tilde{I}_k} g(i/n) \right| \right)^p \right] \leq \int_0^\infty z^p e^{-z} dz < \infty$$

follows. By the Hölder condition g varies at most by Rh^β on each block and thus $\mathbb{E}[(\tau - \min_{i \in \tilde{I}_k} g(i/n))^p] \lesssim (Rh^\beta + (n\lambda h)^{-1})^p$ holds. \square

We need the following interesting self-normalising property. The constant is certainly not optimal.

Lemma A.3. *Suppose that a non-negative random variable X satisfies $P(X \geq x) = e^{-a(x)}$, $x \geq 0$, with a strictly increasing convex function a . Then $\mathbb{E}[X^2] \leq 6(e+1)\mathbb{E}[X]^2$ holds.*

Proof. The property $P(a(X) \geq a(x)) = e^{-a(x)}$ shows that $Y := a(X)$ is $\text{Exp}(1)$ -distributed. Since the inverse a^{-1} of a exists, we may use $a^{-1}(0) = 0$ and the concavity of a^{-1} to calculate

$$\begin{aligned} \mathbb{E}[X]^2 &= \mathbb{E}[a^{-1}(Y)^2] = \int_0^\infty \int_0^\infty a^{-1}(x)a^{-1}(y)e^{-x-y} dy dx \\ &= \int_0^\infty \int_0^z a^{-1}(x)a^{-1}(z-x) dx e^{-z} dz \\ &\geq \int_0^\infty \int_0^z \frac{x}{z} a^{-1}(z) \frac{z-x}{z} a^{-1}(z) dx e^{-z} dz = \int_0^\infty \frac{z}{6} a^{-1}(z)^2 e^{-z} dz. \end{aligned}$$

By monotonicity, we have $\int_1^2 z a^{-1}(z)^2 e^{-z} dz \geq e^{-1} \int_0^1 a^{-1}(z)^2 e^{-z} dz$. This shows

$$6\mathbb{E}[X]^2 \geq \frac{1}{e+1} \int_0^\infty a^{-1}(z)^2 e^{-z} dz = \frac{1}{e+1} \mathbb{E}[X^2]. \quad \square$$

A.2. Proof of Theorem 2.4

For $h_m < h^*$ we infer from the deviation bound in Proposition A.4 below

$$\begin{aligned} P(\hat{h} = h_m) &\leq \sum_{m'=1}^{m-1} (P(|\tilde{\vartheta}_{n,h_{m'}}^{\text{block}} - \vartheta^{(n)}| > \kappa_{m'}) + P(|\tilde{\vartheta}_{n,h_{m+1}}^{\text{block}} - \vartheta^{(n)}| > \kappa_{m+1})) \\ &\leq \sum_{m'=1}^{m-1} (4n^{-2c} + n\bar{F}_\varepsilon((c \log n - 2)/(nh_{m'}))^{nh_{m'}} + n\bar{F}_\varepsilon((c \log n - 2)/(nh_{m+1}))^{nh_{m+1}}) \\ &\lesssim M(n^{-2c} + n^{1-\lambda c}) \end{aligned}$$

for $\underline{\lambda} \in (0, \lambda)$ and n sufficiently large. By Cauchy–Schwarz inequality, we thus infer

$$\mathbb{E}[(\tilde{\vartheta}_n^{\text{block}} - \vartheta^{(n)})^2 \mathbf{1}(\hat{h} < h^*)] \lesssim \mathbb{E}[(\tilde{\vartheta}_n^{\text{block}} - \vartheta^{(n)})^4]^{1/2} M(n^{-2c} + n^{1-\underline{\lambda}c})^{1/2}.$$

From the exponential moment bound (A.1) and Lemma A.2, we obtain that the fourth moment of the error remains bounded (even tends to zero) such that the first inequality follows.

By construction, we have for $h_{\hat{m}} := \hat{h} > h^* =: h_{m^*}$ that $|\tilde{\vartheta}_n^{\text{block}} - \tilde{\vartheta}_{n,h^*}^{\text{block}}| \leq \kappa_{\hat{m}} + \kappa_{m^*}$ holds. Using $H_x(y) \approx xy^2$ and $h_m \gtrsim (\log n)^2 n^{-1}$, we obtain

$$\kappa_{\hat{m}} + \kappa_{m^*} \lesssim \frac{(\log n)^2}{(nh^*)^2} + \frac{\sqrt{\log n}}{n\sqrt{h^*}} \left(1 + \max_{h_m \geq h^*} h_m \sum_{k=0}^{h_m^{-1}-1} \#\{i \in \tilde{I}_{k,h_m} : \mathcal{Y}_i \leq \mathcal{Y}_{k,h}^* + (nh_m)^{-1}\} \right).$$

For each fixed h_m we have by compensation of the block-wise counting process

$$\mathbb{E} \left[\left(\sum_{k=0}^{h_m^{-1}-1} \#\{i \in \tilde{I}_{k,h_m} : \mathcal{Y}_{k,h_m}^* < \mathcal{Y}_i \leq \mathcal{Y}_{k,h_m}^* + (nh_m)^{-1}\} \right)^2 \right] \leq h_m^{-1} \sum_{k=0}^{h_m^{-1}-1} \mathbb{E}[A_k^2 + A_k]$$

with

$$A_k = \sum_{i \in \tilde{I}_{k,h_m}} \int \mathbf{1}(\mathcal{Y}_{k,h_m}^* < s + g(i/n) \leq \mathcal{Y}_{k,h_m}^* + (nh_m)^{-1}) \frac{f_\varepsilon(s)}{\bar{F}_\varepsilon(s)} ds \leq \|f_\varepsilon / \bar{F}_\varepsilon\|_\infty \sim 1.$$

Since by definition $\#\{i \in \tilde{I}_{k,h_m} : \mathcal{Y}_i \leq \mathcal{Y}_{k,h_m}^*\} = 1$ a.s., we have

$$\mathbb{E} \left[\left(h_m \sum_{k=0}^{h_m^{-1}-1} \#\{i \in \tilde{I}_{k,h_m} : \mathcal{Y}_i \leq \mathcal{Y}_{k,h_m}^* + (nh_m)^{-1}\} \right)^2 \right] \lesssim 1.$$

A (crude) bound for the maximum via the sum thus yields the second inequality:

$$\mathbb{E}[(\tilde{\vartheta}_n^{\text{block}} - \tilde{\vartheta}_{n,h^*}^{\text{block}})^2 \mathbf{1}(\hat{h} \geq h^*)] \leq \mathbb{E}[(\kappa_{\hat{m}} + \kappa_{m^*})^2 \mathbf{1}(\hat{h} > h^*)] \lesssim \frac{(\log n)^4}{(nh^*)^4} + \frac{M \log n}{n^2 h^*}.$$

For the asymptotic rate just note that the geometric grid of bandwidths suffices to achieve $h^* \sim n^{-1/(\beta+1)}$ asymptotically such that inserting $\mathbb{E}[(\tilde{\vartheta}_{n,h^*}^{\text{block}} - \vartheta^{(n)})^2] \lesssim n^{-(2\beta+1)/(\beta+1)} + n^{-4\beta/(\beta+1)}$ from Theorem 2.2 and the triangle inequality yield the result, noting that the risk on $\{\hat{h} < h^*\}$ is negligible due to the choice of c . It remains to prove the following deviation inequality.

Proposition A.4. *For any $h, x, \kappa > 0$ with $Rh^\beta \leq (nh)^{-1}$, $2xh^{1/2}\|w\|_\infty < 1$ and $\kappa < (\delta - 2Rh^\beta)nh$ we have with probability at least $1 - 2e^{-2x^2} - h^{-1}\bar{F}_\varepsilon(\kappa/(nh))^{nh}$ the bound*

$$n\lambda h^{1/2} |\tilde{\vartheta}_{n,h}^{\text{block}} - \vartheta^{(n)}| \leq \sum_{i=1}^n \mathbf{1}(\mathcal{Y}_i \leq \tau^{(i)}) H_x(h^{1/2}w(i/n)) + C^2(\kappa + 2)^2 n^{-1} h^{-3/2} \|w\|_1 + x.$$

Proof. We consider the martingale $M(t)$ in (2.4) and the associated stopping rule τ . By the substitution rule [10], Theorem 26.7, we obtain for $\gamma > -1/\|w\|_\infty$ the exponential (local) martingale

$$\begin{aligned} \mathcal{E}(t) &= \exp\left(\sum_{i \in \tilde{I}_k} \mathbf{1}(\mathcal{Y}_i \leq t) \log(1 + \gamma w(i/n)) + \log(\bar{F}_\varepsilon(\mathcal{Y}_i \wedge t - g(i/n))) \gamma w(i/n)\right) \\ &= \exp\left(\gamma M(t) - \sum_{i \in \tilde{I}_k} \mathbf{1}(\mathcal{Y}_i \leq t) (\gamma w(i/n) - \log(1 + \gamma w(i/n)))\right). \end{aligned}$$

We infer from $|\tilde{I}_k| < \infty$ and the fact that $\bar{F}_\varepsilon(\varepsilon_i) \sim U([0, 1])$ has finite p -moments for all $p > -1$ via Lemma A.1 the stopping result

$$\mathbb{E}\left[\exp\left(\gamma M(\tau) - \sum_{i \in \tilde{I}_k} \mathbf{1}(\mathcal{Y}_i \leq \tau) (\gamma w(i/n) - \log(1 + \gamma w(i/n)))\right)\right] = 1. \quad (\text{A.1})$$

Using representation (2.5), the independence among blocks yields $\mathbb{E}[e^{\gamma Z_\gamma}] = 1$ for

$$\begin{aligned} Z_\gamma &:= n\lambda(\tilde{\vartheta}_{n,h}^{\text{block}} - \vartheta^{(n)}) - \sum_{i=1}^n G_\varepsilon(\mathcal{Y}_i \wedge \tau^{(i)} - g(i/n))w(i/n) \\ &\quad - \sum_{i=1}^n \mathbf{1}(\mathcal{Y}_i \leq \tau^{(i)}) \left(w(i/n) - \frac{\log(1 + \gamma w(i/n))}{\gamma}\right). \end{aligned}$$

We choose $\gamma = \pm 2xh^{1/2}$ and obtain by Markov inequality $P(h^{1/2}Z_{2xh^{1/2}} \geq x) \leq e^{-2x^2}$, $P(h^{1/2}Z_{-2xh^{1/2}} \leq -x) \leq e^{-2x^2}$ such that with probability $1 - 2e^{-2x^2}$

$$\begin{aligned} &\left|n\lambda h^{1/2}(\tilde{\vartheta}_{n,h}^{\text{block}} - \vartheta^{(n)}) - h^{1/2} \sum_{i=1}^n G_\varepsilon(\mathcal{Y}_i \wedge \tau^{(i)} - g(i/n))w(i/n)\right| \\ &\leq \sum_{i=1}^n \mathbf{1}(\mathcal{Y}_i \leq \tau^{(i)}) H_x(h^{1/2}w(i/n)) + x. \end{aligned} \quad (\text{A.2})$$

From $|G_\varepsilon(z)| \leq C^2 z^2$ for $z \in [0, \delta]$ and $Rh^\beta \leq (nh)^{-1}$ we infer

$$\begin{aligned} &P\left(\max_{i \in \tilde{I}_k} |G_\varepsilon(\mathcal{Y}_i \wedge \tau - g(i/n))| > C^2(\kappa + 2)^2/(nh)^2\right) \\ &\leq P\left(\max_{i \in \tilde{I}_k} (\mathcal{Y}_i \wedge \tau - g(i/n)) > (\kappa + 2)/(nh)\right) \\ &\leq P\left(\min_{i \in \tilde{I}_k} \varepsilon_i + 2(nh)^{-1} > (\kappa + 2)/(nh)\right) = \bar{F}_\varepsilon(\kappa/(nh))^{nh}. \end{aligned}$$

We thus obtain with probability $1 - \bar{F}_\varepsilon(\kappa/(nh))^{nh}$ the bound

$$\left| \sum_{i \in \tilde{I}_k} G_\varepsilon(\mathcal{Y}_i \wedge \tau - g(i/n))w(i/n) \right| \leq \frac{C^2(\kappa + 2)^2}{(nh)^2} \sum_{i \in \tilde{I}_k} |w(i/n)|.$$

Summing over the h^{-1} blocks implies with probability $1 - h^{-1} \bar{F}_\varepsilon(\kappa/(nh))^{nh}$

$$h^{1/2} \left| \sum_{i=1}^n G_\varepsilon(\mathcal{Y}_i \wedge \tau^{(i)} - g(i/n))w(i/n) \right| \leq C^2(\kappa + 2)^2 n^{-1} h^{-3/2} \|w\|_1.$$

In view of (A.2) this yields the result. □

A.3. Proof of Theorem 3.4

Let $r_n \rightarrow 0$ such that $r_n^3 n \rightarrow \infty$ and $r_n^{-1} \in \mathbb{N}$. On each block $J_l = [lr_n, (l + 1)r_n)$, $l = 0, \dots, r_n^{-1} - 1$, we can define the blockwise $\mathcal{C}^1(R)$ -MLE

$$\hat{g}_l^{\text{MLE}}(x) = \min_{i: X_i \in J_l} (Y_i + R|x - X_i|), \quad x \in J_l.$$

Note that by definition the blockwise MLE is at least as large as the global MLE, that is, $\hat{g}_l^{\text{MLE}} \geq \hat{g}^{\text{MLE}}$. By construction, $(\hat{g}_l^{\text{MLE}})_l$ are independent and each

$$\hat{\vartheta}_l^{\text{MLE}} := \int_{J_l} \hat{g}_l^{\text{MLE}}(x)w(x) dx - \frac{1}{n} \sum_{j \geq 1} \mathbf{1}(X_j \in J_l, \hat{g}_l^{\text{MLE}}(X_j) = Y_j)w(X_j)$$

enjoys the non-asymptotic properties of Theorem 3.2 on J_l , in particular $\mathbb{E}[\hat{\vartheta}_l^{\text{MLE}}] = \int_{J_l} g(x) \times w(x) dx$ and $\text{Var}(\hat{\vartheta}_l^{\text{MLE}}) = \frac{1}{n} \int_{J_l} \mathbb{E}[\hat{g}_l^{\text{MLE}}(x) - g(x)]w(x)^2 dx$. Let us therefore first establish for the blockwise MLE $\tilde{\vartheta}_n := \sum_{l=0}^{r_n^{-1}-1} \hat{\vartheta}_l^{\text{MLE}}$ that

$$\left(\frac{1}{n} \sum_{l=0}^{r_n^{-1}-1} \int_{J_l} \mathbb{E}[\hat{g}_l^{\text{MLE}}(x) - g(x)]w(x)^2 dx \right)^{-1/2} (\tilde{\vartheta}_n - \vartheta) \Rightarrow N(0, 1). \tag{A.3}$$

By independence of $(\hat{\vartheta}_l^{\text{MLE}})$, for the CLT to hold it suffices to check the 4th moment Lyapunov condition

$$\frac{\sum_{l=0}^{r_n^{-1}-1} \mathbb{E}[(\hat{\vartheta}_l - \vartheta_l)^4]}{\text{Var}(\tilde{\vartheta}_n)^2} \rightarrow 0.$$

For each $l = 0, \dots, r_n - 1$ let $(\bar{M}_{l,t})_t$ be the compensated weighted counting process from the proof of Theorem 3.2, restricted to J_l . The (non-predictable) quadratic variation of $(\bar{M}_{l,t})$ is given

by the sum of squared jumps:

$$[\bar{M}_t]_t = \sum_{s \leq t} (\Delta \bar{M}_{t,s})^2 = \sum_{j \geq 1} \mathbf{1}(X_j \in J_l, Y_j \leq t \wedge \min_{i: X_i \in J_l} (Y_i + R|X_j - X_i|)) w(X_j)^2.$$

The Burkholder–Davis–Gundy inequality (e.g., Theorem 26.12 in [10]) then yields by similar arguments as for (\bar{M}_t) above

$$\mathbb{E}[\bar{M}_{t,\infty}^4] \lesssim \mathbb{E}[\bar{M}_t^2] = \mathbb{E}\left[\left(n \int_{J_l} (\hat{g}_l^{\text{MLE}} - g) w^2\right)^2 + n \int_{J_l} (\hat{g}_l^{\text{MLE}} - g) w^4\right].$$

Using Jensen’s inequality, we find

$$\begin{aligned} \sum_{l=0}^{r_n^{-1}-1} \mathbb{E}[(\hat{\vartheta}_l^{\text{MLE}} - \vartheta_l)^4] &= \frac{1}{n^4} \sum_{l=0}^{r_n^{-1}-1} \mathbb{E}[\bar{M}_{l,\infty}^4] \\ &\lesssim \sum_{l=0}^{r_n^{-1}-1} \mathbb{E}\left[n^{-2} \left(\int_{J_l} (\hat{g}_l^{\text{MLE}} - g) w^2\right)^2 + n^{-3} \int_{J_l} (\hat{g}_l^{\text{MLE}} - g) w^4\right] \\ &\leq \sum_{l=0}^{r_n^{-1}-1} \left(n^{-2} r_n \int_{J_l} \mathbb{E}[(\hat{g}_l^{\text{MLE}} - g)^2] w^4 + n^{-3} \int_{J_l} \mathbb{E}[\hat{g}_l^{\text{MLE}} - g] w^4\right). \end{aligned}$$

As in (3.3), we can bound

$$P(\hat{g}_l^{\text{MLE}}(x) - g(x) \geq s) \leq \begin{cases} \exp(-nR(s/2R)^2), & s \in [0, 2Rr_n], \\ \exp(-n(sr_n - Rr_n^2)), & s > 2Rr_n. \end{cases} \tag{A.4}$$

Noting $r_n n^{1/2} \rightarrow \infty$ and $\|w\|_\infty < \infty$, we apply the moment bound of Lemma A.3 to $\hat{g}_l^{\text{MLE}}(x) - g(x)$ with $a(s) = n \int_{J_l} (s - R|\xi - x| + g(x) - g(\xi))_+ dx$ and integrate over s to obtain

$$\sum_{l=0}^{r_n^{-1}-1} \mathbb{E}[(\hat{\vartheta}_l^{\text{MLE}} - \vartheta_l)^4] \lesssim (r_n + n^{-1/2})(n^{-3/2})^2.$$

Hence, in view of $\text{Var}(\tilde{\vartheta}_n) \geq \text{Var}(\hat{\vartheta}_n^{\text{MLE}}) \sim n^{-3/2}$ the Lyapunov condition is satisfied and the CLT (A.3) follows.

In the second step we show that the difference between $\tilde{\vartheta}_n$ and $\hat{\vartheta}_n^{\text{MLE}}$ is of small stochastic order $o_P(n^{-3/4})$. First, we note that the above martingale arguments yield

$$\mathbb{E}[(\tilde{\vartheta}_n - \hat{\vartheta}_n^{\text{MLE}})^2] = \text{Var}(\tilde{\vartheta}_n - \hat{\vartheta}_n^{\text{MLE}}) = n^{-1} \sum_{l=0}^{r_n^{-1}-1} \int_{J_l} \mathbb{E}[\hat{g}_l^{\text{MLE}}(x) - \hat{g}_l^{\text{MLE}}(x)] w(x)^2 dx.$$

Introduce the notation $\hat{g}_{-l}^{\text{MLE}}(x) = \min_{i: X_i \notin J_l} (Y_i + R|x - X_i|^\beta)$ and consider the event

$$\Omega_n = \{\forall l = 0, \dots, r_n^{-1} - 1 \exists x \in J_l : \hat{g}_l^{\text{MLE}}(x) = \hat{g}_{-l}^{\text{MLE}}(x)\}$$

whose complement is given by $\Omega_n^c = \bigcup_l \{\min_{x \in J_l} (\hat{g}_l^{\text{MLE}} - \hat{g}_{-l}^{\text{MLE}})(x) > 0\}$. By independence of \hat{g}_l^{MLE} and $\hat{g}_{-l}^{\text{MLE}}$ and conditioning on the latter we obtain

$$\begin{aligned} &P\left(\min_{x \in J_l} (\hat{g}_l^{\text{MLE}} - \hat{g}_{-l}^{\text{MLE}})(x) > 0\right) \\ &= \mathbb{E}\left[\exp\left(-n \int_{J_l} (\hat{g}_{-l}^{\text{MLE}} - g)(x) dx\right)\right] \\ &\leq \mathbb{E}\left[\exp(-nr_n \min((\hat{g}_{-l}^{\text{MLE}} - g)(lr_n), (\hat{g}_{-l}^{\text{MLE}} - g)((l+1)r_n))\right)]. \end{aligned}$$

Using $\hat{g}_{l'}^{\text{MLE}} \geq \hat{g}_{-l}^{\text{MLE}}$ for $l' \neq l$, the bound (A.4) yields

$$\begin{aligned} P\left(\min_{x \in J_l} (\hat{g}_l^{\text{MLE}} - \hat{g}_{-l}^{\text{MLE}})(x) > 0\right) &\leq 2 \max_{l', x} \mathbb{E}\left[\exp(-nr_n (\hat{g}_{l'}^{\text{MLE}}(x) - g(x)))\right] \\ &\lesssim n^{-1} r_n^{-2}. \end{aligned}$$

We conclude $P(\Omega_n^c) = O(n^{-1} r_n^{-3}) \rightarrow 0$ by a union bound and the choice of r_n .

On the event Ω_n the left-most point L_l in J_l where \hat{g}_l^{MLE} and $\hat{g}_{-l}^{\text{MLE}}$ coincide is well defined and satisfies for $l \geq 1$

$$\begin{aligned} L_l &:= \inf\{x \in J_l \mid \hat{g}_l^{\text{MLE}}(x) = \hat{g}_{-l}^{\text{MLE}}(x)\} \\ &= \inf\{x \in J_l \mid \hat{g}_l^{\text{MLE}}(x) \leq \hat{g}_{l-1}^{\text{MLE}}(lr_n) + R(x - lr_n)\}. \end{aligned}$$

Now $L_l = lr_n$ holds on Ω_n if the corresponding right-most point

$$R_{l-1} := \sup\{x \in J_{l-1} \mid \hat{g}_{l-1}^{\text{MLE}}(x) = \hat{g}_{-l}^{\text{MLE}}(x)\}$$

on J_{l-1} satisfies $R_{l-1} < lr_n$ and vice versa $L_l > lr_n \Rightarrow R_{l-1} = lr_n$. Due to this symmetry we only consider the case $L_l > lr_n$. For $z \in (0, r_n]$ and $l = 1, \dots, r_n^{-1} - 1$ a rough bound yields:

$$\begin{aligned} P(L_l \geq lr_n + z \mid ((X_i, Y_i)\mathbf{1}(X_i < lr_n))_{i \geq 1}) &\leq \exp\left(-n \int_{lr_n}^{lr_n+z} (\hat{g}_{l-1}^{\text{MLE}}(x) - g(x)) dx\right) \\ &\leq \exp(-nz (\hat{g}_{l-1}^{\text{MLE}}(lr_n) - g(lr_n))). \end{aligned}$$

Since $\hat{g}_l^{\text{MLE}}(L_l) = \hat{g}_{l-1}^{\text{MLE}}(L_l)$ holds and both functions are in $\mathcal{C}^1(R)$, we obtain the bound

$$\int_{lr_n}^{L_l} (\hat{g}_l^{\text{MLE}} - \hat{g}_{l-1}^{\text{MLE}})(x) dx \leq \int_{lr_n}^{L_l} 2R(L_l - x) dx \lesssim (L_l - lr_n)^2.$$

By the identity $\mathbb{E}[Z^2] = \int_0^\infty 2zP(Z \geq z) dz$ for non-negative random variables Z and by the above probability bound for L_l we obtain further

$$\begin{aligned} \mathbb{E}\left[\int_{lr_n}^{L_l} (\hat{g}_l^{\text{MLE}} - \hat{g}^{\text{MLE}})\mathbf{1}_{\Omega_n}\right] &\lesssim \mathbb{E}\left[\int_0^{r_n} ze^{-nz(\hat{g}_{l-1}^{\text{MLE}}(lr_n) - g(lr_n))} dz\right] \\ &\lesssim \mathbb{E}\left[\min\left(\frac{1}{n(\hat{g}_{l-1}^{\text{MLE}}(lr_n) - g(lr_n))}, r_n\right)^2\right]. \end{aligned}$$

Using (A.4) on J_{l-1} , we arrive, after suitable substitution inside the integral, at

$$\begin{aligned} \mathbb{E}\left[\int_{lr_n}^{L_l} (\hat{g}_l^{\text{MLE}} - \hat{g}^{\text{MLE}})\mathbf{1}_{\Omega_n}\right] &\lesssim \int_0^\infty \min(u^{-1}n^{-1}, r_n^2)e^{-u} du \\ &\lesssim n^{-1} \log(nr_n^2). \end{aligned}$$

Summing over l , bounding the alternative case $R_{l-1} < lr_n$ by the same estimate and using $\|w\|_\infty < \infty$, we arrive at

$$\mathbb{E}\left[\sum_{l=0}^{r_n^{-1}-1} \int_{J_l} (\hat{g}_l^{\text{MLE}} - \hat{g}^{\text{MLE}})(x)w(x)^2 dx \mathbf{1}_{\Omega_n}\right] \lesssim (nr_n)^{-1} \log(nr_n^2) = o(n^{-1/2}). \tag{A.5}$$

This gives the desired result $\mathbb{E}[(\check{\vartheta}_n - \hat{\vartheta}_n^{\text{MLE}})^2 \mathbf{1}_{\Omega_n}] = o(n^{-3/2})$ with $P(\Omega_n) \rightarrow 1$.

Furthermore, from (A.5) we derive also that

$$\begin{aligned} &\sum_{l=0}^{r_n^{-1}} \int_{J_l} \mathbb{E}[\hat{g}_l^{\text{MLE}}(x) - g(x)]w(x)^2 dx - \int_0^1 \mathbb{E}[\hat{g}^{\text{MLE}}(x) - g(x)]w(x)^2 dx \\ &= o(n^{-1/2}) + O\left(\sup_{l,x \in J_l} \mathbb{E}[(\hat{g}_l^{\text{MLE}}(x) - \hat{g}^{\text{MLE}}(x))\mathbf{1}_{\Omega_n^c}]\right). \end{aligned}$$

By the Cauchy–Schwarz inequality, the last term is at most of order $O(n^{-1/2}P(\Omega_n^c)^{1/2}) = o(n^{-1/2})$. Hence, applying Slutsky’s lemma twice to the CLT (A.3), we arrive at

$$\left(\frac{1}{n} \int_0^1 \mathbb{E}[\hat{g}^{\text{MLE}}(x) - g(x)]w(x)^2 dx\right)^{-1/2} (\hat{\vartheta}_n^{\text{MLE}} - \vartheta) \Rightarrow N(0, 1).$$

By Jensen’s inequality, $\text{Var}(\int_{J_l} (\hat{g}_l^{\text{MLE}} - g)w^2) \leq r_n \int_{J_l} \mathbb{E}[(\hat{g}_l^{\text{MLE}} - g)^2]w^4 \lesssim r_n^2 n^{-2/3}$, which implies

$$\text{Var}\left(\sum_{l=0}^{r_n^{-1}} \int_{J_l} (\hat{g}_l^{\text{MLE}}(x) - g(x))w(x)^2 dx\right) \lesssim r_n n^{-2/3} = o(n^{-2/3}).$$

We infer $\sum_{l=0}^{r_n-1} \int_{J_l} (\hat{g}_l^{\text{MLE}} - g)w^2 \xrightarrow{P} \sum_{l=0}^{r_n-1} \int_{J_l} \mathbb{E}[\hat{g}_l^{\text{MLE}} - g]w^2$. Together with (A.5), this yields the CLT

$$\left(\int_0^1 (\hat{g}^{\text{MLE}}(x) - g(x))w(x)^2 dx \right)^{-1/2} n^{-1/2} (\hat{\vartheta}_n^{\text{MLE}} - \vartheta) \Rightarrow N(0, 1).$$

Now note that the MLE for the functional $\int g w^2$

$$\hat{\vartheta}_n^{\text{MLE}}(w^2) := \int_0^1 \hat{g}^{\text{MLE}}(x)w(x)^2 dx - \frac{1}{n} \sum_{j \geq 1} \mathbf{1}(\hat{g}^{\text{MLE}}(X_j) = Y_j)w(X_j)^2$$

is also unbiased with $\text{Var}(\hat{\vartheta}_n^{\text{MLE}}(w^2))^{1/2} \lesssim n^{-3/4}$. Since by assumption $\int_0^1 \mathbb{E}[\hat{g}^{\text{MLE}} - g]w^2 \sim n^{-1/2}$ is of larger order, Slutsky's lemma permits to replace $\int_0^1 \mathbb{E}[\hat{g}^{\text{MLE}} - g]w^2$ by $\int_0^1 \hat{g}^{\text{MLE}}w^2 - \hat{\vartheta}_n^{\text{MLE}}(w^2)$ in the CLT, which gives the desired self-normalising form.

Acknowledgements

We are grateful for very helpful comments and questions by the referees. Financial support by the DFG through Research Unit FOR 1735 *Structural Inference in Statistics: Adaptation and Efficiency* is acknowledged.

References

- [1] Baldin, N. and Reiß, M. (2015). Unbiased estimation of the volume of a convex body. Preprint. Available at [arXiv:1502.05510](https://arxiv.org/abs/1502.05510).
- [2] Baraud, Y. and Birgé, L. (2014). A new method for estimation and model selection: ρ -estimation. Preprint. Available at [arXiv:1403.6057](https://arxiv.org/abs/1403.6057).
- [3] Bibinger, M., Jirak, M. and Reiß, M. (2015). Volatility estimation under one-sided errors with applications to order books. *Ann. Appl. Probab.* To appear. Preprint. Available at [arXiv:1408.3768](https://arxiv.org/abs/1408.3768).
- [4] Daley, D.J. and Vere-Jones, D. (2008). *An Introduction to the Theory of Point Processes. Vol. II: General Theory and Structure*, 2nd ed. New York: Springer. [MR2371524](https://doi.org/10.1007/978-1-4939-9826-3)
- [5] Gayraud, G. (1997). Estimation of functionals of density support. *Math. Methods Statist.* **6** 26–46. [MR1456645](https://doi.org/10.1007/BF01456645)
- [6] Girard, S. and Jacob, P. (2003). Projection estimates of point processes boundaries. *J. Statist. Plann. Inference* **116** 1–15. [MR1997128](https://doi.org/10.1007/s11222-003-1122-2)
- [7] Groeneboom, P. and Wellner, J.A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation. DMV Seminar 19*. Basel: Birkhäuser. [MR1180321](https://doi.org/10.1007/BF01180321)
- [8] Jacod, J. and Shiryaev, A.N. (1987). *Limit Theorems for Stochastic Processes. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences] 288*. Berlin: Springer. [MR0959133](https://doi.org/10.1007/BF00959133)
- [9] Jirak, M., Meister, A. and Reiß, M. (2014). Adaptive function estimation in nonparametric regression with one-sided errors. *Ann. Statist.* **42** 1970–2002. [MR3262474](https://doi.org/10.1214/13-AOS1247)
- [10] Kallenberg, O. (2002). *Foundations of Modern Probability*, 2nd ed. New York: Springer. [MR1876169](https://doi.org/10.1007/978-1-4939-9826-3)

- [11] Karr, A.F. (1991). *Point Processes and Their Statistical Inference*, 2nd ed. *Probability: Pure and Applied* **7**. New York: Dekker. [MR1113698](#)
- [12] Korostel'ev, A.P., Simar, L. and Tsybakov, A.B. (1995). Efficient estimation of monotone boundaries. *Ann. Statist.* **23** 476–489. [MR1332577](#)
- [13] Korostel'ev, A.P. and Tsybakov, A.B. (1993). *Minimax Theory of Image Reconstruction. Lecture Notes in Statistics* **82**. New York: Springer. [MR1226450](#)
- [14] Kutoyants, Yu.A. (1998). *Statistical Inference for Spatial Poisson Processes. Lecture Notes in Statistics* **134**. New York: Springer. [MR1644620](#)
- [15] Lehmann, E.L. and Romano, J.P. (2006). *Testing Statistical Hypotheses*. New York: Springer. [MR2135927](#)
- [16] Lepskiĭ, O.V. (1990). A problem of adaptive estimation in Gaussian white noise. *Teor. Veroyatn. Primen.* **35** 459–470. [MR1091202](#)
- [17] Mammen, E. and Tsybakov, A.B. (1995). Asymptotical minimax recovery of sets with smooth boundaries. *Ann. Statist.* **23** 502–524. [MR1332579](#)
- [18] Meister, A. and Reiß, M. (2013). Asymptotic equivalence for nonparametric regression with non-regular errors. *Probab. Theory Related Fields* **155** 201–229. [MR3010397](#)
- [19] Rényi, A. and Sulanke, R. (1964). Über die konvexe Hülle von n zufällig gewählten Punkten. II. *Probab. Theory Related Fields* **3** 138–147. [MR0169139](#)
- [20] Rudin, W. (1987). *Real and Complex Analysis*, 3rd ed. New York: McGraw-Hill. [MR0924157](#)
- [21] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes. Springer Series in Statistics*. New York: Springer. [MR1385671](#)
- [22] Wojtaszczyk, P. (1997). *A Mathematical Introduction to Wavelets. London Mathematical Society Student Texts* **37**. Cambridge: Cambridge Univ. Press. [MR1436437](#)
- [23] Zhirar, S., Yuditskiĭ, A. and Nazin, A.V. (2005). L_1 -optimal nonparametric estimation of the boundary of the support by means of linear programming. *Autom. Remote Control* **66** 2000–2018. [MR2196469](#)

Received August 2014 and revised June 2015