

Functional estimation and hypothesis testing in nonparametric boundary models

MARKUS REISS* and MARTIN WAHL**

Institut für Mathematik, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany.
E-mail: *mreiss@math.hu-berlin.de; **martin.wahl@math.hu-berlin.de

Consider a Poisson point process with unknown support boundary curve g , which forms a prototype of an irregular statistical model. We address the problem of estimating non-linear functionals of the form $\int \Phi(g(x)) dx$. Following a nonparametric maximum-likelihood approach, we construct an estimator which is UMVU over Hölder balls and achieves the (local) minimax rate of convergence. These results hold under weak assumptions on Φ which are satisfied for $\Phi(u) = |u|^p$, $p \geq 1$. As an application, we consider the problem of estimating the L^p -norm and derive the minimax separation rates in the corresponding nonparametric hypothesis testing problem. Structural differences to results for regular nonparametric models are discussed.

Keywords: minimax hypothesis testing; non-linear functionals; Poisson point process; support estimation

1. Introduction

Point processes serve as canonical models for dealing with support estimation. Poisson point processes (PPP) appear in the continuous limit of nonparametric regression models with one-sided or irregular error variables, cf. Meister and Reiß [13], and thus form counterparts of the Gaussian white noise (GWN) model. In this paper, we consider the observation of a PPP on $[0, 1] \times \mathbb{R}$ with intensity function

$$\lambda_g(x, y) = n \mathbb{1}(y \geq g(x)), \quad x \in [0, 1], y \in \mathbb{R}, \quad (1.1)$$

where g is an unknown support boundary curve and $n \in \mathbb{N}$. A prototypical regression model corresponding to this PPP model with n replaced by αn is given by

$$Y_i = g(i/n) + \varepsilon_i, \quad i = 1, \dots, n, \quad (1.2)$$

with one-sided i.i.d. error variables $\varepsilon_i \geq 0$, satisfying $\mathbb{P}(\varepsilon_i \leq x) = \alpha x + O(x^2)$ as $x \downarrow 0$, cf. the discussion in Reiß and Selk [15]. The important point to keep in mind is that in these support boundary models, the standard parametric rate is n^{-1} due to the behaviour of extreme value statistics.

In Korostelev and Tsybakov [8], Chapter 8, the problem of estimating functionals of a binary image boundary from noisy observations has been studied. Although the noise is regular, the Hellinger metric is an L^1 -distance exactly as in our PPP model. In both models, the minimax rate of convergence for estimating linear functionals of the form $\langle g, \psi \rangle = \int_0^1 g(x) \psi(x) dx$, $\psi \in$

Table 1. Minimax rates for regularity β in the Poisson point process (PPP) and Gaussian white noise (GWN) model

Rate	PPP	GWN
estimate $g(x)$	$n^{-\beta/(\beta+1)}$	$n^{-\beta/(2\beta+1)}$
estimate $\langle g, \psi \rangle$	$n^{-(\beta+1/2)/(\beta+1)}$	$n^{-1/2}$
estimate $\ g\ _p^p$	$n^{-(\beta+1/2)/(\beta+1)}$	$p = 2: n^{-4\beta/(4\beta+1)} \vee n^{-1/2}$
estimate $\ g\ _p$	$n^{-(\beta+1/(2p))/(\beta+1)}$	$p \text{ even: } n^{-\beta/(2\beta+1-1/p)}$
testing	$n^{-(\beta+1/(2p))/(\beta+1)}$	$n^{-\beta/(2\beta+1/2+(1/2-1/p)_+)}$

$L^2([0, 1])$, is $n^{-(\beta+1/2)/(\beta+1)}$ over the Hölder ball

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

with $\beta \in (0, 1]$ and radius $R > 0$. For the PPP model Reiß and Selk [15] build up a nonparametric maximum-likelihood approach and construct an unbiased estimator achieving this rate. Besides minimax optimality, their estimator has the striking property of being UMVU (uniformly of minimum variance among all unbiased estimators) over $\mathcal{C}^\beta(R)$.

Here, we consider the problem of estimating and testing non-linear functionals of the form

$$F_\Phi(g) = \int_0^1 \Phi(g(x)) dx, \tag{1.3}$$

where $\Phi : \mathbb{R} \rightarrow \mathbb{R}$ is a known weakly differentiable function with derivative $\Phi' \in L^1_{\text{loc}}(\mathbb{R})$ (i.e. $\Phi(u) = \Phi(0) + \int_0^u \Phi'(v) dv$, $u \in \mathbb{R}$, holds). An important class of functionals of the form (1.3) is given by p -th powers $\|g\|_p^p$ of L^p -norms using $\Phi(u) = |u|^p$, $p \geq 1$.

We show that it is still possible to construct an unbiased estimator of $F_\Phi(g)$ which is UMVU over $\mathcal{C}^\beta(R)$. Moreover, under weak assumptions on Φ' , we compute the minimax risk of estimation over small neighbourhoods of g and show that the estimator achieves the local minimax rate of convergence $\|\Phi' \circ g\|_2 n^{-(\beta+1/2)/(\beta+1)}$. For the special case of estimating $\|g\|_p^p$ and the L^p -norm $\|g\|_p$, we prove that the minimax rates of convergence are $n^{-(\beta+1/2)/(\beta+1)}$ and $n^{-(\beta+1/(2p))/(\beta+1)}$, respectively.

Based on these results we consider the testing problem $H_0 : g = g_0$ versus $H_1 : g \in \{g_0 + h \in \mathcal{C}^\beta(R) : \|h\|_p \geq r_n\}$, where the nonparametric alternative is separated by a ball of radius $r_n > 0$ in L^p -norm. We show that the minimax separation rate is $n^{-(\beta+1/(2p))/(\beta+1)}$ and that this rate can be achieved by a plug-in test, using a minimax optimal estimator of the L^p -norm of g . In particular, the minimax rates of testing and estimation coincide, and they are located strictly between the parametric rate n^{-1} and the rate $n^{-\beta/(\beta+1)}$, corresponding to the problem of estimating the function g itself (see e.g. Jirak, Meister and Reiß [5] and the references therein).

These fundamental questions have been studied extensively in the mean regression and Gaussian white noise (GWN) model. In the latter, we observe a realisation of

$$dX(t) = g(t) dt + n^{-1/2} dW(t), \quad t \in [0, 1],$$

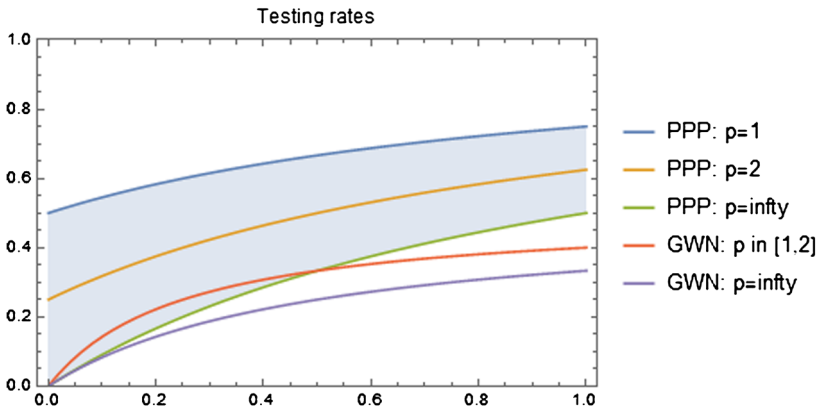


Figure 1. Testing rate exponents for the Poisson point process (PPP) and Gaussian white noise (GWN) model as a function of the regularity β .

where g is the unknown regression function and $(W(t) : t \in [0, 1])$ is a standard Brownian motion. Significant differences appear. Consider, for instance, the case $\Phi(u) = |u|^p$ with $p \in \mathbb{N}$. For p even and β large enough, the smooth functional (1.3) can be estimated with the parametric rate of convergence $n^{-1/2}$, using the method from Ibragimov, Nemirovski and Khasminski [3] (see Table 1 for the case $p = 2$ and the monograph by Nemirovski [14] for more general functionals). Estimation of the L^p -norm has been considered by Lepski, Nemirovski and Spokoiny [11]. For p even, the optimal rate of convergence is $n^{-\beta/(2\beta+1-1/p)}$, while for p odd, the standard non-parametric rate $n^{-\beta/(2\beta+1)}$ can only be improved by $\log n$ factors. In Table 1, we compare these GWN estimation rates with the PPP rates. A structural difference is that for vanishing regularity $\beta \downarrow 0$ the GWN convergence rates become arbitrarily slow, while in the PPP case the rates always remain faster than $n^{-1/2}$ and $n^{-1/(2p)}$, respectively. This phenomenon will be further discussed at the beginning of Section 2. More generally, the PPP rates hold universally for all $1 \leq p < \infty$, while the GWN rates depend on p in a very delicate way, showing that L^p -norm estimation is to some extent a regular estimation problem in the otherwise rather irregular PPP statistical model.

Further differences arise in the testing problem, which for the GWN model is the topic of the monograph by Ingster and Suslina [4]. The testing problem $H_0 : g = 0$ versus $H_1 : g \in \{h \in L^2([0, 1]) : \|h\|_p \geq r_n \text{ and } \|h\|_{\beta,q} \leq R\}$ is considered, where $\|\cdot\|_{\beta,q}$ is a Sobolev or Besov norm with smoothness measured in L^q -norm. For instance, in the case $1 \leq p \leq 2$ and $q = \infty$, the minimax separation rate is $n^{-2\beta/(4\beta+1)}$ which coincides with the minimax rate for estimating the L^p -norm if $p = 2$ but not if $p = 1$. The general minimax GWN separation rates for the case $q \geq p$ are given in the last row of Table 1 (for the cases $1 \leq p \leq 2$, $p \leq q \leq \infty$ and $2 < p = q < \infty$), results for the case $q < p$ can be found in Lepski and Spokoiny [12]. Figure 1 visualises the differences between the GWN and the PPP case by plotting the separation rate exponents for the range of $p \in [1, \infty)$ as a function of the regularity β . In the GWN model the rates become arbitrarily slow when β approaches zero and they do not change for $p \in [1, 2]$ (elbow effect), which is not the case in the PPP case. The absence of an elbow effect in the PPP model may be

understood by a different Hellinger geometry: the Hellinger distance is given by an L^1 -distance between the curves, while it is based on the L^2 -distance in the GWN model.

In the next Section 2, we construct the estimator, compute its mean and variance using the underlying point process geometry and martingale arguments, and we derive the (local) minimax rates of convergence. In Sections 3 and 4, we focus on the special case where $\Phi(u) = |u|^p$ and apply our results to the problem of estimating the L^p -norm and to the corresponding hypothesis testing problem.

2. Estimation of non-linear functionals

2.1. The estimator

Let $(X_j, Y_j)_{j \geq 1}$ be the observed support points of a Poisson point process on $[0, 1] \times \mathbb{R}$ with intensity function given by (1.1). The support boundary curve g is supposed to lie in the Hölder ball $\mathcal{C}^\beta(\mathbb{R})$ with $\beta \in (0, 1]$. The aim is to estimate the functional in (1.3). Similarly to [15], our estimation method can be motivated as follows. Suppose that we know a deterministic function $\bar{g} \in \mathcal{C}^\beta(\mathbb{R})$ with $\bar{g}(x) \geq g(x)$ for all $x \in [0, 1]$. Then the sum

$$\frac{1}{n} \sum_{j \geq 1} \Phi'(Y_j) \mathbb{1}(\bar{g}(X_j) \geq Y_j) \tag{2.1}$$

is a.s. finite, has expectation equal to

$$\frac{1}{n} \int_0^1 \int_{\mathbb{R}} \Phi'(y) \mathbb{1}(\bar{g}(x) \geq y) \lambda_g(x, y) dy dx = \int_0^1 (\Phi(\bar{g}(x)) - \Phi(g(x))) dx$$

and variance equal to

$$\begin{aligned} & \frac{1}{n^2} \int_0^1 \int_{\mathbb{R}} \Phi'(y)^2 \mathbb{1}(\bar{g}(x) \geq y) \lambda_g(x, y) dy dx \\ &= \frac{1}{n} \int_0^1 \int_{\mathbb{R}} \Phi'(y)^2 \mathbb{1}(\bar{g}(x) \geq y \geq g(x)) dy dx, \end{aligned} \tag{2.2}$$

provided the last integral is finite (see, e.g., [9], Lemma 1.1 or [10], Theorem 4.4). Thus,

$$\hat{F}_\Phi^{\text{pseudo}} = \int_0^1 \Phi(\bar{g}(x)) dx - \frac{1}{n} \sum_{j \geq 1} \Phi'(Y_j) \mathbb{1}(\bar{g}(X_j) \geq Y_j)$$

forms an unbiased pseudo-estimator (relying on the knowledge of \bar{g}) of $F_\Phi(g)$ whose variance is given by (2.2). The closer \bar{g} is to g , the smaller the variance. Concerning the rate results for p -th powers of L^p -norms in Table 1 note that already the very minor knowledge of some upper bound of g suffices to construct an estimator with convergence rate $n^{-1/2}$, which explains why in the PPP case even for $\beta \downarrow 0$ estimation and testing rates remain consistent.

The main idea is now to find a data-driven upper bound of g which is as small as possible. A solution to this problem is given by

$$\hat{g}^{\text{MLE}}(x) = \min_{k \geq 1} (Y_k + R|x - X_k|^\beta), \quad x \in [0, 1], \tag{2.3}$$

which is the maximum-likelihood estimator over $\mathcal{C}^\beta(R)$ [15], Section 3. Indeed, \hat{g}^{MLE} is an upper bound for g noting that $Y_k \geq g(X_k)$ and $g(X_k) + R|x - X_k|^\beta \geq g(x)$ for all $k \geq 1$, where the latter follows from $g \in \mathcal{C}^\beta(R)$.

The idea is now that the sum

$$\frac{1}{n} \sum_{j \geq 1} \Phi'(Y_j) \mathbb{1}(\hat{g}^{\text{MLE}}(X_j) \geq Y_j)$$

is a.s. finite and satisfies

$$\begin{aligned} \mathbb{E} \left[\frac{1}{n} \sum_{j \geq 1} \Phi'(Y_j) \mathbb{1}(\hat{g}^{\text{MLE}}(X_j) \geq Y_j) \right] &= \frac{1}{n} \int_0^1 \int_{\mathbb{R}} \Phi'(y) \mathbb{E}[\mathbb{1}(\hat{g}^{\text{MLE}}(x) \geq y)] \lambda_g(x, y) dy dx \\ &= \int_0^1 \mathbb{E}[\Phi(\hat{g}^{\text{MLE}}(x))] dx - \int_0^1 \Phi(g(x)) dx, \end{aligned}$$

provided that the integral in the second line is well-defined. For the first equality observe that

$$\mathbb{1}(\hat{g}^{\text{MLE}}(X_j) \geq Y_j) = \mathbb{1}\left(\min_{k \geq 1: k \neq j} (Y_k + R|X_j - X_k|^\beta) \geq Y_j\right)$$

where the term $j = k$ can be dropped. This implies that the observation (X_j, Y_j) can be integrated out, by following the usual arguments for computing sums with respect to a Poisson process (see, e.g., [10], Theorem 4.4). To summarise, we propose the following estimator

$$\hat{F}_\Phi = \int_0^1 \Phi(\hat{g}^{\text{MLE}}(x)) dx - \frac{1}{n} \sum_{j \geq 1} \Phi'(Y_j) \mathbb{1}(\hat{g}^{\text{MLE}}(X_j) \geq Y_j), \tag{2.4}$$

which is indeed an unbiased estimator of $F_\Phi(g)$ under the appropriate integrability condition.

Proposition 2.1. *Suppose that*

$$\int_0^1 \int_0^\infty |\Phi'(g(x) + u)| \mathbb{P}(\hat{g}^{\text{MLE}}(x) - g(x) \geq u) du dx < \infty. \tag{2.5}$$

Then \hat{F}_Φ from (2.4) is an unbiased estimator of $F_\Phi(g)$.

Remark 2.2. The above argument can be worked out for more general functionals of the form $\int_0^1 \dots \int_0^1 \Phi(x_1, \dots, x_m, g(x_1), \dots, g(x_m)) dx_1 \dots dx_m$, but then involves complex expressions in mixed partial derivatives of Φ . We therefore focus on estimation of the basic functional F_Φ .

Remark 2.3. For the one-sided regression model (1.2) the discrete functional $\hat{F}_\Phi^{(n)} = (1/n) \times \sum_{i=1}^n \Phi(g(i/n))$ can be estimated analogously by

$$\frac{1}{n} \sum_{i=1}^n \Phi(\hat{g}^{\text{rMLE}}(i/n)) - \frac{1}{\alpha n} \sum_{i=1}^n \Phi'(Y_i) \mathbb{1}(\hat{g}^{\text{rMLE}}(i/n) \geq Y_i)$$

with the regression analogue $\hat{g}^{\text{rMLE}}(x) = \min_i (Y_i + R|x - i/n|^\beta)$ of \hat{g}^{MLE} . This estimator can be analysed with the martingale arguments of the next section, compare the results in [15] for the linear case.

2.2. The martingale approach

We pursue a martingale-based analysis of the estimator \hat{F}_Φ in (2.4). The following result extends [15], Theorem 3.2 to non-linear functionals.

Theorem 2.4. *Suppose that the right-hand side in (2.6) below is finite. Then the estimator \hat{F}_Φ is UMVU over $g \in C^\beta(\mathbb{R})$ with variance*

$$\text{Var}(\hat{F}_\Phi) = \frac{1}{n} \int_0^1 \int_0^\infty (\Phi'(g(x) + u))^2 \mathbb{P}(\hat{g}^{\text{MLE}}(x) - g(x) \geq u) du dx. \tag{2.6}$$

Remark 2.5. If the right-hand side in (2.6) is finite, then Condition (2.5) holds since $\mathbb{P}(\hat{g}^{\text{MLE}}(x) - g(x) \geq u)$ is integrable in u , see also (2.8) below.

Proof. We first show the formula for the variance. Let $\lambda = (\lambda_t)$ be the process defined by $\lambda_t = n \int_0^1 \mathbb{1}(g(x) \leq t \leq \hat{g}^{\text{MLE}}(x)) dx$, $t \in \mathbb{R}$. Making a linear change of variables, the right-hand side in (2.6) can be written as

$$n^{-2} \mathbb{E} \left[\int_{t_0}^\infty \Phi'(s)^2 \lambda_s ds \right],$$

where t_0 is a lower bound for g . In the proof of Theorem 3.2 in [15], it is shown that the pure counting process $N = (N_t)$ defined by

$$N_t = \sum_{j \geq 1} \mathbb{1}(Y_j \leq t \wedge \hat{g}^{\text{MLE}}(X_j)), \quad t \geq t_0,$$

has compensator $A = (A_t)$ given by $A_t = \int_{t_0}^t \lambda_s ds$ and that $M = N - A$ is a square-integrable martingale with respect to the filtration $\mathcal{F}_t = \sigma((X_j, Y_j) \mathbb{1}(Y_j \leq t), j \geq 1)$. Its predictable quadratic variation is

$$\langle M \rangle_t = \int_{t_0}^t \lambda_s ds$$

(see also [7], Proposition 2.32). We conclude (e.g., via [6], Theorem 26.2) that

$$(\Phi' \cdot M)_t = \int_{t_0}^t \Phi'(s) dM_s = \sum_{j \geq 1} \Phi'(Y_j) \mathbb{1}(Y_j \leq t \wedge \hat{g}^{\text{MLE}}(X_j)) - \int_{t_0}^t \Phi'(s) \lambda_s ds$$

is an L^2 -bounded martingale with

$$\langle \Phi' \cdot M \rangle_t = \int_{t_0}^t \Phi'(s)^2 \lambda_s ds,$$

noting that $\mathbb{E}[\langle \Phi' \cdot M \rangle_t]$ is bounded by the right-hand side in (2.6), which is finite by assumption. For $t \rightarrow \infty$, the process $(\langle \Phi' \cdot M \rangle_t)$ converges almost surely to

$$\begin{aligned} (\Phi' \cdot M)_\infty &= \sum_{j \geq 1} \Phi'(Y_j) \mathbb{1}(Y_j \leq \hat{g}^{\text{MLE}}(X_j)) - \int_{t_0}^\infty \Phi'(s) \lambda_s ds \\ &= \sum_{j \geq 1} \Phi'(Y_j) \mathbb{1}(Y_j \leq \hat{g}^{\text{MLE}}(X_j)) - n \int_0^1 \Phi(\hat{g}^{\text{MLE}}(x)) dx + n \int_0^1 \Phi(g(x)) dx. \end{aligned}$$

Moreover, the process $(\langle \Phi' \cdot M \rangle_t)$ converges almost surely and in L^1 to

$$\langle \Phi' \cdot M \rangle_\infty = \int_{t_0}^\infty \Phi'(s)^2 \lambda_s ds.$$

Hence, unbiasedness and (2.6) follow from

$$\mathbb{E}[(\Phi' \cdot M)_\infty] = 0 \quad \text{and} \quad \mathbb{E}[(\Phi' \cdot M)_\infty^2 - \langle \Phi' \cdot M \rangle_\infty] = 0, \tag{2.7}$$

which holds due to the L^2 -convergence of $\Phi' \cdot M$ [6], Corollary 6.22.

Finally, the fact that \hat{F}_Φ is UMVU follows from the Lehmann–Scheffé theorem and [15], Proposition 3.1 which says that $(\hat{g}^{\text{MLE}}(x) : x \in [0, 1])$ is a sufficient and complete statistic for $\mathcal{C}^\beta(R)$. □

2.3. Rates of convergence

In this section, we derive convergence rates for the estimator \hat{F}_Φ . Using the argument leading to [15], Equation (3.3), we have the following deviation inequality for $x \in [0, 1]$:

$$\mathbb{P}(\hat{g}^{\text{MLE}}(x) - g(x) \geq u) \leq \begin{cases} \exp\left(-\frac{n\beta(2R)^{-\frac{1}{\beta}} u^{\frac{\beta+1}{\beta}}}{\beta+1}\right), & \text{if } u \in [0, 2R], \\ \exp\left(-n\left(u - \frac{2R}{\beta+1}\right)\right), & \text{if } u > 2R. \end{cases} \tag{2.8}$$

Thus, the right-hand side in (2.6) is finite if $(\Phi')^2$ has at most exponential growth with parameter strictly smaller than n . In particular, this holds for $\Phi(u) = |u|^p$, $p \geq 1$, in which case we have $\Phi'(u) = p|u|^{p-1} \operatorname{sgn}(u)$. A more detailed analysis gives:

Corollary 2.6. *Let $p \geq 1$ be a real number and consider $\Phi(u) = |u|^p$, $g \in \mathcal{C}^\beta(R)$. Then*

$$\hat{F}_p = \int_0^1 |\hat{g}^{\text{MLE}}(x)|^p dx - \frac{1}{n} \sum_{j \geq 1} p|Y_j|^{p-1} \operatorname{sgn}(Y_j) \mathbb{1}(\hat{g}^{\text{MLE}}(X_j) \geq Y_j) \tag{2.9}$$

is an unbiased estimator of $\|g\|_p^p$ with

$$\mathbb{E}[(\hat{F}_p - \|g\|_p^p)^2] \leq C \max(\|g\|_{2p-2}^{2p-2} n^{-\frac{2\beta+1}{\beta+1}}, n^{-\frac{2\beta p+1}{\beta+1}}), \tag{2.10}$$

where C is a constant depending only on R , β and p . Here, we use the notation $\|\cdot\|_q$ also for $q < 1$ with $\|g\|_0^0 := 1$.

Remark 2.7. In the proof, a more precise upper bound is derived in which the dependence on the constant R is explicit, see (2.12). For an asymptotically more precise result see Corollary 2.10 below.

Remark 2.8. Since $\Phi(u) = |u|^p$ is non-negative, the positive part $(\hat{F}_p)_+$ of \hat{F}_p always improves the estimator. This means that \hat{F}_p is not an admissible estimator in the decision-theoretic sense, while $(\hat{F}_p)_+$ on the other hand is no longer unbiased.

Proof. Throughout the proof $C > 0$ denotes a constant depending only on β and p that may change from line to line. By Theorem 2.4 and the discussion above, we have

$$\mathbb{E}[(\hat{F}_p - \|g\|_p^p)^2] = \frac{1}{n} \int_0^1 \int_0^\infty p^2 |u + g(x)|^{2p-2} \mathbb{P}(\hat{g}^{\text{MLE}}(x) - g(x) \geq u) du dx.$$

Applying (2.8) and the inequality $|u + g(x)|^{2p-2} \leq 2^{2p-2}(u^{2p-2} + |g(x)|^{2p-2})$, the last term is bounded by

$$\begin{aligned} & \frac{p^2 2^{2p-2}}{n} \int_0^{2R} (\|g\|_{2p-2}^{2p-2} + u^{2p-2}) \exp\left(-\frac{n\beta(2R)^{-\frac{1}{\beta}} u^{\frac{\beta+1}{\beta}}}{\beta+1}\right) du \\ & + \frac{p^2 2^{2p-2}}{n} \int_{2R}^\infty (\|g\|_{2p-2}^{2p-2} + u^{2p-2}) \exp\left(-n\left(u - \frac{2R}{\beta+1}\right)\right) du \\ & =: \text{(I)} + \text{(II)}. \end{aligned}$$

By a linear substitution, we have for $q \geq 0$

$$\begin{aligned} & \int_0^{2R} u^q \exp\left(-\frac{n\beta(2R)^{-\frac{1}{\beta}} u^{\frac{\beta+1}{\beta}}}{\beta+1}\right) du \\ & \leq \left(\frac{\beta+1}{\beta}\right)^{\frac{\beta(q+1)}{\beta+1}} (2R)^{\frac{q+1}{\beta+1}} n^{-\frac{\beta(q+1)}{\beta+1}} \int_0^\infty v^q \exp(-v^{\frac{\beta+1}{\beta}}) dv \\ & = \left(\frac{\beta+1}{\beta}\right)^{\frac{\beta q-1}{\beta+1}} (2R)^{\frac{q+1}{\beta+1}} \Gamma\left(\frac{\beta(q+1)}{\beta+1}\right) n^{-\frac{\beta(q+1)}{\beta+1}} \end{aligned} \tag{2.11}$$

with the Gamma function Γ . Consequently,

$$(I) \leq CR^{\frac{1}{\beta+1}} \|g\|_{2p-2}^{2p-2} n^{-\frac{2\beta+1}{\beta+1}} + CR^{\frac{2p-1}{\beta+1}} n^{-\frac{2\beta p+1}{\beta+1}}.$$

Next, consider the remainder term (II). We have

$$\int_{2R}^\infty \exp\left(-n\left(u - \frac{2R}{\beta+1}\right)\right) du = n^{-1} e^{-\frac{2\beta Rn}{\beta+1}}$$

and

$$\begin{aligned} \int_{2R}^\infty u^{2p-2} \exp\left(-n\left(u - \frac{2R}{\beta+1}\right)\right) du & \leq \int_{2R}^\infty u^{2p-2} \exp\left(-\frac{n\beta u}{\beta+1}\right) du \\ & \leq \left(\frac{\beta+1}{n\beta}\right)^{2p-1} \int_{2\beta Rn/(\beta+1)}^\infty v^{2p-2} \exp(-v) dv \\ & \leq Cn^{-2p+1} e^{-\frac{\beta Rn}{\beta+1}}. \end{aligned}$$

Note that the last integral can be computed using partial integration. Thus

$$(II) \leq C \|g\|_{2p-2}^{2p-2} n^{-2} e^{-\frac{2\beta Rn}{\beta+1}} + Cn^{-2p} e^{-\frac{\beta Rn}{\beta+1}}.$$

Summarising, we have

$$\begin{aligned} \mathbb{E}[(\hat{F}_p - \|g\|_p^p)^2] & \leq CR^{\frac{1}{\beta+1}} \|g\|_{2p-2}^{2p-2} n^{-\frac{2\beta+1}{\beta+1}} + CR^{\frac{2p-1}{\beta+1}} n^{-\frac{2\beta p+1}{\beta+1}} \\ & \quad + C \|g\|_{2p-2}^{2p-2} n^{-2} e^{-\frac{2\beta Rn}{\beta+1}} + Cn^{-2p} e^{-\frac{\beta Rn}{\beta+1}}, \end{aligned} \tag{2.12}$$

and the claim follows. □

One might wonder whether \hat{F}_p achieves the rate $n^{-(\beta+1/2)/(\beta+1)}$ uniformly over $g \in \mathcal{C}^\beta(R) \cap B_p(R)$ with the L^p -ball $B_p(R) = \{g \in L^p([0, 1]) : \|g\|_p \leq R\}$. For $1 \leq p \leq 2$ this follows from the inclusion $B_p(R) \subseteq B_{2p-2}(R)$. For $p > 2$ this holds as well and is a consequence of the

following useful lemma (with $q = 2p - 2$) providing a simple interpolation result. Results of this type are well known (cf. Bergh and Löfström [1]), but since only Hölder semi-norms appear, we provide a self-contained proof in the Appendix.

Lemma 2.9. *Let $1 \leq p \leq q \leq \infty$ and $f \in \mathcal{C}^\beta(R)$. Then we have*

$$\|f\|_q \leq C \|f\|_p \max(1, R/\|f\|_p)^{\frac{1/p-1/q}{\beta+1/p}},$$

where $C > 0$ is a constant depending only on β , p and q and the right-hand side is understood to be zero for $f = 0$.

Let us come to another corollary of Theorem 2.4 which provides a local asymptotic upper bound for the minimax risk under weak assumptions on the functional:

Corollary 2.10. *Suppose that there is a constant $C > 0$ such that $|\Phi'(u)| \leq C \exp(C|u|)$ for all $u \in \mathbb{R}$. Let $f \in \mathcal{C}^\beta(R)$. Suppose that $\|\Phi' \circ f\|_2 \neq 0$ and that the map $F' : \mathcal{C}^\beta(R) \subseteq L^2([0, 1]) \rightarrow L^2([0, 1])$, $F'(g) = \Phi' \circ g$ is continuous at $g = f$ with respect to the L^2 -norms. Then the estimator $\hat{F}_{\Phi,n} = \hat{F}_\Phi$ satisfies the local asymptotic upper bound*

$$\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} \sup_{\substack{g \in \mathcal{C}^\beta(R): \\ \|f-g\|_2 \leq \delta}} n^{\frac{2\beta+1}{\beta+1}} \mathbb{E}_g[(\hat{F}_{\Phi,n} - F_\Phi(g))^2] \leq \Gamma \left(\frac{\beta}{\beta+1} \right) \left(\frac{2R\beta}{\beta+1} \right)^{\frac{1}{\beta+1}} \|\Phi' \circ f\|_2^2$$

with the Gamma function Γ .

Proof. By Theorem 2.4 and Equation (2.8), we have

$$\begin{aligned} \mathbb{E}_g[(\hat{F}_\Phi - F_\Phi(g))^2] &\leq \frac{1}{n} \int_0^{2R} \|\Phi' \circ (u + g)\|_2^2 \exp\left(-\frac{n\beta(2R)^{-\frac{1}{\beta}} u^{\frac{\beta+1}{\beta}}}{\beta+1}\right) du \\ &\quad + \frac{1}{n} \int_{2R}^\infty \|\Phi' \circ (u + g)\|_2^2 \exp\left(-\frac{n\beta u}{\beta+1}\right) du. \end{aligned}$$

By Lemma 2.9, applied to $f - g$ and with $p = 2$, $q = \infty$, we infer from $g \in \mathcal{C}^\beta(R)$ with $\|f - g\|_2 \leq \delta$ that

$$\|f - g\|_\infty \leq C' R^{1/(2\beta+1)} \delta^{2\beta/(2\beta+1)} \tag{2.13}$$

holds with some constant C' , provided that $\delta \leq R$. Using that Φ' has at most exponential growth, we get that $\|\Phi' \circ (u + g)\|_2^2 \leq C \exp(C|u|)$ uniformly over all $g \in \mathcal{C}^\beta(R)$ with $\|f - g\|_2 \leq \delta$ (adjusting C appropriately). This shows that the second term is of smaller order than n^{-2} and thus asymptotically negligible for our result. Similarly, for every fixed $\delta' > 0$ the first integral from δ' to $2R$ becomes exponentially small in n . Thus, for any $\delta' > 0$ the left-hand side in

Corollary 2.10 is bounded by

$$\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} \sup_{\substack{g \in \mathcal{C}^\beta(R): \\ \|f - g\|_2 \leq \delta}} n^{\frac{\beta}{\beta+1}} \int_0^{\delta'} \|\Phi' \circ (u + g)\|_2^2 \exp\left(-\frac{n\beta(2R)^{-\frac{1}{\beta}} u^{\frac{\beta+1}{\beta}}}{\beta + 1}\right) du. \tag{2.14}$$

By the continuity of F'_Φ at f and the fact that $\|\Phi' \circ f\|_2 \neq 0$, for every $\varepsilon > 0$ there exist $\delta, \delta' > 0$ such that $\|\Phi' \circ (u + g)\|_2 \leq (1 + \varepsilon)\|\Phi' \circ f\|_2$ for all $|u| \leq \delta'$ and $g \in \mathcal{C}^\beta(R)$ with $\|f - g\|_2 \leq \delta$. We conclude that (2.14) is bounded by (using the computation in (2.11) for $q = 0$)

$$\Gamma\left(\frac{\beta}{\beta + 1}\right) \left(\frac{2R\beta}{\beta + 1}\right)^{\frac{1}{\beta+1}} \|\Phi' \circ f\|_2^2,$$

and the claim follows. □

Remark 2.11. By Lemma 2.9 continuity of F'_Φ on $\mathcal{C}^\beta(R)$ with respect to L^2 -norm implies continuity with respect to supremum norm. Under the assumptions of Corollary 2.10, one can indeed show that the functional F_Φ is Fréchet-differentiable in f along $\mathcal{C}^\beta(R)$ with derivative $F'_\Phi(f) = \Phi' \circ f$.

Remark 2.12. Local asymptotic minimax results for estimating smooth functionals in the GWN model can be found in Nemirovski [14], Chapter 7. The rate is different (see the discussion in the Introduction), but the term $\|F'_\Phi(g)\|_2^2$ appears as well. The latter fact can be explained by linearising F_Φ at g .

Remark 2.13. The estimators are non-adaptive in the sense that they rely on the knowledge of the regularity parameters β and R . In [15] the Lepski method has been employed to construct adaptive estimators in the linear case, based on a blockwise estimator. We conjecture that this approach would also give an adaptive rate-optimal estimator here. Note also the restriction $\beta \leq 1$ on the regularity parameter. The reason is that the MLE for $g \in \mathcal{C}^\beta(R)$ with $\beta > 1$ does not necessarily provide a pointwise upper bound for g such that the present approach may fail.

2.4. Lower bounds

In this section, we establish lower bounds corresponding to Corollaries 2.6 and 2.10. We will apply the method of two fuzzy hypotheses (see [16], Chapter 2.7.4) with a prior corresponding to independent non-identical Bernoulli random variables. Our main result states a local asymptotic lower bound in the case that Φ is continuously differentiable. Possible extensions are discussed afterwards.

Theorem 2.14. Let Φ be continuously differentiable and $f \in \mathcal{C}^\beta(\mathbb{R})$ with $\|\Phi' \circ f\|_2 \neq 0$. Then there is a constant $c_1 > 0$, depending only on β , such that

$$\lim_{\delta \rightarrow 0} \liminf_{n \rightarrow \infty} \inf_{\substack{\tilde{F}_n \\ g \in \mathcal{C}^\beta(\mathbb{R}): \\ \|f-g\|_2 \leq \delta}} \sup_n n^{\frac{2\beta+1}{\beta+1}} \mathbb{E}_g[(\tilde{F}_n - F_\Phi(g))^2] > c_1 R^{\frac{1}{\beta+1}} \|\Phi' \circ f\|_2^2.$$

The infimum is taken over all estimators in the PPP model with intensity (1.1).

Proof. We want to apply the χ^2 -version of the method of two fuzzy hypotheses as described in [16], Theorem 2.15. Consider the functions

$$g_\theta = \sum_{k=1}^m \theta_k g_k \quad \text{with } \theta_k \in \{0, 1\}$$

and

$$g_k(x) = cRh^\beta K\left(\frac{x - (k-1)h}{h}\right) = cRh^{\beta+1} K_h(x - (k-1)h)$$

with $h = 1/m$, triangular kernel $K(u) = 4(u \wedge (1-u))\mathbb{1}_{[0,1]}(u)$, $K_h(\cdot) = K(\cdot/h)/h$ and $c > 0$ sufficiently small such that $g_\theta \in \mathcal{C}^\beta(\mathbb{R})$ for all m and θ . Let π_n be the probability measure on $\{0, 1\}^m$ obtained when $\theta_1, \dots, \theta_m$ are independent (non-identical) Bernoulli random variables with success probabilities p_1, \dots, p_m . Let P_g denote the law of the observations in the PPP model with intensity function (1.1). We set $\mathbf{P}_{0,n} = P_f$ and

$$\mathbf{P}_{1,n}(\cdot) = \int P_{f+g_\theta}(\cdot) \pi_n(d\theta).$$

In order to obtain the result, it suffices to find $m \geq 1$ and probabilities p_1, \dots, p_m (both depending on n) as well as a constant $c_1 > 0$, only depending on β , and an absolute constant $c_2 < \infty$, such that

(i) For each fixed $\delta > 0$ the inequality $\|g_\theta\|_2 \leq \delta$ holds for all n sufficiently large and for $n \rightarrow \infty$ the prior satisfies

$$\pi_n(F_\Phi(f + g_\theta) \geq F_\Phi(f) + 2c_1 \|\Phi' \circ f\|_2 R^{\frac{1/2}{\beta+1}} n^{-\frac{\beta+1/2}{\beta+1}}) \rightarrow 1;$$

(ii) $\limsup_{n \rightarrow \infty} \chi^2(\mathbf{P}_{1,n}, \mathbf{P}_{0,n}) \leq c_2$.

We start with the following lemma on the χ^2 -distance.

Lemma 2.15. Suppose that the success probabilities satisfy $\sum_{k=1}^m p_k^2 = 1$. Then

$$\chi^2(\mathbf{P}_{1,n}, \mathbf{P}_{0,n}) = \int \left(\frac{d\mathbf{P}_{1,n}}{d\mathbf{P}_{0,n}} \right)^2 d\mathbf{P}_{0,n} - 1 \leq \exp\left(\exp\left(n \int_{I_1} g_1(x) dx\right) - 1\right) - 1$$

holds, where $I_1 = [0, h)$.

Proof of Lemma 2.15. We abbreviate $\int g_k = \int_{I_k} g_k(x) dx$, where $I_k = [(k-1)h, kh)$ for $k < m$ and $I_m = [1-h, 1]$. Let us first see that

$$\frac{d\mathbf{P}_{1,n}}{d\mathbf{P}_{0,n}} = \prod_{k=1}^m (1 - p_k + p_k e^{n \int g_k} \mathbb{1}(\forall X_j \in I_k : Y_j \geq f(X_j) + g_k(X_j))). \quad (2.15)$$

Indeed, by definition the left-hand side is equal to

$$\begin{aligned} & \sum_{\theta \in \{0,1\}^m} \left(\prod_{k:\theta_k=0} (1 - p_k) \prod_{k:\theta_k=1} p_k \right) \frac{dP_{f+g\theta}}{dP_f} \\ &= \sum_{\theta \in \{0,1\}^m} \left(\prod_{k:\theta_k=0} (1 - p_k) \prod_{k:\theta_k=1} p_k e^{n \int g_k} \mathbb{1}(\forall X_j \in I_k : Y_j \geq f(X_j) + g_k(X_j)) \right) \\ &= \prod_{k=1}^m (1 - p_k + p_k e^{n \int g_k} \mathbb{1}(\forall X_j \in I_k : Y_j \geq f(X_j) + g_k(X_j))), \end{aligned}$$

where we used the formula (see [9], Theorem 1.3 or [15], Section 3)

$$\begin{aligned} \frac{dP_{f+g\theta}}{dP_f} &= e^{n \int g\theta} \mathbb{1}(\forall j : Y_j \geq f(X_j) + g\theta(X_j)) \\ &= \prod_{k:\theta_k=1} e^{n \int g_k} \mathbb{1}(\forall X_j \in I_k : Y_j \geq f(X_j) + g_k(X_j)) \end{aligned}$$

in the first equality. By the defining properties of the PPP, under $\mathbf{P}_{0,n}$, the right-hand side in (2.15) is a product of independent random variables and the corresponding indicators have success probabilities $e^{-n \int g_k}$. Thus, we obtain

$$\begin{aligned} \int \left(\frac{d\mathbf{P}_{1,n}}{d\mathbf{P}_{0,n}} \right)^2 d\mathbf{P}_{0,n} &= \prod_{k=1}^m ((1 - p_k)^2 + 2p_k(1 - p_k) + p_k^2 e^{n \int g_k}) \\ &= \prod_{k=1}^m (1 + p_k^2 (e^{n \int g_k} - 1)) \\ &\leq \prod_{k=1}^m e^{p_k^2 (e^{n \int g_k} - 1)} = e^{e^{n \int g_1} - 1}, \end{aligned}$$

where we used the bound $1 + x \leq e^x$ and the assumption $\sum_{k=1}^m p_k^2 = 1$. □

Using Lemma 2.15 and the identity

$$n \int_{I_1} g_1(x) dx = cRnh^{\beta+1},$$

we get (ii) provided that we choose $m = 1/h$ of size $(Rn)^{1/(\beta+1)}$ and p_1, \dots, p_m such that $\sum_{k=1}^m p_k^2 = 1$. Thus it remains to choose the p_k such that the second convergence in (i) is satisfied.

We first consider the case that $\Phi' \circ f \geq 0$. Let $\varepsilon > 0$ be a small constant to be chosen later. Since Φ' is uniformly continuous on compact intervals, there is a $\delta' > 0$ such that

$$\int_0^1 \Phi(f(x) + g(x)) dx - \int_0^1 \Phi(f(x)) dx \geq \int_0^1 \Phi'(f(x))g(x) dx - \varepsilon \int_0^1 |g(x)| dx$$

for all $g \in \mathcal{C}^\beta(\mathbb{R})$ with $\|f - g\|_2 \leq \delta'$ (using (2.13) above). Thus, for n sufficiently large, we get

$$\begin{aligned} F_\Phi(f + g_\theta) - F_\Phi(f) &\geq \langle \Phi' \circ f, g_\theta \rangle - \varepsilon \langle 1, g_\theta \rangle \\ &= \sum_{k=1}^m \theta_k \langle \Phi' \circ f, g_k \rangle - \varepsilon \sum_{k=1}^m \theta_k \langle 1, g_k \rangle \\ &= cRh^{\beta+1} \left(\sum_{k=1}^m \theta_k \langle \Phi' \circ f, K_h(\cdot - (k-1)h) \rangle - \varepsilon \sum_{k=1}^m \theta_k \right). \end{aligned}$$

Setting $a_k = \langle \Phi' \circ f, K_h(\cdot - (k-1)h) \rangle$, this can be written as

$$F_\Phi(f + g_\theta) - F_\Phi(f) \geq cRh^{\beta+1} \left(\sum_{k=1}^m a_k \theta_k - \varepsilon \sum_{k=1}^m \theta_k \right). \tag{2.16}$$

The first sum is a weighted sum of independent non-identical Bernoulli random variables and the maximising choice for the success probabilities is

$$p_k = \frac{a_k}{\|a\|_2} \tag{2.17}$$

(the a_k satisfy $a_k \geq 0$ since we assumed $\Phi' \circ f \geq 0$). By the mean value theorem and the fact that $\Phi' \circ f$ is continuous, we get $a_k = \Phi'(f(x_k))$ with $x_k \in [(k-1)h, kh]$ and also

$$\frac{1}{m} \|a\|_q^q = \frac{1}{m} \sum_{k=1}^m a_k^q \rightarrow \int_0^1 (\Phi'(f(x)))^q dx = \|\Phi' \circ f\|_q^q \quad \text{as } n \rightarrow \infty \tag{2.18}$$

for each $q \geq 1$. Using the Chebyshev inequality, we get

$$\begin{aligned} \pi_n \left(\sum_{k=1}^m a_k \theta_k < \|a\|_2/2 \right) &= \pi_n \left(\sum_{k=1}^m a_k (\theta_k - p_k) < -\|a\|_2/2 \right) \\ &\leq \frac{4 \sum_{k=1}^m a_k^2 p_k (1 - p_k)}{\|a\|_2^2} \leq 4 \left(\frac{\|a\|_3}{\|a\|_2} \right)^3 \end{aligned}$$

and the latter converges to 0 as $n \rightarrow \infty$ by (2.18). Similarly,

$$\begin{aligned} \pi_n \left(\sum_{k=1}^m \theta_k > 2 \|a\|_1 / \|a\|_2 \right) &= \pi_n \left(\sum_{k=1}^m (\theta_k - p_k) > \|a\|_1 / \|a\|_2 \right) \\ &\leq \frac{\|a\|_2^2 \sum_{k=1}^m p_k (1 - p_k)}{\|a\|_1^2} \leq \frac{\|a\|_2}{\|a\|_1} \end{aligned}$$

and the latter converges to 0 as $n \rightarrow \infty$ by (2.18). Combining these two bounds with (2.16) we get

$$\pi_n \left(F_\Phi(f + g_\theta) - F_\Phi(f) \geq cRh^{\beta+1/2} \left(\frac{1}{2\sqrt{m}} \|a\|_2 - \varepsilon \frac{2}{\sqrt{m}} \frac{\|a\|_1}{\|a\|_2} \right) \right) \rightarrow 1 \tag{2.19}$$

as $n \rightarrow \infty$. This implies (i) if ε is chosen small enough since $\|a\|_2/\sqrt{m}$ and $\|a\|_1/(\sqrt{m}\|a\|_2)$ have non-zero limits by (2.18) and the assumption $\|\Phi' \circ f\|_2 \neq 0$. This completes the proof in the case $\Phi' \circ f \geq 0$.

If $\Phi' \circ f \leq 0$, then we may follow the same line of arguments where (ii) is replaced with a left-deviation inequality (which corresponds to apply the above arguments to the functional $F_{-\Phi}$). Next, if $\Phi' \circ f$ takes both, positive and negative values, then we may choose $p_k = a_{k+}/\|a_+\|_2$ (resp. $p_k = a_{k-}/\|a_-\|_2$) leading to a lower bound with $\|\Phi' \circ f\|_2^2$ replaced by $\|(\Phi' \circ f)_+\|_2^2$ (resp. $\|(\Phi' \circ f)_-\|_2^2$). Summing up both lower bounds gives the claim in the general case. \square

Remark 2.16. If Φ is convex, then we can replace (2.16) by

$$F_\Phi(f + g_\theta) - F_\Phi(f) \geq \langle \Phi' \circ f, g_\theta \rangle = cRh^{\beta+1} \sum_{k=1}^m a_k \theta_k,$$

leading to a shortening of the above proof. In this case, the lower bound also holds without continuity of Φ' . The arguments, however, must be adapted slightly since the convergence in (2.18) may not hold in this case.

Remark 2.17. By making the constants in the proof of Theorem 2.14 explicit, one can also establish non-asymptotic lower bounds which include lower-order terms. Consider for instance $\Phi(u) = |u|^p$, $p \in \mathbb{N}$ and $f \equiv a > 0$. Then we have

$$\begin{aligned} F_\Phi(a + g_\theta) - a^p &= \left(\sum_{k=1}^m \theta_k \right) \sum_{j=1}^p \binom{p}{j} a^{p-j} c^j R^j h^{\beta j+1} \|K\|_j^j \\ &\geq \left(\sum_{k=1}^m \theta_k \right) \max(p a^{p-1} c R h^{\beta+1}, c^p R^p \|K\|_p^p h^{\beta p+1}). \end{aligned} \tag{2.20}$$

We choose

$$p_1 = \dots = p_m = 1/\sqrt{m} \quad \text{and} \quad m = \lfloor 2(cRn)^{1/(\beta+1)} \rfloor.$$

In order to ensure $\|g_\theta\|_2 \leq \delta$, it suffices that $m \geq 1$ and $2cRh^\beta \leq \delta$ hold, which is satisfied if $n \geq c_1$ with c_1 depending only on c, R and δ . Now, by Lemma 2.15 and the choice of m we have $\chi^2(\mathbf{P}_{0,n}, \mathbf{P}_{1,n}) \leq e^{e^{-1}} - 1$. Moreover, using the simplification of Remark 2.16, (2.19) becomes

$$\pi_n \left(F_\Phi(a + g_\theta) \geq a^p + \frac{1}{2} \max(pa^{p-1}cRh^{\beta+1/2}, c^p R^p \|K\|_p^p h^{\beta p+1/2}) \right) \geq 1 - 4/\sqrt{m}.$$

Inserting the value of h and applying [16], Theorem 2.15(iii), we get

$$\begin{aligned} \inf_{\tilde{F}_n} \sup_{\substack{g \in \mathcal{C}^\beta(R): \\ \|a-g\|_2 \leq \delta}} \mathbb{P}_g \left(|\tilde{F}_n - F_\Phi(g)| \geq \max \left(c_2 pa^{p-1} R^{\frac{1/2}{\beta+1}} n^{-\frac{\beta+1/2}{\beta+1}}, c_3 R^{\frac{p-1/2}{\beta+1}} n^{-\frac{\beta p+1/2}{\beta+1}} \right) \right) \\ \geq \frac{1}{4} \exp(-(e^{e^{-1}} - 1)) - 2/\sqrt{m}, \end{aligned}$$

provided that $n \geq c_1$, where c_2 is a constant depending only on β and c_3 is a constant depending only on β and p . Thus we obtain a lower bound which has the form of the upper bound in Corollary 2.6 (resp. (2.12)).

Remark 2.18. In the case of linear functionals, the above proof can be used to obtain the lower bound in [15], Theorem 2.6. Instead of using the method of fuzzy hypothesis, one can also try to apply the method used in Reiß and Selk [15] and Korostelev and Tsybakov [8] which is based on a comparison of the minimax risk with a Bayesian risk. This works for instance for the special case $\Phi(u) = |u|^p$, $p \in \mathbb{N}$, and $f \equiv a > 0$, but it is not clear whether this structurally different prior can produce the correct lower bounds more generally.

3. Hypothesis testing

3.1. Main result

In this section, we use the previous results to address the hypothesis testing problem

$$H_0 : g = g_0 \quad \text{vs.} \quad H_1 : g \in g_0 + \mathcal{G}_n,$$

where g_0 is a known function and

$$\mathcal{G}_n = \mathcal{G}_n(\beta, R, p, r_n) = \{g \in \mathcal{C}^\beta(R) : \|g\|_p \geq r_n\}.$$

In the sequel, we restrict to the case $g_0 = 0$, since the general case can be reduced to this one by a simple shift of the observations. We propose the following plug-in test

$$\psi_{n,p} = \mathbb{1}(\hat{F}_p \geq r_n^p/2), \tag{3.1}$$

with the estimator \hat{F}_p from (2.9). We follow a minimax approach to hypothesis testing, see, for example, [4], Chapter 2.4. Our main result of this section states that $\psi_{n,p}$ achieves the minimax separation rates:

Theorem 3.1. *Let $p \geq 1$ be a real number and*

$$r_n^* = n^{-\frac{\beta+1/(2p)}{\beta+1}}.$$

Then, the following holds as $n \rightarrow \infty$:

(a) *If $r_n/r_n^* \rightarrow \infty$, then the tests $\psi_{n,p}$ from (3.1) satisfy*

$$\mathbb{E}_0[\psi_{n,p}] + \sup_{g \in \mathcal{G}_n} \mathbb{E}_g[1 - \psi_{n,p}] \rightarrow 0.$$

(b) *If $r_n/r_n^* \rightarrow 0$, then we have*

$$\inf_{\psi_n} \left(\mathbb{E}_0[\psi_n] + \sup_{g \in \mathcal{G}_n} \mathbb{E}_g[1 - \psi_n] \right) \rightarrow 1,$$

where the infimum is taken over all tests in the PPP model with intensity (1.1).

3.2. Proof of the upper bound

Throughout the proof, $C > 0$ denotes a constant depending only on R , β and p that may change from line to line. Under the null hypothesis we have, using the Chebyshev inequality and Corollary 2.6,

$$\mathbb{E}_0[\psi_{n,p}] = \mathbb{P}_0(\hat{F}_p \geq r_n^p/2) \leq \frac{4\mathbb{E}_0[\hat{F}_p^2]}{r_n^{2p}} \leq C \frac{n^{-\frac{2\beta p+1}{\beta+1}}}{r_n^{2p}} = C \left(\frac{r_n^*}{r_n} \right)^{2p} \quad (3.2)$$

and by assumption the right-hand side tends to zero as $n \rightarrow \infty$. Next, consider the type-two error $\mathbb{E}_g[1 - \psi_{n,p}]$ with $g \in \mathcal{G}_n$. Let $k \in \mathbb{N}$ be such that $2^{k-1}r_n^p \leq \|g\|_p^p < 2^k r_n^p$ and set $r_{n,k} = 2^{k/p} r_n$. By the Chebyshev inequality, we have

$$\begin{aligned} \mathbb{E}_g[1 - \psi_{n,p}] &= \mathbb{P}_g(\hat{F}_p < r_n^p/2) \\ &= \mathbb{P}_g(\|g\|_p^p - \hat{F}_p > \|g\|_p^p - r_n^p/2) \\ &\leq \mathbb{P}_g(\|g\|_p^p - \hat{F}_p > r_{n,k}^p/4) \\ &\leq \frac{16\mathbb{E}_g[(\hat{F}_p - \|g\|_p^p)^2]}{r_{n,k}^{2p}}. \end{aligned} \quad (3.3)$$

Now, we may restrict ourselves to the case that

$$\|g\|_{2p-2}^{2p-2} n^{-\frac{2\beta+1}{\beta+1}} \geq n^{-\frac{2\beta p+1}{\beta+1}}. \quad (3.4)$$

Indeed, if (3.4) does not hold, then the maximal type-two error is also bounded by $C(r_n^*/r_n)^{2p}$, as can be seen by the same argument as in (3.2). By (3.3), (3.4) and Corollary 2.6, we obtain

$$\mathbb{E}_g[1 - \psi_{n,p}] \leq C \|g\|_{2p-2}^{2p-2} \frac{n^{-\frac{2\beta+1}{\beta+1}}}{r_{n,k}^{2p}}. \tag{3.5}$$

Let us consider the cases $1 \leq p \leq 2$ and $p > 2$ separately. If $1 < p \leq 2$, then we have $\|g\|_{2p-2} \leq \|g\|_p \leq r_{n,k}$ by the Hölder inequality and the definition of k . Thus, for $1 \leq p \leq 2$, we get

$$\mathbb{E}_g[1 - \psi_{n,p}] \leq C \frac{n^{-\frac{2\beta+1}{\beta+1}}}{r_{n,k}^2} \leq C \left(\frac{r_n^*}{r_{n,k}}\right)^2 n^{-\frac{2\beta+1}{\beta+1} + \frac{2\beta+1/p}{\beta+1}} \leq C \left(\frac{r_n^*}{r_n}\right)^2.$$

Taking the supremum over all $g \in \mathcal{G}_n$, the right-hand side tends to zero as $n \rightarrow \infty$. Next, consider the case $p > 2$. Applied with $q = 2p - 2 > p$, Lemma 2.9 gives

$$\|g\|_{2p-2}^{2p-2} \leq C \|g\|_p^{2p-2} \max(1, \|g\|_p^{-1})^{\frac{1-2/p}{\beta+1/p}}. \tag{3.6}$$

If $\|g\|_p > 1$, then the claim follows as in the case $1 \leq p \leq 2$. If $\|g\|_p \leq 1$, then by (3.5) and (3.6), we have

$$\begin{aligned} \mathbb{E}_g[1 - \psi_{n,p}] &\leq C r_{n,k}^{-2 - \frac{1-2/p}{\beta+1/p}} n^{-\frac{2\beta+1}{\beta+1}} \\ &= C \left(\frac{r_n^*}{r_{n,k}}\right)^{\frac{2\beta+1}{\beta+1/p}} n^{\frac{2\beta+1}{\beta+1/p} - \frac{\beta+1/2p}{\beta+1} - \frac{2\beta+1}{\beta+1}} \leq C \left(\frac{r_n^*}{r_n}\right)^{\frac{2\beta+1}{\beta+1/p}}. \end{aligned}$$

Again, taking the supremum over all $g \in \mathcal{G}_n$, the right-hand side tends to zero as $n \rightarrow \infty$. This completes the proof of (i). □

3.3. Proof of the lower bound

We set $\mathbf{P}_{1,n}(\cdot) = \int P_{g_\theta}(\cdot) \pi_n(d\theta)$ and $\mathbf{P}_{0,n} = P_0$ with g_θ and π_n as in the proof of Theorem 2.14 with the choice

$$p_1 = \dots = p_m = 1/\sqrt{m}. \tag{3.7}$$

By [4], Proposition 2.9 and Proposition 2.12, in order that Theorem 3.1(ii) holds, we have to show that as $n \rightarrow \infty$,

- (i) $\pi_n(g_\theta \in \mathcal{G}_n) \rightarrow 1$;
- (ii) $\chi^2(\mathbf{P}_{1,n}, \mathbf{P}_{0,n}) \rightarrow 0$.

For (i), note that

$$\|g_\theta\|_p = \left(\sum_{k=1}^m \theta_k\right)^{1/p} c R h^{\beta+1/p} \|K\|_p.$$

By the Chebyshev inequality, we have

$$\begin{aligned} \pi_n \left(\left(\sum_{k=1}^m \theta_k \right)^{1/p} \leq 2^{-1/p} m^{1/(2p)} \right) &= \pi_n \left(\sum_{k=1}^m (\theta_k - 1/\sqrt{m}) \leq -\sqrt{m}/2 \right) \\ &\leq \frac{4m(1/\sqrt{m})(1 - 1/\sqrt{m})}{m}, \end{aligned}$$

where the right-hand side tends to zero as $m \rightarrow \infty$. Thus (i) holds provided that we choose $m^{-1} = h$ of size

$$c_1 r_n^{\frac{1}{\beta+1/(2p)}}$$

with $c_1 > 0$ depending only on R and p . Moreover, by Lemma 2.15 and (3.7), we have

$$\chi^2(\mathbf{P}_{1,n}, \mathbf{P}_{0,n}) \leq \exp(\exp(c R n h^{\beta+1}) - 1) - 1.$$

Inserting the above choice of h , the last expression goes to zero as $n \rightarrow \infty$, since

$$n r_n^{\frac{\beta+1}{\beta+1/(2p)}} = (r_n/r_n^*)^{\frac{\beta+1}{\beta+1/(2p)}} \rightarrow 0.$$

This completes the proof. □

4. Estimating the L^p -norm

Finally let us consider the problem of estimating the L^p -norm of g . We define the estimator \hat{T} of $\|g\|_p$ by

$$\hat{T} = (\max(\hat{F}_p, 0))^{1/p} = (\hat{F}_p)_+^{1/p}.$$

Our main result of this section is as follows:

Theorem 4.1. *Let $p \geq 1$ be a real number. Then we have*

$$\sup_{g \in C^\beta(R)} \mathbb{E}_g [|\hat{T} - \|g\|_p|] \leq C n^{-\frac{\beta+1/(2p)}{\beta+1}}$$

with a constant $C > 0$ depending only on R, β and p .

On the other hand, we have

$$\liminf_{n \rightarrow \infty} n^{\frac{\beta+1/(2p)}{\beta+1}} \inf_{\tilde{T}_n} \sup_{g \in C^\beta(R)} \mathbb{E}_g [|\tilde{T}_n - \|g\|_p|] > 0,$$

where the infimum is taken over all estimators in the PPP model with intensity (1.1). In particular, the minimax rate of estimation over $C^\beta(R)$ is $n^{-(\beta+1/(2p))/(\beta+1)}$.

Proof. The lower bound follows from the lower bound in Theorem 3.1. To see this, let $r_n^{\text{est}} = \inf_{\tilde{T}_n} \sup_{g \in \mathcal{C}^\beta(R)} \mathbb{E}_g[|\tilde{T}_n - \|g\|_p|]$ be the minimax risk. If the lower bound in Theorem 4.1 was false, then $r_{n_k}^{\text{est}}/r_{n_k}^* \rightarrow 0$ along a subsequence (n_k) . Now construct (r_{n_k}) such that $r_{n_k}/r_{n_k}^* \rightarrow 0$ and $r_{n_k}/r_{n_k}^{\text{est}} \rightarrow \infty$. Using [4], Proposition 2.17 and the fact that $r_{n_k}/r_{n_k}^{\text{est}} \rightarrow \infty$, we would get $\mathbb{E}_0[\psi_{n_k}] + \sup_{g \in \mathcal{G}_{n_k}} \mathbb{E}_g[1 - \psi_{n_k}] \rightarrow 0$ for suitable plug-in tests ψ_{n_k} based on minimax optimal estimators, contradicting the lower bound in Theorem 3.1 and the fact that $r_{n_k}/r_{n_k}^* \rightarrow 0$.

It remains to prove the upper bound. Throughout the proof $C > 0$ denotes a constant depending only on R, β and p that may change from line to line. Since the case $p = 1$ is covered in Corollary 2.6, we restrict to the case $p > 1$. By the convexity of $y \mapsto y^p$, we have (for non-negative real numbers $a \neq b$ the inequality $(b^p - a^p)/(b - a) \geq \max(a, b)^{p-1}$ holds)

$$|\hat{T} - \|g\|_p| \leq \frac{|\hat{T}^p - \|g\|_p^p|}{\|g\|_p^{p-1}}.$$

Hence,

$$\mathbb{E}_g[|\hat{T} - \|g\|_p|] \leq \frac{\mathbb{E}_g[(\hat{T}^p - \|g\|_p^p)^2]^{1/2}}{\|g\|_p^{p-1}} \leq \frac{\mathbb{E}_g[(\hat{F}_p - \|g\|_p^p)^2]^{1/2}}{\|g\|_p^{p-1}}, \tag{4.1}$$

where we also used the fact that $\hat{T}^p = (\hat{F}_p)_+$ improves \hat{F}_p (see also Remark 2.8). On the other hand, we also have $|\hat{T} - \|g\|_p| \leq |\hat{T}| + \|g\|_p$, which leads to

$$\begin{aligned} \mathbb{E}_g[|\hat{T} - \|g\|_p|] &\leq \mathbb{E}_g[\hat{T}^p]^{1/p} + \|g\|_p \\ &\leq \mathbb{E}_g[|\hat{T}^p - \|g\|_p^p|]^{1/p} + 2\|g\|_p \\ &\leq \mathbb{E}_g[(\hat{F}_p - \|g\|_p^p)^2]^{1/(2p)} + 2\|g\|_p, \end{aligned} \tag{4.2}$$

where we applied the Hölder inequality and the concavity of the function $y \mapsto y^{1/p}$ (for non-negative real numbers $a \neq b$ the inequality $(a + b)^{1/p} \leq a^{1/p} + b^{1/p}$ holds).

If $\|g\|_p \leq n^{-(\beta+1/(2p))/(\beta+1)}$, then by (4.2) and Corollary 2.6 it suffices to show

$$\max(\|g\|_{2p-2}^{2p-2} n^{-\frac{2\beta+1}{\beta+1}}, n^{-\frac{2\beta p+1}{\beta+1}})^{1/(2p)} \leq C n^{-\frac{\beta+1/(2p)}{\beta+1}},$$

which itself follows from $\|g\|_{2p-2} \leq C n^{-\beta/(\beta+1)}$. For $p \leq 2$ the latter holds because of $\|g\|_{2p-2} \leq \|g\|_p \leq n^{-(\beta+1/(2p))/(\beta+1)}$. For $p > 2$ this is implied by Lemma 2.9:

$$\|g\|_{2p-2} \leq C \max(\|g\|_p, \|g\|_p^{(\beta+1/(2p-2))/(\beta+1/p)}) \leq C \|g\|_p^{\beta/(\beta+1/(2p))} \leq C n^{-\beta/(\beta+1)},$$

using first $\|g\|_p \leq 1$ and then $1/(2p - 2) \geq 1/(2p)$.

In the opposite case $\|g\|_p > n^{-(\beta+1/(2p))/(\beta+1)}$ we apply (4.1), Corollary 2.6 and obtain the result if

$$\max(\|g\|_{2p-2}^{p-1} n^{-\frac{\beta+1/2}{\beta+1}}, n^{-\frac{\beta p+1/2}{\beta+1}}) \leq C \|g\|_p^{p-1} n^{-(\beta+1/(2p))/(\beta+1)}.$$

For $p \leq 2$ this follows again by $\|g\|_{2p-2} \leq \|g\|_p$. For $p > 2$ Lemma 2.9 yields the bound

$$\|g\|_{2p-2}^{p-1} \leq C \|g\|_p^{p-1} \max(1, \|g\|_p^{(1/2-(p-1)/p)/(\beta+1/p)}) \leq C \|g\|_p^{p-1} n^{(1/2-1/(2p))/(\beta+1)},$$

using $((p-1)/p - 1/2)(\beta + 1/(2p)) = (1/2 - 1/p)(\beta + 1/(2p)) < (1/2 - 1/(2p))(\beta + 1/p)$. Inserting the bound thus gives the result also for $p > 2$. \square

Remark 4.2. For the problem of estimating g in L^∞ -norm, Drees, Neumeier and Selk [2] established the rate $(n^{-1} \log n)^{\beta/(\beta+1)}$ (in a boundary regression model). This result is then used to analyse goodness-of-fit tests for parametric classes of error distributions.

Remark 4.3. Note that we can consider the minimax risk over the whole Hölder class $\mathcal{C}^\beta(R)$ in the case of estimating the norm $\|g\|_p$. In distinction to Corollary 2.6, the upper bound does not depend on any L^q -norm of g .

Inspecting the proof, we see more precisely that the minimax rate is driven by functions whose L^p -norm is smaller than $n^{-(\beta+1/(2p))/(\beta+1)}$. For functions which have a substantially larger norm we get the rate of convergence $n^{-(\beta+1/2)/(\beta+1)}$ corresponding to a smooth functional. This is explained by the fact that the L^p -norm is a non-smooth functional at $g = 0$.

Remark 4.4. There is a close connection between Theorem 4.1 and Theorem 3.1. First, the upper bound in Theorem 3.1 follows from Theorem 4.1 by using for example, [4], Proposition 2.17. Second, the lower bound in Theorem 4.1 is a consequence of the lower bound in Theorem 3.1.

Appendix: Proof of Lemma 2.9

Let us first show that the general case can be deduced from the special case $q = \infty$ and suppose that

$$\|f\|_\infty \leq C \|f\|_p \max(1, R/\|f\|_p)^{\frac{1/p}{\beta+1/p}} \tag{A.1}$$

holds. Clearly, we have

$$\|f\|_q^q \leq \|f\|_\infty^{q-p} \|f\|_p^p. \tag{A.2}$$

Now, if $\|f\|_p > R$, then (A.1) and (A.2) give $\|f\|_q \leq C^{1-p/q} \|f\|_p$. On the other hand, if $\|f\|_p \leq R$, then (A.1) and (A.2) give

$$\|f\|_q^q \leq C^{q-p} \|f\|_p^q (R/\|f\|_p)^{\frac{(q-p)/p}{\beta+1/p}}$$

and thus

$$\|f\|_q \leq C^{1-p/q} \|f\|_p (R/\|f\|_p)^{\frac{1/p-1/q}{\beta+1/p}}.$$

It remains to prove (A.1). Using the definition of $\mathcal{C}^\beta(R)$, we get

$$\|f\|_p^p = \int_0^1 |f(x)|^p dx \geq \int_0^{\min(1, (\|f\|_\infty/R)^{1/\beta})} (\|f\|_\infty - Rx^\beta)^p dx.$$

Setting $a = \|f\|_\infty$ and $b = (\|f\|_\infty/R)^{1/\beta}$, we obtain

$$\begin{aligned} \int_0^1 |f(x)|^p dx &\geq \int_0^{\min(1,b)} (a - a(x/b)^\beta)^p dx \\ &= a^p \int_0^{\min(1,b)} (1 - (x/b)^\beta)^p dx \\ &\geq a^p \min(1, b) \int_0^1 (1 - y^\beta)^p dy, \end{aligned}$$

where we make the substitution $x = by$ if $b \leq 1$ and use the inequality $1 - (x/b)^\beta \geq 1 - x^\beta$ if $b > 1$. Thus we have proven

$$\|f\|_p \geq \|f\|_\infty \min(1, \|f\|_\infty/R)^{\frac{1}{\beta p}} \|1 - y^\beta\|_p,$$

which gives (A.1).

Acknowledgements

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

References

- [1] Bergh, J. and Löfström, J. (1976). *Interpolation Spaces. An Introduction*. Berlin: Springer. Grundlehren der Mathematischen Wissenschaften, No. 223. [MR0482275](#)
- [2] Drees, H., Neumeyer, N. and Selk, L. Estimation and hypotheses testing in boundary regression models. *Bernoulli*. To appear.
- [3] Ibragimov, I.A., Nemirovskii, A.S. and Khas'minskii, R.Z. (1986). Some problems of nonparametric estimation in Gaussian white noise. *Teor. Veroyatn. Primen.* **31** 451–466. [MR0866866](#)
- [4] Ingster, Yu.I. and Suslina, I.A. (2003). *Nonparametric Goodness-of-Fit Testing Under Gaussian Models. Lecture Notes in Statistics* **169**. New York: Springer. [MR1991446](#)
- [5] Jirak, M., Meister, A. and Reiß, M. (2014). Adaptive function estimation in nonparametric regression with one-sided errors. *Ann. Statist.* **42** 1970–2002. [MR3262474](#)
- [6] Kallenberg, O. (2002). *Foundations of Modern Probability*, 2nd ed. *Probability and Its Applications (New York)*. New York: Springer. [MR1876169](#)
- [7] Karr, A.F. (1991). *Point Processes and Their Statistical Inference*, 2nd ed. *Probability: Pure and Applied* **7**. New York: Dekker. [MR1113698](#)
- [8] Korostelev, A.P. and Tsybakov, A.B. (1993). *Minimax Theory of Image Reconstruction. Lecture Notes in Statistics* **82**. New York: Springer. [MR1226450](#)
- [9] Kutoyants, Yu.A. (1998). *Statistical Inference for Spatial Poisson Processes. Lecture Notes in Statistics* **134**. New York: Springer. [MR1644620](#)
- [10] Last, G. and Penrose, M. (2018). *Lectures on the Poisson Process. Institute of Mathematical Statistics Textbooks* **7**. Cambridge: Cambridge Univ. Press. [MR3791470](#)

- [11] Lepski, O., Nemirovski, A. and Spokoiny, V. (1999). On estimation of the L_r norm of a regression function. *Probab. Theory Related Fields* **113** 221–253. [MR1670867](#)
- [12] Lepski, O.V. and Spokoiny, V.G. (1999). Minimax nonparametric hypothesis testing: The case of an inhomogeneous alternative. *Bernoulli* **5** 333–358. [MR1681702](#)
- [13] Meister, A. and Reiß, M. (2013). Asymptotic equivalence for nonparametric regression with non-regular errors. *Probab. Theory Related Fields* **155** 201–229. [MR3010397](#)
- [14] Nemirovski, A. (2000). Topics in non-parametric statistics. In *Lectures on Probability Theory and Statistics (Saint-Flour, 1998)*. *Lecture Notes in Math.* **1738** 85–277. Berlin: Springer. [MR1775640](#)
- [15] Reiß, M. and Selk, L. (2017). Efficient estimation of functionals in nonparametric boundary models. *Bernoulli* **23** 1022–1055. [MR3606758](#)
- [16] Tsybakov, A.B. (2009). *Introduction to Nonparametric Estimation*. *Springer Series in Statistics*. New York: Springer. [MR2724359](#)

Received August 2017 and revised June 2018