Asymptotic equivalence for nonparametric regression with non-regular errors

Alexander Meister · Markus Reiß

Abstract   Asymptotic equivalence in Le Cam's sense for nonparametric regression experiments is extended to the case of non-regular error densities, which have jump discontinuities at their endpoints. We prove asymptotic equivalence of such regression models and the observation of two independent Poisson point processes which contain the target curve as the support boundary of its intensity function. The intensity of the point processes is of order of the sample size n and involves the jump sizes as well as the design density. The statistical model significantly differs from regression problems with Gaussian or regular errors, which are known to be asymptotically equivalent to Gaussian white noise models.

Keywords   Extreme value statistics · Frontier estimation · Le Cam distance · Le Cam equivalence · Poisson point processes

Mathematics Subject Classification (2010)   62B15 · 62G08 · 62M30

1 Introduction

The goal of transforming nonparametric regression models into asymptotically equivalent statistical experiments, which describe continuous observations of a stochastic process, has stimulated considerable research activity in mathematical statistics.

A. Meister (✉)
Institut für Mathematik, Universität Rostock, Ulmenstraße 69,
18051 Rostock, Germany
e-mail: alexander.meister@uni-rostock.de

M. Reiß
Institut für Mathematik, Humboldt-Universität zu Berlin,
Unter den Linden 6, 10099 Berlin, Germany
e-mail: mreiss@math.hu-berlin.de

The continuous design in these limiting models simplifies the asymptotic analysis and makes statistical procedures more transparent because in the regression case the discrete design points generate distracting approximation errors. Most papers so far establish asymptotic equivalence of certain nonparametric regression models with nonparametric Gaussian shift experiments. In that Gaussian white noise experiment, a process is observed which contains the target function in its drift and a blurring Wiener process which is scaled with a factor of order $n^{-1/2}$, where n denotes the original sample size. The basic equivalence result for standard Gaussian regression with deterministic design has been established by [1]. Afterwards, many important extensions have been achieved. The case of random design for univariate design has been treated by [2]. In [6] the case of unknown error variance and design density is considered; and [5,26] extend the results to the multivariate setting. Recently, the model with dependent regression errors has been investigated in [7]. The work by [11] is the first to consider the important case of non-Gaussian errors which are, however, supposed to be included in an exponential family. Such classes of error distributions are also studied in [3] where the regression error is supposed to be non-additive. General regular distributions for the additive error variables are covered in [12] where only slightly more than standard Hellinger differentiability is required for the error density.

On the other hand, when allowing for jump discontinuities of the error density, the situation changes completely. Standard examples include uniform or exponential error densities. These types of error distributions are non-regular and we know from parametric theory that better rates of convergence and non-Gaussian limit distributions can be expected. The faster convergence rates are attained only by specific estimators, e.g. employing extreme value statistics in their construction instead of local averaging statistics. The Nadaraja–Watson estimator and the local polynomial estimators are procedures of that latter type, which can be improved significantly under non-regular errors. Improved minimax rates for regression functions which satisfy some Hölder condition are established by [24]. In [13] a rigorous theory for the optimal convergence rates is established for nonparametric regression under non-regular errors and smoothness constraints up to regularity one on the target regression function. Their nonparametric minimax rates in dimension one are of the form $n^{-s/(s+1)}$ for Hölder regularity s, which is faster than the usual $n^{-s/(2s+1)}$-rate for regular regression, but slower than $n^{-2s/(2s+1)}$, the squared regular rate in analogy with the parametric rates. At first sight, this is counter-intuitive, but may be explained by a Poisson instead of Gaussian limiting law. Many applications of non-regular regression models occur in the field of econometrics, see [8] for an overview and a precise asymptotic investigation of the parametric likelihood ratio process. Irregular regression problems are also closely related to nonparametric boundary estimation in image reconstruction, see the monograph of [18]. Considerable interest has also found the problem of frontier estimation, see [10] and the references therein.

In [15] weak asymptotic equivalence of the extreme order statistics of a one-dimensional localization problem with non-regular errors and a Poisson point process model is derived in a parametric setup. Other results on the asymptotic theory for non-regular parametric experiments can be found in [14]. Also for the precise asymptotic analysis of regression experiments with non-regular errors the use of Poisson point processes and random measures turn out to be useful, see e.g. [17] for parametric linear

models and [8] for general parametric regression, yet a precise and nonparametric statement lacks. We intend to fill this gap by rigorously proving asymptotic equivalence of nonparametric regression experiments with non-regular errors with a Poisson point process (PPP) model. Therein the target parameter occurs as the boundary curve of the intensity function. Hence, the Gaussian structure of the process experiment is not kept; nor is the scaling factor $n^{-1/2}$ which will be changed into n^{-1} in agreement with the parametric rate. For a comprehensive review on PPP and their statistical inference we refer to [16,19]. They discuss image reconstruction from laser radar as a practical application of support estimation of the intensity function of a PPP, which corresponds to identifying the target parameter in our PPP experiment. The asymptotic equivalence result therefore links interesting inference questions in both models which might prove useful in both directions.

For the basic concept of asymptotic equivalence of statistical experiments we refer to [20,21]. To grasp the impact let us just mention that asymptotic equivalence between two sequences of statistical models transfers asymptotical risk bounds for any inference problem from one model to the other, at least for bounded loss functions. Moreover, asymptotic equivalence remains valid for the sub-experiments obtained by restricting the parameter class so that we shall also cover smoother nonparametric or just parametric regression problems.

The paper is organized as follows. In Sect. 2 we introduce our models, state our main result in Theorem 2.1 and give a constructive description of the equivalence maps. In Sect. 3 we construct pilot estimators of the target functions which will be employed to localize the model in Sects. 4 and 6. The findings of Sect. 5 yield asymptotic equivalence of the PPP experiment and the regression model when the target functions are changed into approximating step functions. In Sect. 7 all the results are combined to complete the proof of Theorem 2.1. Section 8 discusses limitations and extensions of the results and gives a geometric explanation of the unexpected nonparametric minimax rate for Hölder classes.

2 Model and main result

In this section we specify the statistical experiments under consideration. First we define the joint parameter space Θ of both the regression and the PPP experiment, imposing standard smoothness constraints on the target function.

Definition 2.1 For some constants $C_\Theta > 0$ and $\alpha \in (0, 1]$ the parameter set Θ consists of all functions $\vartheta : [0, 1] \to \mathbb{R}$ which are twice continuously differentiable on $[0, 1]$ with $\|\vartheta\|_\infty \leq C_\Theta$ and $\|\vartheta''\|_\infty \leq C_\Theta$ and where the second derivative satisfies the Hölder condition

$$|\vartheta''(x) - \vartheta''(y)| \leq C_\Theta |x - y|^\alpha, \quad \forall x, y \in [0, 1].$$

In the regression model Θ represents the collection of all admitted regression functions. This parameter space will remain unchanged for all experiments considered here.

Definition 2.2 We define the statistical experiment $\mathcal{A}_n$ in which the data $Y_{j,n}$, $j = 1, \ldots, n$, with

$$Y_{j,n} = \vartheta(x_{j,n}) + \varepsilon_{j,n} \tag{2.1}$$

are observed. The deterministic design points $x_{1,n}, \ldots, x_{n,n} \in [0, 1]$ are assumed to satisfy

$$x_{j,n} = F_D^{-1}((j-1)/(n-1)), \tag{2.2}$$

where the distribution function $F_D : [0, 1] \to [0, 1]$ possesses a Lipschitz continuous Lebesgue density f_D which is uniformly bounded away from zero.

The conditions on the design are adopted from [1]. They imply that

$$d^{-1}/n \leq x_{j+1,n} - x_{j,n} \leq d/n, \tag{2.3}$$

for all $n \in \mathbb{N}$, $j = 1, \ldots, n$ and a finite positive constant d.

The error model describes the class of densities which are supported on $[-1, 1]$, regular within $(-1, 1)$ and which have jumps at their left and right endpoints. Note that by constant extrapolation the density f_ε on $[-1, 1]$ can always be written as

$$f_\varepsilon(x) = 1_{[-1,1]}(x) \cdot \varphi(x),$$

with a strictly positive Lipschitz continuous function $\varphi : \mathbb{R} \to \mathbb{R}$ satisfying for some constant $C_\varepsilon > 0$

$$\sup_{t \neq s} \frac{|\varphi(t) - \varphi(s)|}{|t - s|} + \sup_t |\varphi(t)| \leq C_\varepsilon. \tag{2.4}$$

Instead of constant extrapolation, φ may alternatively be continued such that $\varphi \in L_1(\mathbb{R})$ holds in addition.

Hence, experiment $\mathcal{A}_n$ describes a non-regular nonparametric regression model. We believe that the regularity condition on f_ε in the interior $(-1, 1)$ can be substantially relaxed, but at the cost of more involved estimation techniques. We have restricted our consideration to the specific interval $[-1, 1]$ for convenience.

In the PPP model the target function ϑ occurs as upper and lower boundary curves of the intensity functions of two independent Poisson point processes X_1 and X_2.

Definition 2.3 For functions $\vartheta \in \Theta$, the design density f_D and the noise density f_ε from above we define the experiment $\mathcal{B}_n$ in which we observe two independent Poisson point processes X_j, $j = 1, 2$, on the rectangle $S = [0, 1] \times [-C_\Theta - 1, C_\Theta + 1] \subset \mathbb{R}^2$ with respective intensity functions

$$\begin{aligned}
\lambda_1(x, y) &= f_D(x) \cdot 1_{[-C_\Theta - 1, \vartheta(x)]}(y) \cdot n f_\varepsilon(1), \\
\lambda_2(x, y) &= f_D(x) \cdot 1_{[\vartheta(x), C_\Theta + 1]}(y) \cdot n f_\varepsilon(-1),
\end{aligned} \tag{2.5}$$
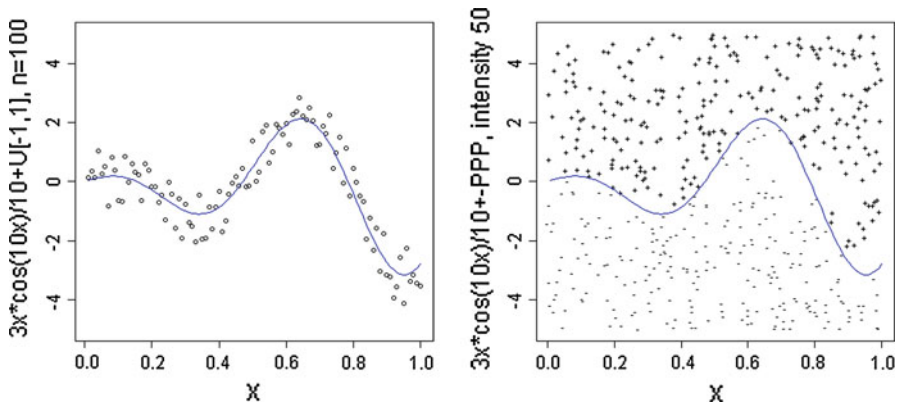
for all $(x, y) \in S$.

Fig. 1 *Left* Regression model $\mathcal{A}_n$ with uniform $U[-1, 1]$ errors. *Right* Asymptotically equivalent Poisson point process model $\mathcal{B}_n$

Each realisation X_j represents a measure mapping from the Borel subsets of S to $\mathbb{N} \cup \{0\}$. Equivalently, $X_j(\cdot)/X_j(S)$ may be characterized by a two-dimensional discrete probability distribution, see [16,19] for more details on PPP.

Figure 1 shows on the left the regression function $\vartheta(x) = \frac{3}{10}x\cos(10x)$ and corresponding $n = 100$ equidistant observations of $\mathcal{A}_n$ corrupted by uniform noise on $[-1, 1]$. A realisation of the asymptotically equivalent PPP model $\mathcal{B}_n$ is shown on the right, with '+', '−' indicating point masses of X_2 and X_1, respectively.

We may conceive X_j as the random point measure $\sum_{k=1}^{N_j} \delta_{(x_k^j, y_k^j)}$ where N_j is drawn from a Poisson-distribution with intensity $\|\lambda_j\|_{L^1(S)}$ and the (x_k^j, y_k^j) are drawn according to the bivariate density $\lambda_j/\|\lambda_j\|_{L^1(S)}$. The vertical bounds $\pm(C_\Theta+1)$ for the domain S are non-informative for $\vartheta \in \Theta$, but the boundedness avoids technicalities. The equivalent unbounded PPP can be described by infinite random point measures $\sum_{k=1}^{\infty} \delta_{(x_k^j, y_k^j)}$ where the x_k^j are drawn according to the density f_D and

$$y_k^1 = \vartheta\left(x_k^1\right) - (nf_\varepsilon(1))^{-1}\sum_{l=1}^{k} z_l^1, \quad y_k^2 = \vartheta(x_k^2) + (nf_\varepsilon(-1))^{-1}\sum_{l=1}^{k} z_l^2$$

holds with exponentially distributed (z_k^j) of mean one (all independent). In this form, the PPP already appears in [17], yielding the limiting law for parametric estimators in the nonregular linear model.

We present the main result of this work in the following theorem.

Theorem 2.1 *The statistical experiments $\mathcal{A}_n$ and $\mathcal{B}_n$ are asymptotically equivalent in Le Cam's sense as $n \to \infty$.*

This asymptotic equivalence is achieved constructively by consecutive invertible (in law) and parameter-independent mappings of the data, which generate new experiments where the observation laws are shown to be asymptotically close (uniformly over ϑ in total variation norm). In order to highlight the main ideas in the subsequent

proof and to indicate how to use our theoretical result in practice, let us give an algo-
rithmic description of these equivalence mappings leading from experiment $\mathcal{A}_n$ to
experiment $\mathcal{B}_n$ (in the version with unbounded domain).

1. Take the data $Y_{j,n}$, $j = 1, \ldots, n$, from experiment $\mathcal{A}_n$.
2. Split the data and bin one part: consider the odd indices $J_n := \{1, 3, \ldots,$
   $2\lceil n/2 \rceil - 1\}$ and intervals $I_k = [k/m, (k+1)/m)$ with some appropriate m.
   Put $\mathbf{X}_1 = (Y_{j+1,n})_{j \in J_n \setminus \{n\}}$ and $\bar{\mathbf{Z}} = (\bar{Z}_j)_{j \in J_n}$ with

$$\bar{Z}_j = Y_{j,n} - \hat{\vartheta}_1(\xi_j) - \hat{\vartheta}_1'(\xi_j)(x_j - \xi_j), \quad j \in J_n,$$

   where ξ_j is the centre of that interval I_k with $x_{j,n} \in I_k$ and where $\hat{\vartheta}_1$ is a (good)
   estimator of ϑ based on the data $\mathbf{X}_1$.
3. Consider the local extremes in $\bar{\mathbf{Z}}$, i.e. $s_k = \min\{\bar{Z}_j : x_j \in I_k\}$, $S_k = \max\{\bar{Z}_j : x_j \in I_k\}$, $k = 0, \ldots, m - 1$.
4. Use $\hat{\vartheta}$ on the data $\mathbf{X}_1$ again to transform $s_k'' = s_k + \hat{\vartheta}_1(\tilde{\xi}_k) + 1$, $S_k'' = S_k + \hat{\vartheta}_1(\tilde{\xi}_k) - 1$
   where $\tilde{\xi}_k$ denotes the centre of the interval I_k.
5. Randomization to build PPP X_l, X_u: on each interval I_k generate (x_k^l, y_k^l) with
   x_k^l having the density $f_k = f_D 1_{I_k} / \int_{I_k} f_D$ independent of everything else and
   $y_k^l = S_k'' - \hat{\vartheta}_1'(x_k^l)(\tilde{\xi}_k - x_k^l)$; define the PPP X_l where independently on each I_k we
   observe a point measure in (x_k^l, y_k^l) plus independently (conditionally on S_k'', $\hat{\vartheta}_1'$)
   a PPP with intensity

$$\frac{n}{2} f_\varepsilon(1) \left(m \int_{I_k} f_D \right) \mathbf{1}\{x \in I_k, \ y \leq S_k'' - \hat{\vartheta}_1'(x)(\tilde{\xi}_k - x)\};$$

   analogously generate x_k^u with the density f_k independently, $y_k^u = s_k'' - \hat{\vartheta}_1'(x_k^u)(\tilde{\xi}_k - x_k^u)$ and use the intensity

$$\frac{n}{2} f_\varepsilon(-1) \left(m \int_{I_k} f_D \right) \mathbf{1}\{x \in I_k, \ y \geq s_k'' - \hat{\vartheta}_1'(x)(\tilde{\xi}_k - x)\}$$

   to build X_u independently conditionally on s_k'', $\hat{\vartheta}_1'$.
6. Use a (good) estimator $\hat{\vartheta}_2$ based on the PPP data $\mathbf{X}_2 = (X_l, X_u)$ and redo steps (2)–
   (5) to transform $\mathbf{X}_1$ via $\bar{Z}_{j+1} = Y_{j+1,n} - \hat{\vartheta}_2(\xi_{j+1}) - \hat{\vartheta}_2'(\xi_{j+1})(x_{j+1} - \xi_{j+1})$, $j \in J_n$, to another couple (X_l', X_u') of PPP; the final PPP are obtained by $X_1 = X_l + X_l'$, $X_2 = X_u + X_u'$.

   In this algorithmic description we could do without substracting and adding the
pilot estimator itself (i.e., only use the derivative) in steps (2) and (4), but in the proof
this localization permits an easy sufficiency argument for the local extremes. Put in a
nutshell, the asymptotic equivalence is achieved by considering block-wise extreme

values in the regression experiment, in conjunction with a pre- and post-processing procedure (localization step) performing a linear correction on each block. The easier block-wise constant approximation approach by [1] does not work here since we need a much higher approximation order.

Throughout we shall write const. for a generic positive constant which may change its value from line to line and does not depend on the parameter ϑ nor on the sample size n. Similarly, the Landau symbols O, o and the asymptotic order symbol $\asymp$ will denote uniform bounds with respect to ϑ and n.

3 Pilot estimators

In order to prove Theorem 2.1 a localization strategy is required as in [25] for the density estimation problem. To that end we construct pilot estimators of the target function ϑ and its derivative in both, experiments $\mathcal{A}_n$ and $\mathcal{B}_n$.

Let us fix the estimation point $x_0 \in [0, 1]$ and apply a local polynomial estimation approach. We introduce the neighbourhood $U_h = [x_0 - h, x_0 + h]$ for $x_0 \in [h, 1 - h]$ and the one-sided analogue $U_h = [0, 2h]$ for $x_0 \in [0, h)$, $U_h = [1 - 2h, 1]$ for $x_0 \in (1 - h, 1]$. We introduce the set $\Pi := \Pi_2(U_h)$ of quadratic polynomials on U_h. Standard approximation theory (by a Taylor series argument) gives for $h \downarrow 0$

$$\gamma_h := \sup_{\vartheta \in \Theta} \min_{p \in \Pi} \max_{x \in U_h} (h^{-(2+\alpha)} |\vartheta(x) - p(x)| + h^{-(1+\alpha)} |\vartheta'(x) - p'(x)|) \leq \text{const.} < \infty,$$

where the constant does not depend on h.

Definition 3.1 We call $\hat{\vartheta} \in \Pi$ in experiment $\mathcal{A}_n$ locally admissible at x_0 if

$$\max_{j : x_{j,n} \in U_h} |Y_{j,n} - \hat{\vartheta}(x_{j,n})| \leq 1 + \gamma_h h^{2+\alpha}$$

holds. Similarly, in experiment $\mathcal{B}_n$ we call $\hat{\vartheta} \in \Pi$ locally admissible at x_0 if

$$X_1(\{x \in U_h, \ y > \hat{\vartheta}(x) + \gamma_h h^{2+\alpha}\}) = 0 \quad \text{and} \quad X_2(\{x \in U_h, \ y < \hat{\vartheta}(x) - \gamma_h h^{2+\alpha}\}) = 0$$

hold. Our estimator $\hat{\vartheta}_{n,h}(x_0)$ is just any locally admissible $\hat{\vartheta}_{n,h} \in \Pi$, evaluated at x_0 and selected as a measurable function of the data (by the measurable selection theorem).

Note that the by γ_h enlarged band size guarantees that $\hat{\vartheta}_{n,h}$ exists since the minimizer $\vartheta_h \in \Pi$ in the definition of γ_h is eligible. The following result gives the pointwise risk bounds for the regression function and its derivative with orders $O(n^{-s/(s+1)})$ and $O(n^{-(s-1)/(s+1)})$, respectively, where $s = 2 + \alpha$ denotes the regularity in a Hölder class. As an application of our asymptotic equivalence we shall show in Sect. 8.2 below the optimality of these rates in a minimax sense. The upper bound proof relies on entropy arguments and norm equivalences for polynomials and could be easily extended to more general local polynomial estimation and L^p-loss functions.

Proposition 3.1 *Select the bandwidth h such that $h \asymp n^{-1/(3+\alpha)}$. Then we have in experiment $\mathcal{A}_n$ as well as in experiment $\mathcal{B}_n$*

$$\sup_{\vartheta \in \Theta} \sup_{x_0 \in [0,1]} E_\vartheta (n^{2(2+\alpha)/(3+\alpha)}|\hat{\vartheta}_{n,h}(x_0) - \vartheta(x_0)|^2 + n^{2(1+\alpha)/(3+\alpha)}|\hat{\vartheta}'_{n,h}(x_0) - \vartheta'(x_0)|^2)$$
$$\leq const.$$

Proof of Proposition 3.1 We shall need the following bounds in $\Pi = \Pi_2(U_h)$ from [9]: $\|p\|_{L^\infty(U_h)} \leq 8h^{-1}\|p\|_{L^1(U_h)}$ (their Theorem IV.2.6); $\|p'\|_{L^\infty(U_h)} \leq c_0 h^{-1}\|p\|_{L^\infty(U_h)}$ (their Thm. IV.2.7); their proof of Thm. IV.2.6 establishes $|p(x)| \geq (1 - 4(x - x_M)/h)\|p\|_\infty$ for $x_M := \text{argmax}_{x \in U_h}|p(x)|$ and $x_M \leq x < x_M + h/4$, assuming without loss of generality that x_M lies in the left half of U_h, such that uniformly over x_0

$$\|p\|_{n,h,1} := \frac{1}{nh} \sum_{x_{j,n} \in U_h} |p(x_{j,n})| \geq \text{const.} \cdot |p(x_M)| = \text{const.} \cdot \|p\|_{L^\infty(U_h)}$$

is derived.

Let us start with considering the regression experiment $\mathcal{A}_n$. We apply a standard chaining argument in the finite-dimensional space Π together with an approximation argument. From above we have $\|p\|_{L^\infty(U_h)}/\|p\|_{n,h,1} \asymp 1$ as well as $\|p\|_{n,h,1} \geq c_1|p(x_0)|$ with some $c_1 > 0$ uniformly in $p \in \Pi$. Fix $R > 2$. For every $\delta > 0$ we can find elements $(p_l)_{l \geq 1}$ that form a δ-net in $\Pi \cap \{\|p\|_{n,h,1} \geq c_1 \max(1, c_0)(R-1)\gamma_h h^{2+\alpha}\}$ with respect to the $L^\infty(U_h)$-norm satisfying $\|p_l\|_{n,h,1} \asymp \delta l^{1/3}$ as $l \to \infty$; for this note that, by the above norm equivalences, $\Pi \cap \{\|p\|_{n,h,1} \geq c_1 \max(1, c_0)(R - 1)\gamma_h h^{2+\alpha}\}$ with maximum norm is isometric to $\mathbb{R}^3 \cap \{|x| \geq c_1 \max(1, c_0)(R - 1)\gamma_h h^{2+\alpha}\}$ with the Euclidean metric uniformly for $h \to 0$ and $nh \to \infty$ and use standard coverings of Euclidean balls, e.g. Lemma 2.5 in [29]. We obtain

$$P_\vartheta \left(\exists p \in \Pi : \max_{j:x_{j,n} \in U_h} |Y_{j,n} - p(x_{j,n})| \leq 1 + \gamma_h h^{2+\alpha}, \right.$$
$$\left. \max(h^{-(2+\alpha)}|p(x_0) - \vartheta(x_0)|, h^{-(1+\alpha)}|p'(x_0) - \vartheta'(x_0)|) \geq R\gamma_h \right)$$

$$= P_\vartheta \left(\exists p \in \Pi : \max_{j:x_{j,n} \in U_h} |\varepsilon_{j,n} - (p(x_{j,n}) - \vartheta(x_{j,n}))| \leq 1 + \gamma_h h^{2+\alpha}, \right.$$
$$\left. \max(h^{-(2+\alpha)}|p(x_0) - \vartheta(x_0)|, h^{-(1+\alpha)}|p'(x_0) - \vartheta'(x_0)|) \geq R\gamma_h \right)$$

$$\leq P_\vartheta \left(\exists p \in \Pi : \max_{j:x_{j,n} \in U_h} |\varepsilon_{j,n} - (p(x_{j,n}) - \vartheta_h(x_{j,n}))| \leq 1 + 2\gamma_h h^{2+\alpha}, \right.$$
$$\left. \|p - \vartheta_h\|_{n,h,1} \geq \max(1, c_0)c_1(R - 1)\gamma_h h^{2+\alpha} \right)$$

$$\leq P_\vartheta \left(\exists l \geq 1 : \max_{j:x_{j,n} \in U_h} |\varepsilon_{j,n} - p_l(x_{j,n})| \leq 1 + 2\gamma_h h^{2+\alpha} + \delta \right)$$

$$\leq \sum_{l \geq 1} P_\vartheta \left(\max_{j:x_{j,n} \in U_h} |\varepsilon_{j,n} - p_l(x_{j,n})| \leq 1 + 2\gamma_h h^{2+\alpha} + \delta \right).$$

From $f_\varepsilon(-1) > 0$, $f_\varepsilon(+1) > 0$ and the Lipschitz continuity of f_ε within $[-1, 1]$ we infer that any $\varepsilon_{j,n}$ satisfies

$$\min(P(\varepsilon_{j,n} \geq 1 - \kappa), P(\varepsilon_{j,n} \leq -1 + \kappa)) \geq c\kappa$$

for some constant $c > 0$ and all $\kappa \in (0, 1)$. We derive an exponential inequality for any $f : U_h \to \mathbb{R}$ and $\Delta > 0$:

$$P \left(\max_{j:x_{j,n} \in U_h} |\varepsilon_{j,n} - f(x_{j,n})| \leq 1 + \Delta \right)$$

$$\leq \prod_{j:x_{j,n} \in U_h} \left(1 - \min \left(P(\varepsilon_{j,n} > 1 + \Delta - |f(x_{j,n})|), P(\varepsilon_{j,n} < -1 - \Delta + |f(x_i)|) \right) \right)$$

$$\leq \exp \left(\sum_{j:x_{j,n} \in U_h} \log(1 - c(|f(x_i)| - \Delta)_+) \right)$$

$$\leq \exp \left(-c \sum_{j:x_{j,n} \in U_h} (|f(x_i)| - \Delta)_+ \right)$$

$$\leq \exp(-cnh(\|f\|_{n,h,1} - \Delta)),$$

using $\log(1 + h) \leq h$. We therefore choose $\delta = R\gamma_h h^{2+\alpha}$ and arrive at

$$P_\vartheta (\exists p \in \Pi : p \text{ is locally admissible},$$
$$\max(h^{-(2+\alpha)}|p(x_0) - \vartheta(x_0)|, h^{-(1+\alpha)}|p'(x_0) - \vartheta'(x_0)|) \geq R\gamma_h)$$
$$\leq \sum_{l \geq 1} \exp(-\text{const.} \cdot nh(\delta + \gamma_h h^{2+\alpha})l^{1/3}) = O(\exp(-\text{const.} \cdot Rnh^{3+\alpha})).$$

We conclude, substituting $h \asymp n^{-1/(3+\alpha)}$, that uniformly over $R \geq 2$

$$P_\vartheta (h^{-(2+\alpha)}|\hat{\vartheta}_{n,h}(x_0) - \vartheta(x_0)| \geq R\gamma_h) = O(\exp(-\text{const.} \cdot R)),$$
$$P_\vartheta (h^{-(1+\alpha)}|\hat{\vartheta}'_{n,h}(x_0) - \vartheta'(x_0)| \geq R\gamma_h) = O(\exp(-\text{const.} \cdot R)).$$

Integrating out these exponential tail bounds yields the desired moment bound in experiment $\mathcal{A}_n$.

All the results obtained so far remain valid for the PPP experiment $\mathcal{B}_n$ when the empirical norm $\|\cdot\|_{n,h,1}$ is replaced by the rescaled $L_1(U_h)$-norm $\|g\|_{1,U_h} := \frac{1}{h} \int_{U_h} |g|$,

the admissibility conditions are exchanged and the following (easier) exponential inequality is used:

$$P_\vartheta(X_1(\{x \in U_h, \ y > \vartheta(x)+f(x)-\Delta\})=0, \ X_2(\{x \in U_h, \ y < \vartheta(x)+f(x)+\Delta\})=0)$$
$$= P_0(X_1(\{x \in U_h, \ y > f(x) - \Delta\}) = 0) P_0(X_2(\{x \in U_h, \ y < f(x) + \Delta\}) = 0)$$

$$= \exp\left(-nf_\varepsilon(1) \int_{U_h} (f(x) - \Delta)_+ f_D(x)\, dx\right)$$

$$\cdot \exp\left(-nf_\varepsilon(-1) \int_{U_h} (-f(x) - \Delta)_+ f_D(x)\, dx\right)$$

$$\leq \exp(-c'nh(\|f\|_{1,U_h} - \Delta))$$

with some constant $c' > 0$. $\qquad\square$

4 Design adjustment for the regression experiment

We use a piecewise constant approximation strategy and introduce the intervals

$$I_{k,n} = [k/m, (k + 1)/m), \quad k = 0, \ldots, m - 2, \ \text{and} \ I_{m-1,n} = [(m - 1)/m, 1] \tag{4.1}$$

for some integer m. For any design point $x_{j,n} \in I_{k,n}$ we introduce the centre of the interval

$$\xi_{j,n} := (k + 1/2)/m \quad \text{for } x_{j,n} \in I_{k,n}. \tag{4.2}$$

Now we apply a sample splitting scheme and write J_n for the collection of odd $j \in \{1, \ldots, n\}$. The experiment $\mathcal{A}_n$ is considered as the totality of the two independent data sets $\mathbf{X} = (Y_{j+1,n})_{j \in J_n, j < n}$ and $\mathbf{Y}' = (Y_{j,n})_{j \in J_n}$.

Subsequently, we shall not touch upon $\mathbf{X}$ to establish asymptotic equivalence, but just assume the existence of sufficiently good estimators based on the data $\mathbf{X}$. Therefore, we forget about the specific definition of $\mathbf{X}$ and write $\mathbf{X}^*$ instead.

Definition 4.1 Let $\mathbf{X}^*$ be an arbitrary observation in a Polish space, which is independent of $\mathbf{Y}'$. We generalize the experiment $\mathcal{A}_n$ to $\mathcal{A}_n^*$, which consists of the data $\mathbf{Y}'$ and $\mathbf{X}^*$.

The original experiment $\mathcal{A}_n$ is still included by putting $\mathbf{X}^* = \mathbf{X}$. This enables us to repeatedly use the following results later also when $\mathbf{X}^*$ will denote a PPP observation.

In a first step we show asymptotic equivalence for the regression experiment $\mathcal{A}_n^*$ with the same experiment, but where for $j \in J_n$ the regression function is observed at the interval centres $\xi_{j,n}$.

Definition 4.2 In experiment $\mathcal{C}_n$ we observe independently the vectors $\mathbf{X}^*$ as under experiment $\mathcal{A}_n{}^*$ and, independently, the vector $\mathbf{Z}$ with the components

$$Z_{j,n} = \vartheta(\xi_{j,n}) + \varepsilon_{j,n}, \quad j \in J_n.$$

Lemma 4.1 *Choose* $m \in \mathbb{N}$ *such that* $m^{-1} = o(n^{-1/2})$ *holds and assume that an estimator* $\hat{\vartheta}'$ *can be constructed based on the data set* $\mathbf{X}^*$ *with*

$$\sup_{\vartheta \in \Theta} \sup_{x \in [0,1]} E_\vartheta |\hat{\vartheta}'(x) - \vartheta'(x)| = o(mn^{-1}).$$

Then the experiments $\mathcal{A}_n{}^*$ *and* $\mathcal{C}_n$ *are asymptotically equivalent.*

Proof of Lemma 4.1 The observations $\mathbf{Y}'$ from the experiment $\mathcal{A}_n{}^*$ are transformed into the data set $\tilde{\mathbf{Y}}$ with the components

$$\begin{aligned}
\tilde{Y}_{j,n} &= Y_{j,n} - \hat{\vartheta}'(\xi_{j,n})(x_{j,n} - \xi_{j,n}) \\
&= \vartheta(x_{j,n}) - \vartheta'(\xi_{j,n})(x_{j,n} - \xi_{j,n}) - [\hat{\vartheta}'(\xi_{j,n}) - \vartheta'(\xi_{j,n})](x_{j,n} - \xi_{j,n}) + \varepsilon_{j,n},
\end{aligned}$$

for all $j \in J_n$. The data set $\mathbf{X}^*$ is not affected by this transformation. As $\hat{\vartheta}'$ is based on the data $\mathbf{X}^*$, this transformation is invertible so that the original data are uniquely reconstructable from the transformed ones; and observing $(\mathbf{X}^*, \mathbf{Y}')$ on the one hand and $(\mathbf{X}^*, \tilde{\mathbf{Y}})$ on the other hand is equivalent. Therefore, for any measurable functional R with $\|R\|_\infty \leq 1$ we observe that

$$|E_\vartheta R(\mathbf{X}^*, \tilde{\mathbf{Y}}) - E_\vartheta R(\mathbf{X}^*, \mathbf{Z})| \leq E_\vartheta |E_\vartheta\{R(\mathbf{X}^*, \tilde{\mathbf{Y}})|\mathbf{X}^*\} - E_\vartheta\{R(\mathbf{X}^*, \mathbf{Z})|\mathbf{X}^*\}| \tag{4.3}$$

$$\leq \sum_{j \in J_n} E_\vartheta \|f_{\tilde{Y}_{j,n}|X^*} - f_{Z_{j,n}|X^*}\|_1, \tag{4.4}$$

where $\|\cdot\|_1$ denotes the $L_1(\mathbb{R})$-norm; in general, $f_{Y|X}$ stands for the conditional density of Y given X. The conditional independence of the $\tilde{Y}_{j,n}$ and the $Z_{j,n}$ given $\mathbf{X}^*$ as well as an elementary telescopic sum argument with respect to the $L_1(\mathbb{R})$-distance of the multivariate conditional densities of $\tilde{\mathbf{Y}}$ and $\mathbf{Z}$ given $\mathbf{X}^*$ have been exploited. We obtain by the Lipschitz continuity of φ

$$\|f_{\tilde{Y}_{j,n}|X^*} - f_{Z_{j,n}|X^*}\|_1 \leq 2\|\varphi\|_\infty \cdot |\Delta_{1,j,n}| + \int_{-1}^{1} |\varphi(x + \Delta_{1,j,n}) - \varphi(x)|dx$$

$$\leq 4C_\varepsilon \cdot |\Delta_{1,j,n}|, \tag{4.5}$$

where

$$\Delta_{1,j,n} = \vartheta(x_{j,n}) - \vartheta(\xi_{j,n}) - \vartheta'(\xi_{j,n})(x_{j,n} - \xi_{j,n}) - [\hat{\vartheta}'(\xi_{j,n}) - \vartheta'(\xi_{j,n})](x_{j,n} - \xi_{j,n}).$$

We conclude that the total variation distance between $(\mathbf{X}^*, \tilde{\mathbf{Y}})$ and $(\mathbf{X}^*, \mathbf{Z})$ is bounded from above by

$$\text{const.} \cdot \sum_{j \in J_n} E_\vartheta (|\Delta_{1,j,n}|).$$

By the Hölder constraints imposed on the parameter class Θ we derive that

$$|\Delta_{1,j,n}| \leq \text{const.} \cdot (m^{-2} + |\hat{\vartheta}'(x_{j,n}) - \vartheta'(x_{j,n})| m^{-1}).$$

Using $m^{-2} = o(n^{-1})$ and the convergence rate of $\hat{\vartheta}'$, we conclude that the Le Cam distance between the experiments $\mathcal{A}_n^*$ and $\mathcal{C}_n$ tends to zero uniformly in ϑ, which gives the assertion of the lemma.   □

Usually, the bound on the total variation of product measures which is used in the proof is suboptimal, but here the order is optimal due to the jumps of the density. Note also that the data $Z_{j,n}$ may be viewed as random responses drawn from a regression function which is locally constant on the intervals $I_{k,n}$ with the values $\vartheta(\xi_{j,n})$ when $x_{j,n} \in I_{k,n}$.

5 Asymptotic equivalence for step functions

We revisit the experiment $\mathcal{C}_n$ from Definition 4.2. The data $Z_{j,n}$ may be transformed into

$$\tilde{Z}_{j,n} = Z_{j,n} - \hat{\vartheta}(\xi_{j,n}),$$

where $\hat{\vartheta}$ denotes a preliminary estimator of ϑ which is based on the data from $\mathbf{X}^*$ as contained in the experiment $\mathcal{C}_n$. Again this transformation is invertible so that the experiment $\mathcal{C}_n$ is equivalent to the experiment $\mathcal{C}_n{}'$ under which one observes the data $\mathbf{X}^*$ and the vector $\tilde{\mathbf{Z}} = (\tilde{Z}_{j,n})_{j \in J_n}$. The $\tilde{Z}_{j,n}$, $j \in J_n$, are conditionally independent given $\mathbf{X}^*$ and have the conditional densities

$$f_\varepsilon(x - \Delta_{0,j,n}) = \varphi(x - \Delta_{0,j,n}) 1_{[\Delta_{0,j,n}-1, \Delta_{0,j,n}+1]}(x) \quad \text{with } \Delta_{0,j,n}$$
$$= \vartheta(\xi_{j,n}) - \hat{\vartheta}(\xi_{j,n}). \tag{5.1}$$

The next key step is to replace these densities by those with unshifted φ where local minima and maxima will turn out to be sufficient statistics.

Definition 5.1 Let $W_{j,n}$, $j \in J_n$, conditionally on $\mathbf{X}^*$ be independent random variables with respective densities

$$f_{W,j}(x) = \varphi(x) \left(\int_{\Delta_{0,j,n}-1}^{\Delta_{0,j,n}+1} \varphi(t)dt \right)^{-1} 1_{[\Delta_{0,j,n}-1, \Delta_{0,j,n}+1]}(x), \quad j \in J_n,$$

where $\Delta_{0,j,n}$ is given in (5.1). The experiment in which $\mathbf{X}^*$ and the $W_{j,n}$, $j \in J_n$, are observed for $\vartheta \in \Theta$ is denoted by $\mathcal{D}_n$.

Lemma 5.1 *Suppose that an estimator $\hat{\vartheta}$ of ϑ can be constructed based on the data set $\mathbf{X}^*$ such that*

$$\sup_{\vartheta \in \Theta} \sup_{x \in [0,1]} E_\vartheta |\hat{\vartheta}(x) - \vartheta(x)|^2 = O(n^{-1-\delta}), \tag{5.2}$$

for some $\delta > 0$. Then the experiments $\mathcal{C}_n$ and $\mathcal{D}_n$ are asymptotically equivalent.

Proof of Lemma 5.1 By Le Cam's inequality and the subadditivity of the squared Hellinger distance H for product measures (cf. Section 2.4 in [28]) we deduce that for any measurable functional R with $\|R\|_\infty \leq 1$ we have

$$|E_\vartheta R(\mathbf{X}^*, \tilde{\mathbf{Z}}) - E_\vartheta R(\mathbf{X}^*, \mathbf{W})| \leq E_\vartheta \int \cdots \int \left| \prod_{j \in J_n} f_\varepsilon(y_j - \Delta_{0,j,n}) - \prod_{j \in J_n} f_{W,j}(y_j) \right| d\mathbf{y}$$

$$\leq 2 \sum_{j \in J_n} E_\vartheta H^2(f_{W,j}, f_\varepsilon(\cdot - \Delta_{0,j,n})), \tag{5.3}$$

where the expectation is taken over $\Delta_{0,j,n}$. Hence, it remains to be shown that the sum converges to zero uniformly with respect to $\vartheta \in \Theta$. That sum equals

$$\sum_{j \in J_n} E_\vartheta \int_{\Delta_{0,j,n}-1}^{\Delta_{0,j,n}+1} \left(\sqrt{\varphi(x)} \left(\int_{\Delta_{0,j,n}-1}^{\Delta_{0,j,n}+1} \varphi(t) dt \right)^{-1/2} - \sqrt{\varphi(x - \Delta_{0,j,n})} \right)^2 dx$$

$$\leq 4 C_\varepsilon^2 \left(2 + \left\{ \inf_{|x| \leq 1} \varphi(x) \right\}^{-1} \right) \sum_{j \in J_n} E_\vartheta \Delta_{0,j,n}^2,$$

since φ is strictly positive, continuous and satisfies the condition (2.4). The imposed convergence rate of the estimator $\hat{\vartheta}$ yields that the supremum taken over $\vartheta \in \Theta$ tends to zero at the rate $O(n^{-\delta})$ and the proof is complete. □

The conditional joint density of the $W_{j,n}$, $j \in J_n$, given $\mathbf{X}^*$ from the experiment $\mathcal{D}_n$ can be represented by

$$f_W(\mathbf{w}) = \prod_{j \in J_n} f_{W,j}(w_j) = \left(\prod_{j \in J_n} \varphi(w_j) \right) \left(\prod_{j \in J_n} \int_{\Delta_{0,j,n}-1}^{\Delta_{0,j,n}+1} \varphi(t) dt \right)^{-1}$$

$$\cdot \left(\prod_{k=0}^{m-1} 1(\min\{w_j : x_{j,n} \in I_{k,n}\} \geq \Delta_{0,j(k),n} - 1) \cdot 1(\max\{w_j : x_{j,n} \in I_{k,n}\} \right.$$

$$\left. \leq \Delta_{0,j(k),n} + 1) \right), \tag{5.4}$$

where the $I_{k,n}$ are as in Sect. 4 and $j(k) = \min\{l \in J_n : x_{l,n} \in I_{k,n}\}$, $\mathbf{w} = (w_j)_{j \in J_n}$. Note that the parameter ϑ is included in the term $\Delta_{0,j(k),n}$.

Definition 5.2 In experiment $\mathcal{E}_n$ only the data $(\mathbf{X}^*, s_{k,n}, S_{k,n})$, $k = 0, \ldots, m-1$, with

$$s_{k,n} = \min\{W_{j,n} : x_{j,n} \in I_{k,n}\},$$
$$S_{k,n} = \max\{W_{j,n} : x_{j,n} \in I_{k,n}\},$$

are observed for $\vartheta \in \Theta$.

An inspection of (5.4) yields that $(\mathbf{X}^*, s_{k,n}, S_{k,n})$, $k = 0, \ldots, m-1$, provides a sufficient statistic for the whole empirical information contained in $(\mathbf{X}^*, \{W_{j,n} : j \in J_n\})$ by the Fisher–Neyman factorization theorem. Sufficiency implies equivalence (e.g. Lemma 3.2 in [1]) and we have

Lemma 5.2 *Experiments $\mathcal{D}_n$ and $\mathcal{E}_n$ are equivalent.*

In the following we study the conditional distribution of $(s_{k,n}, S_{k,n})$ given $\mathbf{X}^*$. Note that, conditionally on $\mathbf{X}^*$, the $(s_{k,n}, S_{k,n})$ are independent for $k = 0, \ldots, m-1$ as the intervals $I_{k,n}$ are disjoint. We derive that

$$P[s_{k,n} > x, \ S_{k,n} \le y | \mathbf{X}^*] = P[W_{j,n} \in (x, y], \ \forall j \in J_n \text{ with } x_{j,n} \in I_{k,n} | \mathbf{X}^*]$$
$$= \left(\int_x^y f_{W,j(k)}(t)dt \right)^{l_{k,n}},$$

for $y > x$ where $l_{k,n} := \#\{j \in J_n : x_{j,n} \in I_{k,n}\}$. Thus we obtain the conditional joint density of $(s_{k,n}, S_{k,n})$ via

$$f_{(s_{k,n},S_{k,n})}(x, y) = -\frac{\partial^2}{\partial x \partial y} P[s_{k,n} > x, \ S_{k,n} \le y | \mathbf{X}^*]$$
$$= A_{k,n}(x, y) \cdot l_{k,n}(l_{k,n} - 1) f_{W,j(k)}(x) f_{W,j(k)}(y) 1_{\{y \ge x\}},$$

where

$$A_{k,n}(x, y) = \left(1 - \int_{\Delta_{0,j(k),n}-1}^x f_{W,j(k)}(t)dt - \int_y^{\Delta_{0,j(k),n}+1} f_{W,j(k)}(t)dt \right)^{l_{k,n}-2}.$$

Definition 5.3 Consider for each k two conditionally on $\mathbf{X}^*$ independent random variables $s'_{k,n}$ and $S'_{k,n}$ with conditional exponential densities

$$f_{s'_{k,n}}(x) = (l_{k,n} - 2) f_{W,j(k)}(\Delta_{0,j(k),n} - 1) \exp(-(l_{k,n} - 2) f_{W,j(k)}(\Delta_{0,j(k),n} - 1)$$
$$\cdot (x - \Delta_{0,j(k),n} + 1)) \mathbf{1}_{[\Delta_{0,j(k),n}-1,\infty)}(x),$$
$$f_{S'_{k,n}}(x) = (l_{k,n} - 2) f_{W,j(k)}(\Delta_{0,j(k),n} + 1) \exp(-(l_{k,n} - 2) f_{W,j(k)}(\Delta_{0,j(k),n} + 1)$$
$$\cdot (-x + \Delta_{0,j(k),n} + 1)) \mathbf{1}_{(-\infty, \Delta_{0,j(k),n}+1]}(x),$$

and the conditional joint density $f_{(s'_{k,n}, S'_{k,n})}$. Then the experiment $\mathcal{F}_n$ is obtained by observing $\mathbf{X}^*$ as well as conditionally on $\mathbf{X}^*$ independent tuples $(s'_{k,n}, S'_{k,n})$, $k = 0, \ldots, m-1$.

Lemma 5.3 *Assume that $m \leq \text{const.} \cdot n^{1-\delta}$ for some $\delta > 0$ and that*

$$\sup_{k=0,\ldots,m-1} |\Delta_{0, j(k), n}| \leq 2C_\Theta, \quad a.s., \quad \forall \vartheta \in \Theta. \tag{5.5}$$

Conditionally on the data set $\mathbf{X}^*$, the squared Hellinger distance between $f_{(s'_{k,n}, S'_{k,n})}$ and $f_{(s_{k,n}, S_{k,n})}$ satisfies

$$H^2(f_{(s'_{k,n}, S'_{k,n})}, f_{(s_{k,n}, S_{k,n})}) \leq \text{const.} \cdot \{\log(n/m)\}^4 (m/n)^2,$$

where const. is uniform with respect to n, $\mathbf{X}^*$, ϑ and k.

Remark 5.1 This approximation result together with the ensuing corollary tells us that we need to choose the number m of intervals of polynomially smaller order than $n^{2/3}$. To see that we cannot hope for a better approximation order, note that already in the most simple univariate case where $s := \min(U_i, i = 1, \ldots, I)$ with U_i i.i.d. uniform on $[0, 1]$ and s' exponentially distributed with intensity $I \in \mathbb{N}$, we have for $I \to \infty$

$$H^2(f_s, f_{s'}) \geq \int\limits_0^{1/I} \left(\sqrt{I(1-x)^{I-1}} - \sqrt{I \exp(-Ix)} \right)^2 dx$$

$$\approx ((1 - 1/I)^{(I-1)/2} - \exp(-1/2))^2 \asymp I^{-2}.$$

Corollary 5.1 *Assume that an estimator $\hat{\vartheta}$ of ϑ can be constructed from the data $\mathbf{X}^*$ such that (5.5) holds. Then for $m = O(n^{2/3-\delta})$ with some $\delta > 0$ as $n \to \infty$ the experiments $\mathcal{E}_n$ and $\mathcal{F}_n$ are asymptotically equivalent.*

Proof of Corollary 5.1 Focussing on the total variation distance between the distributions of the data $(\mathbf{X}^*, \{(s'_{k,n}, S'_{k,n}) : k = 0, \ldots, m-1\})$ and $(\mathbf{X}^*, \{(s_{k,n}, S_{k,n}) : k = 0, \ldots, m-1\})$ we consider for any measurable functional R on an appropriate domain and $\|R\|_\infty \leq 1$ that

$$|E_\vartheta R(\mathbf{X}^*, s_{0,n}, S_{0,n}, \ldots, s_{m-1,n}, S_{m-1,n}) - E_\vartheta R(\mathbf{X}^*, s'_{0,n}, S'_{0,n}, \ldots, s'_{m-1,n}, S'_{m-1,n})|$$

$$\leq 2 \sum_{k=0}^{m-1} E_\vartheta H^2(f_{(s_{k,n}, S_{k,n})}, f_{(s'_{k,n}, S'_{k,n})})$$

$$\leq \text{const.} \cdot n^{-3\delta} \log^4 n,$$

using the conditional independence of the $(s_{k,n}, S_{k,n})$, $k = 0, \ldots, m-1$, on the one hand and the $(s'_{k,n}, S'_{k,n})$, $k = 0, \ldots, m-1$, on the other hand and arguments as in the proof of Lemma 5.1; as well as Lemma 5.3 in the last line. Thus the total variation distance between the distributions of the data $(\mathbf{X}^*, \{(s'_{k,n}, S'_{k,n}) : k = 0, \ldots, m-1\})$

and $(\mathbf{X}^*, \{(s_{k,n}, S_{k,n}) : k = 0, \ldots, m - 1\})$ converges to zero as $n \to \infty$, which proves the claim of the corollary. $\square$

Proof of Lemma 5.3 First we mention that, although the arguments of the Hellinger distance are most usually densities, its definition $H^2(f, g) = \int(\sqrt{f}(x) - \sqrt{g}(x))^2 dx$ may easily be extended to all nonnegative functions $f, g \in L_1(\mathbb{R})$. This fact will be used in the sequel. Moreover, note that $l_{k,n} \asymp n/m \geq \text{const.} \cdot n^\delta$ holds uniformly over k by our design assumption (2.3). We set

$$f_{1,k,n}(x, y) = \frac{(l_{k,n} - 2)^2}{l_{k,n}(l_{k,n} - 1)} f_{(s_{k,n}, S_{k,n})}(x, y),$$

so that

$$H^2(f_{1,k,n}, f_{(s_{k,n}, S_{k,n})}) \leq \frac{(4 - 3l_{k,n})^2}{l_{k,n}(l_{k,n} - 1)(l_{k,n} - 2)^2} \asymp l_{k,n}^{-2}. \tag{5.6}$$

Note that the support of $f_{(s_{k,n}, S_{k,n})}$ and hence of $f_{1,k,n}$ is included in the square $Q_{k,n} = [\Delta_{0,j(k),n} - 1, \Delta_{0,j(k),n} + 1]^2$. A sub-square is defined by

$$Q_{1,k,n} = [\Delta_{0,j(k),n} - 1, \Delta_{0,j(k),n} - 1 + a_{k,n}]$$
$$\times [\Delta_{0,j(k),n} + 1 - a_{k,n}, \Delta_{0,j(k),n} + 1] \subseteq Q_{k,n},$$

which will contain most probability masses, and we set $Q_{2,k,n} = Q_{k,n} \backslash Q_{1,k,n}$ where $a_{k,n} = d_0 l_{k,n}^{-1} \log l_{k,n}$ with a constant $d_0 > 0$ for n sufficiently large. We split the Hellinger distance into integrals over disjoint domains so that

$$H^2(f_{1,k,n}, f_{(s'_{k,n}, S'_{k,n})}) \leq \int_{Q_{1,k,n}} \left(\sqrt{f_{1,k,n}}(x, y) - \sqrt{f_{(s'_{k,n}, S'_{k,n})}}(x, y)\right)^2 dx\, dy$$

$$+ 2 \int_{Q_{2,k,n}} f_{1,k,n}(x, y) dx\, dy$$

$$+ 2P[s'_{k,n} > \Delta_{0,j(k),n} - 1 + a_{k,n}|\mathbf{X}^*]$$

$$+ 2P[S'_{k,n} < \Delta_{0,j(k),n} + 1 - a_{k,n}|\mathbf{X}^*]$$

$$=: T_1 + T_2 + T_3 + T_4. \tag{5.7}$$

The conditions (2.4) and (5.5) combined with the positivity of φ imply that $\|f_{W,j(k)}\|_\infty \leq \text{const.}$ and that

$$\int_{\Delta_{0,j(k),n}-1}^{x} f_{W,j(k)}(t) dt \geq \text{const.} \cdot (x - \Delta_{0,j(k),n} + 1),$$

$$\forall x \in [\Delta_{0,j(k),n} - 1, \Delta_{0,j(k),n} + 1],$$

$$\int\limits_{y}^{\Delta_{0,j(k),n}+1} f_{W,j(k)}(t)dt \geq \text{const.} \cdot (\Delta_{0,j(k),n} + 1 - y),$$

$$\forall y \in [\Delta_{0,j(k),n} - 1, \Delta_{0,j(k),n} + 1].$$

As the Lebesgue measure of $Q_{k,n}$ is equal to 4, thus bounded, we deduce by the definition of $f_{1,k,n}$ and $f_{(S_{k,n}, S_{k,n})}$ that

$$T_2 \leq c_\nu n^{-\nu},$$

for each $\nu > 0$ when selecting the constant d_0 in the definition of $a_{k,n}$ sufficiently large where c_ν denotes a finite constant which depends on neither the data $\mathbf{X}^*$, ϑ nor x, y.

Concerning terms T_3 and T_4, easy calculations yield that these terms are equal to $2 \exp\{-a_{k,n}(l_{k,n} - 2) f_{W,j(k)}(\Delta_{0,j(k),n} \mp 1)\}$, respectively. We may use (2.4), (5.5) and $\varphi > 0$ to show that $f_{W,j(k)}(\Delta_{0,j(k),n} \mp 1) \geq \text{const.}$ Again choosing the constant d_0 sufficiently large implies that $\max\{T_3, T_4\} \leq c'_\nu n^{-\nu}$, for any $\nu > 0$ with a constant c'_ν which has the same properties as c_ν.

Let us focus on the main term T_1. For $(x, y) \in Q_{1,k,n}$, we have

$$\log A_{k,n}(x, y) = (l_{k,n} - 2) \left(-\int\limits_{\Delta_{0,j(k),n}-1}^{x} f_{W,j(k)}(t)dt - \int\limits_{y}^{\Delta_{0,j(k),n}+1} f_{W,j(k)}(t)dt \right)$$
$$+ R_{1,k,n}(x, y),$$

where $\sup_{(x,y)\in Q_{1,k,n}} \max_{k=0,\ldots,m-1} |R_{1,k,n}(x, y)| \leq \text{const.} \cdot l_{k,n} a_{k,n}^2 \asymp l_{k,n}^{-1} \log^2 l_{k,n}$ by the Taylor expansion of the logarithm. Furthermore, the functions to be integrated are locally approximated by constant functions,

$$-\int\limits_{\Delta_{0,j(k),n}-1}^{x} f_{W,j(k)}(t)dt - \int\limits_{y}^{\Delta_{0,j(k),n}+1} f_{W,j(k)}(t)dt$$
$$= -f_{W,j(k)}(\Delta_{0,j(k),n} - 1) \cdot (x - \Delta_{0,j(k),n} + 1)$$
$$- f_{W,j(k)}(\Delta_{0,j(k),n} + 1) \cdot (-y + \Delta_{0,j(k),n} + 1) + R_{2,k,n}(x, y),$$

where $\sup_{(x,y)\in Q_{1,k,n}} \max_{k=0,\ldots,m-1} |R_{2,k,n}(x, y)| \leq \text{const.} \cdot l_{k,n}^{-2} \log^2 l_{k,n}$, using the Lipschitz continuity of φ.

We introduce $B_{k,n}(x, y) := A_{k,n}(x, y) f_{W,j(k)}(x) f_{W,j(k)}(y)(l_{k,n} - 2)^2$ so that $B_{k,n}(x, y)$ coincides with $f_{1,k,n}(x, y)$ on its restriction to $(x, y) \in Q_{1,k,n}$ for n large enough, as well as

$$\tilde{B}_{k,n}(x, y) := f_{(S'_{k,n}, S'_{k,n})}(x, y) \frac{f_{W,j(k)}(x) f_{W,j(k)}(y)}{f_{W,j(k)}(\Delta_{0,j(k),n} - 1) f_{W,j(k)}(\Delta_{0,j(k),n} + 1)}.$$

We obtain

$$B_{k,n}^{1/2}(x, y) = \tilde{B}_{k,n}^{1/2}(x, y) \, \exp(R_{1,k,n}(x, y)/2 + (l_{k,n} - 2) R_{2,k,n}(x, y)/2)$$
$$= \tilde{B}_{k,n}^{1/2}(x, y) + \tilde{B}_{k,n}^{1/2}(x, y) R_{3,k,n}(x, y),$$

where $\sup_{(x,y) \in Q_{1,k,n}} \max_{k=0,\dots,m-1} |R_{3,k,n}(x, y)| \leq \text{const.} \cdot l_{k,n}^{-1} \log^2 l_{k,n}$ so that

$$\tilde{B}_{k,n}^{1/2}(x, y) = f_{(s'_{k,n}, S'_{k,n})}^{1/2}(x, y) + f_{(s'_{k,n}, S'_{k,n})}^{1/2}(x, y) R_{4,k,n}(x, y),$$

where

$$|R_{4,k,n}(x, y)| \leq \text{const.} \cdot (|f_{W,j(k)}(x) - f_{W,j(k)}(\Delta_{0,j(k),n} - 1)|$$
$$+ |f_{W,j(k)}(y) - f_{W,j(k)}(\Delta_{0,j(k),n} + 1)|)$$
$$\leq \text{const.} \cdot a_{k,n} \asymp l_{k,n}^{-1} \log l_{k,n},$$

where the conditions (5.5), (2.4) and their consequences have been used. We conclude that

$$B_{k,n}^{1/2}(x, y) = f_{(s'_{k,n}, S'_{k,n})}^{1/2}(x, y) + f_{(s'_{k,n}, S'_{k,n})}^{1/2}(x, y) R_{5,k,n}(x, y),$$

where $\sup_{(x,y) \in Q_{1,k,n}} \max_{k=0,\dots,m-1} |R_{5,k,n}(x, y)| \leq \text{const.} \cdot l_{k,n}^{-1} \log^2 l_{k,n}$. Hence, the term T_1 is bounded from above by

$$T_1 \leq \int_{Q_{1,k,n}} R_{5,k,n}^2(x, y) f_{(s'_{k,n}, S'_{k,n})}(x, y) \, dx \, dy \leq \text{const.} \cdot (\log^4 l_{k,n}) l_{k,n}^{-2},$$

as the density $f_{(s'_{k,n}, S'_{k,n})}$ integrates to one. By inserting the upper bounds on $T_1, \dots, T_4$ into (5.7) and combining that result with (5.6), we complete the proof. $\qquad \square$

Definition 5.4 In experiment $\mathcal{G}_n$ we observe the data $(\mathbf{X}^*, (d_{k,n}, D_{k,n})_{k=0,\dots,m-1})$ for $\vartheta \in \Theta$ where $d_{0,n}, D_{0,n}, \dots, d_{m-1,n}, D_{m-1,n}$ are independent random variables, also independent of $\mathbf{X}^*$, with densities

$$f_{d_{k,n}}(x) = \rho_{k,n} \varphi(-1) \exp(-\rho_{k,n} \varphi(-1) \cdot [x - \vartheta(\xi_{j(k),n})]) \, \mathbf{1}_{[\vartheta(\xi_{j(k),n}), \infty)}(x),$$
$$f_{D_{k,n}}(x) = \rho_{k,n} \varphi(1) \exp(\rho_{k,n} \varphi(1) \cdot [x - \vartheta(\xi_{j(k),n})]) \, \mathbf{1}_{(-\infty, \vartheta(\xi_{j(k),n})]}(x),$$

where $\rho_{k,n} = (n/2) \int_{I_{k,n}} f_D(t) dt$ with f_D as in (2.2).

Lemma 5.4 *We select m such that $m = o(n^{2/3})$. Also we assume the existence of an estimator $\hat{\vartheta}$ of ϑ based on $\mathbf{X}^*$ such that (5.5) and*

$$\sup_{\vartheta \in \Theta} \sup_{x \in [0,1]} E_{\vartheta} |\hat{\vartheta}(x) - \vartheta(x)|^2 = o(m^{-1})$$

are fulfilled. Then the experiments $\mathcal{F}_n$ and $\mathcal{G}_n$ are asymptotically equivalent as $n \to \infty$.

Proof of Lemma 5.4 As the estimator $\hat{\vartheta}$ is based on the data set $\mathbf{X}^*$ the transformation $\mathcal{T}$ which maps the observations $(\mathbf{X}^*, \{(s'_{k,n}, S'_{k,n}) : k = 0, \ldots, m - 1\})$ to $(\mathbf{X}^*, \{(s''_{k,n}, S''_{k,n}) : k = 0, \ldots, m - 1\})$ with $s''_{k,n} = s'_{k,n} + \hat{\vartheta}(\xi_{j(k),n}) + 1$ and $S''_{k,n} = S'_{k,n} + \hat{\vartheta}(\xi_{j(k),n}) - 1$ is invertible. Therefore, the experiment under which the data $(\mathbf{X}^*, \{(s''_{k,n}, S''_{k,n}) : k = 0, \ldots, m - 1\})$ are observed is equivalent to the experiment $\mathcal{F}_n$.

The squared Hellinger distance between the exponential densities with the same endpoint and the scaling parameters μ_1 and μ_2 turns out to be $2(\mu_1 - \mu_2)^2(\mu_1 + \mu_2)^{-1}(\sqrt{\mu_1} + \sqrt{\mu_2})^{-2}$.

Also, (2.2) implies that $|l_{k,n} - \rho_{k,n}| \le 3$ for all $k = 0, \ldots, m - 1$. We may set $\mu_{1,\pm} = \rho_{k,n}\varphi(\pm 1) \int_{\Delta_{0,j(k),n}-1}^{\Delta_{0,j(k),n}+1} \varphi(t)dt$ and $\mu_{2,\pm} = (l_{k,n} - 2)\varphi(\Delta_{0,j(k),n} \pm 1)$. Hence,

$$H^2(f_{S''_{k,n}}, f_{D_{k,n}}) + H^2(f_{s''_{k,n}}, f_{d_{k,n}}) \le \text{const.} \cdot \{l_{k,n}^{-2} + \Delta_{0,j(k),n}^2\},$$

where the constant does not depend on $\mathbf{X}^*$. Therein we have utilized condition (5.5) as well as the Lipschitz continuity, positivity and boundedness of φ. We take the expectation of the sum of these terms over $k = 0, \ldots, m - 1$ which converges to zero uniformly in $\vartheta \in \Theta$ by the assumption on m and the imposed convergence rates of the estimator $\hat{\vartheta}$. Then the asymptotic equivalence is evident by the argument (5.3) from the proof of Lemma 5.1 when replacing the data sets $\tilde{Z}$ and $\mathbf{W}$ by the data samples $(d_{k,n}, D_{k,n})_{k=0,\ldots,m-1}$ and $(s''_{k,n}, S''_{k,n})_{k=0,\ldots,m-1}$, respectively, and inserting the conditional densities of their components given $\mathbf{X}^*$. The sum is, of course, to be taken over $k = 0, \ldots, m - 1$ instead of $j \in J_n$. □

Now we go over to experiments involving Poisson point processes (PPP).

Definition 5.5 In experiment $\mathcal{H}_n$ we observe $\mathbf{X}^*$ and independently two independent Poisson point processes X_l and X_u whose domain is the Borel σ-algebra of $\mathbb{R}^2$ and whose intensity functions equal

$$\lambda_l(x, y) = m\varphi(1) \sum_{k=0}^{m-1} \rho_{k,n} \mathbf{1}_{I_{k,n}}(x) \mathbf{1}_{[-C_\Theta - 1, \vartheta(\xi_{j(k),n})]}(y),$$

$$\lambda_u(x, y) = m\varphi(-1) \sum_{k=0}^{m-1} \rho_{k,n} \mathbf{1}_{I_{k,n}}(x) \mathbf{1}_{[\vartheta(\xi_{j(k),n}), C_\Theta + 1]}(y),$$

and are hence locally constant. We recall that C_Θ is the uniform upper bound on $|\vartheta|$ in the parameter set Θ.

We define the extreme points of X_l and X_u in the strip $I_{k,n} \times \mathbb{R}$ by

$$X_{l,k} = \inf\{y \in \mathbb{R} : X_l(I_{k,n} \times [y, \infty)) = 0\},$$
$$X_{u,k} = \sup\{y \in \mathbb{R} : X_u(I_{k,n} \times (-\infty, y]) = 0\}.$$

Lemma 5.5 (a) *The statistic* $(X_{l,k}, X_{u,k})$, $k = 0, \ldots, m-1$, *is sufficient for the whole empirical information contained in* X_l *and* X_u.
(b) *The distribution functions of* $X_{l,k}$ *and* $X_{u,k}$ *are equal to those of* $\max\{-C_\Theta - 1, D_{k,n}\}$ *and* $\min\{C_\Theta + 1, d_{k,n}\}$, *respectively where* $d_{k,n}$ *and* $D_{k,n}$ *are as in experiment* $\mathcal{G}_n$. *Moreover, all* $X_{l,k}$, $k = 0, \ldots, m-1$, *on the one hand and all* $X_{u,k}$, $k = 0, \ldots, m-1$ *on the other hand are independent.*

Proof of Lemma 5.5 (a) Let X_0 denote the PPP with the intensity function $\lambda_0 = \mathbf{1}_{[0,1] \times [-C_\Theta - 1, C_\Theta + 1]}$. The probability measures generated by X_0, X_l, X_u are denoted by $\mathbf{P}_0, \mathbf{P}_l, \mathbf{P}_u$, respectively. As the functions $\lambda_0, \lambda_l, \lambda_u$ are piecewise constant and the support of λ_l and λ_u is included in that of λ_0 the measure $\mathbf{P}_0$ dominates $\mathbf{P}_l$ and $\mathbf{P}_u$ and the corresponding Radon–Nikodym derivatives are equal to

$$\frac{d\mathbf{P}_l}{d\mathbf{P}_0}(X) = \exp\left\{\int \log \frac{\lambda_l(x,y)}{\lambda_0(x,y)} dX(x,y) - \int \left(\frac{\lambda_l(x,y)}{\lambda_0(x,y)} - 1\right) \lambda_0(x,y)\, dx\, dy\right\},$$

$$\frac{d\mathbf{P}_u}{d\mathbf{P}_0}(X) = \exp\left\{\int \log \frac{\lambda_u(x,y)}{\lambda_0(x,y)} dX(x,y) - \int \left(\frac{\lambda_u(x,y)}{\lambda_0(x,y)} - 1\right) \lambda_0(x,y)\, dx\, dy\right\},$$

see e.g. Theorem 1.3 in [19] which apparently goes back to [4]. Therein X may be viewed as an arbitrary counting process on the Borel σ-algebra of $[0, 1] \times [-C_\Theta - 1, C_\Theta + 1]$. We write $\Gamma_\vartheta = \bigcup_{k=0}^{m-1} I_{k,n} \times (\vartheta(\xi_{j(k),n}), C_\Theta + 1]$ and $\Phi = \bigcup_{k=0}^{m-1} I_{k,n} \times [-C_\Theta - 1, \tilde{X}_{l,k}]$ where $\tilde{X}_{l,k}$ equals $X_{l,k}$ except that X_l is changed into the general process X in the definition. Then $d\mathbf{P}_l/d\mathbf{P}_0$ is equal to

$$\frac{d\mathbf{P}_l}{d\mathbf{P}_0}(X) = \mathbf{1}_{\{\emptyset\}}(\Gamma_\vartheta \cap \Phi) \cdot \exp\left\{\sum_{k=0}^{m-1} \log[\rho_{k,n} m \varphi(1)] X(I_{k,n} \times [-C_\Theta - 1, C_\Theta + 1])\right\}$$

$$\cdot \exp\left\{-\sum_{k=0}^{m-1} (\vartheta(\xi_{j(k),n}) + C_\Theta + 1)\rho_{k,n}\varphi(1)\right\} \exp(2C_\Theta + 2),$$

where we have used that $X(I_{k,n} \times [-C_\Theta - 1, C_\Theta + 1]) = X(I_{k,n} \times [-C_\Theta - 1, \vartheta(\xi_{j(k),n})])$ whenever $X(\Gamma_\vartheta) = 0$; and that Γ_ϑ and Φ are disjoint if and only if $X(\Gamma_\vartheta) = 0$. It follows from the Fisher–Neyman factorization theorem that the $X_{l,k}$, $k = 0, \ldots, m-1$ represent a sufficient statistic for X_l. The corresponding assertion for the $X_{u,r}$ is proved analogously.

(b) We consider for $x \in [-C_\Theta - 1, \vartheta(\xi_{j(k),n})]$ that

$$P[X_{l,k} \le x] = P[X_l(I_{k,n} \times (x, \infty)) = 0] = \exp(-(\vartheta(\xi_{j(k),n}) - x)\rho_{k,n}\varphi(1))$$
$$= P[D_{k,n} \le x].$$

Clearly we have $P[X_{l,k} > \vartheta(\xi_{j(k),n})] = P[D_{k,n} > \vartheta(\xi_{j(k),n})] = 0$ and $P[X_{l,k} < -C_\Theta - 1] = 0$ so that the distribution functions of $X_{l,k}$ and $\max\{-C_\Theta - 1, D_{k,n}\}$ coincide. The claim that $X_{u,k}$ and $\min\{C_\Theta + 1, d_{k,n}\}$ are identically distributed follows analogously. Finally the independence of the data $X_{l,k}$, $k = 0, \ldots, m-1$ as well as

of the data $X_{u,k}$, $k = 0, \ldots, m - 1$ follows from the fact that $X(A_0), \ldots, X(A_{m-1})$ are independent for all $A_k \subseteq I_{k,n} \times [-C_\Theta - 1, C_\Theta + 1]$ by the definition of the PPP. □

Lemma 5.6 *For $m = O(n^{1-\delta})$, $\delta > 0$, the total variation distance between the distributions of $((\min\{C_\Theta + 1, d_{k,n}\}, \max\{-C_\Theta - 1, D_{k,n}\}) : k = 0, \ldots, m - 1)$ and $((d_{k,n}, D_{k,n}) : k = 0, \ldots, m - 1)$ converges to zero.*

Proof of Lemma 5.6 Due to the independence of the data the desired total variation distance is bounded from above by the sum of the total variation distances between the distributions of $d_{k,n}$ and $\min\{C_\Theta + 1, d_{k,n}\}$ plus the corresponding distances between the distributions of $D_{k,n}$ and $\max\{-C_\Theta - 1, D_{k,n}\}$ where $k = 0, \ldots, m - 1$. The total variation distance between $d_{k,n}$ and $\min\{C_\Theta + 1, d_{k,n}\}$ is bounded by

$$2P[d_{k,n} \geq C_\Theta + 1] \leq 2\exp(-\text{const.} \cdot n/m),$$

so that, because of $m \leq \text{const.} \cdot n^{1-\delta}$, the sum of these terms for $k = 0, \ldots, m - 1$ tends to zero exponentially fast. The distributions of $\max\{-C_\Theta - 1, D_{k,n}\}$ and $D_{k,n}$ are treated in the same way. □

Combining these two lemmata we obtain directly asymptotic equivalence.

Corollary 5.2 *Experiments $\mathcal{G}_n$ and $\mathcal{H}_n$ are asymptotically equivalent for m as in Lemma 5.6.*

We observe that the choice $m \asymp n^{2/3-\delta}$ for some $\delta \in (0, 1/6)$ meets all requirements imposed on m so far and we summarize our results.

Proposition 5.1 *Select $m \asymp n^{2/3-\delta}$ for some $\delta \in (0, 1/6)$ and suppose that there is an estimator $\hat{\vartheta}$, based on the data $\mathbf{X}^*$ alone, which satisfies (5.5) and*

$$\sup_{\vartheta \in \Theta} \sup_{x \in [0,1]} E_\vartheta |\hat{\vartheta}(x) - \vartheta(x)|^2 = O(n^{-1-\delta}).$$

Then we have asymptotic equivalence between experiments $\mathcal{C}_n$ and $\mathcal{H}_n$. Moreover, if we have additionally

$$\sup_{\vartheta \in \Theta} \sup_{x \in [0,1]} E_\vartheta |\hat{\vartheta}'(x) - \vartheta'(x)| = o(n^{-1/3-\delta}),$$

for an estimator $\hat{\vartheta}'$, which is not necessarily the derivative of $\hat{\vartheta}$, then also $\mathcal{A}_n^*$ and $\mathcal{H}_n$ are asymptotically equivalent.

6 Localization of the PPP model

The processes X_l and X_u in the experiment $\mathcal{H}_n$ have step functions as their intensity boundaries which approximate continuous functions as m tends to infinity. Therefore we consider now the experiment where $\mathbf{X}^*$ and independently two PPP with boundary function ϑ are observed.

Definition 6.1 In experiment $\mathcal{I}_n$ we observe $\mathbf{X}^*$ and independently two independent PPP $X_{1,0}$ and $X_{2,0}$ with intensities

$$
\begin{aligned}
\lambda_{1,0}(x, y) &= (n/2) f_\varepsilon(1) f_D(x) \mathbf{1}_{[-C_\Theta-1, \vartheta(x)]}(y), \\
\lambda_{2,0}(x, y) &= (n/2) f_\varepsilon(-1) f_D(x) \mathbf{1}_{[\vartheta(x), C_\Theta+1]}(y).
\end{aligned}
\tag{6.1}
$$

Proposition 6.1 *We impose the conditions of Lemma* 4.1 *and, in addition, that for all* $\vartheta \in \Theta$, *we have*

$$
\sup_{x \in [0,1]} |\hat{\vartheta}'(x)| \le 2 \sup_{\vartheta \in \Theta} \sup_{x \in [0,1]} |\vartheta'(x)|, \quad a.s.
\tag{6.2}
$$

Then the experiments $\mathcal{H}_n$ and $\mathcal{I}_n$ are asymptotically equivalent.

Proof of Proposition 6.1 First, we show asymptotic equivalence of the experiment $\mathcal{H}_n$ with the experiment $\mathcal{H}_n'$ in which one observes the data $(\mathbf{X}^*, \tilde{X}_1, \tilde{X}_2)$ where $\tilde{X}_1$ and $\tilde{X}_2$ are PPP with the intensity functions

$$
\begin{aligned}
\tilde{\lambda}_1(x, y) &= (n/2) f_\varepsilon(1) f_D(x) \mathbf{1}_{[-C_\Theta-1, \vartheta(x)+\hat{\vartheta}'(x)(\xi(x)-x)]}(y), \\
\tilde{\lambda}_2(x, y) &= (n/2) f_\varepsilon(-1) f_D(x) \mathbf{1}_{[\vartheta(x)+\hat{\vartheta}'(x)(\xi(x)-x), C_\Theta+1]}(y),
\end{aligned}
$$

conditionally on $\mathbf{X}^*$, respectively. As $\hat{\vartheta}'$ we may choose the pilot estimator from Lemma 4.1 based on the data set $\mathbf{X}^*$ under appropriate truncation such that (6.2) is satisfied; and we write $\xi(x)$ for the centre of that interval $I_{k,n}$ which contains the element x.

By a similar argument as in (4.3), it suffices to show that the expected Hellinger distance between the distribution of $\tilde{X}_1$ and X_l on the one hand and $\tilde{X}_2$ and X_u on the other hand converges to zero. We shall now employ a general formula bounding the Hellinger distance between two PPP laws P_1, P_2 with respective intensities λ_1, λ_2 by the (generalized) Hellinger distance of the intensities ; when P denotes the law of the PPP with intensity $\lambda = \lambda_1 + \lambda_2$, we derive from the likelihood expression

$$
\begin{aligned}
H^2(P_1, P_2) &= 2 \left(1 - \exp\left(-\int \left(\sqrt{\lambda_1} - \sqrt{\lambda_2} \right)^2 / 2 \right) \right) \\
&\le \int \left(\sqrt{\lambda_1} - \sqrt{\lambda_2} \right)^2,
\end{aligned}
\tag{6.3}
$$

see e.g. Theorem 3.2.1 in [27] or [22,23]. Thus we bound the Hellinger distance between the intensities of $\tilde{X}_1$ and X_l by

$$
\int \left(\sqrt{\lambda_l} - \sqrt{\tilde{\lambda}_1} \right)^2 \le \text{const.} \cdot n \sum_{k=0}^{m-1} \left\{ \int_{I_{k,n}} |\vartheta(\xi(x)) - \vartheta(x) - \hat{\vartheta}'(x)(\xi(x) - x)| dx \right.
$$

$$+ \int\limits_{I_{k,n}} \left| f_D(x) - m \int\limits_{I_{k,m}} f_D(y) dy \right|^2 dx \Bigg\},$$

where the constant does not depend on $\mathbf{X}^*$. As f_D is assumed to be Lipschitz on $[0, 1]$ the latter term contributes to the asymptotic order by the deterministic upper bound $O(nm^{-2})$ independently of ϑ. Then we apply the expectation to the above expression and we obtain

$$O(nm^{-2}) + \text{const.} \cdot nm^{-1} \sup_{\vartheta \in \Theta} \sup_{x \in [0,1]} E_\vartheta |\hat{\vartheta}'(x) - \vartheta'(x)| = o(1),$$

as a uniform upper bound. Together with the same bound for the Hellinger distance, conditionally on $\mathbf{X}^*$, between the intensities of $\tilde{X}_2$ and X_u this implies asymptotic equivalence between $\mathcal{H}_n$ and $\mathcal{H}_n'$ again by arguments as in (5.3).

For any two-dimensional Borel set B let us define the pointwise shifted version

$$\hat{B} = \{(x, y) \in \mathbb{R}^2 : (x, y + \hat{\vartheta}'(x)[\xi(x) - x]) \in B\},$$

and the processes $\overline{X}_j(B) = \tilde{X}_j(\hat{B})$, $j = 1, 2$, conditionally on the data set $\mathbf{X}^*$. Note that $\hat{B}$ is a Borel set as well as the shift function $\hat{\vartheta}'(\cdot)[\xi(\cdot) - \cdot]$ is measurable. Then $\overline{X}_j$ represents a PPP with the shifted intensity function

$$\overline{\lambda}_1(x, y) = (n/2)\varphi(1) f_D(x) \mathbf{1}_{[-C_\Theta - 1 - \hat{\vartheta}'(x)(\xi(x) - x), \vartheta(x)]}(y)$$
$$\overline{\lambda}_2(x, y) = (n/2)\varphi(-1) f_D(x) \mathbf{1}_{[\vartheta(x), C_\Theta + 1 - \hat{\vartheta}'(x)(\xi(x) - x)]}(y).$$

Note that this transformation is invertible as long as the data set $\mathbf{X}^*$ is available. Therefore, the experiment $\mathcal{H}_n''$ of observing $\mathbf{X}^*$ and $\overline{X}_j$, $j = 1, 2$ independently is equivalent to the experiment $\mathcal{H}_n'$.

By the imposed upper bound on the estimator $\hat{\vartheta}'$ we may assume that

$$\sup_{x \in [0,1]} |\hat{\vartheta}'(x)||\xi(x) - x| \leq 1/2, \quad P_\vartheta\text{-a.s. for all } \vartheta \in \Theta,$$

for m sufficiently large. Hence, the observation of $\overline{X}_j$, $j = 1, 2$, is equivalent with the observation of two conditionally independent Poisson processes $\overline{X}_{j,1}$ and $\overline{X}_{j,2}$ with the intensity functions

$$\overline{\lambda}_{1,1}(x, y) = (n/2)\varphi(1) f_D(x) \mathbf{1}_{[-C_\Theta - 1/2, \vartheta(x)]}(y),$$
$$\overline{\lambda}_{1,2}(x, y) = (n/2)\varphi(1) f_D(x) \mathbf{1}_{[-C_\Theta - 1 - \hat{\vartheta}'(x)(\xi(x) - x), -C_\Theta - 1/2)}(y),$$
$$\overline{\lambda}_{2,1}(x, y) = (n/2)\varphi(-1) f_D(x) \mathbf{1}_{[\vartheta(x), C_\Theta + 1/2]}(y),$$
$$\overline{\lambda}_{2,2}(x, y) = (n/2)\varphi(-1) f_D(x) \mathbf{1}_{(C_\Theta + 1/2, C_\Theta + 1 - \hat{\vartheta}'(x)(\xi(x) - x))]}(y),$$

Thus all processes $\overline{X}_{j,i}$, $i, j = 1, 2$, are independent. Also we realize that the processes $\overline{X}_{1,2}$ and $\overline{X}_{2,2}$ represent conditionally ancillary statistics given the data set $\mathbf{X}^*$ as $\overline{\lambda}_{1,2}$ and $\overline{\lambda}_{2,2}$ do not explicitly depend on ϑ, but are fixed by knowledge of $\mathbf{X}^*$ for n sufficiently large. Therefore, the observation of $\mathbf{X}^*$ and $\overline{X}_{j,1}$, $j = 1, 2$ is sufficient for complete empirical information contained in experiment $\mathcal{H}_n''$. On the other hand we may also add two independent PPP $\overline{X}_{j,3}$, $j = 1, 2$ with the intensity functions

$$\overline{\lambda}_{1,3}(x, y) = (n/2)\varphi(1)f_D(x)\mathbf{1}_{[-C_\Theta - 1, -C_\Theta - 1/2)}(y),$$
$$\overline{\lambda}_{2,3}(x, y) = (n/2)\varphi(-1)f_D(x)\mathbf{1}_{(C_\Theta + 1/2, C_\Theta + 1]}(y),$$

which are totally uninformative. Combining the independent processes $\overline{X}_{j,1}$ and $\overline{X}_{j,3}$ whose intensity functions are supported on (almost) disjoint domains for both $j = 1, 2$, the considered experiment is equivalent to the experiment $\mathcal{I}_n$. □

7 Final proof

In this section, we combine all results derived in the previous sections in order to complete the proof of Theorem 2.1. For simplicity we suppose that n is even. By Proposition 3.1 with sample size $n/2$, there exists an estimator $\hat{\vartheta}$ based on the data $\mathbf{X} = \mathbf{X}^*$ from experiment $\mathcal{A}_n$ which satisfies the conditions of Proposition 5.1, e.g. by choosing $\delta = \alpha/2$. Therefore, experiments $\mathcal{A}_n$ and $\mathcal{I}_n$ are asymptotically equivalent by Propositions 5.1 and 6.1. The conditions (5.5) and (6.2) are satisfied when truncating the range of $\hat{\vartheta}$ and $\hat{\vartheta}'$ suitably without losing validity of Proposition 3.1. Therein, note that the uniform upper bounds on $\vartheta \in \Theta$ as well as on its derivative are known. Then we set $\mathcal{A}_n^* = \mathcal{I}_n$ by using the processes $X_{1,0}$ and $X_{2,0}$ as the data set $\mathbf{X}^*$ and let $\mathbf{X}$ take the role of the data $\mathbf{Y}'$ from experiment $\mathcal{A}_n$. Note that all of our arguments from the previous sections remain valid when transforming the responses with even instead of odd observation number. Applying Propositions 5.1 and 6.1 again, we obtain asymptotic equivalence of the experiments $\mathcal{I}_n$ and $\mathcal{J}_n$ where the latter model just consists of $X_{1,0}$ and $X_{2,0}$ and two independent copies $X_{1,0}^*$ and $X_{2,0}^*$. The likelihood process of experiment $\mathcal{J}_n$ and experiment $\mathcal{B}_n$ turns out to be the same, using Theorem 1.3 in [19] as in the proof of Lemma 5.5, such that $\mathcal{J}_n$ and $\mathcal{B}_n$ are equivalent experiments. The concrete equivalence mapping is given by looking at the sum of the processes $X_j = X_{j,0} + X_{j,0}^*$, $j = 1, 2$, in one direction and by splitting the point masses in X_j randomly and independently with probability one half into point masses for $X_{j,0}$ and $X_{j,0}^*$ (*thinning* of a PPP) for the other equivalence direction.

8 Discussion

8.1 General remarks

We have shown asymptotic equivalence of nonparametric regression with non-regular additive errors and the observation of two specific independent PPP. Our result also yields that those nonparametric regression models are asymptotically equivalent

to each other as long as the corresponding error densities have the same jump sizes at -1 and $+1$ and are Lipschitz continuous and positive within the interval $(-1, 1)$ – regardless of the specific shape of the density inside its support. This unifies the asymptotic theory for these experiments and properties such as asymptotic minimax bounds, adaptation, superefficiency can be studied simultaneously for those models. At least after suitable linear correction by a pilot estimator, local minima and maxima are asymptotically sufficient for inference in these models.

The limiting Poisson point process model $\mathcal{B}_n$ exhibits a fascinating new geometric structure. According to (6.3), the squared Hellinger distances between observations with parameters $\vartheta_1, \vartheta_2 \in \Theta$ is given by

$$H^2(P_{\vartheta_1}, P_{\vartheta_2}) = 2\left(1 - \exp\left(-\frac{n}{2}(f_\varepsilon(-1) + f_\varepsilon(+1))\int |\vartheta_1(x) - \vartheta_2(x)| f_D(x)\, dx\right)\right).$$

Setting $\|g\|_{L^1_X} := \int |g(x)| f_D(x)\, dx$, the squared Hellinger distance is thus equivalent to an L^1-distance

$$H^2(P_{\vartheta_1}, P_{\vartheta_2}) \asymp n\{f_\varepsilon(-1) + f_\varepsilon(+1)\}\|\vartheta_1 - \vartheta_2\|_{L^1_X}. \tag{8.1}$$

In contrast, for nonparametric regression with regular errors the continuous limit model is a Gaussian shift where the corresponding squared Hellinger distance is equivalent to $n\sigma^{-2}\|\vartheta_1 - \vartheta_2\|_{L^2_X}^2$ with $\sigma^2 = \mathrm{Var}(\varepsilon_{j,n})$. While it is well known that the standard parametric rate improves from $n^{-1/2}$ to n^{-1}, the nonparametric view reveals that we face here an L^1_X-topology instead of the usual Hilbert space L^2_X-structure. As discussed below, this different Banach space geometry is even visible at the level of minimax rates, which are in general worse than for regular nonparametric regression with sample size n^2. A boundary behaviour of the error density f_ε other than finite jumps will imply a different Hellinger topology, in particular the whole range of L^p_X-geometries, $p \in (0, \infty)$, might arise, whose statistical consequences will be far-reaching and remain to be explored in detail.

8.2 A nonparametric lower bound

Let us apply the asymptotic equivalence result to study nonparametric lower bounds for all models in $\mathcal{A}_n$ and for $\mathcal{B}_n$, simultaneously. We restrict our investigation to rate results, but we track explicitly the dependence on the total jump size $J := f_\varepsilon(-1) + f_\varepsilon(1)$ and the design density f_D.

Proposition 8.1 *In the PPP model $\mathcal{B}_n$, but with ϑ from the parameter space*

$$\Theta_{s,L} := \{\vartheta \in C^s([0, 1]) \mid \|\vartheta\|_s \le L\}, \quad s, L > 0$$

with generalized Hölder norm

$$\|g\|_s := \max_{k=0,1,\ldots,\lfloor s \rfloor} \|g^{(k)}\|_\infty + \sup_{x \neq y} \frac{|g(x) - g(y)|}{|x - y|^{s - \lfloor s \rfloor}}$$

the following lower bound for the pointwise loss in estimating ϑ and its derivatives at $x_0 \in [0, 1]$ holds uniformly in $J := f_\varepsilon(-1) + f_\varepsilon(1)$, x_0 and $f_D(x_0)$

$$\liminf_{n \to \infty} \inf_{\hat{\vartheta}_n} \sup_{\vartheta \in \Theta_{s,L}} P_\vartheta \left(|\hat{\vartheta}_n^{(k)}(x_0) - \vartheta^{(k)}(x_0)| \geq c_0 \frac{L^{(k+1)/(s+1)}}{(nJf_D(x_0))^{(s-k)/(s+1)}} \right)$$

$$\geq \frac{2 - \sqrt{3}}{4} > 0$$

with $c_0 > 0$, where the infimum is taken over all estimators in $\mathcal{B}_n$ and $k = 0, 1, \ldots, \lfloor s \rfloor$.

By asymptotic equivalence and the boundedness of the involved loss function $\mathbb{1}\{|\hat{\vartheta}_n^{(k)}(x_0) - \vartheta^{(k)}(x_0)| \geq cL^{(k+1)/(s+1)}(nJf_D(x_0))^{-(s-k)/(s+1)}\}$, this result immediately generalizes to the regression experiments $\mathcal{A}_n$ provided the regularity s is larger than two. Moreover, by Markov's inequality it also applies to pth moment risk. We thus have:

Corollary 8.1 *For estimators $\hat{\vartheta}_n$ in experiment $\mathcal{A}_n$ with $\vartheta \in \Theta_{s,L} \subset \Theta$ and $s > 2$, $L > 0$ we have for all $p > 0$, $k = 0, 1, \ldots, \lfloor s \rfloor$ the lower bound*

$$\liminf_{n \to \infty} L^{-(k+1)/(s+1)}(nJf_D(x_0))^{(s-k)/(s+1)} \inf_{\hat{\vartheta}_n} \sup_{\vartheta \in \Theta_{s,L}} \left(E_\vartheta |\hat{\vartheta}_n^{(k)}(x_0) - \vartheta^{(k)}(x_0)|^p \right)^{1/p}$$

$$\geq c_1$$

for some constant $c_1 > 0$.

Proof of the Proposition 8.1 Let us fix $k \in \{0, 1, \ldots \lfloor s \rfloor\}$. By Theorem 2.2(ii) in [28] it suffices to find $\vartheta_1, \vartheta_2 \in \Theta_{s,L}$ with

$$|\vartheta_1^{(k)}(x_0) - \vartheta_2^{(k)}(x_0)| \geq L^{(k+1)/(s+1)}(nJf_D(x_0))^{-(s-k)/(s+1)}$$

and Hellinger distance of the corresponding observation laws satisfying $H(P_{\vartheta_1}, P_{\vartheta_2}) \leq 1$.

We choose some kernel function $K \in \Theta_{s,1}$ with $\int_{-1}^1 K(x)\,dx = 1$, $K^{(k)}(0) > 0$ and support in $[-1/2, 1/2]$ and we set $\vartheta_1(x) = 0$, $\vartheta_2(x) = Lh^s K((x - x_0)/h)$ with $h = (LnJf_D(x_0))^{-1/(s+1)}$ (using one-sided kernel versions near the boundary). Then for n sufficiently large we have $\vartheta_1, \vartheta_2 \in \Theta_{s,L}$ and moreover by (8.1)

$$H^2(P_{\vartheta_1}, P_{\vartheta_2}) = (1 + o(1))nJ \int_{-1}^1 |\vartheta_1(x) - \vartheta_2(x)| f_D(x)\,dx$$

and the integral satisfies $\int_{-1}^{1}|\vartheta_2(x)|f_D(x)\,dx = (L+o(1))h^{s+1}f_D(x_0)$ as $h \to 0$. We conclude that $H(P_{\vartheta_1}, P_{\vartheta_2})$ converges to one for $n \to \infty$. The result therefore follows from

$$|\vartheta_2^{(k)}(x_0) - \vartheta_2^{(k)}(x_0)| = K^{(k)}(0)L^{(k+1)/(s+1)}(nJf_D(x_0))^{-(s-k)/(s+1)}.$$

$\square$

The rate $L^{(k+1)/(s+1)}n^{-(s-k)/(s+1)}$ instead of $L^{(k+1/2)/(s+1/2)}\sqrt{n}^{-(s-k)/(s+1/2)}$ for regular nonparametric regression is obviously due to the L_X^1-bound on ϑ_2 instead of the squared L_X^2-bound. Let us mention that a careful study of our upper bound proof in Proposition 3.1 will also yield the same dependence on $L = C_\Theta$ for regularity $s = 2+\alpha$ and $k \in \{0, 1\}$. More geometrically, we can establish a lower bound for estimating a linear functional $L(\vartheta)$ by maximising $L(\vartheta)$ over $\vartheta \in \Theta_{s,L}$ with $\|\vartheta\|_{L_X^1} \le 1/(nJ)$. In the scale of Besov spaces $B_{p,p}^\alpha$ with norms $\|\cdot\|_{\alpha,p}$, $\alpha \in \mathbb{R}$, $1 \le p \le \infty$, we have $\|\vartheta\|_{L^1} \ge \|\vartheta\|_{-1,\infty}$ by duality from $\|\vartheta\|_{L^\infty} \le \|\vartheta\|_{1,1}$. Here, we can therefore expect to maximise $L(\vartheta) = \vartheta^{(k)}(x_0)$ as far as the interpolation inequality

$$\|\vartheta\|_{k,\infty} \le \|\vartheta\|_{-1,\infty}^{(s-k)/(s+1)}\|\vartheta\|_{s,\infty}^{(k+1)/(s+1)} \le \text{const.}(nJ)^{-(s-k)/(s+1)}L^{(k+1)/(s+1)}$$

permits. This is in fact achieved by the choice of ϑ_2 above, involving also the localized value $f_D(x_0)$. In the corresponding regular nonparametric regression model the Hellinger constraint is given by $\|\vartheta\|_{L_X^2}^2 \le \sigma^2/n$ and we use $\|\vartheta\|_{L^2} \ge \|\vartheta\|_{-1/2,\infty}$ by duality from $\|\vartheta\|_{L^2} \le \|\vartheta\|_{1/2,1}$ to obtain the interpolation inequality

$$\|\vartheta\|_{k,\infty} \le \|\vartheta\|_{-1/2,\infty}^{(s-k)/(s+1/2)}\|\vartheta\|_{s,\infty}^{(k+1/2)/(s+1/2)}$$
$$\le \text{const.}(\sigma^{-2}n)^{-(s-k)/(2s+1)}L^{(k+1/2)/(s+1/2)},$$

which similarly reveals the minimax rate in the regular case. Very roughly, we might therefore say that the PPP noise induces a regularity -1 in the Hölder scale, while the Gaussian white noise leads to the higher regularity $-1/2$. In analogy with $\sigma/\sqrt{n}$ in the regular case we might call $1/(nJ)$ the *noise level* for the regression problem with irregular noise and $nJf_D(x_0)$ the *effective local sample size* at x_0.

8.3 One-sided frontier estimation

In many of the applications mentioned in the introduction, the noise density f_ε has just one jump and not two as in our model $\mathcal{A}_n$. We want to stress that our proof of asymptotic equivalence can also cover the one-jump case. To make the analogy clear, let us assume that f_ε is still a density on $[-1, 1]$ with $f_\varepsilon(-1) > 0$ and $f_\varepsilon(1) = 0$. Instead of positivity and Lipschitz continuity, we now require f_ε to be Lipschitz continuous and Hellinger differentiable on $[-1, 1]$, i.e. $\sqrt{f_\varepsilon}$ is weakly differentiable with derivative in $L^2([-1, 1])$. Note that f_ε can then be extended to a function φ on the real line with the same local properties. All other properties of the model $\mathcal{A}_n$ are kept the same.

For the pilot estimator in this model we can obtain the same convergence rates when we select that admissible local polynomial which is the smallest at x_0. Lemma 4.1 remains the same, while in Definition 5.1 of experiment $\mathcal{D}_n$ we adjust only the left boundary of the density and set

$$f_{W,j}(x) = \varphi(x) \left(\int_{\Delta_{0,j,n}-1}^{\Delta_{0,j,n}+1} \varphi(t)dt \right)^{-1} 1_{[\Delta_{0,j,n}-1,\infty)}(x), \quad j \in J_n.$$

Lemma 5.1 then remains true as well, using the Hellinger differentiability in the proof instead of the uniform positivity. From the form of the density of W we conclude this time that the local minima $s_{k,n} = \min\{W_{j,n} : x_{j,n} \in I_{k,n}\}$, $k = 0, \ldots, m-1$, are conditionally sufficient. Then the remaining results remain all valid if we just consider $s_{k,n}$ instead of $(s_{k,n}, S_{k,n})$ and merely the upper PPP model. Consequently, this establishes asymptotic equivalence with the PPP X_2 of experiment $\mathcal{B}_n$. In this PPP model the regression function ϑ appears as the lower frontier of a Poisson point process with intensity $f_D(x)nf_\varepsilon(-1)$ on its epigraph. Frontier estimation where the support of f_ε is on $[-1, \infty)$ or $(-\infty, 1]$, respectively, can be treated analogously.

8.4 Counterexample for regularity one

We give a short argument that for equidistant design $x_{j,n} = \frac{j-1}{n-1}$ and parameter classes Θ where the target function $\vartheta \in \Theta$ is required to satisfy $\|\vartheta'\| \leq C$ for some $C > 0$ the experiments $\mathcal{A}_n$ and $\mathcal{B}_n$ are not asymptotically equivalent. Whether Hölder classes of order $1 + \alpha$ instead of $2 + \alpha$ suffice as parameter sets for establishing asymptotic equivalence remains a challenging open question.

Let us consider the function $f_n(x) = C(\pi(n-1))^{-1} \sin(\pi(n-1)x)$ so that $\|f_n'\|_\infty = C$ holds for all $n \geq 1$. Now observe that f_n satisfies $f_n(x_{j,n}) = 0$ for all $j = 1, \ldots, n$. This means in particular that in the regression experiment $\mathcal{A}_n$ the observations with regression function f_n cannot be distinguished from those with zero regression function. In experiment $\mathcal{B}_n$, however, a test between $H_0 : \vartheta = 0$ and $H_1 : \vartheta = f_n$ of the form $T_n = 1\{X_1([0,1] \times \mathbb{R}^+) > 0 \text{ or } X_2([0,1] \times \mathbb{R}^-) > 0\}$ satisfies $P_0(T_n = 0) = 1$ and

$$P_{f_n}(T_n = 1) = 1 - \exp\left(-n \int_0^1 |f_n(x)|dx\right) = 1 - \exp(-2C\pi^{-2}n(n-1)^{-1})$$

$$\to 1 - \exp(-2C/\pi^2) > 0,$$

for $n \to \infty$. Consequently, testing between H_0 and H_1 in experiment $\mathcal{B}_n$ is possible with non-trivial power uniformly over n. This implies that experiments $\mathcal{A}_n$ and $\mathcal{B}_n$ are asymptotically non-equivalent.

References

1. Brown, L.D., Low, M.: Asymptotic equivalence of nonparametric regression and white noise. Ann. Stat. **24**, 2384–2398 (1996)
2. Brown, L., Cai, T., Low, M., Zhang, C.-H.: Asymptotic equivalence theory for nonparametric regression with random design. Ann. Stat. **30**, 688–707 (2002)
3. Brown, L., Cai, T., Zhou, H.H.: Nonparametric regression in exponential families. Ann. Stat. **38**, 2005–2046 (2010)
4. Brown, M.: Discrimination of Poisson processes. Ann. Math. Stat. **42**, 773–776 (1971)
5. Carter, A.: A continuous Gaussian approximation to a nonparametric regression in two dimensions. Bernoulli **12**, 143–156 (2006)
6. Carter, A.: Asymptotic approximation of nonparametric regression experiments with unknown variances. Ann. Stat. **35**, 1644–1673 (2007)
7. Carter, A.: Asymptotically sufficient statistics in nonparametric regression experiments with correlated noise. J. Prob. Stat. **2009**, 19 (2009) (ID 275308)
8. Chernozhukov, V., Hong, H.: Likelihood estimation and inference in a class of nonregular econometric models. Econometrica **72**, 1445–1480 (2004)
9. DeVore, R.A., Lorentz, G.G.: Constructive Approximation, Grundlehren Series, vol. 303. Springer, Berlin (1993)
10. Gijbels, I., Mammen, E., Park, B., Simar, L.: On estimation of monotone and concave frontier functions. J. Am. Stat. Assoc. **94**, 220–228 (1999)
11. Grama, I., Nussbaum, M.: Asymptotic equivalence for nonparametric generalized linear models. Prob. Theor. Rel. Fields **111**, 167–214 (1998)
12. Grama, I., Nussbaum, M.: Asymptotic equivalence for nonparametric regression. Math. Methods Stat. **11**(1), 1–36 (2002)
13. Hall, P., van Keilegom, I.: Nonparametric "regression" when errors are positioned at end-points. Bernoulli **15**, 614–633 (2009)
14. Ibragimov, I.A., Hasminskii, R.Z.: Statistical Estimation. Asymptotic Theory. Springer, New York (1981) (translated from the Russian by S. Kotz)
15. Janssen, A., Marohn, D.M.: On statistical information of extreme order statistics, local extreme value alternatives and Poisson point processes. J. Multivar. Anal. **48**, 1–30 (1994)
16. Karr, A.F.: Point Processes and Their Statistical Inference. 2nd edn. Marcel Dekker, New York (1991)
17. Knight, K.: Limiting distributions of linear programming estimators. Extremes **4**, 87–103 (2001)
18. Korostelev, A.P., Tsybakov, A.B.: Minimax Theory of Image Reconstruction. In: Lecture Notes in Statistics, vol. 82. Springer, New York (1993)
19. Kutoyants, Y.A.: Statistical Inference for Spatial Poisson Processes. In: Lecture Notes in Statistics, vol. 134. Springer, New York (1998)
20. Le Cam, L.M.: Sufficiency and approximate sufficiency. Ann. Math. Stat. **35**, 1419–1455 (1964)
21. Le Cam, L.M., Yang, G.L.: Asymptotics in Statistics, Some Basic Concepts. 2nd edn. Springer, New York (2000)
22. Liese, F.: Eine informationstheoretische Bedingung für die Äquivalenz unbegrenzt teilbarer Punktprozesse. Math. Nachr. **70**, 183–196 (1975)
23. Liese, F., Vajda, I.: Convex Statistical Distances. Teubner, Leipzig (1987)
24. Müller, U.U., Wefelmeyer, W.: Estimation in nonparametric regression with nonregular errors. Comm. Stat. Theor. Methods **39**, 1619–1629 (2010)
25. Nussbaum, M.: Asymptotic equivalence of density estimation and Gaussian white noise. Ann. Stat. **24**, 2399–2430 (1996)
26. Reiß, M.: Asymptotic equivalence for nonparametric regression with multivariate and random design. Ann. Stat. **36**, 1957–1982 (2008)
27. Reiß, R.-D.: A Course on Point Processes. Springer, New York (1993)
28. Tsybakov, A.B.: Introduction to Nonparametric Estimation. Springer Series in Statistics, Berlin (2009)
29. van de Geer, S.A.: Empirical Processes in M-Estimation. Reprint, Cambridge University Press, New York (2006)