DISPERSAL DENSITY ESTIMATION ACROSS SCALES

Marc Hoffmann, Mathias Trabs

DISPERSAL DENSITY ESTIMATION ACROSS SCALES

By Marc Hoffmann[†] and Mathias Trabs[*,†]

Université Paris-Dauphine and Karlsruhe Institute of Technology

We consider a space structured population model generated by two point clouds: a homogeneous Poisson process M with intensity $n \to \infty$ as a model for a parent generation together with a Cox point process N as offspring generation, with conditional intensity given by the convolution of M with a scaled dispersal density $\sigma^{-1} f(\cdot/\sigma)$. Based on a realisation of M and N, we study the nonparametric estimation of f and the estimation of the physical scale parameter $\sigma > 0$ simultaneously for all regimes $\sigma = \sigma_n$. We establish that the optimal rates of convergence do not depend monotonously on the scale and we construct minimax estimators accordingly whether σ is known or considered as a nuisance, in which case we can estimate it and achieve asymptotic minimaxity by plug-in. The statistical reconstruction exhibits a competition between a direct and a deconvolution problem. Our study reveals in particular the existence of a least favourable intermediate inference scale, a phenomenon that seems to be new.

1. Introduction.

1.1. *Statistical inference across scales.* Data may behave differently at different scales. Depending on the interplay between the information parameter (the number of observations, the inverse of a noise level or the time length of measurement) and the physical scale of the data, the structure of the underlying statistical model may exhibit different properties. We may encode this idea by viewing rather a statistical experiment as family

$$\mathcal{E} = \big\{ \mathsf{P}^n_{f,\sigma} : f \in \Theta \big\}_{n \geq 1, \sigma > 0} \tag{1}$$

where the probability measures $\mathsf{P}^n_{f,\sigma}$ are simultaneously indexed by the information parameter $n \geq 1$ and a physical scale $\sigma > 0$, and that we shall refer to as a (family of) statistical experiment(s) across scales. Depending on the choice of the scale $\sigma = \sigma_n$ varying with the information n, the statistical geometrical properties of $\mathcal{E}$ (such as LAN type conditions or asymptotic equivalence features Le Cam (2012); van der Vaart (2002)) may differ. In particular, the choice of an optimal procedure may be dictated by different regimes governed by σ_n.

It is therefore desirable to understand the larger picture given by (1) *simultaneously for all subsequences* σ_n. In an asymptotic setting, we may attempt to realise the following program:

[*]Corresponding author.

[†]We thank our colleagues Marie Doumic and Alexander Goldenshluger for helpful discussions. We are grateful to the comments of three referees that convinced us to extend the results of a former version to the case of an unknown scale parameter. M.T. has been financially supported by DFG via the Heisenberg grant TR 1349/4-1.

AMS 2000 subject classifications: Primary 62G05; secondary 62G07, 62M30, 60G57

Keywords and phrases: Nonparametric estimation and minimax theory, point processes, statistical inference across scales, dispersal models, deconvolution

- Identify the optimal estimation for f (in an asymptotic minimax sense for a given loss function) for an arbitrary (but known) $\sigma = \sigma_n$.
- Considering $\sigma = \sigma_n$ as unknown, estimate simultaneously σ and f and achieve optimality for f in this setting.

In this paper, we build a family of statistical experiments across spatial scales that exhibit nontrivial behaviours at certain critical levels and for which different estimation procedures with different rates of convergence enter into competition as the scale varies. This can be of crucial importance in practice, and is in stark contrast with the results in Duval and Hoffmann (2011); Nickl et al. (2016); Chorowski (2018) where some robustness of estimation methods and of the minimax rates of convergence is observed across time scales for Lévy and diffusion processes.

1.2. *A model for dispersal estimation.*

Informal description. We start with two random points $X, Y \in \mathbb{R}^d$, where $X \in \mathcal{O}$ for some domain $\mathcal{O} \subseteq \mathbb{R}^d$ represents the trait of a parent in a spatially structured population, and $Y \in \mathbb{R}^d$ is the location of (one of) its children. We are interested in recovering the dispersal distribution of $Y - X$. This means that $Y - X$ has a density function

$$f_\sigma(z) = \sigma^{-d} f(z/\sigma), \qquad z \in \mathbb{R}^d, \tag{2}$$

with a physical (dispersal) scale parameter $\sigma > 0$ which determines the order of $\mathbb{E}[|Y-X|^2]^{1/2}$, where $|\cdot|$ denotes the Euclidean distance. The parameter of interest is the density function f. If we observe an n-sample $(X_i, Y_i)_{1\le i\le n}$, the $Y_i - X_i$ have common distribution f_σ, and we are in a classical density estimation framework; the scale σ is irrelevant. Assume now that we are rather given two point clouds $\mathcal{X}$ and $\mathcal{Y}$ in $\mathbb{R}^d$, with

$$\mathcal{X} = \{X_i : i = 1, \dots, n\} \text{ and } \mathcal{Y} = \{Y_j : j = 1, \dots n\},$$

i.e. we do not know the match between a parent and its offspring, hence we do not observe the variables $Y_i - X_i$ anymore. This is the topic of the paper.

The scale parameter σ now becomes crucial. Heuristically, if $\sigma \ll n^{-1/d}$, *i.e.* the dispersal scale is small with respect to the typical distance between the locations of the parents population $\mathcal{X}$, then we may guess the parents-offspring match by a nearest distance procedure, *i.e.* take $X_{(j)}$ solution to

$$|Y_j - X_{(j)}| = \min\{|Y_j - X_i| : i = 1, \dots, n\},$$

and proceed as if the $Y_j - X_{(j)}$ were an n-sample with distribution f_σ, up to controlling the mismatch error. However, if $\sigma \gg n^{-1/d}$, the mismatch error explodes and alternative methods need to be found. For instance, in dimension $d = 1$, for a child trait Y_j with parent trait X_{i_j}, writing $Y_j = X_{i_j} + \sigma D_j$, we see that D_j has density f, therefore, if the parent distribution p is known, then the Y_j have common distribution $p * f_\sigma$, where $*$ denotes convolution. We may then implement a deconvolution approach to recover f and ignore the potential information given by the point cloud of the parent traits $\mathcal{X}$. This has some price, namely the ill-posedness of an inverse problem, and has to be assessed with some care. Our objective is to formalise

this model and these approaches in order to encompass potential applications as described in Section 1.3 below. In particular, we need not impose that $\mathcal{X}$ and $\mathcal{Y}$ have the same size, allowing for a random number of parents and children. To provide a complete and transparent picture, we will greatly simplify our approach by restricting ourselves to the one-dimensional case $d = 1$, with $\mathcal{O} = [0, 1]$. Extensions to more general domains $\mathcal{O}$ for the state space of the parents as well as in higher dimension $d > 1$ are available and discussed in Section 4.

Formal construction of the model. Random point clouds are equivalently represented by random finite point measures. The location traits of the parent generation, *i.e.* the point cloud $\mathcal{X} \subseteq \mathcal{O} = [0, 1]$ is represented by the realisation of a homogeneous Poisson point process

$$M(\mathrm{d}x) = \sum_j \delta_{X_j}(\mathrm{d}x)$$

on the unit interval $[0, 1]$ with intensity measure $m(\mathrm{d}x) = n\lambda \mathrm{d}x$, where $n \to \infty$ and $\lambda > 0$ is fixed. Note that the size $|\mathcal{X}|$ is random, with $\mathbb{E}[|\mathcal{X}|] = n\lambda$. Given a realisation of M, the point cloud $\mathcal{Y} \subseteq \mathbb{R}$ that represent the traits of the offspring is generated by a Cox point process

$$N(\mathrm{d}y) = \sum_j \delta_{Y_j}(\mathrm{d}y)$$

with (conditional) intensity measure

$$\mu(M * f_\sigma)(y)\mathrm{d}y = \sum_i \mu f_\sigma(y - X_i)\mathrm{d}y,$$

where the *dispersal density* is $f_\sigma = \sigma^{-1} f(\cdot/\sigma)$ as in (2) with *dispersal scale* parameter $\sigma > 0$. The parameter $\mu > 0$ represents the average number of an offspring given one parent. Hence, f_σ describes the distribution of the random variable $Y_j - X_{i_j}$ (when the child j has parent i_j). The distance between the traits of the children and the trait of their parent is of order σ. The expected size of the offspring population (*i.e.* the average size of $\mathcal{Y}$) is thus $n\lambda\mu$. In Figure 1 we simulate a realisation of the (M, N) process, for different values of $\sigma = \sigma_n$ depending on n.

Keeping up with Section 1.1, we study a statistical experiment of the form (1), generated by the observation of $(\mathcal{X}, \mathcal{Y})$ or equivalently (M, N), with information n and scale σ. The unknown parameter is f and, for simplicity, λ, μ are assumed to be known. Our aim is to reconstruct f asymptotically in a minimax sense as $n \to \infty$, for all scaling regimes $\sigma = \sigma_n$.

1.3. *Dispersal inference in applications.* We briefly present some specific application domains compatible with the approach of dispersal inference as described in Section 1.2. Admittedly, further adjustments may be needed in order to be directly applicable in a given application; in particular, there might be a natural one-to-one correspondence between parents and children or not.

Example 1: Service time estimation in M/G/∞ queuing models. Dating back to Brown (1970), certain M/G/∞ queuing models are embedded in our approach, see in particular the recents results of Goldenshluger (2016, 2018); Goldenshluger and Koops (2019). Here, the state space of the parents $\mathcal{O} \subseteq \mathbb{R}$ represents time. The parent location trait is identified with an input arrival of a request to a server, according to Poisson arrivals at rate λ. Once the

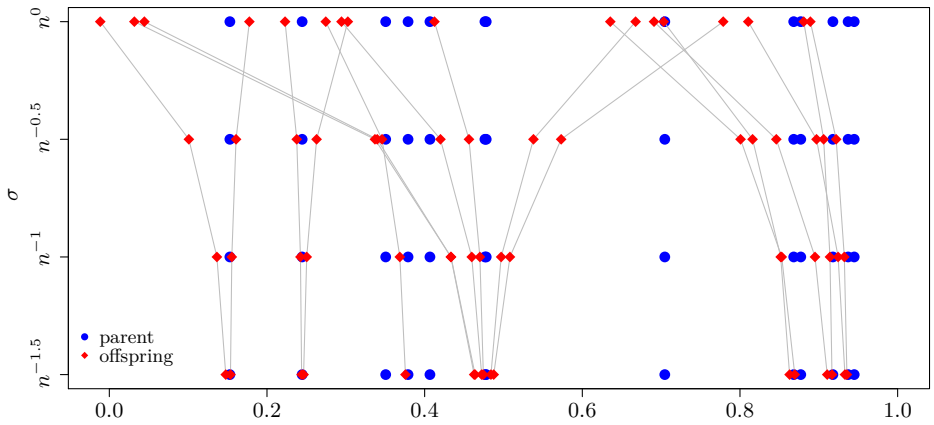

FIG 1. A realisation of (M, N) for different values of $\sigma = \sigma_n = n^{-a}$, with $a = 0, 0.5, 1, 1.5$. The match between parents traits (blue points) and their offspring traits (red diamonds) is graphically obvious for small $\sigma_n = n^{-1.5}$ but becomes more difficult if not impossible as σ_n increases. In the statistical experiment generated by (M, N), we are only given one horizontal line at a scale σ_n.

(random) service time of the request is fullfilled, an output is observed, that corresponds to the location trait of an offspring. See for instance Baccelli et al. (2009) where the emphasis is put on queueing systems where the service time cannot be observed. A small σ compared to $1/\lambda$ indicates that the service time is small compared to the order of magnitude of a typical interarrival between two queries, in which case one may take the time between an input and an output as a proxy for the service time. Otherwise, this is no longer true and alternative methods must be sought. Most aforementioned studies assume $\sigma\lambda = 1$. The case where σ is larger than $1/\lambda$ has been adressed by Blanghaps et al. (2013) where it is still required that $\lambda\sigma$ is bounded.

The goal is to estimate the density of service time, that matches exactly with the dispersal density f of our model. However, in the M/G/∞, model, to an incoming call, one associate one output exactly, which is slightly more stringent than having $\mu = 1$ only. See also Hall and Park (2004) and Section 4.1 for a more specific discussion in that direction.

Example 2: Poisson random convolution in functional genomics. This is actually the application that originated our approach, following informal discussions with our colleague Marie Doumic, that are formulated in Hunt et al. (2018). See also the recent work by Bonnet et al. (2022). The objective is to propose a model of distance interaction between motifs (or occurences of transcription regulatory elements) along DNA sequences. Related literature using point processes alternatives is developed for instance in Gusto and Schbath (2005); Carstensen et al. (2010).

Dispersal inference proposes an alternative approach compatible with Hunt et al. (2018), at least at a conceptual level: along the DNA sequence, transcription binding sites are observed according to a Poisson rate λ and serve as a parent generation model. Conditional to their parent location, transcription start sites (TSS) along the sequence are drawn via a random distribution f that we wish to infer, the dispersal distribution. Depending on the dispersal scale σ, we are back to our original problem and we obtain a continuous nonparametric alternative to the model described in Hunt et al. (2018).

Example 3: Dispersal distance in plants genetics. Introgression from cultivated to wild plants is a challenging problem for evolutionary ecology, especially in the context of genetically engineered crops. The study of gene flow from crops to wild relatives starts with understanding the typical dispersal distribution – in a spatial sense – between plants and their offspring. Although our model is too simple to account for various heterogeneity in natural environment, we emphasise some encouraging similarities: in the study of Arnaud et al. (2003), plants of interest and their offspring are distributed along a river bank. This accounts for a dispersal density a state space understood as a one-dimensional manifold (a curve), which is similar (and rate equivalent) to estimating a one-dimensional density, cf. Berenfeld and Hoffmann (2021).

Beyond the specific case of measuring dispersal along such idealised geometric features, the problem of estimating the distance between parents and saplings (accounting for seed dispersal from maternal or paternal parents plus pollen movement) is explicitly addressed in Isagi et al. (2000) via microsatellite analysis. Other plant based dispersal issues such as seed versus pollen dispersal from spatial genetic structure are discussed in Heuertz et al. (2003), see also Lavorel et al. (1995) and the references therein.

Example 4: Estimating diffusivity based on counting occupation numbers of particles. In a suspension of particles in a fluid, a Poissonian number of particles is recorded as they enter a fixed domain $\mathcal{A}$ and likewise when they leave $\mathcal{A}$. Applications in fluctuation spectroscopy enables one to infer the diffusivity (or other parameters) from such counting data, assuming that the particles have velocity $(V_t)_{t\geq 0}$ with random dynamics governed by a diffusion process

$$\mathrm{d}V_t = -\beta V_t \mathrm{d}t + \sqrt{2\beta D}\mathrm{d}W_t$$

where $(W_t)_{t\geq 0}$ is a Wiener process and $\beta > 0$ is a thermal relaxation parameter. There exist explicit formulae that relate the sojourn time of a particle within $\mathcal{A}$ and the diffusivity D, when the process is at equilibrium, see in particular Bingham and Dunham (1997). We thus have a typical dispersal inference problem, where the dispersal density corresponds here to the sojourn time of the particles. See also the recent paper by Goldenshluger and Jacobovic (2021).

1.4. *Results and organisation of the paper.* We first analyse the interaction between parents $\mathcal{X}$ and children $\mathcal{Y}$ via the correlation structure between the measures $M(\mathrm{d}x)$ and $N(\mathrm{d}y)$. In Proposition 1 in Section 2.1 below, building on the approach of Goldenshluger (2018), we establish the formula

$$\frac{1}{n\lambda\mu}\mathbb{E}\Big[\frac{M(\mathrm{d}x)N(\mathrm{d}y)}{\mathrm{d}x\mathrm{d}y}\Big] = n\lambda(f_\sigma * p)(y) + f_\sigma(x-y), \tag{3}$$

where $p = \mathbb{1}_{[0,1]}$ denotes the density function of the parent distribution. Formula (3) reveals the competition between a direct approach and a convolution problem, as mentioned above. From the observation of (M, N), we have access to empirical averages of the form

$$\sum_{i,j}\varphi(X_i, Y_j) = \int_{[0,1]\times\mathbb{R}}\varphi(x,y)M(\mathrm{d}x)N(\mathrm{d}y),$$

for test functions φ. We can take advantage of the information given by the first term in the right-hand side of (3) by picking φ of the form $\varphi(x,y) = \psi(y)$ and thus ignoring the

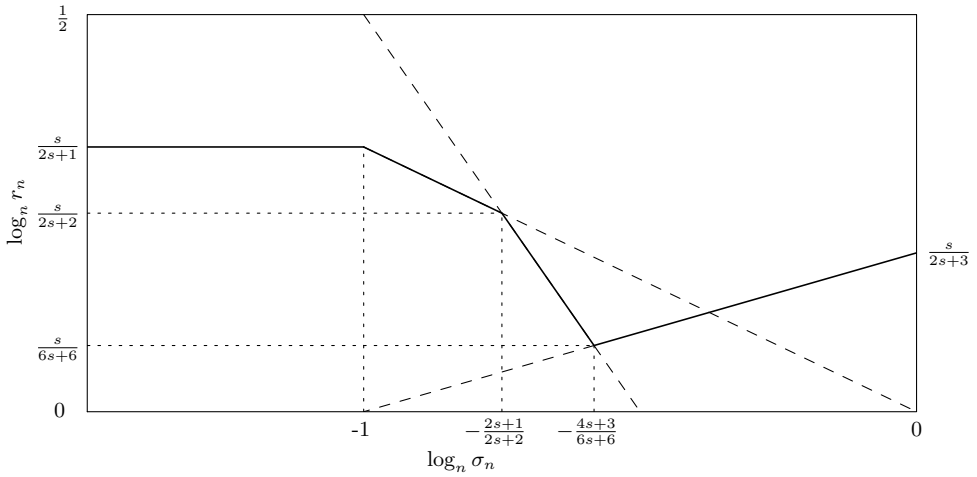

FIG 2. Dependence of the optimal convergence rate r_n on the dispersal rate σ_n in a $(\log_n \sigma_n, \log_n r_n)$-plot.

information given by the parents generation. For the second integral, we pick φ of the form $\varphi(x,y) = \psi((x-y)/\sigma))$ and we can take benefit from the interplay between the parent generation and its offspring. This results in generic estimators of the form

$$\sum_{i,j} \varphi^\star(Y_j, (X_i - Y_j)/\sigma) \tag{4}$$

for a specific choice of $\varphi^\star$. Whereas these heuristics give an overall flavour of the statistical model structure, the general situation is more subtle. In Section 2.1, we elaborate on the properties of the point process (M, N) to obtain an estimator of $f(z_0)$ for an arbitrary point $z_0 \in \mathbb{R}$. It takes the form

$$\widehat{f}^\star_{h_1,h_2}(z_0) = \begin{cases} \dfrac{1}{n\lambda h_1}\widehat{f}_{h_1,h_2}(z_0), & \text{for large scales,} \\ \dfrac{1}{h_2}\widehat{f}_{h_1,h_2}(z_0) - \sigma n\lambda, & \text{for small scales} \end{cases}$$

where

$$\widehat{f}_{h_1,h_2}(z_0) = \frac{1}{n\lambda\mu\sigma h_1}\sum_{i,j} \psi'\Big(\frac{z_0}{h_2} - \frac{Y_j}{\sigma h_2}\Big)\psi\Big(\frac{z_0}{h_1} - \frac{Y_j - X_i}{\sigma h_1}\Big)$$

is inspired by (4) for a suitable kernel ψ (and ψ' its derivative). For a suitable choice of the bandwidths h_1 and h_2, we prove in Theorem 6 that

$$\sup_f \mathbb{E}\Big[\big(\widehat{f}^\star_{h_1,h_2}(z_0) - f(z_0)\big)^2\Big]^{1/2} \lesssim r_n,$$

where the notation $A \lesssim B$ is equivalent to the Landau notation $A = O(B)$, the supremum is taken over Hölder balls of regularity $s > 0$ locally around z_0, and

$$r_n = \begin{cases} n^{-s/(2s+1)}, & \text{if } \sigma \leqslant n^{-1} \\ \sigma^{s/(2s+1)}, & \text{if } \sigma \in [n^{-1}, n^{-(2s+1)/(2s+2)}), \\ \sigma\sqrt{n}, & \text{if } \sigma \in [n^{-(2s+1)/(2s+2)}, n^{-(4s+3)/(6s+6)}), \\ (n\sigma)^{-s/(2s+3)}, & \text{if } \sigma \in [n^{-(4s+3)/(6s+6)}, 1]. \end{cases} \tag{5}$$

The shape of the rate of convergence $r_n = r_n(\sigma)$ as $\sigma = \sigma_n$ varies is illustrated in Figure 2. We prove in Theorem 7 that this result is indeed optimal:

$$\liminf_{n\to\infty} \inf_{\widehat{\vartheta}} \sup_{f} r_n^{-1} \mathbb{E}_f\Big[\big(\widehat{\vartheta} - f(z_0)\big)^2\Big]^{1/2} > 0,$$

where the infimum is taken over all estimators $\widehat{\vartheta}$ built upon the point clouds $\mathcal{X}$ and $\mathcal{Y}$, and the supremum is taken over Hölder balls of regularity $s > 0$ locally around z_0.

As illustrated in Figure 3, a direct estimation regime with minimax rate $r_n = n^{-s/(2s+1)}$ dominates for $\sigma \ll n^{-1}$ (the far left side of the picture), whereas for fixed σ, we have $r_n = n^{-s/(2s+3)}$, *i.e.* the minimax rate of convergence of an inverse problem of order one (the far right side of the picture). However, when σ_n slowly goes to 0, the inverse problem minimax rates deteriorates to $(n\sigma_n)^{-s/(2s+3)}$. Surprisingly, other regimes appear in the intermediate regime $\sigma_n \in [n^{-1}, n^{-(4s+3)/(6s+6)}]$. In particular, we find a worst case region, around the scale $\sigma_n \approx n^{-(4s+3)/(6s+6)}$ that yields the exotic minimax rate $n^{-s/(6s+6)}$. We discuss this phenomenon in detail in Section 2.2. In Section 3, we consider the case of an unknown scale $\sigma = \sigma_n$. We first show that it is possible to estimate σ so that we can ultimately decide whether $n\sigma$ is sufficiently large to apply the $n\sigma \to \infty$ asymptotics. We establish in particular in Section 3.1 a bound for the relative error $(\widehat{\sigma}-\sigma)/\sigma$ of our estimator $\widehat{\sigma}$. This is the gateway for a plug-in strategy to estimate f optimally when σ is unknown and considered as a nuisance parameter as we demonstrate in Section 3.2. The sensitivity of the plug-in estimator $\widehat{f}^{\star}_{h_1,h_2}(z_0) = \widehat{f}^{\star}_{h_1,h_2}(z_0)(\widehat{\sigma})$ is controlled via the smoothness of the process $\sigma \mapsto \widehat{f}^{\star}_{h_1,h_2}(z_0)(\sigma)$ via a chaining argument based on Kolmogorov-Chentsov criterion. We show that the optimal rates are achievable in probability.

The rest of the paper is organised as follows: In Section 2.1 we construct an estimator of $f(z_0)$ that takes $\mathcal{X}$ and $\mathcal{Y}$ as inputs and that adjusts to the scale $\sigma = \sigma_n$. Convergence rates for the estimator and matching minimax lower bounds are given in Section 2.2 and Section 2.3, respectively. The estimation of σ is studied in Section 3.1 while the estimation of f when σ is unknown via plug-in is undertaken in Section 3.2. A discussion with possible extensions is the content of Section 4. A short numerical simulation study is proposed in Section 5. All proofs are postponed to Section 6.

2. Main results.

2.1. *Construction of estimators across scales.*

The correlation structure of (M, N). The starting point of our estimation approach lies in the analysis of the correlation structure between M and N, that builds upon the approach by Goldenshluger (2018). In our study, the fact that the parent data $\mathcal{X}$ are distributed on a bounded interval has a considerable impact on the correlation structure.

PROPOSITION 1. *Let* $(A_i)_{1\le i\le I}$ *and* $(B_j)_{1\le j\le J}$ *be two families of disjoint subsets of* $[0,1]$ *and* $\mathbb{R}$*, respectively. Then for any* $(\eta_1,\dots,\eta_I) \in \mathbb{R}^I$ *and* $(\xi_1,\dots,\xi_J) \in \mathbb{R}^J$ *we have*

$$\log \mathbb{E}\Big[\exp\big(\sum_{i=1}^{I} \eta_j M(A_i) + \sum_{j=1}^{J} \xi_j N(B_j)\big)\Big] \tag{6}$$

$$= n\lambda \sum_{i=1}^{I} (e^{\eta_i} - 1)|A_i| + n\lambda \int_0^1 \Big(\exp\big(\mu \sum_{j=1}^{J} (e^{\xi_j} - 1) \int_{B_j} f_\sigma(y-x)\mathrm{d}y\big) - 1\Big)\mathrm{d}x$$
$$+ n\lambda \sum_{i=1}^{I} (e^{\eta_i} - 1) \int_{A_i} \Big(\exp\Big(\mu \sum_{j=1}^{J} (e^{\xi_j} - 1) \int_{B_j} f_\sigma(y-x)\mathrm{d}y\big) - 1\Big)\mathrm{d}x,$$

where $|A|$ denotes the Lebesgue measure of $A \subseteq \mathbb{R}$.

The proof relies on Campbell's exponential formula and is given in Section 6.1. Differentiating the result of Proposition 1, we obtain the following explicit representation of the correlation structure of M and N.

COROLLARY 2. *For any $A \subseteq [0,1]$ and any bounded $B \subseteq \mathbb{R}$ we have*

$$\mathbb{E}\big[M(A)N(B)\big] = n^2\lambda^2\mu|A| \int_0^1 \int_B f_\sigma(y-x)\,\mathrm{d}y\,\mathrm{d}x + n\lambda\mu \int_A \int_B f_\sigma(y-x)\,\mathrm{d}y\,\mathrm{d}x.$$

Corollary 2 reveals the infinitesimal correlation structure

$$\mathbb{E}[M(\mathrm{d}x)N(\mathrm{d}y)] = n\lambda\mu\big(n\lambda(f_\sigma * p)(y) + f_\sigma(y-x)\big)\mathrm{d}y\,\mathrm{d}x,$$

with $p = \mathbb{1}_{[0,1]}$. Applied to a well-behaved test function $\varphi\colon [0,1]\times\mathbb{R} \to \mathbb{R}$ we obtain

$$\frac{1}{n\lambda\mu}\mathbb{E}\Big[\int_{[0,1]\times\mathbb{R}} \varphi(x,y)M(\mathrm{d}x)N(\mathrm{d}y)\Big]$$
$$= n\lambda \int_0^1 \int_{\mathbb{R}} \varphi(x,y)(f_\sigma * p)(y)\mathrm{d}y\mathrm{d}x + \int_0^1 \int_{\mathbb{R}} \varphi(x, x+\sigma z) f(z)\mathrm{d}z\mathrm{d}x.$$

In order to obtain information on f from the first integral, $\varphi(x,y)$ should depend on y solely, while in the second integral, $\varphi(x,y)$ should rather depend on $(y-x)/\sigma$. We therefore pick a test function of the form

$$\varphi^\star(x,y) = \psi_1(y)\psi_2\big((y-x)/\sigma\big).$$

In the limit $n \to \infty$, we expect the empirical mean to be close to its expectation so that the approximation

$$\sum_{i,j} \varphi^\star(X_i, Y_j) \approx \mathbb{E}\big[\sum_{i,j} \varphi^\star(X_i, Y_j)\big]$$

is valid. Hence, we asymptotically have access to

$$\frac{1}{n\lambda\mu}\mathbb{E}\Big[\int_{[0,1]\times\mathbb{R}} \psi_1(y)\psi_2\big((y-x)/\sigma\big)M(\mathrm{d}x)N(\mathrm{d}y)\Big] = \sigma n\lambda\,\mathcal{U}_\sigma(f*p) + \mathcal{V}_\sigma(f), \tag{7}$$

where

$$\mathcal{U}_\sigma(f*p) = \int_{\mathbb{R}} \psi_1(y)(\psi_2 * \mathbb{1}_{[0,1/\sigma]})(y/\sigma)(f_\sigma * p)(y)\mathrm{d}y$$

and

$$\mathcal{V}_\sigma(f) = \int_{\mathbb{R}} (\psi_1 * \mathbb{1}_{[-1,0]})(\sigma z)\psi_2(z)f(z)\mathrm{d}z.$$

It is noteworthy that

$$|\mathcal{U}_\sigma(f*p)| \leqslant \|\psi_1\|_{L^1}\|\psi_2\|_{L^1} \text{ and } |\mathcal{V}_\sigma(f)| \leqslant \|f\|_\infty\|\psi_1\|_{L^1}\|\psi_2\|_{L^1},$$

showing that **1)** the functionals $\mathcal{U}_\sigma$ and $\mathcal{V}_\sigma$ are not sensitive to the order of magnitude of σ, and **2)** the influence of the test functions ψ_i is bounded in L^1-norm, hence they can be subsequently chosen as kernels that weakly converge to a Dirac mass as $n \to \infty$.

The deconvolution approach via $\mathcal{U}_\sigma(f*p)$. Pick $\psi_2 = 1$ as a constant function in (7) to obtain

$$\begin{aligned}\mathbb{E}[|\mathcal{X}|\sum_j \psi_1(Y_j)] &= n^2\lambda^2\mu\int_{\mathbb{R}}\psi_1(y)(f_\sigma * p)(y)\mathrm{d}y + n\lambda\int_0^1\int_{\mathbb{R}}\psi_1(x+\sigma z)f(z)\mathrm{d}z\mathrm{d}x\\ &= n^2\lambda^2\mu\Big(\int_{\mathbb{R}}\psi_1(y)(f_\sigma * p)(y)\mathrm{d}y + O(n^{-1})\Big),\end{aligned}$$

where $|\mathcal{X}| = M([0,1]) = \lambda n + O_{\mathbb{P}}(n^{1/2})$ is the total (random) number of parents. Ignoring remainder terms and using the fact that $|\mathcal{Y}| = N(\mathbb{R}) = n\lambda\mu + O_{\mathbb{P}}(n^{1/2})$ we also have the approximation

$$\frac{1}{|\mathcal{Y}|}\sum_j \psi_1(Y_j) \approx \int_{\mathbb{R}}\psi_1(y)(f_\sigma * p)(y)\mathrm{d}y. \tag{8}$$

The empirical estimate (8) is transparent: each child with trait Y_j has a parent with trait X_{i_j} such that $Y_j = X_{i_j} + \sigma D_j$, where D_j is distributed according to the dispersal density f. With $X_{i_j} \sim p$, we obtain $Y_j \sim f_\sigma * p$. However, the parent trait distribution p is uniform on $[0,1]$, its Fourier transform oscillates and vanishes on a discrete set, hence a classical deconvolution estimators based on spectral approaches cannot be readily applied. While there are some general constructions in the literature to overcome this problem (see *e.g.* Meister (2007); Delaigle and Meister (2011); Belomestny and Goldenshluger (2021) and the references therein), we take a more explict route.

We elaborate on the approach of Groeneboom and Jongbloed (2003), relying on the specific structure of a uniform $p = \mathbb{1}_{[0,1]}$. (The case of more general parent trait distribution is discussed in Section 4.2). Denoting by F the cumulative distribution of D_{ij} and writing $g_\sigma = f_\sigma * p$, we have

$$g_\sigma(y) = \int_{\mathbb{R}}\mathbb{1}_{[0,1]}(y-z)f_\sigma(z)\mathrm{d}z = \int_{\mathbb{R}}\mathbb{1}_{[y-1,y]}(z)f_\sigma(z)\mathrm{d}z = F\big(\tfrac{y}{\sigma}\big) - F\big(\tfrac{y-1}{\sigma}\big).$$

Hence, the representation

$$F\big(\tfrac{y}{\sigma}\big) = g_\sigma(y) + F\big(\tfrac{y-1}{\sigma}\big) = g_\sigma(y) + g_\sigma(y-1) + F\big(\tfrac{y-2}{\sigma}\big) = \cdots = \sum_{\ell\geq 0} g_\sigma(y-\ell) \tag{9}$$

is valid for all $y \in \mathbb{R}$. Based on the observation $\mathcal{Y}$, the density g_σ can be estimated at $z_0 \in \mathbb{R}$ by a kernel density estimator with kernel K and bandwidth $h > 0$

$$\widehat{g}_{\sigma,h}(z_0) = \frac{1}{|\mathcal{Y}|}\sum_j \frac{1}{h}K\Big(\frac{\sigma z_0 - Y_j}{h}\Big)$$

and then use representation (9) to obtain an estimator of $F(z_0)$ via

$$\widehat{F}_h(z_0) = \frac{1}{|\mathcal{Y}|}\sum_j\sum_{\ell\geq 0}\frac{1}{h}K\Big(\frac{\sigma z_0 - \ell - Y_j}{h}\Big). \tag{10}$$

Note that for compactly supported kernels the sum in ℓ is finite. For simplicity, we further consider the case where f is compactly supported and will adjust our assumptions accordingly. With no loss of generality, we assume

$$\operatorname{Supp} f \subseteq \Big[-\frac{1}{2}, \frac{1}{2}\Big]$$

so that only the term for $\ell = 0$ appears in the above sum.

Finally, we obtain a nonparametric estimator of $f(z_0)$ by taking the derivative of $\widehat{F}_h(z_0)$ in (10). We take a bandwidth $h = \sigma h_1$ that scales with σ that will prove technically convenient. We obtain a *deconvolution estimator* of $f(z_0)$ by setting

$$\widehat{f}^{\mathtt{dec}}_{h_1}(z_0) = \frac{1}{|\mathcal{Y}|}\sum_j \frac{1}{\sigma h_1^2} K'\Big(\frac{z_0}{h_1} - \frac{Y_j}{\sigma h_1}\Big).$$

We recover the representation (8) with

$$\psi_1 = \frac{1}{\sigma h_1^2} K'\Big(\frac{z_0}{h_1} - \frac{\cdot}{\sigma h_1}\Big). \tag{11}$$

Note that the convolution term in (7) is uninformative if M is a homogeneous Poisson point process on whole real line as in Goldenshluger (2018) or on the torus as in Hunt et al. (2018).

The interaction approach via $\mathcal{V}_\sigma(f)$. While the deconvolution approach ignores the information of the parents, an estimator based on the interaction of parents and their offspring can be constructed via ψ_2, taking now $\psi_1 = 1$ to be constant. From (7), we obtain

$$\begin{aligned}
&\mathbb{E}\Big[\sum_{i,j}\psi_2\Big(\frac{Y_j - X_i}{\sigma}\Big)\Big] \\
&= n\lambda\mu\int_{\mathbb{R}}\psi_2(z)f(z)\mathrm{d}z + n^2\sigma\lambda^2\mu\int_{\mathbb{R}}(\psi_2 * \mathbb{1}_{[0,1/\sigma]})(y/\sigma)(f_\sigma * p)(y)\mathrm{d}y \\
&= n\lambda\mu\Big(\int_{\mathbb{R}}\psi_2(y)f(z)\mathrm{d}z + O(n\sigma)\Big).
\end{aligned}$$

The bias is small only if $n\sigma$ is small, a result which is consistent with the heuristics of Section 1.2. Beyond that scale, as soon as $\sigma \approx n^{-1}$ the situation is a bit more involved.

More specifically, when $\sigma \ll n^{-1}$, the offspring traits concentrate around their parents: we expect roughly to have $n\sigma h$ parental traits in a σh-neighbourhood of the trait of a child $Y_j = X_{i_j} + \sigma D_j$. With overwhelming probability only the true parent trait X_{i_j} of Y_j is actually present in this neighbourhood. Then the sum in i over all parents in vanishes and we expect the approximation

$$\sum_{i,j}\psi_2\big((Y_j - X_i)/\sigma\big) \approx \sum_j \psi_2(D_j)$$

to be valid, while the sum is of order $n\lambda\mu$. More precisely, we expect

$$\frac{1}{|\mathcal{Y}|}\sum_{i,j}\psi_2\big((Y_j - X_i)/\sigma\big) \approx \int_{\mathbb{R}}\psi_2(z)f(z)\mathrm{d}z. \tag{12}$$

Applying kernel density estimator with kernel K with bandwidth $h_2 > 0$, we obtain an *interaction estimator* of $f(z_0)$ by setting

$$\widehat{f}^{\mathtt{int}}_{h_2}(z_0) = \frac{1}{|\mathcal{Y}|}\sum_{i,j}\frac{1}{h_2}K\Big(\frac{z_0}{h_2} - \frac{Y_j - X_i}{\sigma h_2}\Big). \tag{13}$$

We obtain the representation (12) with

$$\psi_2 = \frac{1}{h_2}K\Big(\frac{z_0}{h_2} - \frac{\cdot}{h_2}\Big). \tag{14}$$

Whenever $n\sigma \gtrsim 1$, the relevance of a procedure like (13) is less obvious. In particular, it is not clear whether the parent traits in the neighbourhood of an offspring can be used to estimate f. For $n\sigma = 1$ (and $\lambda = \mu = 1$) Brown (1970) constructed an estimator based on an explicit formula that relates f to the distribution function of the distance of an offspring to its nearest parents. An interaction estimator was also applied by Goldenshluger (2018) in a setting where the intensity measure of N is the Lebesgue measure on whole real line and which corresponds to $n\sigma = 1$. As we will see below the interaction estimator is still applicable if $\sigma > 1/n$ as long as σ is not too large. However, the interaction estimator then requires to incorporate a non-trivial kernel ψ_1 and a bias correction. A naive estimator could rely on simply counting the number of parents in a neighbourhood of the trait of a children, as already mentioned in Section 1.2.

An estimator across scales. Thanks to the heuristics developed for the construction of $\widehat{f}^{\mathtt{dec}}_{h_1}(z_0)$ and $\widehat{f}^{\mathtt{int}}_{h_2}(z_0)$, we are ready to implement an estimator across all scales σ, when the scale σ is known. We will treat the case of an unknown σ in Section 3 below. We first define

$$\widehat{f}_{h_1,h_2}(z_0) = \frac{1}{n\lambda\mu\sigma h_1}\sum_{i,j} K'\Big(\frac{z_0}{h_1} - \frac{Y_j}{\sigma h_1}\Big)K\Big(\frac{z_0}{h_2} - \frac{Y_j - X_i}{\sigma h_2}\Big),$$

where K is a smooth compactly supported kernel with derivative K'. We formally retrieve the preceding representation $\widehat{f}_{h_1,h_2}(z_0) = \frac{h_1h_2}{n\lambda\mu}\sum_{i,j}\psi_1(Y_j)\psi_2\big(\frac{Y_j - X_i}{\sigma}\big)$ with ψ_1 defined in (11) and ψ_2 in (14). Next, we elaborate on the properties we require for the kernel function K:

Assumption 3. *The function $K\colon \mathbb{R} \to \mathbb{R}$ is differentiable, symmetric, bounded and satisfies*

$$\mathrm{Supp}(K) \subseteq [-1,1], \;\; K(z) = 1 \;\; \textit{for } |z| \le \tfrac{1}{4},$$
$$\int_{[-1,1]} z^\ell K(z)\mathrm{d}z = \mathbb{1}_{\{\ell=0\}} \;\; \textit{for } \; \ell = 0,\dots,\ell_K$$

for some $\ell_K \ge 0$ (the order of the kernel K).

For $\ell_K = 0$ or 1, which is generally sufficient in practice, Assumption 3 is simply obtained from any suitably dilated and translated compactly supported symmetric (even) density function, see Section 5. Finally we define the appropriate normalisations and bias corrections that need to be tuned depending on which scale we work. This is done as follows:

Definition 4. *Let K be a kernel satisfying Assumption 3. We define the following estimators across scales:*

(i) (Deconvolution or large scales.) For $h_1 \in [(\sigma n)^{-1}, 1]$ and $h_2 = 8/\sigma$ set

$$\begin{aligned}\widehat{f}^{(1)}_{h_1}(z_0) &= \frac{1}{n\lambda h_1}\widehat{f}_{h_1,8/\sigma}(z_0)\\ &= \frac{1}{\sigma n\lambda\mu h_1^2}\sum_j K'\Big(\frac{z_0}{h_1} - \frac{Y_j}{\sigma h_1}\Big)\Big(\frac{1}{n\lambda}\sum_i K\Big(\frac{\sigma z_0}{8} - \frac{Y_j - X_i}{8}\Big)\Big).\end{aligned} \tag{15}$$

(ii) (direct or small scales.) Let $\sigma < 1/8$, $h_1 = 1/(2\sigma)$. We set for $h_2 \in (0,1]$:

$$\begin{aligned}\widehat{f}^{(2)}_{h_2}(z_0) &= \frac{1}{h_2}\widehat{f}_{1/(2\sigma),h_2}(z_0) - \sigma n\lambda\\ &= \frac{1}{n\lambda\mu h_2}\sum_{i,j} 2K'(2(\sigma z_0 - Y_j))K\Big(\frac{z_0}{h_2} - \frac{Y_j - X_i}{\sigma h_2}\Big) - \sigma n\lambda.\end{aligned} \tag{16}$$

For the deconvolution estimator $\widehat{f}^{(1)}_{h_1}(z_0)$ we could also set $\psi_2 = 1$. In this case the second factor in the right-hand side of (15) equals $\frac{|\mathcal{X}|}{n\lambda} \approx 1$ and we recover $\widehat{f}^{\mathsf{dec}}_{h_1}(z_0)$ from above. A similar simplication for $\widehat{f}^{(2)}_{h_2}(z_0)$ is not possible across all scales. As soon as $\sigma n \to \infty$, the small scales estimator crucially profits from the specific structure of ψ_1 which excludes all offspring traits Y_j outside of an annulus with radius of order $1/\sigma$, see Proposition 19*(ii)* for details.

2.2. *Rates of convergence.* Recall that, given some neighbourhood U_{z_0} of $z_0 \in \mathbb{R}$, the function $f\colon \mathbb{R} \to \mathbb{R}$ belongs to the local Hölder class $\mathcal{H}^s(z_0)$ with $s > 0$ if f is $\lfloor s \rfloor$ times continuously differentiable for every $z, z' \in U_{z_0}$ and

$$|f^{(\lfloor s \rfloor)}(z) - f^{(\lfloor s \rfloor)}(z')| \le C|z - z'|^{s - \lfloor s \rfloor} \tag{17}$$

where $\lfloor s \rfloor$ is the largest integer *stricty smaller* than s, and $f^{(n)}$ denotes n-fold derivation (with $f^{(0)} = f$). The definition depends on U_{z_0}, further omitted in the notation. We obtain a semi-norm $|f|_{\mathcal{H}^s(z_0)}$ by taking the smallest constant C for which (17) holds. Moreover, as explained in Section 2.1, we assume for technical convenience that f is bounded and supported in $[-\frac{1}{2}, \frac{1}{2}]$ which yields the following nonparametric class of densities[1]

$$\mathcal{G}^s(z_0, L) := \Big\{ f\colon \mathbb{R} \to [0,\infty) : |f|_{\mathcal{H}^s(z_0)} \leqslant L, \|f\|_\infty \leqslant L, \mathrm{Supp}(f) \subseteq [-\tfrac{1}{2}, \tfrac{1}{2}], \int f(z)\mathrm{d}z = 1 \Big\}.$$

We first exhibit rates of convergence for $\widehat{f}^{(1)}_{h_1}(z_0)$ and $\widehat{f}^{(2)}_{h_2}(z_0)$ of Definition 4 built upon $\widehat{f}_{h_1,h_2}(z_0)$.

Proposition 5.

(i) If $h_1 \leqslant 1 \leqslant \sigma n$, then we have for any $z_0 \in (-\frac{1}{2}, \frac{1}{2})$

$$\sup_{f \in \mathcal{G}^s(z_0, L)} \mathbb{E}\Big[\big(\widehat{f}^{(1)}_{h_1}(z_0) - f(z_0) \big)^2 \Big]^{1/2} \lesssim h_1^s + \big(n\sigma h_1^3\big)^{-1/2}, \tag{18}$$

up to a constant that depends on L, s, K and z_0. Choosing $h_1 = (n\sigma)^{-1/(2s+3)}$, we obtain the optimised rate

$$\mathbb{E}\Big[\big(\widehat{f}^{(1)}_{h_1}(z_0) - f(z_0) \big)^2 \Big]^{1/2} \lesssim (n\sigma)^{-s/(2s+3)}.$$

(ii) Let $\sigma < 1/8$. For any $h_2 \in (0, 1]$ and $z_0 \in (-\frac{1}{2}, \frac{1}{2})$, we have

$$\sup_{f \in \mathcal{G}^s(z_0, L)} \mathbb{E}\Big[\big(\widehat{f}^{(2)}_{h_2}(z_0) - f(z_0) \big)^2 \Big]^{1/2} \lesssim h_2^s + \max\big((nh_2)^{-1/2}, \sigma^{1/2} h_2^{-1/2}, n^{1/2}\sigma \big), \tag{19}$$

up to a constant that depends on L, s, K and z_0. Choosing $h_2 = (n \wedge \sigma^{-1})^{-1/(2s+1)}$, we obtain the optimised rate

$$\mathbb{E}\Big[\big(\widehat{f}^{(2)}_{h_2}(z_0) - f(z_0) \big)^2 \Big]^{1/2} \lesssim \max\big((n \wedge \sigma^{-1})^{-s/(2s+1)}, n^{1/2}\sigma \big).$$

[1]We omit a slight ambiguity in our definition: the neighbourhood U_{z_0} in the definition of $|f|_{\mathcal{H}^s(z_0)}$ is implicitly taken independently of f.

Some remarks are in order: **1)** The rate $(n\sigma)^{-2s/(2s+3)}$ in *(i)* reflects the ill-posedness of degree one due to the convolution with the indicator function. Moreover, we see that the rate is determined by $n\sigma$ instead of n solely: the information about f is concentrated at the boundary $[-\frac{\sigma}{2}, \frac{\sigma}{2}] \cup [1-\frac{\sigma}{2}, 1+\frac{\sigma}{2}]$ of the support of the parent distribution since in the interior we have $\mathbb{1}_{[0,1]} * f_\sigma(y) = 1$ for all $y \in (\frac{\sigma}{2}, 1-\frac{\sigma}{2})$. Since the number of children in this boundary is of order $n\sigma$, the latter can be understood as effective sample size. In particular, the convergence rate deteriorates for $\sigma \to 0$ and the deconvolution estimator is only consistent as long as $n\sigma \to \infty$. **2)** In *(ii)* we obtain the classical rate of convergence $n^{-2s/(2s+1)}$ for nonparametric density estimation as long as $\sigma n \lesssim 1$. For large scaling factors the bias correction $-\sigma n\lambda$ becomes crucial and the rate gets slower, *i.e.* the local interaction between parents and children becomes less informative. We obtain the rate $\sigma^{2s/(2s+1)} \vee (n\sigma^2)$. In particular, the interaction approach is only consistent as long as $\sigma = o(n^{-1/2})$. This limitation is a consequence of the non-negligible correlations between two different offspring traits in a σ-neighborhood of a parent. **3)** Interestingly, there is an intermediate regime $\sigma \in [n^{-1}, n^{-1/2}]$ where both approaches are applicable and we can choose the estimator with the faster rate.

We wrap together the results of Proposition 5 to obtain our main result:

THEOREM 6. *Let $s > 0$. For any $z_0 \in (-1/2, 1/2)$, there exists an estimator $\widehat{f}(z_0)$ depending on σ, λ, μ and s, explicitly obtained from Proposition 5 above such that*

$$\sup_{f \in \mathcal{G}^s(z_0, L)} \mathbb{E}\Big[\big(\widehat{f}(z_0) - f(z_0)\big)^2\Big]^{1/2} \lesssim r_n,$$

up to a constant that depends on L, s, K and y, and with rate of convergence from (5).

Some remarks again: **1)** The graph of $\log r_n$ as a function of τ for $\sigma = n^{-\tau}$ is illustrated in Figure 3. Quite surprisingly, the dependence of the convergence rate on the scaling parameter σ is not monotonic which is a consequence of (7): The information on f in the deconvolution term decreases if σ gets smaller, while the second information based on interaction decreases if σ gets larger. The elbows between the regimes correspond to the points where

$$\sigma = n^{-(2s+1)/(2s+2)} \quad i.e. \quad \sqrt{n}\sigma = \sigma^{s/(2s+1)}$$

and

$$\sigma = n^{-(4s+3)/(6s+6)} \quad i.e. \quad \sqrt{n}\sigma = (n\sigma)^{-s/(2s+3)}.$$

In particular, the best estimator uses the deconvolution approach if $\sigma > n^{-(4s+3)/(6s+6)}$ and the interaction approach otherwise. **2)** For the construction of the estimator, we need to know λ, μ and σ. A canonical estimator for λ is given by $\widehat{\lambda} = n^{-1}|\mathcal{X}| = n^{-1}M([0,1]) \sim n^{-1}\mathrm{Poiss}(\lambda n)$ satisfying $\mathbb{E}[|\widehat{\lambda}/\lambda - 1|^2] = (n\lambda)^{-1}$. Hence, it should not be a problem to replace λ by $\widehat{\lambda}$ in the weighting of $\widehat{f}_{h_1,h_2}$. Similarly, we can estimate μ. The scaling parameter σ is more critical, because even the parametric accuracy is not sufficient to construct an estimator which is adaptive in σ: We have to decide whether $\sigma > n^{-(4s+3)/(6s+6)}$ or not and the boundary $n^{-(4s+3)/(6s+6)}$ is $o(n^{-1/2})$ as soon as $s > 2$ (additionally the boundary dependence on the typically unknown regularity s). Since the usual construction principles for adaptive estimators rely on a monotonic dependence of the rate, more precisely of an upper bound for the stochastic error, the observed dependence on σ might complicate the construction of an adaptive estimator considerably.

2.3. *Minimax optimality.* For $\sigma \leqslant 1/n$ the rate $n^{-s/(2s+1)}$ is optimal: a lower bound is obtained by noting that it is more informative to observe the point cloud of the parent traits $\mathcal{Y}$ and the dispersal realisation (D_j) via a Poisson point process with intensity $n\lambda\mu f$. The offspring point N process can then the be constructed via uniformly distributing the children around the parents. Observing (D_j), the classical minimax rate for estimating $f(z_0)$ is $n^{-s/(2s+1)}$. Less obviously, for $n\sigma \to \infty$, the rate r_n is optimal too:

THEOREM 7. *Let $z_0 \in (-1/2, 1/2)$, $s > 0$ and $L > 0$. Suppose $\sigma n |\log \sigma|^{-1} \to \infty$ for $\sigma = \sigma_n \in (0,1)$ as $n \to \infty$. Then we have*

$$\liminf_{n\to\infty} \inf_{\widehat{\vartheta}} \sup_{f \in \mathcal{G}^s(z_0,L)} r_n^{-1} \mathbb{E}_f \Big[\big(\widehat{\vartheta} - f(z_0)\big)^2 \Big]^{1/2} > 0,$$

where the infimum is taken over all estimators $\widehat{\vartheta}$ built upon $\mathcal{X}$ and $\mathcal{Y}$, with r_n given by (5).

Hence, r_n is the minimax rate of convergence in squared pointwise error for nonparametric dispersal estimation. In order to prove a sharp lower bound, we have to show that the parents indeed become uninformative if $n\sigma \to \infty$ such that ignoring the X_i is then the best we can do. This argument is based on the following insight: Given a number of parents $|\mathcal{X}| \sim \mathrm{Poiss}(\lambda n)$ the distribution of a child trait Y, conditional on the parents satisfies:

$$\begin{aligned}\mathbb{P}(Y \in \mathrm{d}y \,|X_1, \dots, X_{|\mathcal{X}|}) &= \frac{1}{|\mathcal{X}|} \sum_{i=1}^{|\mathcal{X}|} f_\sigma(y - X_i)\mathrm{d}y \\ &= \int_0^1 f_\sigma(y-x)\mathrm{d}x\mathrm{d}y + \Big(\frac{1}{|\mathcal{X}|}\sum_{i=1}^{|\mathcal{X}|} f_\sigma(y - X_j) - \mathbb{E}[f_\sigma(y - X_1)]\Big)\mathrm{d}y,\end{aligned}$$

where, conditional on $|\mathcal{X}|$,

$$\mathrm{Var}\Big(\frac{1}{|\mathcal{X}|}\sum_{i=1}^{|\mathcal{X}|} f_\sigma(y - X_i)\Big||\mathcal{X}|\Big) = \frac{1}{|\mathcal{X}|}\mathrm{Var}(f_\sigma(y - X_1)) \leqslant \frac{1}{|\mathcal{X}|\sigma}\|f\|_{L^2}^2.$$

Since $\mathbb{E}[|\mathcal{X}|] = \lambda n$, the influence of the parent traits becomes uninformative if $n\sigma \to \infty$. See Section 6.3 for a rigorous proof.

3. The case of unknown scale parameter σ. In practice, it well may be the case that the scale σ is unknown itself. We address this issue in a two-steps strategy: first, we study the estimation of σ as a statistical problem in its own right. In particular, we need to be able to distinguish from the data in which regime we stand ($n\sigma \to \infty$ versus $n\sigma$ bounded). Also, we need an accurate estimator $\widehat{\sigma}$ of σ with respect to the relative error $\widehat{\sigma}/\sigma - 1$ since σ itself may vanish as $n \to \infty$. This is done in Section 3.1 below. Second, we use the estimator $\widehat{\sigma}$ and the associated decision rules to determine the underlying scale to cook-up a σ-adaptive and final estimator $\widetilde{f}(z_0)$ by plug-in that proves to be optimal in all regimes (in probability, for simplicity). This is done in Section 3.2 below.

3.1. *Estimation of σ.* In view of the different regimes for the estimation procedure the first question to settle is how to decide whether $n\sigma$ is sufficiently large to apply $n\sigma \to \infty$ asymptotics or not. To quantify $n\sigma$ empirically, we define

$$\widehat{T} := N(\mathbb{R} \setminus [0,1]).$$

Since the support of the offspring point process N is given by $[-\frac{\sigma}{2}, 1+\frac{\sigma}{2}]$ and the intensity of N is of the order n, we expect that $\widehat{T}$ is indeed of order $n\sigma$.

LEMMA 8. *Let* $\operatorname{Supp} f \subseteq [-\frac{1}{2}, \frac{1}{2}]$ *and set* $I_f := \int_{\mathbb{R}} |x| f(x) \mathrm{d}x$. *For any* $\sigma \in (0,1]$ *and* $n \in \mathbb{N}$ *we have*

$$\mathbb{E}[\widehat{T}] = n\sigma\lambda\mu I_f \qquad \text{and} \qquad \operatorname{Var}(\widehat{T}) \leqslant n\sigma\lambda(\mu+\mu^2).$$

In particular, Chebychev's inequality shows for any $\kappa > 0$ that

$$\mathbb{P}(\widehat{T} \leqslant \kappa) \to \begin{cases} 0, & \text{if } n\sigma \to \infty, \\ 1, & \text{if } n\sigma \leqslant \frac{\kappa}{2\lambda\mu I_f}. \end{cases}$$

We cannot use $\widehat{T}$ to estimate σ since the quantity I_f is unknown, but we can further exploit the support $[-\frac{\sigma}{2}, 1+\frac{\sigma}{2}]$ of the offspring location traits Y_j. Namely, we can construct a boundary type estimator for σ. We focus on left boundary, but the method can be easily modified to the right boundary or a combination of both. For $l \in \{1, \dots, |\mathcal{Y}|\}$ the order statistics are denoted by $Y_{(l)}$. A naive estimator for σ is thus $-2Y_{(1)} = -2\min_j Y_j$. We actually need to improve this estimator by taking the parent location traits near the left boundary into account. The resulting estimator is defined as

$$\widehat{\sigma}^{(1)} := -2Y_{(1)} + 2\overline{X}_{\widehat{(l)}} \qquad \text{with} \qquad \overline{X}_{(l)} := \frac{1}{l}\sum_{j=1}^{l} X_{(j)}, \qquad \widehat{l} := \kappa_n\sqrt{\widehat{T}}.$$

If f is bounded away from zero on its support $[-\frac{1}{2}, \frac{1}{2}]$, the corresponding c.d.f. F admits at least a linear growth at the boundary.

PROPOSITION 9. *Suppose the dispersal c.d.f.* F *satisfies*

$$F(z - \frac{1}{2}) \geqslant (\gamma z) \vee 0 \qquad \text{for } z \leqslant 1 \tag{20}$$

for some constant $\gamma > 0$. *If* $\sigma n \to \infty$ *and* $\widehat{\sigma}^{(1)}$ *is specified with some sequence* $\kappa_n \to \infty$ *that can be taken arbitrarily slowly diverging, then*

$$\frac{\widehat{\sigma}^{(1)}}{\sigma} - 1 = O_{\mathbb{P}}(\frac{\kappa_n}{\sqrt{\sigma n}}).$$

While for constant σ we obtain the typical parametric rate, the error bound is considerably improved if $\sigma \to 0$. Most importantly, the relative estimation error $\frac{\widehat{\sigma}^{(1)}-\sigma}{\sigma}$ is small as soon as $\sigma n \to \infty$.

The estimator $\widehat{\sigma}^{(1)}$ works well for large scale parameters. On the other side, when $\sigma \ll n^{-1}$, we almost can guess the relationship $Y_j = X_{i_j} + \sigma D_j$ between an offspring trait Y_j and its parent trait X_{i_j} such that we can estimate σ by a local boundary estimation approach around the distinct parent traits. To use this local information we specifically use the kernel

$$\psi^\dagger := (1-|x|)_+ = \mathbb{1}_{[-1/2,1/2]} * \mathbb{1}_{[-1/2,1/2]}$$

and write

$$\frac{1}{\mu\lambda n}\sum_{i,j} \psi^\dagger((Y_j - X_i)/h) =: \mathbb{P}(\sigma|D_1| \leqslant \frac{h}{2}) + n\lambda h + \xi(h),$$

where the stochastic noise term $\xi(h)$ defined via the last display satisfies

$$\begin{aligned}\mathbb{E}[\xi(h)] &= \mathbb{E}\Big[\frac{1}{\mu\lambda n}\sum_{i,j}\psi^\dagger\big((Y_j - X_i)/h\big)\Big] - \mathbb{P}\big(\sigma|D_1| \leqslant \frac{h}{2}\big) - n\lambda h \\ &= O(nh(\sigma+h)+h),\end{aligned}$$

cf. Step 1 in the proof of Lemma 24. Due to Proposition 20 we moreover have for $h \leqslant \sigma$:

$$\begin{aligned}\mathrm{Var}\left(\xi(h)\right) &\lesssim \frac{1}{n}\big((n\sigma + n^2\sigma^2)\frac{h^2}{\sigma^2} + (n\sigma+1)\frac{h}{\sigma}\big) \\ &\lesssim \frac{h^2}{\sigma} + nh^2 + h + \frac{h}{n\sigma} \lesssim nh^2 + \frac{1}{n}. \qquad (21)\end{aligned}$$

Since $h \mapsto \mathbb{P}(\sigma|D_1| \leqslant \frac{h}{2})$ is increasing and equals one as soon as h reaches σ, we define for some sequence $\kappa_n > 0$:

$$\widehat{\sigma}^{(2)} := \min\Big\{h > 0 : \frac{1}{\mu\lambda n}\sum_{i,j}\psi^\dagger\big((Y_j - X_i)/h\big) \geqslant n\lambda h + 1 - \sqrt{nh^2 + n^{-1}}\kappa_n\Big\}.$$

Proposition 10. *Suppose $n\sigma^{3/2} \to 0$ and (20). Then we have for some $\kappa_n \to \infty$ (arbitrarily slowly) that*

$$\frac{\widehat{\sigma}^{(2)}}{\sigma} - 1 = O_{\mathbb{P}}\big(\kappa_n\sqrt{n\sigma^2 + n^{-1}}\big).$$

Note that the condition $n\sigma^{3/2} \to 0$ exactly characterises the regime where the rate of $\widehat{\sigma}^{(2)}$ is faster than the rate of $\widehat{\sigma}^{(1)}$. Combining the estimators $\widehat{\sigma}^{(1)}$ and $\widehat{\sigma}^{(2)}$ with the decision rule $\{\widehat{T} > \kappa_n\}$ that enables us to decide whether we are in the regime $n\sigma \to \infty$ or not, we define our final estimator for σ as:

$$\widehat{\sigma} := \begin{cases} \widehat{\sigma}^{(1)} & \text{on} \quad \{\widehat{T} > \kappa_n\} \cap \{\widehat{\sigma}^{(1)} > n^{-2/3}\} \\ \widehat{\sigma}^{(2)} & \text{otherwise.}\end{cases} \qquad (22)$$

We conclude:

Theorem 11. *Under the boundary condition (20) the estimator $\widehat{\sigma}$ defined above with $\kappa_n = \log n$ satisfies*

$$\frac{\widehat{\sigma}}{\sigma} - 1 = O_{\mathbb{P}}\Big((\log n)\big(\sqrt{n\sigma^2 + n^{-1}} \wedge \frac{1}{\sqrt{n\sigma}}\big)\Big).$$

The performance of $\widehat{\sigma}$ in terms of its fluctuations in relative error are shown in Figure 3. They will be sufficient to implement an optimal scale adaptive plug-in strategy for the estimation of f, as developed in the next section.

3.2. *Estimation of f when σ is unknown.* Recall that the optimal rate of convergence is achieved by the interaction estimator or the deconvolution estimator depending whether

$$\sigma \leqslant n^{-(4s+3)/(6s+6)} \qquad \text{or} \qquad \sigma > n^{-(4s+3)/(6s+6)}$$

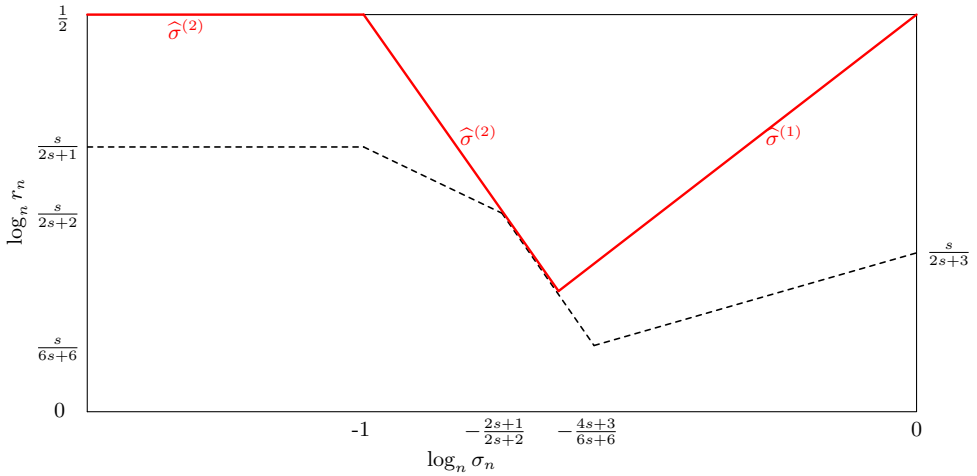

FIG 3. The rates of convergence of $\widehat{\sigma} = \widehat{\sigma}^{(1)}$ or $\widehat{\sigma}^{(2)}$ depending on the decision rule (22), as a function the dispersal rate σ on a log-log plot (in solid red). The minimax rate r_n in a log-log plot (in dashed black). The red curve always dominates the black one.

respectively. Theorem 11 implies that $\widehat{\sigma} n^{(4s+3)/(6s+6)} = (1+o_{\mathbb{P}}(1))\sigma n^{(4s+3)/(6s+6)}$ in all regimes for σ. In turn, we can decide for the best estimator in a data-driven way by setting

$$\widetilde{f_n}(z_0) = \begin{cases} \widetilde{f}^{(1)}_{\widehat{\sigma}}(z_0) & \text{on} \qquad \{\widehat{\sigma} n^{(4s+3)/(6s+6)} \geq 1\} \\ \\ \widetilde{f}^{(1)}_{\widehat{\sigma}}(z_0) & \text{otherwise.} \end{cases}$$

where $\widehat{\sigma}$ is specified in Theorem 11, and we use the plug-in counterparts to the deconvolution estimator from (15) and the interaction estimator from (16), respectively, given by

$$\widetilde{f}^{(1)}_{\widehat{\sigma}}(z_0) := \frac{1}{\widehat{\sigma}\widehat{h}_1^2|\mathcal{Y}|} \sum_j K'\Big(\frac{z_0}{\widehat{h}_1} - \frac{Y_j}{\widehat{\sigma}\widehat{h}_1}\Big)\Big(\frac{1}{|\mathcal{X}|}\sum_i K\Big(\frac{\widehat{\sigma} z_0}{9} - \frac{Y_j - X_i}{9}\Big)\Big),$$

specified by $\widehat{h}_1 = (n\widehat{\sigma})^{-1/(2s+3)}$, and

$$\widetilde{f}^{(2)}_{\widehat{\sigma}}(z_0) = \frac{1}{|\mathcal{Y}|\widehat{h}_2} \sum_{i,j} 2K'(2(\widehat{\sigma} z_0 - Y_j))K\Big(\frac{z_0}{\widehat{h}_2} - \frac{Y_j - X_i}{\widehat{\sigma}\widehat{h}_2}\Big) - \widehat{\sigma}|\mathcal{X}|,$$

specified by $\widehat{h}_2 = (n \wedge \widehat{\sigma}^{-1})^{-1/(2s+1)}$.

THEOREM 12. *Let $s, L > 0, z_0 \in (-\frac{1}{2}, \frac{1}{2})$ and suppose K fulfills Assumption 3 with order $\ell_K \geq \lfloor s \rfloor$. The following holds uniformly for $f \in \mathcal{G}^s(z_0, L)$ with property (20):*

(i) If $n\sigma/(\log n)^2 \to \infty$, we have

$$\widetilde{f}^{(1)}_{\widehat{\sigma}}(z_0) - f(z_0) = O_{\mathbb{P}}(u_n), \quad \text{with } u_n = (\sigma n)^{-s/(2s+3)}.$$

(ii) If $n\sigma^2 \to 0$ and $s > 3/2$, we have

$$\widetilde{f}^{(2)}_{\widehat{\sigma}}(z_0) - f(z_0) = O_{\mathbb{P}}(v_n), \quad \text{with } v_n = (n \wedge \sigma^{-1})^{-s/(2s+1)} + \sqrt{n}\sigma.$$

In particular, we achieve σ-adaptation in the following sense:

$$\widetilde{f_n}(z_0) - f(z_0) = O_{\mathbb{P}}(r_n),$$

where r_n is the minimax rate for the estimation of $f(z_0)$ in squared error loss, given in (5)*, according to Theorems 6 and 7.*

Our final Theorem 12 shows that under our set of assumptions, it is possible to achieve optimality for the pointwise estimation of the dispersal density $f(z_0)$ across scales without any prior knowledge of the scale σ. The proof is based upon the study of the smoothness of the interaction and convolution estimators as random processes indexed by σ together with sharp estimation rates for $\widehat{\sigma}$ provided by Theorem 11. For technical reason, we have the additional restriction $s > 3/2$ for the smoothness of f locally around z_0 and our bounds are in the probability and not expectation.

4. Discussion.

4.1. *One-to-one correspondence between parents and children.* In some applications, *e.g.* Example 1 in Section 1.3, it is desirable to impose a one-to-one correspondence between parents and their children. Each parent has exactly one child and in particular $|\mathcal{X}| = |\mathcal{Y}|$. The point process N of the offspring generation should be modified as follows. For $M = \sum_i \delta_{X_i}$ as parent generating process, the offspring of a specific parent trait X_i is given by

$$Y_j = X_j + \sigma D_j \tag{23}$$

for independent random variables $(D_j)_{j\geq 1}$ distributed according to the dispersal density f and with the scaling parameter $\sigma \in (0, 1]$. The offspring point process is then simply given by

$$N(\mathrm{d}y) = \sum_j \delta_{Y_j}(\mathrm{d}y) = \sum_j \delta_{X_j+\sigma D_j}(\mathrm{d}y).$$

Let us compare this one-to-one model with the original model of Section 1.2: Here, we have an urn model without replacement and a fixed number $|\mathcal{X}|$ of draws while in Section 1.2 we have an urn model with replacement and random number of draws. In this modified case we can proceed analogously to Goldenshluger (2018, Proposition 1) and we obtain the following counterpart to Proposition 1 with $\mu = 1$:

Proposition 13. *Let $(A_i)_{1\leq i\leq I}$ and $(B_j)_{1\leq j\leq J}$ be two families of disjoint subsets of $[0, 1]$ and $\mathbb{R}$, respectively. Then for any $(\eta_1, \dots, \eta_I) \in \mathbb{R}^I$ and $(\xi_1, \dots, \xi_J) \in \mathbb{R}^J$ we have*

$$\begin{aligned}&\log \mathbb{E}\Big[\exp\Big(\sum_{i=1}^{I} \eta_i M(A_i) + \sum_{j=1}^{J} \xi_j N(B_j)\Big)\Big] \\ &\quad = n\lambda \sum_{i=1}^{I} (e^{\eta_i} - 1)|A_i| + n\lambda \sum_{j=1}^{J} (e^{\xi_k} - 1) Q_\sigma([0, 1], B_j) + n\lambda \sum_{i=1}^{I}\sum_{j=1}^{J} (e^{\eta_i} - 1)(e^{\xi_j} - 1) Q_\sigma(A_i, B_j),\end{aligned}$$

where $|A|$ denotes the Lebesgue measure of A and $Q_\sigma(A, B) = \int_A \int_B f_\sigma(y - x)\,\mathrm{d}y\,\mathrm{d}x$.

Note that this modified exponential formula coincides with the result of Proposition 1 if we apply a first order approximation of the exponential functions on the right-hand side of (6) and set $\mu = 1$. As a consequence, differentiation yields the same form of the intensity measure

$$\frac{1}{n\lambda}\mathbb{E}\Big[\frac{M(\mathrm{d}x)N(\mathrm{d}y)}{\mathrm{d}x\mathrm{d}y}\Big] = n\lambda(f_\sigma * \mathbb{1}_{[-1,1]})(y) + f_\sigma(y-x)$$

as in Proposition 1 while second order properties differ. In fact, higher order integrals in the one-to-one setting are a bit simpler, see Remark 17 below. We can thus apply exactly the same estimator as before and Theorem 6 remains true in the one-to-one setting.

Corollary 14. *Let $z_0 \in (-1/2, 1/2)$, and $s, L > 0$. Based on the observations $M = \sum_i \delta_{X_i}$ and $N = \sum_j \delta_{Y_j} = \sum_j \delta_{X_j+\sigma D_j}$ with D_j from (23) there is an estimator $\widehat{f}(z_0)$ depending on λ, σ and s such that*

$$\sup_{f\in\mathcal{G}^s(z_0,L)} \mathbb{E}\Big[\big(\widehat{f}(z_0) - f(z_0)\big)^2\Big]^{1/2} \lesssim r_n,$$

where the rate of convergence is given by (5).

4.2. *The multidimensional case with arbitrary parent distributions.* We investigate briefly in this section two essential extensions of our approach and the results of Theorem 6:

1) How will the rate change in the deconvolution regime if we consider multidimensional observations, *i.e.* when $\mathcal{X} \subseteq \mathcal{O} \subseteq \mathbb{R}^d$ and $\mathcal{Y} \subseteq \mathbb{R}^d$ with $d > 1$?
2) How does the parent distribution affects the problem when p is not uniform over $\mathcal{O}$?

For the first question, we argue that Theorem 6 and 7 generalise to a parent point process in $\mathbb{R}^d$ with intensity measure $\lambda\mathbb{1}_\mathcal{O}$ for a rectangular set $\mathcal{O} \subseteq \mathbb{R}^d$. In general, the smoothing properties of a convolution with $p = |\mathcal{O}|^{-1}\mathbb{1}_\mathcal{O}$ for a bounded set $\mathcal{O} \subseteq \mathbb{R}^d$ considerably depends on the geometry of $\mathcal{O}$ and its boundary $\partial\mathcal{O}$ in particular, see *e.g.* Randol (1969). More specifically, a more regular boundary results in a faster decay of the characteristic function of the uniform distribution on $\mathcal{O}$. As a consequence, the statistical deconvolution problems depends on the geometry, too. To investigate the impact of the regularity properties of the parent distribution p, we assume in this section that the characteristic function of the parent distribution is bounded away from zero. In this case the classical spectral approach is applicable and allows for a transparent analysis of statistical estimation, even in the multidimensional case $d > 1$.

Let $M = \sum_i \delta_{X_i}(\mathrm{d}x)$ be Poisson point process with intensity $\lambda n p(x)\mathrm{d}x$ on $\mathbb{R}^d$, where $p\colon \mathbb{R}^d \to [0,\infty)$ is a bounded probability density function. As before the point process N on $\mathbb{R}^d$ of offspring traits has conditional intensity $\mu(M * f_\sigma(y))\mathrm{d}y$ with f_σ from (2). The decomposition (7) generalises to

$$\frac{1}{n\lambda\mu}\mathbb{E}\Big[\int_{\mathcal{O}\times\mathbb{R}^d} \psi_1(y)\psi_2((y-x)/\sigma)M(\mathrm{d}x)N(\mathrm{d}y)\Big] = \sigma^d n\lambda\overline{\mathcal{U}}_\sigma(f*p) + \overline{\mathcal{V}}_\sigma(f), \tag{24}$$

with

$$\overline{\mathcal{U}}_\sigma(f*p) = \int_{\mathbb{R}^d} \psi_1(y)(\psi_2 * p(\sigma\cdot))(y/\sigma)(f_\sigma * p)(y)\mathrm{d}y$$

and

$$\overline{\mathcal{V}}_\sigma(f) = \int_{\mathbb{R}^d} (\psi_1 * p(-\cdot))(\sigma z)\psi_2(z) f(z)\mathrm{d}z.$$

To deconvolve $f_\sigma * p$ in $\overline{\mathcal{U}}_\sigma(f*p)$, we denote the characteristic function of p by $\varphi_p(u) = \mathcal{F}[p](u) = \int_{\mathbb{R}^d} e^{iu^\top x} p(x)\mathrm{d}x$, $u \in \mathbb{R}^d$, and assume that $\varphi_p(u) \neq 0$ for all $u \in \mathbb{R}^d$. Then we can choose the spectral deconvolution kernel

$$\psi_1 = \mathcal{F}^{-1}\Big[\frac{\mathcal{F}K(h_1 u)}{\varphi_p(u/\sigma)}\Big](z_0 - \cdot/\sigma), \qquad z_0 \in \mathbb{R}^d,$$

with inverse Fourier transform $\mathcal{F}^{-1}[h(u)](x) = (2\pi)^{-d}\int_{\mathbb{R}^d} e^{-iu^\top x} h(u)\mathrm{d}u$ for any $h \in L^1(\mathbb{R}^d)$ and where $K\colon \mathbb{R}^d \to \mathbb{R}$ is a band limited kernel with bandwidth $h_1 > 0$. Plancherel's identity and $\mathcal{F}[(f*p)(\sigma\cdot)](u) = \sigma^{-d}\mathcal{F}[f](u)\varphi_p(u/\sigma)$ indeed yields

$$\begin{aligned}\int_{\mathbb{R}^d} \psi_1(z)(f_\sigma * p)(z)\mathrm{d}z &= \frac{\sigma^d}{(2\pi)^d}\int_{\mathbb{R}^d} e^{-iu^\top z}\frac{\mathcal{F}K(h_1u)}{\varphi_p(u/\sigma)}\mathcal{F}\big[(f*p(\sigma\cdot))\big](u)\mathrm{d}u \\ &= \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} e^{-iu^\top z}\mathcal{F}[K_{h_1}](u)\mathcal{F}[f](u)\mathrm{d}u = (K_{h_1} * f)(z). \qquad (25)\end{aligned}$$

Using ψ_1 from above and $\psi_2 = 1$, we define the following *spectral deconvolution estimator* on $\mathbb{R}^d$:

$$\widehat{f}^{\mathtt{sd}}_{h_1}(z_0) = \frac{1}{n\lambda\mu}\sum_j \mathcal{F}^{-1}\Big[\frac{\mathcal{F}K(h_1u)}{\varphi_p(u/\sigma)}\Big]\Big(z_0 - \frac{Y_j}{\sigma}\Big).$$

If the parent distribution is unknown, then we can profit from the observations $\mathcal{X}$ by replacing φ_p with its empirical counterpart $\widehat{\varphi}_p(u) = |\mathcal{X}|^{-1}\sum_j e^{iX_j^\top u}$ as demonstrated in the classical (univariate) deconvolution literature, see *e.g.* Neumann (2007); Comte and Lacour (2011); Dattner et al. (2016).

The rate of convergence will be determined by the decay of $\varphi_p(u)$. Note that $\varphi_p(u)$ should decay at least as $|u|^{-d}$ in order to allow for a bounded density $h = \mathcal{F}^{-1}\varphi_p$. In the multivariate case, the extensions $\mathcal{H}^s_d(z_0)$ and $\mathcal{G}^s_d(z_0, L)$ of the local Hölder classes $\mathcal{H}^s(z_0)$ and the class of Hölder regular, bounded densities $\mathcal{G}^s(z_0, L)$, respectively, from the univariate case to $\mathbb{R}^d$ is straightfroward by considering partial derivatives. A kernel K of order $\ell_K \geqslant \lfloor s\rfloor$ in dimension d can be constructed, for instance by tensorisation of the one dimensional case. We keep-up with the notation $|\cdot|$ to denote the Euclidean norm on $\mathbb{R}^d$.

THEOREM 15. *Let $z \in \mathbb{R}^d$ and $f \in \mathcal{G}^s_d(z_0, L)$ for some $s > 0$ and let p be a bounded probability density on $\mathbb{R}^d$ with $\varphi_p(u) \neq 0$ for all $u \in \mathbb{R}^d$. If K is a kernel or order $\ell_K \geqslant \lfloor s\rfloor$ that satisfies* $\operatorname{Supp}\mathcal{F}K \subseteq \{u \in \mathbb{R}^d : |u| \leqslant 1\}$*, then we have*

$$\sup_{f\in\mathcal{G}^s_d(z_0,L)} \mathbb{E}\Big[\big|\widehat{f}^{\mathtt{sp}}_{h_1}(z_0) - f(z_0)\big|^2\Big]^{1/2} \lesssim h^s + \frac{\sigma^d}{n^{1/2}}\Big(\int_{\{u\in\mathbb{R}^d:|u|\leqslant 1/(\sigma h)\}} |\varphi_p(u)|^{-2}du\Big)^{1/2}.$$

In the mildly ill-posed case with $|\varphi_p(u)| \gtrsim (1+|u|^2)^{-t/2}$ for some $t \geqslant d$, we obtain

$$\sup_{f\in\mathcal{G}^s_d(z_0,L)} \mathbb{E}\Big[\big|\widehat{f}^{\mathtt{sp}}_{h_1}(z_0) - f(z_0)\big|^2\Big] \lesssim (n\sigma^{2t-d})^{-2s/(2s+2t+d)}.$$

for the choice $h_1 = (n\sigma^{2t-d})^{1/(2s+2t+d)}$.

In the severely ill-posed case $|\varphi_p(u)| \gtrsim e^{-\gamma|u|^\beta}$ *for some* $\gamma, \beta > 0$*, we obtain*

$$\sup_{f\in\mathcal{G}_d^s(z_0,L)} \mathbb{E}\Big[\big|\widehat{f}_{h_1}^{(d)}(z_0) - f(z)\big|^2\Big]^{1/2} \lesssim \sigma^{-s}(\log n)^{-s/\beta}$$

for the choice $h_1 = \sigma^{-1}(\frac{1}{4\gamma}\log n)^{-1/\beta}$.

Several remarks are in order: **1)** for $d = 1$, the uniform distribution corresponds to the degree of ill-posedness $t = 1$ for which we indeed recover the rate $(n\sigma)^{2s/(2s+3)}$. **2)** For more regular distributions with $t > 1$ the dependence of the deconvolution rate on the scaling parameter σ is even more severe. For $t = \frac{3}{2}d$ the deconvolution estimator is only consistent if $n\sigma^{2d} \to \infty$. Since the analysis of the variance of an interaction approach in the general setting reveals a term of order $n\sigma^{2d}$, *cf.* Remark 22, we conjecture that there is a regime where f cannot be estimated consistently if $t > \frac{3}{2}d$.

To discuss the behaviour of an interaction estimator similiar to (13), we note first that our variance estimates in Section 6.2 can be generalised to other parent distributions with bounded densities and to higher dimensions, see in particular Remark 22 at the end of Section 6.2. A soon as the bias due to $\overline{\mathcal{U}_\sigma(f*p)}$ in the interaction regime can be controlled, one can in principle build an estimator $\widehat{f}_h^{\mathtt{int}}(z_0)$ with mean squared-error of order

$$\mathbb{E}\Big[\big(\widehat{f}_h^{\mathtt{int}}(z_0) - f(z_0)\big)^2\Big]^{1/2} \lesssim h^s + \max\Big(\frac{1}{n^{1/2}h^{d/2}}, \frac{\sigma^{d/2}}{h^{d/2}}, n^{1/2}\sigma^d\Big)$$

for $f \in \mathcal{G}_d^s(z_0, L)$. An optimised choice of $h = (n \wedge \sigma^{-d})^{-1/(2s+d)}$ then yields

$$\mathbb{E}\Big[\big(\widehat{f}_h^{\mathtt{int}}(z_0) - f(z_0)\big)^2\Big]^{1/2} \lesssim \max\big((n \wedge \sigma^{-d})^{-s/(2s+d)}, n^{1/2}\sigma^d\big).$$

However, the analysis of the bias due to $\overline{\mathcal{U}_\sigma(f*p)}$ is quite delicate and we do not have a clear understanding of its behaviour at the moment. Note also that the analysis of the interaction estimator is applicable to a generating parent trait point process with intensity $\lambda n \mathbb{1}_{\mathcal{O}}(x)\mathrm{d}x$ for any Borel set $\mathcal{O} \subseteq \mathbb{R}^d$ without additional difficulties.

5. A numerical example. In order to illustrate the main results, we will apply the estimators from Definition 4 together with the pure deconvolution estimator and the interaction estimator from (11) and (14), respectively, on simulated observations.

We choose $n = 1000$, $\lambda = \mu = 1$ and consider the Beta$(2, 3)$-distribution (shifted by $-1/2$) for the dispersal, *i.e.*

$$f(z) = \frac{1}{12}\Big(\frac{1}{2} + z\Big)\Big(\frac{1}{2} - z\Big)^2 \mathbb{1}_{[-1/2,1/2]}(z), \qquad z \in \mathbb{R}.$$

For the estimators we choose the kernel

$$K(z) := \begin{cases} 1, & |z| \leqslant \frac{1}{4}, \\ \big((\frac{32}{15}(|z| - \frac{1}{4})\big)^2 - 1\big)^2, & |z| \in (\frac{1}{4}, \frac{23}{32}], \\ 0, & \text{otherwise}, \end{cases}$$

which is continuously differentiable, non-negative and satisfies Assumption 3 with order $\ell_K = 1$. The bandwidths are chosen as $h_1 = 0.7(n\sigma)^{-1/7}$ and $h_2 = 0.7\min(n, \sigma^{-1})^{-1/5}$ according

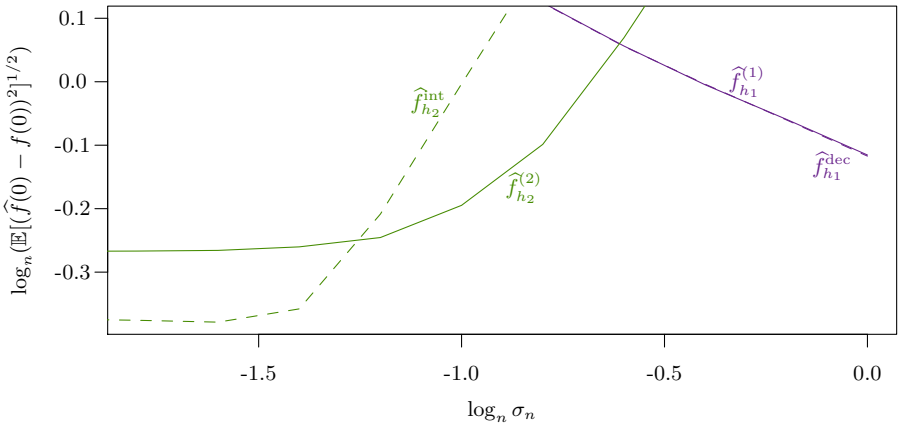

FIG 4. Logarithmic plot of the root mean squared error of $\widehat{f}^{(1)}_{h_1}$ (purple), $\widehat{f}^{(2)}_{h_2}$ (green) *as well as* $\widehat{f}^{\mathrm{dec}}_{h_1}$ (purple, dashed), $\widehat{f}^{\mathrm{int}}_{h_2}$ (green, dashed) *at* $z_0 = 0$ *depending on* σ_n *based on a Monte Carlo simulation.*

to Proposition 5. In the numerical experiments we note a considerable sensitivity of the small scale estimator $\widehat{f}^{(2)}_{h_2}$ to the choice of h_1. While the estimator achieves the optimal rate with $h_1 = 1/(2\sigma)$, the proofs reveal that the conditions $h_1 + h_2 < \sigma^{-1}$ and $h_1 \geqslant 4$ are sufficient for the bias analysis and the variance grows by the factor $(\sigma h_1)^{-1}$. Hence, h_1 should be as large as possible and we choose $h_1 = \max(4, \sigma^{-1} - 1.1\, h_2)$ for $\widehat{f}^{(2)}_{h_2}$.

A Monte Carlo simulation confirms our theoretical findings. Figure 4 shows the root mean squared error at point $z_0 = 0$ based on a Monte Carlo simulation with 500 Monte Carlo iterations. In each of these iterations the same random variables are drawn to define the point clouds $\mathcal{X}$ and $\mathcal{Y}$ along a grid of scaling parameters $\sigma_n = n^\tau$, $\tau \in \{-2, -1.8, \dots, -0.2, 0\}$. We see that $\widehat{f}^{(1)}_{h_1}$ is much better for large scales, but its error increases as σ decreases. For $\sigma < n^{-0.6} \approx 0.016$ the direct estimator $\widehat{f}^{(2)}_{h_2}$ is better and its error improves when σ decreases further. As we can see in this figure $\widehat{f}^{(1)}_{h_1}$ and $\widehat{f}^{\mathrm{dec}}_{h_1}$ behave similarly. In contrast, there is a notable difference between $\widehat{f}^{(2)}_{h_2}$ and $\widehat{f}^{\mathrm{int}}_{h_2}$. More precisely, our numerical experiments indicate that the stochastic error of $\widehat{f}^{\mathrm{int}}_{h_2}$ is smaller across all scales, but even before $\sigma = n^{-1}$ the bias effect drastically kicks in.

For a more detailed impression on the behaviour of our main estimators Figure 5 shows 20 realisations of the estimators $\widehat{f}^{(1)}_{h_1}$ and $\widehat{f}^{(2)}_{h_2}$ for two different choices the scaling parameter $\sigma \in \{n^{-1/20}, n^{-19/20}\}$. While the deconvolution estimator $\widehat{f}^{(1)}_{h_1}$ fluctuates around the true density for the rather large scale $\sigma = n^{-1/20} \approx 0.71$, its result is completely useless at the small scale $\sigma = n^{-19/20} \approx 0.0014$. The direct estimator $\widehat{f}^{(2)}_{h_2}$ reveals the opposite behaviour. At the small scale $\widehat{f}^{(2)}_{h_2}$ concentrates around f, but its shape is unrelated to the dispersal density if σ is large.

6. Proofs.

6.1. *The covariance structure of* (M, N).

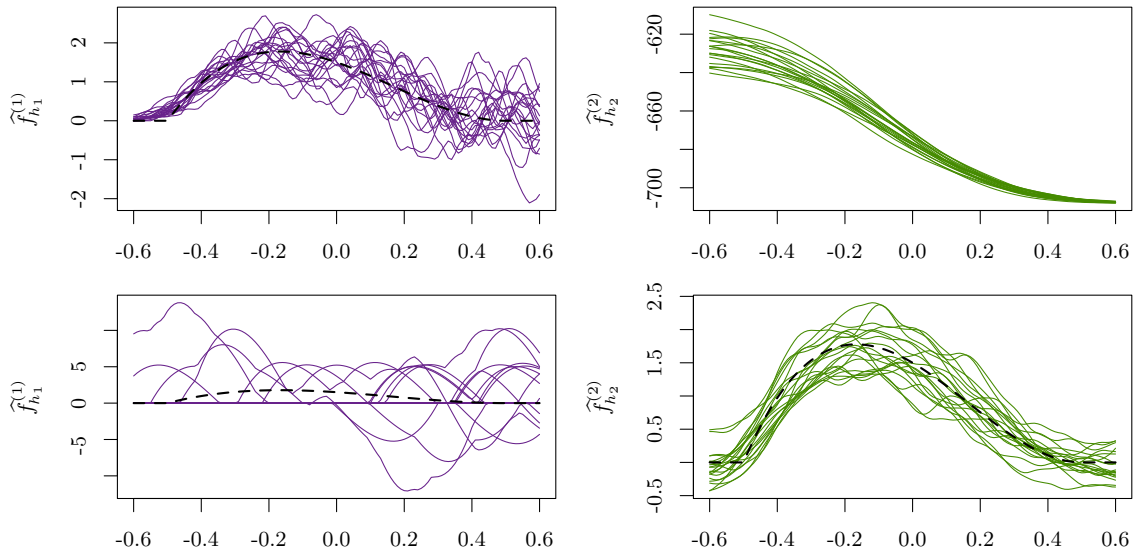

FIG 5. 20 realisations of $\widehat{f}^{(1)}_{h_1}$ (left) *and* $\widehat{f}^{(2)}_{h_2}$ (right) *and true density (*dashed*) in the point process model with parameters* $n = 1000$, $\lambda = \mu = 1$ *and scaling parameters* $\sigma = n^{-1/20} \approx 0.71$ *(*top*) and* $\sigma = n^{-19/20} \approx 0.0014$ *(*bottom*).*

PROOF OF PROPOSITION 1. Conditional on M we obtain from Campbell's formula

$$\begin{aligned}
&\mathbb{E}\Big[\exp\Big(\sum_{j=1}^{J}\eta_j M(A_j)+\sum_{k=1}^{K}\xi_k N(B_k)\Big)\Big|(X_j)\Big]\\
&\quad=\exp\Big(\sum_{j=1}^{J}\eta_j M(A_j)\Big)\mathbb{E}\Big[\exp\Big(\sum_{k=1}^{K}\xi_k N(B_k)\Big)\Big|(X_j)\Big]\\
&\quad=\exp\Big(\sum_{j=1}^{J}\eta_j M(A_j)\Big)\exp\Big(\int(e^{\sum_k \xi_k \mathbb{1}_{B_k}(y)}-1)\mu\sum_l f_\sigma(y-X_l)\mathrm{d}y\Big)\\
&\quad=\exp\Big(\sum_{j=1}^{J}\eta_j M(A_j)+\sum_{k=1}^{K}\int_{B_k}(e^{\xi_k}-1)\mu\sum_l f_\sigma(y-X_l)\mathrm{d}y\Big)=\exp\Big(\int g\mathrm{d}M\Big)
\end{aligned}$$

for

$$g(x):=\sum_{j=1}^{J}\eta_j\mathbb{1}_{A_j}(x)+h(x),\qquad h(x):=\mu\sum_{k=1}^{K}(e^{\xi_k}-1)\int_{B_k}f_\sigma(y-x)\mathrm{d}y.$$

Applying again Campbell's formula yields

$$\log\mathbb{E}\Big[\exp\Big(\sum_{j=1}^{J}\eta_j M(A_j)+\sum_{k=1}^{K}\xi_k N(B_k)\Big)\Big]=n\lambda\int_0^1(e^{g(x)}-1)\mathrm{d}x.$$

If we plug in $g(x)$, we obtain

$$n\lambda\int_0^1(e^{g(x)}-1)\mathrm{d}x=n\lambda\sum_{j=1}^{J}\int_{A_j}(e^{\eta_j+h(x)}-1)\mathrm{d}x+n\lambda\int_{(\bigcup_j A_j)^c}(e^{h(x)}-1)\mathrm{d}x$$

$$= n\lambda \sum_{j=1}^{J} (e^{\eta_j} - 1) \int_{A_j} (e^{h(x)} - 1) \mathrm{d}x + n\lambda \sum_{j=1}^{J} (e^{\eta_j} - 1)|A_j| + n\lambda \int_0^1 (e^{h(x)} - 1) \mathrm{d}x.$$

We conclude by differentiating the previous exponential formula. □

Proof of Corollary 2. It suffices to note that

$$\begin{aligned}\Psi(\eta, \xi) :=& \mathbb{E}\Big[e^{\eta M(A) + \xi N(B)}\Big] \\ &= \exp\Big(n\lambda(e^\eta - 1)|A| + n\lambda \int_0^1 (e^{\psi_B(\xi,x)} - 1)\mathrm{d}x + n\lambda(e^\eta - 1) \int_A (e^{\psi_B(\xi,x)} - 1)\mathrm{d}x\Big),\end{aligned}$$

where $\psi_B(\xi, x) := \mu(e^\xi - 1) \int_B f_\sigma(y - x)\mathrm{d}y$, satisfies

$$\partial_\xi \Psi(\eta, \xi) = \Psi(\eta, \xi)\Big(n\lambda \int_0^1 e^{\psi_B(\xi,x)} \partial_\xi \psi_B(\xi, x)\mathrm{d}x + n\lambda(e^\eta - 1) \int_A e^{\psi_B(\xi,x)} \partial_\xi \psi_B(\xi, x)\mathrm{d}x\Big)$$

and

$$\begin{aligned}\partial_\eta \partial_\xi \Psi(\eta, \xi) = \Psi(\eta, \xi)&\Big(n\lambda e^\eta |A| + n\lambda e^\eta \int_A (e^{\psi_B(\xi,x)} - 1)\mathrm{d}x\Big) \\ &\times \Big(n\lambda \int_0^1 e^{\psi_B(\xi,x)} \partial_\xi \psi_B(\xi, x)\mathrm{d}x + n\lambda(e^\eta - 1) \int_A e^{\psi_B(\xi,x)} \partial_\xi \psi_B(\xi, x)\mathrm{d}x\Big) \\ &+ \Psi(\eta, \xi) n\lambda e^\eta \int_A e^{\psi_B(\xi,x)} \partial_\xi \psi_B(\xi, x)\mathrm{d}x.\end{aligned}$$

Hence, due to $\Psi(0, 0) = 1$, $\psi_B(0, x) = 0$ and $\partial_\xi \psi_B(0, x) = \mu \int_B f_\sigma(y - x)\mathrm{d}y$, the claimed formula is given by $\partial_\eta \partial_\xi \Psi(0, 0)$. □

The previous proof also shows that

$$\mathbb{E}[N(B)] = n\lambda\mu \int_0^1 \int_B f_\sigma(y - x)\, \mathrm{d}y\mathrm{d}x, \qquad B \subseteq \mathbb{R}, \tag{26}$$

by calculating $\partial_\xi \Psi(0, 0)$. While Corollary 2 determines the mean of linear functionals of M and N the following lemma investigates the covariance structure.

Lemma 16. *For*

$$Q_\sigma(A, B) := \int_A \int_B f_\sigma(y - x)\, \mathrm{d}y\, \mathrm{d}x, \quad Q_\sigma^2(A, B_1, B_2) := \int_A \int_{B_1} \int_{B_2} f_\sigma(y_1 - x) f_\sigma(y_2 - x)\, \mathrm{d}y_2\, \mathrm{d}y_1\, \mathrm{d}x$$

we have:

(i) If $B_1, B_2 \subseteq \mathbb{R}$ are intervals such that $B_1 \cap B_2 = \emptyset$, then

$$\begin{aligned}\mathbb{E}[N(B_1)N(B_2)] &= n^2\lambda^2 Q_\sigma([0, 1], B_1) Q_\sigma([0, 1], B_2) + n\lambda\mu^2 Q_\sigma^2([0, 1], B_1, B_2) \\ &= \mathbb{E}[N(B_1)]\mathbb{E}[N(B_2)] + n\lambda\mu^2 Q_\sigma^2([0, 1], B_1, B_2).\end{aligned} \tag{27}$$

(ii) If $A_1, A_2 \subseteq [0, 1]$ and $B \subseteq \mathbb{R}$ are intervals such that $A_1 \cap A_2 = \emptyset$, then

$$\begin{aligned}\mathbb{E}[M(A_1)M(A_2)N(B)] = n^3\lambda^3\mu|A_1||A_2|Q_\sigma([0, 1], B) + n^2\lambda^2\mu|A_2|Q_\sigma(A_1, B) \\ + n^2\lambda^2\mu|A_1|Q_\sigma(A_2, B)\end{aligned} \tag{28}$$

(iii) If $A \subseteq [0,1]$ and $B_1, B_2 \subseteq \mathbb{R}$ are intervals such that $B_1 \cap B_2 = \emptyset$, then

(29)
$$\begin{aligned}\mathbb{E}[M(A)N(B_1)N(B_2)] = {} & n^3\lambda^3\mu^2|A|Q_\sigma([0,1],B_1)Q_\sigma([0,1],B_2) \\ & + n^2\lambda^2\mu^2 Q_\sigma(A,B_1)Q_\sigma([0,1],B_2) + n^2\lambda^2\mu^2 Q_\sigma([0,1],B_1)Q_\sigma(A,B_2) \\ & + n^2\lambda^2\mu^2|A_1|Q_\sigma^2([0,1],B_1,B_2) + n\lambda\mu^2 Q_\sigma^2(A,B_1,B_2).\end{aligned}$$

(iv) For $A_1, A_2 \subseteq [0,1]$ and $B_1, B_2 \subseteq \mathbb{R}$ with $A_1 \cap A_2 = \emptyset$ and $B_1 \cap B_2 = \emptyset$ we have

$$\begin{aligned}&\mathbb{E}[M(A_1)M(A_2)N(B_1)N(B_2)] \\ &= \mathbb{E}[M(A_1)N(B_1)]\mathbb{E}[M(A_2)N(B_2)] \\ &\quad + n^3\lambda^3\mu^2\Big(|A_1|Q_\sigma(A_2,B_1)Q_\sigma([0,1],B_2) + |A_2|Q_\sigma([0,1],B_1)Q_\sigma(A_1,B_2) \\ &\qquad\qquad + |A_1||A_2|Q_\sigma^2([0,1],B_1,B_2)\Big) \\ &\quad + n^3\lambda^2\mu^2\Big(Q_\sigma(A_2,B_1)Q_\sigma(A_1,B_2) + |A_1|Q_\sigma^2(A_2,B_1,B_2) + |A_2|Q_\sigma^2(A_1,B_1,B_2)\Big).\end{aligned}$$

The proof is similiar to the proof of Corollary 2, but using forth order partial derivatives (6). We postpone the details to the appendix.

REMARK 17. A modified proof shows that in the one-to-one case the formulas in Lemma 16 remain true if we set $Q_\sigma^2(A) = 0$.

6.2. *Bias and variance bounds.* The two integrals $\mathcal{U}_\sigma(f*p)$ and $\mathcal{V}_\sigma(f)$ from (7) are analysed in the following two propositions.

PROPOSITION 18. *For $s, L > 0$ let $f \in \mathcal{G}^s(z_0, L)$. Suppose that ψ_1 and ψ_2 are given by (11) and (14), respectively, with K satisfying Assumption 3 with $\ell_K \geq \lfloor s \rfloor$.*

(i) If $\sigma h_2 \geqslant 8$ for $\sigma \in (0,1]$, then we have for all $z_0 \in (-1/2, 1/2)$, and $h_1 \in (0, \frac{h_2}{8}]$

$$\mathcal{U}_\sigma(f*p) = \frac{1}{\sigma h_2} f(z_0) + O(\frac{h_1^s}{\sigma h_2}).$$

(ii) If $h_1 \in [4, \sigma^{-1})$ for $\sigma \in (0, 1/4)$, then we have for all $y \in (-1/2, 1/2)$, and $h_2 \in (0, \sigma^{-1} - h_1]$

$$\mathcal{U}_\sigma(f*p) = \frac{1}{h_1}.$$

PROOF. *(i)* Noting that $(f_\sigma * \mathbb{1}_{[0,1]})(y) = F(\frac{y}{\sigma}) - F(\frac{y-1}{\sigma})$ for the cumulative distribution function F of f, we plug in the choice of ψ_1 and substitute $z = z_0 - \frac{x}{\sigma}$ to obtain

$$\begin{aligned}\mathcal{U}_\sigma(f*p) &= \frac{1}{\sigma h_1^2}\int_{\mathbb{R}} (\psi_2 * \mathbb{1}_{[0,1/\sigma]})(x/\sigma) K'(\frac{z_0}{h_1} - \frac{x}{\sigma h_1})(F(\frac{x}{\sigma}) - F(\frac{x-1}{\sigma}))\mathrm{d}x \\ &= \frac{1}{h_1^2}\int_{\mathbb{R}} (\psi_2 * \mathbb{1}_{[0,1/\sigma]})(z_0 - z) K'(\frac{z}{h_1})(F(z_0 - z) - F(z_0 - z - \frac{1}{\sigma}))\mathrm{d}z. \qquad (30)\end{aligned}$$

Moreover, we have

$$(\psi_2 * \mathbb{1}_{[0,\frac{1}{\sigma}]})(z_0 - z) = \frac{1}{h_2}\int_{\mathbb{R}} K(\frac{z_0 - x}{h_2})\mathbb{1}_{[0,\frac{1}{\sigma}]}(z_0 - z - x)\mathrm{d}x = \frac{1}{h_2}\int_{\mathbb{R}} K(\frac{x}{h_2})\mathbb{1}_{[0,\frac{1}{\sigma}]}(x - z)\mathrm{d}x.$$

Denoting the anti-derivative of K by $K^{(-1)}(z) := \int_{-\infty}^{z} K(x)\mathrm{d}x$, we obtain

$$(\psi_2 * \mathbb{1}_{[0,1/\sigma]})(z_0 - z) = K^{(-1)}(\frac{z}{h_2} + \frac{1}{\sigma h_2}) - K^{(-1)}(\frac{z}{h_2}). \tag{31}$$

On the assumptions that $K(x) = 1$ for $|x| \leqslant \frac{1}{4}$, $h_1 \leqslant \frac{h_2}{8}$ and $\sigma h_2 \geqslant 8$, we have $\frac{z+1/\sigma}{h_2}, \frac{z}{h_2} \in [-\frac{1}{4}, \frac{1}{4}]$ for any $|z| \leqslant h_1$ and thus $(\psi_2 * \mathbb{1}_{[0,1/\sigma]})(z_0 - z) = \frac{1}{\sigma h_2}$. Since the boundary terms vanish by the compact support of K, we conclude from (30) together with integration by parts

$$\begin{aligned}
\mathcal{U}_\sigma(f * p) &= \frac{1}{\sigma h_2 h_1^2} \int_{\mathbb{R}} K'(\frac{z}{h_1})(F(z_0 - z) - F(z_0 - z - \sigma^{-1}))\mathrm{d}z \\
&= \frac{1}{\sigma h_1 h_2} \int_{\mathbb{R}} K(\frac{z}{h_1})(f(z_0 - z) - f(z_0 - z - \sigma^{-1}))\mathrm{d}z \\
&= \frac{1}{\sigma h_2}\Big((\frac{1}{h_1}K(\frac{\cdot}{h_1}) * f)(z_0) - (\frac{1}{h_1}K(\frac{\cdot}{h_1}) * f)(z_0 - \sigma^{-1})\Big) \\
&= \frac{1}{\sigma h_2}(f(z_0) + f(z_0 - \sigma^{-1})) + O(\frac{h_1^s}{\sigma h_2}),
\end{aligned}$$

where the last bound is due to the usual bias estimate based on the Hölder regularity of f. Note that $f(z_0 - \sigma^{-1}) = 0$ since $\sigma \leqslant 1$ and $|z_0| < 1/2$ and, especially, f has arbitrary Hölder regularity in a small neighborhood arround $z_0 x - \sigma^{-1}$.

(ii) If $h_1 + h_2 < \frac{1}{\sigma}$, then $\frac{z}{h_2} + \frac{1}{\sigma h_2} > 1$ for any $|z| \leqslant h_1$ and thus (31) reads as

$$(\psi_2 * \mathbb{1}_{[0,1/\sigma]})(z_0 - z) = 1 - K^{(-1)}(\frac{z}{h_2}).$$

If $\sigma < 1/2$, we moreover have $F(z_0 - z - \sigma^{-1}) = 0$ for all $z_0 \in (-1/2, 1/2)$ and $z \in [-h_1, h_1] \subseteq [-(2\sigma)^{-1}, (2\sigma)^{-1}]$. Using that $K(x) = 1$ for $|x| \leqslant 1/4$ and using $F(x) = 0$ for $x < -1/2$ and $F(x) = 1$ for $x \geqslant 1/2$, we obtain from (30) for $h_1 \geqslant 4$

$$\begin{aligned}
\mathcal{U}_\sigma(f * p) &= \frac{1}{h_1^2} \int_{|z| > h_1/4} (1 - K^{(-1)}(z/h_2))K'(\frac{z}{h_1})F(z_0 - z)\mathrm{d}z \\
&= \frac{1}{h_1^2} \int_{-\infty}^{-h_1/4} (1 - K^{(-1)}(z/h_2))K'(\frac{z}{h_1})\mathrm{d}z \\
&= \frac{1}{h_1^2} \int_{-\infty}^{0} (1 - K^{(-1)}(z/h_2))K'(\frac{z}{h_1})\mathrm{d}z \\
&= \frac{1}{h_1}\Big(1 - K^{(-1)}(0))K(0) + \frac{1}{h_2}\int_{-\infty}^{0} K(\frac{z}{h_2})K(\frac{z}{h_1})\mathrm{d}z\Big) \\
&= \frac{1}{2h_2} + \frac{1}{h_1 h_2}\int_{-\infty}^{0} K(\frac{z}{h_1})K(\frac{z}{h_2})\mathrm{d}z
\end{aligned}$$

where we have used integration by parts and symmetry of K. Since $K(\cdot/h_1)$ is constant one on $[-\frac{h_1}{4}, \frac{h_1}{4}]$, the last line simplifies for $h_2 \leqslant \frac{h_1}{4}$ to

$$\mathcal{U}_\sigma(f * p) = \frac{1}{2h_2} + \frac{1}{h_1 h_2}\int_{-\infty}^{0} K(\frac{z}{h_1})\mathrm{d}z = \frac{1}{h_1}. \qquad \square$$

PROPOSITION 19. *For $s, L > 0$ let $f \in \mathcal{G}^s(z_0, L)$. Suppose that ψ_1 and ψ_2 are given by* (11) *and* (14)*, respectively, with K satisfying Assumption 3 with $\ell_K \geq \lfloor s \rfloor$. Let $\sigma \in (0, 1]$.*

(i) If $h_2 \leqslant \frac{h_1}{4}$ *and* $h_1 + h_2 < 1/\sigma$*, then*

$$\mathcal{V}_\sigma(f) = h_1^{-1} f(z_0) + O(h_1^{-1} h_2^s).$$

(ii) We have for all $\sigma, h_1, h_2 > 0$

$$|\mathcal{V}_\sigma(f)| \leqslant h_1^{-1} \|K\|_{L^1} \|K'\|_{L^1} \|f\|_\infty.$$

Proof. (i) We use

$$\begin{aligned}(\psi_1 * \mathbb{1}_{[-1,0]})(\sigma z) &= \frac{1}{\sigma h_1^2} \int_{-1}^{0} K'\big(\frac{z_0 - z}{h_1} + \frac{t}{\sigma h_1}\big) \mathrm{d}t \\ &= \frac{1}{h_1} \int_{-1/(\sigma h_1)}^{0} K'\big(\frac{z_0 - z}{h_1} + t\big) \mathrm{d}t = \frac{1}{h_1} \Big(K\big(\frac{z_0 - z}{h_1}\big) - K\big(\frac{z_0 - z}{h_1} - \frac{1}{\sigma h_1}\big) \Big).\end{aligned}$$

Noting that $z \in [z_0 - h_2, z_0 + h_2]$ by the support of ψ_2 and using $h_1 + h_2 < 1/\sigma$, we have $|z_0 - z - \frac{1}{\sigma}| \geqslant \frac{1}{\sigma} - h_2 > h_1$ and thus $K\big(\frac{z_0 - z}{h_1} - \frac{1}{\sigma h_1}\big) = 0$. Since $K\big(\frac{z_0 - z}{h_1}\big) = 1$ for $|\frac{z_0 - z}{h_1}| \leqslant \frac{h_2}{h_1} \leqslant \frac{1}{4}$, we obtain

$$\begin{aligned}\int (\psi_1 * \mathbb{1}_{[-1,0]})(\sigma z) \psi_2(z) f(z) \mathrm{d}z &= \frac{1}{h_1 h_2} \int_{\mathbb{R}} K\big(\frac{z_0 - z}{h_1}\big) K\big(\frac{z_0 - z}{h_2}\big) f(z) \mathrm{d}z. \\ &= \frac{1}{h_1 h_2} \int_{\mathbb{R}} K\big(\frac{z_0 - z}{h_2}\big) f(z) \mathrm{d}z. \qquad (32)\end{aligned}$$

Applying again the usual bias estimates on $(h_2^{-1} K(\cdot/h_2) * f)(z_0)$, we conclude

$$\int (\psi_1 * \mathbb{1}_{[-1,0]})(\sigma z) \psi_2(z) f(z) \mathrm{d}z = h_1^{-1} f(z_0) + O(h_1^{-1} h_2^s).$$

(ii) The second bound easily follows form Young's inequality:

$$\begin{aligned}\Big| \int (\psi_1 * \mathbb{1}_{[-1,0]})(\sigma z) \psi_2(z) f(z) \mathrm{d}z \Big| &\leqslant \|\psi_1 * \mathbb{1}_{[-1,0]}\|_\infty \|\psi_2\|_{L^1} \|f\|_\infty \\ &\leqslant \|\psi_1\|_{L^1} \|\psi_2\|_{L^1} \|f\|_\infty \leqslant h_1^{-1} \|K\|_{L^1} \|K'\|_{L^1} \|f\|_\infty. \qquad \square\end{aligned}$$

The next step is to investigate the variance based on Lemma 16.

Proposition 20. *If* f *is bounded and* $\varphi^\star(x, y) := \psi_1(y) \psi_2\big(\frac{y - x}{\sigma}\big)$ *for some kernels* $\psi_1 \in L^1 \cap L^2, \psi_2 \in L^2$*, then there is some* $C > 0$ *such that*

$$\begin{aligned}\mathrm{Var}\Big(\sum_{j,k} \varphi^\star(X_j, Y_k) \Big) &\leqslant C n \lambda \mu (1 \vee \|f\|_\infty) \\ &\quad \times \Big((\mu + 1)(n\lambda\sigma + n^2 \lambda^2 \sigma^2) \|\psi_2\|_{L^1}^2 + (n\lambda\sigma + \mu + 1) \|\psi_2\|_{L^2}^2 \Big) \|\psi_1\|_{L^2}^2.\end{aligned}$$

Proof. We decompose

$$\begin{aligned}\mathrm{Var}\Big(\sum_{j,k} \varphi^\star(X_j, Y_k) \Big) &= \mathbb{E}\Big[\Big(\sum_{j,k} \varphi^\star(X_j, Y_k) \Big)^2 \Big] - \mathbb{E}\Big[\sum_{j,k} \varphi^\star(X_j, Y_k) \Big]^2 \\ &= \mathbb{E}\Big[\sum_{j,k} \varphi^\star(X_j, Y_k)^2 \Big] + \mathbb{E}\Big[\sum_{j_1 \neq j_2, k} \varphi^\star(X_{j_1}, Y_k) \varphi^\star(X_{j_2}, Y_k) \Big]\end{aligned}$$

$$
\begin{aligned}
&+\mathbb{E}\Big[\sum_{j,k_1\neq k_2}\varphi^\star(X_j,Y_{k_1})\varphi^\star(X_j,Y_{k_2})\Big]\\
&+\Big(\mathbb{E}\Big[\sum_{j_1\neq j_2,k_1\neq k_2}\varphi^\star(X_{j_1},Y_{k_1})\varphi^\star(X_{j_2},Y_{k_2})\Big]-\mathbb{E}\Big[\sum_{j,k}\varphi^\star(X_j,Y_k)\Big]^2\Big)\\
&=: J_1+J_2+J_3+J_4.
\end{aligned}
$$

Due to (7) and Young's inequality we have

$$
\begin{aligned}
J_1 &= n^2\lambda^2\mu\int_0^1\int_{\mathbb{R}}\psi_2^2(\frac{y-x}{\sigma})\psi_1^2(y)(f_\sigma*\mathbb{1}_{[0,1]})(y)\mathrm{d}y\mathrm{d}x+n\lambda\mu\int_0^1\int_{\mathbb{R}}\psi_2^2(z)\psi_1^2(x+\sigma z)f(z)\mathrm{d}z\mathrm{d}x\\
&= n^2\lambda^2\mu\sigma\int_{\mathbb{R}}(\psi_2^2*\mathbb{1}_{[0,1/\sigma]})(y/\sigma)\psi_1^2(y)(f_\sigma*\mathbb{1}_{[0,1]})(y)\mathrm{d}y\\
&\qquad+n\lambda\mu\int(\psi_1^2*\mathbb{1}_{[-1,0]})(\sigma z)\psi_2^2(z)f(z)\mathrm{d}z.\\
&\leqslant n^2\lambda^2\mu\sigma\|\psi_2^2*\mathbb{1}_{[0,1/\sigma]}\|_\infty\|\psi_1\|_{L^2}^2\|f_\sigma*\mathbb{1}_{[0,1]}\|_\infty+n\lambda\mu\|\psi_1^2*\mathbb{1}_{[-1,0]}\|_\infty\|\psi_1\|_{L^2}^2\|f\|_\infty\\
&\leqslant \mu(n^2\lambda^2\sigma+n\lambda\|f\|_\infty)\|\psi_1\|_{L^2}^2\|\psi_2\|_{L^2}^2.
\end{aligned}
$$

From Lemma 16(ii) and (iii), we deduce for $x_1\neq x_2$ in $[0,1]$ and $y\in\mathbb{R}$

$$
\mathbb{E}[\mathrm{d}M(x_1)\mathrm{d}M(x_2)\mathrm{d}N(y)]=n^2\lambda^2\mu\Big(n\lambda(f_\sigma*\mathbb{1}_{[0,1]})(y)+f_\sigma(y-x_1)+f_\sigma(y-x_2)\Big)\mathrm{d}y\,\mathrm{d}x_1\mathrm{d}x_2
$$

as well as for $x\in[0,1]$ and $y_1\neq y_2$

$$
\begin{aligned}
\mathbb{E}[\mathrm{d}M(x)\mathrm{d}N(y_1)\mathrm{d}N(y_2)] = n^2\lambda^2\mu^2\Big(&n\lambda(f_\sigma*\mathbb{1}_{[0,1]})(y_1)(f_\sigma*\mathbb{1}_{[0,1]})(y_2)\\
&+f_\sigma(y_1-x)(f_\sigma*\mathbb{1}_{[0,1]})(y_2)+f_\sigma(y_2-x)(f_\sigma*\mathbb{1}_{[0,1]})(y_1)\\
&+\int_0^1 f_\sigma(y_1-t)f_\sigma(y_2-t)\mathrm{d}t+\lambda f_\sigma(y_1-x)f_\sigma(y_2-x)\Big)\mathrm{d}y_1\mathrm{d}y_2\mathrm{d}x.
\end{aligned}
$$

Therefore,

$$
J_2=\mathbb{E}\Big[\int_0^1\int_0^1\int_{\mathbb{R}}\varphi^\star(x_1,y)\varphi^\star(x_2,y)\mathbb{1}_{\{x_1\neq x_2\}}M(\mathrm{d}x_1)M(\mathrm{d}x_2)N(\mathrm{d}y)\Big]=\mu(J_{2,1}+2J_{2,2})
$$

with

$$
\begin{aligned}
J_{2,1} &:= n^3\lambda^3\int_0^1\int_0^1\int_{\mathbb{R}}\psi_2(\frac{y-x_1}{\sigma})\psi_2(\frac{y-x_2}{\sigma})\psi_1^2(y)(f_\sigma*\mathbb{1}_{[0,1]})(y)\mathrm{d}y\,\mathrm{d}x_1\mathrm{d}x_2\\
&= n^3\lambda^3\sigma^2\int_{\mathbb{R}}(\psi_2*\mathbb{1}_{[0,1/\sigma]})^2(y/\sigma)\psi_1^2(y)(f_\sigma*\mathbb{1}_{[0,1]})(y)\mathrm{d}y\\
&\leqslant n^3\lambda^3\sigma^2\|\psi_1\|_{L^2}^2\|\psi_2\|_{L^1}^2 \qquad \text{and}\\
J_{2,2} &:= n^2\lambda^2\int_0^1\int_0^1\int_{\mathbb{R}}\psi_2(\frac{y-x_1}{\sigma})\psi_2(\frac{y-x_2}{\sigma})\psi_1^2(y)f_\sigma(y-x_1)\mathrm{d}y\,\mathrm{d}x_1\mathrm{d}x_2\\
&= n^2\lambda^2\sigma\int_0^1\int_{\mathbb{R}}(\psi_2*\mathbb{1}_{[0,1/\sigma]})(z+\frac{x_1}{\sigma})\psi_1^2(x_1+\sigma z)\psi_2(z)f(z)\mathrm{d}z\,\mathrm{d}x_1\\
&\leqslant n^2\lambda^2\sigma\|\psi_2\|_{L^1}\|\psi_1\|_{L^2}^2\int_{\mathbb{R}}|\psi_2(z)f(z)|\mathrm{d}z=n^2\lambda^2\sigma\|\psi_1\|_{L^2}^2\|\psi_2\|_{L^1}^2\|f\|_\infty.
\end{aligned}
$$

For the third term we have

$$J_3 = \mathbb{E}\Big[\int_0^1 \int_{\mathbb{R}}\int_{\mathbb{R}} \varphi^\star(x,y_1)\varphi^\star(x,y_2)\mathbb{1}_{\{y_1\neq y_2\}}M(\mathrm{d}x)N(\mathrm{d}y_1)N(\mathrm{d}y_2)\Big] = \mu^2(J_{3,1}+2J_{3,2}+J_{3,3}+J_{3,4}),$$

where

$$\begin{aligned}
J_{3,1} &:= n^3\lambda^3 \int_{[0,1]\times\mathbb{R}^2} \psi_2(\frac{y_1-x}{\sigma})\psi_2(\frac{y_2-x}{\sigma})\psi_1(y_1)\psi_1(y_2)(f_\sigma * \mathbb{1}_{[0,1]})(y_1)(f_\sigma * \mathbb{1}_{[0,1]})(y_2)\mathrm{d}y_1\mathrm{d}y_2\mathrm{d}x\\
&= n^3\lambda^3 \int_0^1 \big(\psi_2(-\cdot/\sigma) * (\psi_1(f_\sigma * \mathbb{1}_{[0,1]}))\big)(x)^2\mathrm{d}x\\
&\leqslant n^3\lambda^3 \|\psi_2(\cdot/\sigma)\|_{L^1}^2\|\psi_1\|_{L^2}^2 = n^3\lambda^3\sigma^2\|\psi_1\|_{L^2}^2\|\psi_2\|_{L^1}^2,\\
J_{3,2} &:= n^2\lambda^2 \int_{[0,1]\times\mathbb{R}^2} \psi_2(\frac{y_1-x}{\sigma})\psi_2(\frac{y_2-x}{\sigma})\psi_1(y_1)\psi_1(y_2)f_\sigma(y_1-x)(f_\sigma * \mathbb{1}_{[0,1]})(y_2)\mathrm{d}y_1\mathrm{d}y_2\mathrm{d}x\\
&= n^2\lambda^2 \int_0^1 \big((\psi_2(-\cdot/\sigma)f_\sigma) * \psi_1\big)(x)\big(\psi_2(-\cdot/\sigma) * (\psi_1(f_\sigma * \mathbb{1}_{[0,1]}))\big)(x)\mathrm{d}x\\
&\leqslant n^2\lambda^2\|\psi_2(-\cdot/\sigma)f_\sigma\|_{L^1}\|\psi_1\|_{L^2}^2\|\psi_2(-\cdot/\sigma)\|_{L^1}\\
&= n^2\lambda^2\sigma\|\psi_2(-\cdot)f\|_{L^1}\|\psi_2\|_{L^1}\|\psi_1\|_{L^2}^2 \leqslant n^2\lambda^2\sigma\|f\|_\infty\|\psi_1\|_{L^2}^2\|\psi_2\|_{L^1}^2,\\
J_{3,3} &:= n^2\lambda^2 \int_{[0,1]^2\times\mathbb{R}^2} \psi_2(\frac{y_1-x}{\sigma})\psi_2(\frac{y_2-x}{\sigma})\psi_1(y_1)\psi_1(y_2)f_\sigma(y_1-t)f_\sigma(y_2-t)\mathrm{d}y_1\mathrm{d}y_2\,\mathrm{d}t\mathrm{d}x\\
&= n^2\lambda^2 \int_{[0,1]^2\times\mathbb{R}^2} \psi_2(z_1+\frac{t-x}{\sigma})\psi_2(z_2+\frac{t-x}{\sigma})\psi_1(\sigma z_1+t)\psi_1(\sigma z_2+t)f(z_1)f(z_2)\mathrm{d}z_1\mathrm{d}z_2\,\mathrm{d}t\mathrm{d}x\\
&\leqslant n^2\lambda^2\sigma \int_{\mathbb{R}^4} \big|\psi_2(z_1+x)\psi_2(z_2+x)\psi_1(\sigma z_1+t)\psi_1(\sigma z_2+t)\big|f(z_1)f(z_2)\mathrm{d}z_1\mathrm{d}z_2\,\mathrm{d}x\mathrm{d}t\\
&= n^2\lambda^2\sigma \int_{\mathbb{R}^2} (|\psi_2| * |\psi_2|)(z_1-z_2)(|\psi_1| * |\psi_1|)(\sigma(z_1-z_2))f(z_1)f(z_2)\mathrm{d}z_1\mathrm{d}z_2\\
&\leqslant n^2\lambda^2\sigma\|\psi_1\|_{L^2}^2 \int_{\mathbb{R}^2} (|\psi_2| * |\psi_2|)(z_1)f(z_1+z_2)f(z_2)\mathrm{d}z_1\mathrm{d}z_2 \leqslant n^2\lambda^2\sigma\|\psi_1\|_{L^2}^2\|\psi_2\|_{L^1}^2\|f\|_\infty,\\
J_{3,4} &:= n\lambda \int_{[0,1]\times\mathbb{R}^2} \psi_2(\frac{y_1-x}{\sigma})\psi_2(\frac{y_2-x}{\sigma})\psi_1(y_1)\psi_1(y_2)f_\sigma(y_1-x)f_\sigma(y_2-x)\mathrm{d}y_1\mathrm{d}y_2\,\mathrm{d}x\\
&= n\lambda \int_0^1 \big((f_\sigma\psi_2(\cdot/\sigma)) * \psi_1\big)(x)^2\mathrm{d}x\\
&\leqslant n\lambda\|\psi_1\|_{L^2}^2\|f_\sigma\psi_2(\cdot/\sigma)\|_{L^1}^2\\
&\leqslant n\lambda\|\psi_1\|_{L^2}^2\|f\|_\infty\|f\|_{L^1}\|\psi_2\|_{L^2}^2 \leqslant n\lambda\|f\|_\infty\|\psi_1\|_{L^2}^2\|\psi_2\|_{L^2}^2.
\end{aligned}$$

Finally, we have due to Lemma 16(iv) for $x_1 \neq x_2$ and $y_1 \neq y_2$ that

$$\begin{aligned}
&\mathbb{E}[\mathrm{d}M(x_1)\mathrm{d}M(x_2)\mathrm{d}N(y_1)\mathrm{d}N(y_2)] - \mathbb{E}[\mathrm{d}M(x_1)\mathrm{d}N(y_1)]\mathbb{E}[\mathrm{d}M(x_2)\mathrm{d}N(y_2)]\\
&\quad= n^2\lambda^2\mu^2\Big(n\lambda f_\sigma(y_1-x_2)(f_\sigma * \mathbb{1}_{[0,1]})(y_2) + n\lambda f_\sigma(y_2-x_1)(f_\sigma * \mathbb{1}_{[0,1]})(y_1)\\
&\qquad + n\lambda\int_0^1 f_\sigma(y_1-t)f_\sigma(y_2-t)\mathrm{d}t + f_\sigma(y_1-x_2)f_\sigma(y_2-x_1)\\
&\qquad + f_\sigma(y_1-x_1)f_\sigma(y_2-x_1) + f_\sigma(y_1-x_2)f_\sigma(y_2-x_2)\Big)\mathrm{d}y_1\mathrm{d}y_2\mathrm{d}x_1\mathrm{d}x_2
\end{aligned}$$

and thus $J_4 = \mu^2(2J_{4,1} + J_{4,2} + J_{4,3} + 2J_{4,4})$ with

$$J_{4,1} := n^3\lambda^3 \int_{[0,1]^2\times\mathbb{R}^2} \psi_2(\frac{y_1-x_1}{\sigma})\psi_2(\frac{y_2-x_2}{\sigma})\psi_1(y_1)\psi_1(y_2)$$

$$\begin{aligned}
&\qquad\times f_\sigma(y_1-x_2)(f_\sigma*\mathbb{1}_{[0,1]})(y_2)\mathrm{d}y_1\mathrm{d}y_2\mathrm{d}x_1\mathrm{d}x_2\\
&=n^3\lambda^3\int_{[0,1]\times\mathbb{R}}(\psi_2(\frac{\cdot}{\sigma})*\mathbb{1}_{[0,1]})(y_1)(\psi_2(-\frac{\cdot}{\sigma})*((f_\sigma*\mathbb{1}_{[0,1]})\psi_1))(x_2)\psi_1(y_1)f_\sigma(y_1-x_2)\mathrm{d}y_1\mathrm{d}x_2\\
&=n^3\lambda^3\int_0^1(f_\sigma*((\psi_2(\cdot/\sigma)*\mathbb{1}_{[0,1]})\psi_1))(x_2)(\psi_2(\cdot/\sigma)*((f_\sigma*\mathbb{1}_{[0,1]})\psi_1))(x_2)\mathrm{d}x_2\\
&\leqslant n^3\lambda^3\|\psi_1\|_{L^2}^2\|\psi_2(\cdot/\sigma)\|_{L^1}^2\\
&=n^3\lambda^3\sigma^2\|\psi_1\|_{L^2}^2\|\psi_2\|_{L^1}^2,\\
J_{4,2}&:=n^3\lambda^3\int_{[0,1]^3\times\mathbb{R}^2}\psi_2(\frac{y_1-x_1}{\sigma})\psi_2(\frac{y_2-x_2}{\sigma})\psi_1(y_1)\psi_1(y_2)f_\sigma(y_1-t)f_\sigma(y_2-t)\mathrm{d}y_1\mathrm{d}y_2\mathrm{d}t\mathrm{d}x_1\mathrm{d}x_2\\
&=n^3\lambda^3\sigma^2\int_{[0,1]\times\mathbb{R}^2}(\psi_2*\mathbb{1}_{[0,1/\sigma]})(z_1+t/\sigma)(\psi_2*\mathbb{1}_{[0,1/\sigma]})(z_2+t/\sigma)\psi_1(\sigma z_1+t)\psi_1(\sigma z_2+t)\\
&\qquad\times f(z_1)f(z_2)\mathrm{d}z_1\mathrm{d}z_2\mathrm{d}t\\
&=n^3\lambda^3\sigma^2\int_0^1(((\psi_2*\mathbb{1}_{[0,1/\sigma]})\psi_1(\sigma\cdot))*f(-\cdot))(t/\sigma)^2\mathrm{d}t\\
&\leqslant n^3\lambda^3\sigma^3\|((\psi_2*\mathbb{1}_{[0,1/\sigma]})\psi_1(\sigma\cdot))\|_{L^2}^2\\
&\leqslant n^3\lambda^3\sigma^2\|\psi_1\|_{L^2}^2\|\psi_2\|_{L^1}^2,\\
J_{4,3}&:=n^2\lambda^2\int_{[0,1]^2\times\mathbb{R}^2}\psi_2(\frac{y_1-x_1}{\sigma})\psi_2(\frac{y_2-x_2}{\sigma})\psi_1(y_1)\psi_1(y_2)\\
&\qquad\times f_\sigma(y_1-x_2)f_\sigma(y_2-x_1)\mathrm{d}y_1\mathrm{d}y_2\mathrm{d}x_1\mathrm{d}x_2\\
&\leqslant n^2\lambda^2\int_{\mathbb{R}^4}\big|\psi_2(z_1-\frac{x_1}{\sigma})\psi_2(z_2+\frac{x_1}{\sigma})\\
&\qquad\times\psi_1(x_2)\psi_1(x_1+x_2+\sigma(z_2-z_1))\big|f(z_1)f(z_2)\mathrm{d}z_1\mathrm{d}z_2\mathrm{d}x_1\mathrm{d}x_2\\
&=n^2\lambda^2\int_{\mathbb{R}^2}(f*|\psi_2|)(\frac{x_1}{\sigma}-z_1)|\psi_2|(\frac{x_1}{\sigma})(|\psi_1|*|\psi_1|)(\sigma z_1-x_1)f(z_1)\mathrm{d}z_1\mathrm{d}x_1\\
&=n^2\lambda^2\sigma\int_{\mathbb{R}}(f*((f*|\psi_2|)(|\psi_1|*|\psi_1|)(\sigma\cdot)))(x_1)|\psi_2|(x_1)\mathrm{d}x_1\\
&\leqslant n^2\lambda^2\sigma\|\psi_1\|_{L^2}^2\|\psi_2\|_{L^1}^2\|f\|_\infty,\\
J_{4,4}&:=n^2\lambda^2\int_{[0,1]^2\times\mathbb{R}^2}\psi_2(\frac{y_1-x_1}{\sigma})\psi_2(\frac{y_2-x_2}{\sigma})\psi_1(y_1)\psi_1(y_2)f_\sigma(y_1-x_1)f_\sigma(y_2-x_1)\mathrm{d}y_1\mathrm{d}y_2\mathrm{d}x_1\mathrm{d}x_2\\
&=n^2\lambda^2\sigma\int_0^1((\psi_2f)*\psi_1(\sigma\cdot))(-x_1/\sigma)(f*((\psi_2*\mathbb{1}_{[0,1/\sigma]})\psi_1(\sigma\cdot)))(-x_1/\sigma)\mathrm{d}x_1\\
&\leqslant n^2\lambda^2\sigma^2\|(\psi_2f)*\psi_1(\sigma\cdot)\|_{L^2}\|f*((\psi_2*\mathbb{1}_{[0,1/\sigma]})\psi_1(\sigma\cdot))\|_{L^2}\\
&\leqslant n^2\lambda^2\sigma\|\psi_1\|_{L^2}^2\|\psi_2\|_{L^1}^2\|f\|_\infty.
\end{aligned}$$

Combining all estimates yields for some $C>0$

$$\begin{aligned}
\mathrm{Var}\Big(\sum_{j,k\in\mathbb{Z}}\varphi^\star(X_j,Y_k)\Big)\leqslant&Cn\lambda\mu(1+\|f\|_\infty)\|\psi_1\|_{L^2}^2\\
&\times\Big((\mu+1)(n\lambda\sigma+n^2\lambda^2\sigma^2)\|\psi_2\|_{L^1}^2+(n\lambda\sigma+1+\mu)\|\psi_2\|_{L^2}^2\Big).\quad\square
\end{aligned}$$

If the test function only depends on Y_k, we obtain the following simplified version:

LEMMA 21. *We have* $\mathrm{Var}\Big(\sum_j\psi_1(Y_j)\Big)\leqslant n\lambda(\mu+\mu^2)\|\psi_1\|_{L^2}^2$.

PROOF. We decompose

$$\begin{aligned}\operatorname{Var}\Big(\sum_j \psi_1(Y_j)\Big) &= \mathbb{E}\Big[\sum_j \psi_1(Y_j)^2\Big] + \mathbb{E}\Big[\sum_{j_1\neq j_2} \psi_1(Y_{j_1})\psi_1(Y_{j_2})\Big] - \mathbb{E}\Big[\sum_j \psi_1(Y_j)\Big]^2 \\ &= n\lambda\mu \int_{\mathbb{R}} \psi_1(y)^2 (f_\sigma * \mathbb{1}_{[0,1]})(y)\mathrm{d}y \\ &\quad + n\lambda\mu^2 \int_0^1 \int_{\mathbb{R}}\int_{\mathbb{R}} \psi_1(y_1)\psi_1(y_2) f_\sigma(y_1-x) f_\sigma(y_2-x)\mathrm{d}y_1\mathrm{d}y_2\mathrm{d}x \\ &=: J_1 + J_2.\end{aligned}$$

These terms can be bounded by

$$\begin{aligned}&J_1 \leqslant n\lambda\mu \|\psi_1\|_{L^2}^2 \|f_\sigma * \mathbb{1}_{[0,1]}\|_\infty \leqslant n\lambda\mu\|\psi_1\|_{L^2}^2 \qquad \text{and} \\ &J_2 \leqslant n\lambda\mu^2 \|(f_\sigma*\psi_1)^2 \mathbb{1}_{[0,1]}\|_{L^1} \leqslant n\lambda\mu^2 \|f_\sigma*\psi_1\|_{L^2}^2 \leqslant n\lambda\mu^2\|\psi_1\|_{L^2}^2.\end{aligned}$$ □

REMARK 22. With only minor modifications the same proofs applies to point processes M and N in $\mathbb{R}^d$ where M has intensity $n\lambda p$ with bounded probability density function p on $\mathbb{R}^d$ and where N has conditional intensity $\mu(N * f_\sigma)$ as before. In this case all indicator functions $\mathbb{1}_{[0,1]}$ have to be replaced by p, all integrals of the type $\int_0^1 \ldots \mathrm{d}x$ have to be replaced by $\int_{\mathbb{R}^d} \ldots p(x)\mathrm{d}x$ and the factors σ in the above estimates have to be replaced by σ^d. We obtain the following variance bounds:

$$\begin{aligned}\operatorname{Var}\Big(\sum_{j,k} \varphi^\star(X_j, Y_k)\Big) \leqslant& C n\lambda\mu(1+\|f\|_\infty)(\|p\|_\infty + \|p\|_\infty^3) \\ &\times \Big((\mu+1)(n\lambda\sigma^d + n^2\lambda^2\sigma^{2d})\|\psi_2\|_{L^1}^2 + (n\lambda\sigma^d + 1 + \mu)\|\psi_2\|_{L^2}^2\Big)\|\psi_1\|_{L^2}^2, \qquad (33) \\ \operatorname{Var}\Big(\sum_k \varphi^\star(Y_k)\Big) \leqslant& n\lambda(\mu+\mu^2)\|p\|_\infty \|\psi_1\|_{L^2}^2.\end{aligned}$$

6.3. *Proof of upper and lower bounds.* Based on the previous bounds, we can prove our main results.

Proof of Theorem 6. The theorem is an immediate consequence of Proposition 5.

PROOF OF PROPOSITION 5. From (7) and Propositions 18 and 19, we conclude

$$\begin{aligned}\mathbb{E}\Big[\frac{1}{n\lambda h_1}\widehat{f}_{h_1,h_2}(z_0)\Big] &= \sigma h_2 \mathcal{U}_\sigma(f*p) + \frac{h_2}{n\lambda}\mathcal{V}_\sigma(f) \\ &= f(z_0) + O(h_1^s) + O\Big(\frac{h_2}{nh_1}\Big), \quad \text{for } h_1 \leqslant \frac{h_2}{8}, \sigma h_2 \geqslant 8, \sigma \leqslant 1, \qquad (34) \\ \mathbb{E}\Big[\frac{1}{h_2}\widehat{f}_{h_1,h_2}(z_0)\Big] &= h_1 \mathcal{V}_\sigma(f) + \sigma n\lambda h_1 \mathcal{U}_\sigma(f*p) \\ &= f(z_0) + O(h_2^s) + \sigma n\lambda, \quad \text{for } h_2 \leqslant \min(1, \frac{1}{\sigma} - h_1), h_1 \in [4, \frac{1}{\sigma}], \sigma < \frac{1}{4} \qquad (35)\end{aligned}$$

Since the kernels from (11) and (14) satisfy

$$\|\psi_1\|_{L^2}^2 = O(\sigma^{-1}h_1^{-3}), \qquad \|\psi_2\|_{L^1}^2 = O(1), \quad \|\psi_2\|_{L^2}^2 = O(h_2^{-1}),$$

Proposition 20 yields with a constant μ

$$\mathrm{Var}\Big(\frac{1}{n\lambda h_1}\widehat{f}_{h_1,h_2}(z_0)\Big) \lesssim \frac{h_2^2\sigma^2}{n\sigma h_1^3}\Big(1+\frac{1}{n\sigma}+\frac{1}{h_2 n\sigma}+\frac{1}{h_2(n\sigma)^2}\Big),$$
$$\mathrm{Var}\Big(\frac{1}{h_2}\widehat{f}_{h_1,h_2}(z_0)\Big) \lesssim \frac{1}{nh_2}\Big(1+n\sigma+h_2 n\sigma+h_2(n\sigma)^2\Big)\frac{1}{\sigma h_1}$$

with the constants depending on $\|f\|_\infty, \lambda$ and μ. Combining these bounds, we conclude:

(i) If $n\sigma \geqslant 1 \geqslant h_1$ and $h_2 = 8/\sigma$ we obtain

$$\mathbb{E}\Big[\big(\widehat{f}^{(1)}_{h_1}(z_0)-f(z_0)\big)^2\Big] = \mathbb{E}\Big[\big(\frac{1}{n\lambda h_1}\widehat{f}_{h_1,8/\sigma}(z_0)-f(z_0)\big)^2\Big]$$
$$\lesssim h_1^{2s}+\frac{(h_2\sigma)^2}{(n\sigma h_1)^2}+\frac{(h_2\sigma)^2}{n\sigma h_1^3}\Big(1+\frac{1}{h_2 n\sigma}\Big) \lesssim h_1^{2s}+\frac{1}{n\sigma h_1^3}.$$

(ii) We have for $h_1 = \frac{1}{2\sigma}$ and $h_2 \in (0,1]$ and $\sigma \leqslant 1/8$

$$\mathbb{E}\Big[\big(\widehat{f}^{(2)}_{h_2}(z_0)-f(z_0)\big)^2\Big] = \mathbb{E}\Big[\big(\frac{1}{h_2}\widehat{f}_{h_1,h_2}(z_0)-f(z_0)-\sigma n\lambda\big)^2\Big] \lesssim h_2^{2s}+\frac{1}{nh_2}\vee\frac{\sigma}{h_2}\vee n\sigma^2. \quad \square$$

Proof of Theorem 7. Without loss of generality let $z_0 = 0$. Consider the density $f_0(z) = 6(\frac{1}{4}-z^2)\mathbb{1}_{[-1/2,1/2]}(z) \in \mathcal{G}^s(0,L)$. Let $K \in C^{s+1}\cap C^2(\mathbb{R})$ be a function with $\mathrm{Supp}\,K \subseteq [-1/2,1/2]$ and $K'(0) > 0$. Set for some $\varepsilon, h > 0$

$$f_1(z) := f_0(z) + \varepsilon h^s K'(z/h), \qquad z \in \mathbb{R},$$

where K' denotes the derivative of K. Since the compact support of K implies $\int K'(z)\mathrm{d}z = 0$ and because K' is uniformly bounded, the function f_1 is a density supported on $[-1/2,1/2]$ if h is small enough. Due to

$$\|f_1\|_{C^s} \leqslant \|f_0\|_{C^s} + \varepsilon h^s \|K'(x/h)\|_{C^s} = O(\|f_0\|_{C^s}+\varepsilon\|K'\|_{C^s}) = O(\|f_0\|_{C^s}+\varepsilon\|K\|_{C^{s+1}}),$$

we also conclude $|f_1|_{\mathcal{H}^s(0)} \leqslant L$. Therefore, we have constructed two alternatives $f_0, f_1 \in \mathcal{G}^s(0,L)$ satisfying

$$f_0(0) - f_1(0) = h^s K'(0) \gtrsim h^s.$$

Hence, the lower bounds on the point-wise loss follow from Tsybakov (2009, Thm. 2.2), if the total variation distance of the corresponding observations laws remains bounded for the choices

$$h = \begin{cases} \sigma^{1/(2s+1)}, & \text{if } n\sigma^{(2s+2)/(2s+1)} < 1, \\ (\sigma\sqrt{n})^{1/s}, & \text{if } n\sigma^{(2s+2)/(2s+1)} \geqslant 1 \text{ and } \sigma < n^{-(4s+3)/(6s+6)}, \\ (n\sigma)^{-1/(2s+3)} & \text{if } \sigma \geqslant n^{-(4s+3)/(6s+6)}. \end{cases} \tag{36}$$

We denote by $\mathbb{P}_{0,n}$ and $\mathbb{P}_{1,n}$ the joint distribution of the point processes M and N where the coniditional intensity measure of N is given by $M * f_{0,\sigma}$ and $M * f_{1,\sigma}$, respectively. By conditioning on the number $n_x = M([0,1]) \sim \mathrm{Poiss}(\lambda n)$ of parents, the number $n_y = N(\mathbb{R}) \sim \mathrm{Poiss}(\mu n_x)|n_x$ of children and the location of the parents traits $(X_i)_{i=1,\dots,n_x} \overset{i.i.d}{\sim} \mathrm{U}([0,1])|n_x$ we obtain

$$\|\mathbb{P}_{1,n}-\mathbb{P}_{0,n}\|_{TV} = \sum_{n_x\geqslant 1}\frac{(n\lambda)^{n_x}}{n_x!}e^{-\lambda n}\sum_{n_y\geqslant 0}\frac{(\mu n_x)^{n_y}}{n_y!}e^{-\mu n_x}d(\mathbb{P}_{1,n},\mathbb{P}_{0,n}|n_x,n_y) \qquad \text{where} \tag{37}$$

$$d(\mathbb{P}_{1,n},\mathbb{P}_{0,n}|n_x,n_y):=\int_{[0,1]^{n_x}}\Big\|\mathbb{P}_{1,n}^{Y_1,\dots,Y_{n_y}|X_1=x_1,\dots,X_{n_x}=x_{n_x}}-\mathbb{P}_{0,n}^{Y_1,\dots,Y_{n_y}|X_1=x_1,\dots,X_{n_x}=x_{n_x}}\Big\|_{TV}\mathrm{d}x,$$

taking into account that on the event $\{M([0,1])=0\}$ the conditional distributions coincide. Conditional on $X_j=x_j, j=1,\dots,n_x$, the offspring (Y_i) are i.i.d. Estimating the total variation distance by the χ^2-distance, we thus have

$$\begin{aligned}
d(\mathbb{P}_{1,n},\mathbb{P}_{0,n}|n_x,n_y)^2 &\leqslant \int_{[0,1]^{n_x}}\|\mathbb{P}_{1,n}^{Y_1,\dots,Y_{n_y}|X_1=x_1,\dots,X_{n_x}=x_{n_x}}-\mathbb{P}_{0,n}^{Y_1,\dots,Y_{n_y}|X_1=x_1,\dots,X_{n_x}=x_{n_x}}\|_{TV}^2\mathrm{d}x\\
&\leqslant \int_{[0,1]^{n_x}}\chi^2(\mathbb{P}_{1,n}^{Y_1,\dots,Y_{n_y}|X_1=x_1,\dots,X_{n_x}=x_{n_x}},\mathbb{P}_{0,n}^{Y_1,\dots,Y_{n_y}|X_1=x_1,\dots,X_{n_x}=x_{n_x}})\wedge 1\,\mathrm{d}x\\
&= \int_{[0,1]^{n_x}}\Big(\Big(1+\chi^2(\mathbb{P}_{1,n}^{Y_1|X_1=x_1,\dots,X_{n_x}=x_{n_x}},\mathbb{P}_{0,n}^{Y_1|X_1=x_1,\dots,X_{n_x}=x_{n_x}})\Big)^{n_y}-1\Big)\wedge 1\,\mathrm{d}x\\
&\leqslant e\int_{[0,1]^{n_x}}\Big(n_y\chi^2(\mathbb{P}_{1,n}^{Y_1|X_1=x_1,\dots,X_{n_x}=x_{n_x}},\mathbb{P}_{0,n}^{Y_1|X_1=x_1,\dots,X_{n_x}=x_{n_x}})\Big)\wedge 1\,\mathrm{d}x.
\end{aligned}$$

Under $\mathbb{P}_{k,n}^{Y_1|X_1=x_1,\dots,X_{n_x}=x_{n_x}}$, $k\in\{0,1\}$, Y_1 is distributed according to the density

$$g_{n,\sigma}^{(k)}(y|x)=\frac{1}{n_x}\sum_{j=1}^{n_x}f_{k,\sigma}(y-x_j).$$

Therefore,

$$\begin{aligned}
(\mathbb{P}_{1,n}^{Y_1|X_1=x_1,\dots,X_{n_x}=x_{n_x}},\mathbb{P}_{0,n}^{Y_1|X_1=x_1,\dots,X_{n_x}=x_{n_x}}) &= \int_{g_{n,\sigma}^{(0)}>0}\frac{(g_{n,\sigma}^{(1)}(y|x)-g_{n,\sigma}^{(0)}(y|x))^2}{g_{n,\sigma}^{(0)}(y|x)}\mathrm{d}y\\
&= \frac{\varepsilon^2h^{2s}}{\sigma^2}\int_{g_{n,\sigma}^{(0)}>0}\Big(\frac{1}{n_x}\sum_{j=1}^{n_x}K'\Big(\frac{y}{\sigma h}-\frac{x_j}{\sigma h}\Big)\Big)^2\frac{\mathrm{d}y}{g_{n,\sigma}^{(0)}(y|x)}.
\end{aligned}$$

In order to estimate the previous integral, we need a lower bound for the denominator $g_{n,\sigma}^{(0)}$ on the support

$$\mathrm{Supp}\sum_{j=1}^{n_x}K'\Big(\frac{y}{\sigma h}-\frac{x_j}{\sigma h}\Big)\subseteq\Big[-\sigma h/2,1+\sigma h/2\Big].$$

Defining the event $A:=\{\forall y\in[-\sigma h/2,1+\sigma h/2]: g_{n,\sigma}^{(0)}(y|X)\geqslant c\}$, we obtain

$$\begin{aligned}
d(\mathbb{P}_{1,\lambda},\mathbb{P}_{0,\lambda}|m,n)^2 &\leqslant e\mathbb{E}_n\Big[\Big(m\chi^2(\mathbb{P}_{1,1}^{Y_1|X_1,\dots,X_n},\mathbb{P}_{0,1}^{Y_1|X_1,\dots,X_n})\Big)\wedge 1\Big]\\
&\leqslant \frac{e}{c^2}\mathbb{E}_n\Big[\frac{m\varepsilon^2h^{2s}}{\sigma^2}\int\Big(\frac{1}{n}\sum_{j=1}^{n}K'\Big(\frac{y}{\sigma h}-\frac{X_j}{\sigma h}\Big)\Big)^2\mathrm{d}y\Big]+e\mathbb{P}_n(A^c).
\end{aligned}$$

where $X_1,\dots,X_{n_x}\overset{i.i.d.}{\sim}\mathrm{U}([0,1])$ under $\mathbb{P}_n$ and $\mathbb{E}_n$ denotes the expectation with respect to $\mathbb{P}_n$. Applying Lemma 23 from below, we have $e\mathbb{P}_n(A^c)\leqslant C\sqrt{\frac{\log\sigma^{-1}}{n_x\sigma}}=:r_n$. Hence,

$$d(\mathbb{P}_{1,\lambda},\mathbb{P}_{0,\lambda}|n_x,n_y)^2\leqslant\frac{e}{c^2}\varepsilon\frac{n_yh^{2s+1}}{\sigma}\int\mathbb{E}_n\Big[\Big(\frac{1}{n_x}\sum_{j=1}^{n_x}K'\Big(y-\frac{X_j}{\sigma h}\Big)\Big)^2\Big]\mathrm{d}y+r_n. \tag{38}$$

To obtain a sharp upper bound, we will use two different approaches to estimate the previous display. While the first one will use a stochastic integral approximation of $\int_0^{1/(\sigma h)} K'(y-x)\mathrm{d}x$, the second approach relies on a numerical approximation.

In the first case we represent (38) via

$$d(\mathbb{P}_{1,n},\mathbb{P}_{0,n}|n_x,n_y)^2 \leqslant \frac{e}{c^2}\varepsilon^2\frac{n_y h^{2s+1}}{\sigma}\int \mathbb{E}_n\Big[\frac{1}{n_x}\sum_{j=1}^{n_x} K'\Big(y-\frac{X_j}{\sigma h}\Big)\Big]^2 + \mathrm{Var}_n\Big(\frac{1}{n_x}\sum_{j=1}^{n_x} K'\Big(y-\frac{X_j}{\sigma h}\Big)\Big)\mathrm{d}y + r_n. \tag{39}$$

Owing to

$$\mathbb{E}\Big[\frac{1}{n_x}\sum_{j=1}^{n_x} K'\Big(y-\frac{X_j}{\sigma h}\Big)\Big] = \mathbb{E}\Big[K'\Big(y-\frac{X_1}{\sigma h}\Big)\Big] = \sigma h\int_0^{1/(\sigma h)} K'(y-x)\mathrm{d}x,$$

the first term is bounded by

$$\begin{aligned}\frac{n_y h^{2s+1}}{\sigma}\int \mathbb{E}_n\Big[\frac{1}{n_x}\sum_{j=1}^{n_x} K'\Big(y-\frac{X_j}{\sigma h}\Big)\Big]^2\mathrm{d}y &= n_y\sigma h^{2s+3}\int\Big(\int_0^{1/(\sigma h)} K'\Big(y-x\Big)\mathrm{d}x\Big)^2\mathrm{d}y\\ &= n_y\sigma h^{2s+3}\int\big(K\big(y-\frac{1}{\sigma h}\big)-K(y)\big)^2\mathrm{d}y\\ &\leqslant 2n_y\sigma h^{2s+3}\|K\|_{L^2}^2. \end{aligned}\tag{40}$$

The variance term in (39) can be estimated by

$$\mathrm{Var}_n\Big(\frac{1}{n_x}\sum_{j=1}^{n_x} K'\Big(y-\frac{X_j}{\sigma h}\Big)\Big) = \frac{1}{n_x}\mathrm{Var}_n\Big(K'\Big(y-\frac{X_1}{\sigma h}\Big)\Big) \leqslant \frac{1}{n_x}\mathbb{E}\Big[\Big(K'\Big(y-\frac{X_1}{\sigma h}\Big)^2\Big].$$

Therefore,

$$\frac{n_y h^{2s+1}}{\sigma}\int \mathrm{Var}_n\Big(\frac{1}{n_x}\sum_{j=1}^{n_x} K'\Big(y-\frac{X_j}{\sigma h}\Big)\Big)\mathrm{d}y \leqslant \frac{n_y}{n_x}\frac{h^{2s+1}}{\sigma}\mathbb{E}_n\Big[\int\Big(K'\Big(y-\frac{X_1}{\sigma h}\Big)^2\mathrm{d}y\Big] = \frac{n_y}{n_x}\frac{h^{2s+1}}{\sigma}\|K'\|_{L^2}^2.$$

Together with (39) and (40) we conclude for some constant $C>0$.

$$d^2(\mathbb{P}_{1,n},\mathbb{P}_{0,n}|n_y,n_x) \leqslant C\varepsilon^2\Big(n_y\sigma h^{2s+3}+\frac{n_y}{n_x}\frac{h^{2s+1}}{\sigma}\Big)+r_n. \tag{41}$$

In the second and third regime we need a different bound. Applying a Riemann sum motivated approximation, we decompose

$$\begin{aligned}\frac{1}{n_x}\sum_{j=1}^{n_x} K'\Big(y-\frac{X_j}{\sigma h}\Big) &= \sum_{j=1}^{n_x}\sum_{k=1}^{n_x}\mathbb{1}_{[(k-1)/n_x,k/n_x)}(X_j)\int_{(k-1)/n_x}^{k/n_x} K'\Big(y-\frac{X_j}{\sigma h}\Big)\mathrm{d}x\\ &= \sum_{j=1}^{n_x}\sum_{k=1}^{n_x}\mathbb{1}_{[(k-1)/n_x,k/n_x)}(X_j)\int_{(k-1)/n_x}^{k/n_x} K'\Big(y-\frac{x}{\sigma h}\Big)\mathrm{d}x\\ &\quad+\sum_{j=1}^{n_x}\sum_{k=1}^{n_x}\mathbb{1}_{[(k-1)/n_x,k/n_x)}(X_j)\int_{(k-1)/n_x}^{k/n_x}\Big(K'\Big(y-\frac{X_j}{\sigma h}\Big)-K'\Big(y-\frac{x}{\sigma h}\Big)\Big)\mathrm{d}x\end{aligned}$$

$$=: I_1(y) + I_2(y).$$

Therefore, we obtain an alternative bound for (38):

(42)
$$d(\mathbb{P}_{1,n}, \mathbb{P}_{0,n}|n_y, n_x)^2 \leqslant \frac{e\varepsilon^2}{c^2}\frac{n_y h^{2s+1}}{\sigma}\int \mathbb{E}_n\Big[I_1(y)^2\Big]\,\mathrm{d}y + \frac{e\varepsilon^2}{c^2}\frac{n_y h^{2s+1}}{\sigma}\int \mathbb{E}_n\Big[I_2(y)^2\Big]\,\mathrm{d}y + r_n.$$

For the first term, we calculate

$$\begin{aligned}
\mathbb{E}_n\Big[I_1(y)^2\Big] &= \sum_{j,j'}\sum_{k,k'} \underbrace{\mathbb{P}_n\Big(X_j \in \Big[\frac{k-1}{n_x}, \frac{k}{n_x}\Big), X_{j'} \in \Big[\frac{k'-1}{n_x}, \frac{k'}{n_x}\Big)\Big)}_{=n^{-2} \text{ if } j\neq j' \text{ and } =n^{-1} \text{ if } j=j', k=k'} \\
&\qquad \times \int_{(k-1)/n_x}^{k/n_x}\int_{(k'-1)/n_x}^{k'/n_x} K'\Big(y - \frac{x}{\sigma h}\Big)K'\Big(y - \frac{x'}{\sigma h}\Big)\mathrm{d}x'\mathrm{d}x \\
&= \sum_{k,k'}\Big(\frac{n_x^2 - n_x}{n_x^2} + \frac{n_x}{n_x}\Big)\int_{(k-1)/n_x}^{k/n_x}\int_{(k'-1)/n_x}^{k'/n_x} K'\Big(y - \frac{x}{\sigma h}\Big)K'\Big(y - \frac{x'}{\sigma h}\Big)\mathrm{d}x'\mathrm{d}x \\
&= \Big(2 - \frac{1}{n_x}\Big)\Big(\int_0^1 K'\Big(y - \frac{x}{\sigma h}\Big)\mathrm{d}x\Big)^2 \\
&\leqslant 2(\sigma h)^2\Big(\int_0^{1/\sigma h} K'(y-x)\mathrm{d}x\Big)^2 = 2(\sigma h)^2\big(K(y) - K(y - 1/(\sigma h))\big)^2.
\end{aligned}$$

Hence,

$$\frac{n_y h^{2s+1}}{\sigma}\int \mathbb{E}_n\Big[I_1(y)^2\Big]\mathrm{d}y \leqslant 2n_y\sigma h^{2s+3}\int \big(K(y) - K(y - 1/(\sigma h))\big)^2\mathrm{d}y \leqslant 4\|K\|_{L^2}^2 n_y\sigma h^{2s+3}.$$

The second term in (42) can be bounded as follows:

$$\begin{aligned}
&\mathbb{E}_n\Big[I_2(y)^2\Big] = \mathbb{E}_n\Big[\Big(\sum_{j=1}^{n_x}\sum_{k=1}^{n_x}\mathbb{1}_{[(k-1)/n_x, k/n_x)}(X_j)\int_{(k-1)/n_x}^{k/n_x}\Big(K'\Big(y - \frac{X_j}{\sigma h}\Big) - K'\Big(y - \frac{x}{\sigma h}\Big)\Big)\mathrm{d}x\Big)^2\Big] \\
&\leqslant \mathbb{E}_n\Big[\Big(\sum_{j=1}^{n_x}\sum_{k=1}^{n_x}\mathbb{1}_{[(k-1)/n_x, k/n_x)}(X_j)\int_{(k-1)/n_x}^{k/n_x}\int_{[(x\wedge X_j)/(\sigma h), (x\vee X_j)/(\sigma h)]}|K''(y-z)|\mathrm{d}z\mathrm{d}x\Big)^2\Big] \\
&\leqslant \mathbb{E}_n\Big[\Big(\sum_{j=1}^{n_x}\sum_{k=1}^{n_x}\mathbb{1}_{[(k-1)/n_x, k/n_x)}(X_j)\int_{(k-1)/n_x}^{k/n_x}\int_{(k-1)/(n_x\sigma h)}^{k/(n_x\sigma h)}|K''(y-z)|\mathrm{d}z\mathrm{d}x\Big)^2\Big] \\
&= \mathbb{E}_n\Big[\Big(\frac{1}{n_x}\sum_{j=1}^{n_x}\sum_{k=1}^{n_x}\mathbb{1}_{[(k-1)/n_x, k/n_x)}(X_j)\int_{(k-1)/(n_x\sigma h)}^{k/(n_x\sigma h)}|K''(y-z)|\mathrm{d}z\Big)^2\Big].
\end{aligned}$$

With an analogous calculation as for $\mathbb{E}[I_1(x)^2]$ we obtain

$$\begin{aligned}
\frac{n_y h^{2s+1}}{\sigma}\int \mathbb{E}_n\Big[I_2(y)^2\Big]\,\mathrm{d}y &\leqslant 2\frac{n_y h^{2s+1}}{n_x^2\sigma}\int\Big(\int_0^{1/(\sigma h)}|K''(y-z)|\mathrm{d}z\Big)^2\mathrm{d}y \\
&\leqslant 2\frac{n_y h^{2s+1}}{n_x^2\sigma}\|K''\|_{L^1}\int_0^{1/(\sigma h)}\int|K''(y-z)|\,\mathrm{d}y\mathrm{d}z \leqslant 2\|K''\|_{L^1}^2\frac{n_y h^{2s}}{n_x^2\sigma^2}.
\end{aligned}$$

Therefore, we conclude from (42) for some constant $C' > 0$

$$d(\mathbb{P}_{1,n}, \mathbb{P}_{0,n}|n_y, n_x)^2 \leqslant C'\varepsilon(n_y\sigma h^{2s+3} + \frac{n_y h^{2s}}{n_x^2\sigma^2}) + r_n.$$

In combination with (41) we obtain for some constant $C''>0$

$$d(\mathbb{P}_{1,n},\mathbb{P}_{0,n}|n_y,n_x)^2 \leqslant C''\min\Big(n_y\sigma h^{2s+3}+\frac{n_y}{n_x}\frac{h^{2s+1}}{\sigma}, n_y\sigma h^{2s+3}+\frac{n_y h^{2s}}{n_x^2\sigma^2}\Big)+C\Big(\frac{\log\sigma^{-1}}{n_x\sigma}\Big)^{1/2}.$$

If we plug this estimate into (37) and we deduce :

$$\begin{aligned}\|\mathbb{P}_{1,n}-\mathbb{P}_{0,n}\|_{TV}^2 \leqslant C''\varepsilon\min\Big(&\mathbb{E}[N(\mathbb{R})]\sigma h^{2s+3}+\mathbb{E}\Big[\frac{N(\mathbb{R})}{M([0,1])}\Big]\frac{h^{2s+1}}{\sigma},\\ &\mathbb{E}[N(\mathbb{R})]\sigma h^{2s+3}+\mathbb{E}\Big[\frac{N(\mathbb{R})}{M([0,1])^2}\mathbb{1}_{\{M([0,1])>0\}}\Big]\frac{h^{2s}}{\sigma^2}\Big)\\ &+C\Big(\frac{\log\sigma^{-1}}{\sigma}\mathbb{E}\Big[\frac{1}{M([0,1])}\mathbb{1}_{\{M([0,1])>0\}}\Big]\Big)^{1/2}.\end{aligned}$$

Using that $\mathbb{E}[N(\mathbb{R})|M([0,1])]=\mu M([0,1])$, the remaining expectations are given by

$$\mathbb{E}[N(\mathbb{R})]=\mu\lambda n,\qquad \mathbb{E}\Big[\frac{N(\mathbb{R})}{M([0,1])}\Big]=\mu,$$

$$\mathbb{E}\Big[\frac{N(\mathbb{R})}{M([0,1])^2}\mathbb{1}_{\{M([0,1])>0\}}\Big]=\mu\mathbb{E}\Big[\frac{\mathbb{1}_{\{M([0,1])>0\}}}{M([0,1])}\Big]=\mu e^{-\lambda}\sum_{n\geqslant 1}\frac{\lambda^n}{n\cdot n!}\leqslant\frac{2\mu}{\lambda}e^{-\lambda}\sum_{n\geqslant 1}\frac{\lambda^{n+1}}{(n+1)!}\leqslant\frac{2\mu}{\lambda n}.$$

Therefore,

$$\|\mathbb{P}_{1,n}-\mathbb{P}_{0,n}\|_{TV}^2\leqslant C'''\varepsilon^2\min\Big(\sigma n\lambda\mu h^{2s+3}+\frac{\mu h^{2s+1}}{\sigma},\sigma n\lambda\mu h^{2s+3}+\frac{\mu h^{2s}}{\lambda n\sigma^2}\Big)+C'''\Big(\frac{\log\sigma^{-1}}{\lambda n\sigma}\Big)^{1/2}$$

where the last term is $o(1)$ by assumption. Based on this estimate, the theorem follows by verifying that this upper bound remains bounded for h from (36). □

Lemma 23. *For $g_{n,\sigma}^{(0)}(y|x)=\frac{1}{n}\sum_{j=1}^n f_{0,\sigma}(y-x_j)$ and $f_0(z)=6(\frac{1}{4}-z^2)\mathbb{1}_{[-1/2,1/2]}(z)$ there is some $C>0$ such that the event $A:=\Big\{\forall y\in[-\sigma h/2,1+\sigma h/2]: g_{n,\sigma}^{(0)}(y|X)\geqslant 1/14\Big\}$ for $X_1,\dots,X_n\overset{i.i.d.}{\sim}\mathrm{U}([0,1])$ satisfies $\mathbb{P}(A)\geqslant 1-C\sqrt{\frac{\log\sigma^{-1}}{n\sigma}}$.*

Proof. We first bound the expectation

$$\mathbb{E}[g_{n,\sigma}^{(0)}(y|X)])=\int_0^1 f_{0,\sigma}(y-x)\mathrm{d}x=\int_0^{1/\sigma}f_0(\frac{y}{\sigma}-x)\mathrm{d}x=f_0*\mathbb{1}_{[0,1/\sigma]}(\frac{y}{\sigma})$$

uniformly from below: For any $h\in(0,1/2)$ we have

$$\begin{aligned}\inf_{y\in[-\sigma h/2,1+\sigma h/2]}f_0*\mathbb{1}_{[0,1/\sigma]}(\frac{y}{\sigma})&=\inf_{y\in[-\sigma h/2,1+\sigma h/2]}6\int_{-1/2}^{1/2}(\frac{1}{4}-z^2)\mathbb{1}_{[0,1/\sigma]}(\frac{y}{\sigma}-z)\mathrm{d}z\\ &=6\int_{h/2}^{1/2}(\frac{1}{4}-z^2)\mathrm{d}z\geqslant\frac{1}{7}.\end{aligned}$$

By continuity of f_0 we deduce

$$\mathbb{P}(A^c)=\mathbb{P}\Big(\inf_{y\in[-\sigma h/2,1+\sigma h/2]\cap\mathbb{Q}}g_{n,\sigma}^{(0)}(y|X)<\frac{1}{14}\Big)$$

$$\leqslant \mathbb{P}\Big(\sup_{y\in[-\sigma h/2,1+\sigma h/2]\cap\mathbb{Q}} \Big|\frac{1}{n\sigma}\sum_{j=1}^{n} f_0((y-X_j)/\sigma) - \mathbb{E}[f_0((y-X_j)/\sigma)]\Big| > \frac{1}{14}\Big)$$
$$\leqslant \frac{14}{\sigma\sqrt{n}}\mathbb{E}\Big[\max_{y\in[-\sigma h/2,1+\sigma h/2]\cap\mathbb{Q}} \Big|\frac{1}{\sqrt{n}}\sum_{j=1}^{n} f_0((y-X_j)/\sigma) - \mathbb{E}[f_0((y-X_j)/\sigma)]\Big|\Big].$$

To bound the previous expectation we will apply an entropy bound: Since f_0 is of bounded variation, the transition class

$$\mathcal{F} = \{[0,1] \ni x \mapsto f_0((y-x)/\sigma) | y \in [-\sigma h/2, 1+\sigma h/2]\cap\mathbb{Q}\}$$

is of Vapnik-Cervonenkis type satisfying the covering number bound $N(\mathcal{F}, L^2(\mathbb{W}), \varepsilon) \leqslant (A/\varepsilon)^{2w}$ for any probability measure $\mathbb{W}$, any $w > 3$ and some constant A which does not depend on the dialation parameter σ (Giné and Nickl, 2016, Proposition 3.6.12). Moreover, $\mathcal{F}$ admits the envelope $F_\sigma := \frac{3}{2}\mathbb{1}_{[-1,2]}(\cdot/\sigma)$ since $\sup_z f_0(z/\sigma) = \frac{3}{2}$ and $\operatorname{Supp} f_0((y-x)/\sigma) \subseteq \operatorname{Supp} F_\sigma$ for any $y \in [-\sigma h/2, 1+\sigma h/2]$. Theorem 3.5.4 and Remark 3.5.5 by Giné and Nickl (2016) thus yield for some $C > 0$

$$\mathbb{E}\Big[\max_{y\in[-\sigma h/2,1+\sigma h/2]\cap\mathbb{Q}} \Big|\frac{1}{\sqrt{n}}\sum_{j=1}^{n} f_0((y-X_j)/\sigma) - \mathbb{E}[f_0((y-X_j)/\sigma)]\Big|\Big]$$
$$\leqslant 8\sqrt{2}\|F_\sigma\|_{L^2(\mathbb{P}^{X_1})}\int_0^1 \sup_{\mathbb{W}} \sqrt{\log 2N(\mathcal{F}, L^2(\mathbb{W}), \tau\|F_\sigma\|_{L^2(\mathbb{P}^{X_1})})}d\tau$$
$$\leqslant C\|F_\sigma\|_{L^2(\mathbb{P}^{X_1})}|\log(\|F_\sigma\|_{L^2(\mathbb{P}^{X_1})})|^{1/2}.$$

Since $\|F_\sigma\|^2_{L^2(\mathbb{P}^{X_1})} = \frac{9}{4}\int_0^1 \mathbb{1}_{[-1,2]}(x/\sigma)\mathrm{d}x \leqslant \frac{27}{4}\sigma$, we conclude $\mathbb{P}(A^c) \leqslant C\sqrt{\frac{\log\sigma^{-1}}{n\sigma}}$. □

Proof of Theorem 15. Due to (26) and (25), we have $\mathbb{E}[\widehat{f}^{\mathrm{sd}}_{h_1}(z_0)] = K_{h_1} * f(z_0)$. Together with the bias-variance decomposition and a standard bias estimate, we obtain

$$\mathbb{E}[|\widehat{f}^{\mathrm{sd}}_{h_1}(z_0) - f(z_0)|^2] \lesssim h_1^{2s} + \operatorname{Var}\left(\widehat{f}^{\mathrm{sd}}_{h_1}(z_0)\right).$$

A bound for the variance of $\widehat{f}^{\mathrm{sp}}_{h_1}(z_0) = (n\lambda\mu)^{-1}\sum_j \psi_1(Y_j)$ is given in Remark 22. Applying Plancherel's identity we have

$$\operatorname{Var}\left(\widehat{f}^{\mathrm{sd}}_{h_1}(z_0)\right) \lesssim \frac{1}{n}\|\psi_1\|^2_{L^2} = \frac{1}{n}\|\mathcal{F}\psi_2\|^2_{L^2}$$
$$\lesssim \frac{\sigma^d}{n}\int_{\mathbb{R}^d}\Big|\frac{e^{-iu^\top y}\mathcal{F}K(hu)}{\varphi_p(u/\sigma)}\Big|^2\mathrm{d}u \lesssim \frac{\sigma^{2d}}{n}\|K\|^2_{L^1}\int_{|u|\leqslant 1/(\sigma h)}|\varphi_p(u)|^{-2}\mathrm{d}u$$

We conlcude the first inequality in Theorem 15.

In the mildly ill-posed case $|\varphi_p(u)| \gtrsim (1+|u|^2)^{-t/2}$ we have $\int_{\{u\in\mathbb{R}^d:|u|\leqslant 1/(\sigma h)\}}|\varphi_p(u)|^{-2}\mathrm{d}u \lesssim (\sigma h)^{-2t-d}$. Therefore,

$$\mathbb{E}[|\widehat{f}^{(d)}_{1,h}(z_0) - f(z_0)|^2] \lesssim h^{2s} + \frac{\sigma^{2d}}{n(h\sigma)^{2t+d}} = h^{2s} + \frac{1}{n\sigma^{2t-d}h^{2t+d}}.$$

For $h = (n\sigma^{2t+d-2})^{1/(2s+2t+d)}$ we thus obtain the asserted rate of convergence. In the severely ill-posed case $|\varphi_p(u)| \gtrsim e^{-\gamma|u|^\beta}$ we obtain

$$\mathbb{E}[|\widehat{f}_h(z_0) - f(z_0)|^2] \lesssim h^{2s} + \frac{\sigma^{2d}}{nh\sigma}e^{2\gamma(h\sigma)^{-\beta}}$$

which yields the claimed rate for $h = \sigma^{-1}(\frac{1}{4\gamma}\log n)^{-1/\beta}$. □

6.4. *Proofs for the scaling estimators.*

PROOF OF LEMMA 8. We denote the c.d.f. corresponding to f_σ by F_σ. Using (26) and integration by parts, we obtain for some random variable $D \sim f$

$$\begin{aligned}
\mathbb{E}[\widehat{T}] = \mathbb{E}[N(\mathbb{R}\setminus[0,1])] &= n\lambda\mu\Big(\int_0^1\int_{-\infty}^0 f_\sigma(y-x)\mathrm{d}y\mathrm{d}x + \int_0^1\int_1^\infty f_\sigma(y-x)\mathrm{d}y\mathrm{d}x\Big)\\
&= n\lambda\mu\Big(\int_0^1 F_\sigma(-x)\mathrm{d}x + 1 - \int_0^1 F_\sigma(1-x)\mathrm{d}x\Big)\\
&= n\sigma\lambda\mu\Big(\int_0^{1/\sigma} F(-x)\mathrm{d}x + \int_0^{1/\sigma}(1-F(x))\mathrm{d}x\Big)\\
&= n\sigma\lambda\mu\int_0^{1/\sigma}\mathbb{P}(D\leqslant -x)+\mathbb{P}(D>x)\mathrm{d}x\\
&= n\sigma\lambda\mu\int_0^{1/\sigma}\mathbb{P}(|D|>x)\mathrm{d}x\\
&= n\sigma\lambda\mu\mathbb{E}[|D|]
\end{aligned}$$

where the last equality exploits that $\mathbb{P}(|D|>x)=0$ for any $x>1/2$ and $1/\sigma>1/2$. The bound for the variance of $\widehat{T}$ follows from Lemma 21 with $\psi_1 := \mathbb{1}_{[-\sigma/2,0]}+\mathbb{1}_{[1,1+\sigma/2]}$. □

PROOF OF PROPOSITION 9. First we note, that $l\mapsto \overline{X}_{(l)}$ is increasing. Indeed, we have

$$\overline{X}_{(l+1)} \geqslant \overline{X}_{(l)} \quad\Longleftrightarrow\quad \sum_{j=1}^{l+1} X_{(j)} \geqslant \frac{l+1}{l}\sum_{j=1}^{l} X_{(j)} \quad\Longleftrightarrow\quad X_{(l+1)} \geqslant \frac{1}{l}\sum_{j=1}^{l} X_{(j)}$$

which holds true since $X_{(j)}$ are ordered increasingly. Furthermore, Lemma 8 yields

$$\begin{aligned}
\mathbb{P}\Big(\Big|\frac{\widehat{l}}{\kappa_n\mathbb{E}[\widehat{T}]^{1/2}}-1\Big|>\frac{1}{2}\Big) &= \mathbb{P}\Big(\Big|\frac{\widehat{T}}{\mathbb{E}[\widehat{T}]}-1\Big|>\frac{1}{2}\Big|\frac{\widehat{T}}{\mathbb{E}[\widehat{T}]}+1\Big|\Big)\\
&\leqslant \mathbb{P}\Big(\Big|\frac{\widehat{T}}{\mathbb{E}[\widehat{T}]}-1\Big|>\frac{1}{2}\Big) \leqslant 2\frac{\mathrm{Var}(\widehat{T})}{\mathbb{E}[\widehat{T}]} \lesssim \frac{4}{n\sigma}\to 0.
\end{aligned}$$

Consequently,

$$\Lambda := \{l\leqslant \widehat{l}\leqslant 3l\} \qquad \text{with} \qquad l := \frac{\kappa_n}{2}\mathbb{E}[\widehat{T}]^{1/2}$$

satisfies $\mathbb{P}(\Lambda)\to 1$ and l is of order $\kappa_n\sqrt{\sigma n}$.

For some large $C>0$ set

$$\varepsilon := C\frac{l}{n} \qquad \text{and} \qquad R := X_{(l+1)} - \overline{X}_{(l)} \in [0,1].$$

We calculate

$$\begin{aligned}
\mathbb{P}\big(|\widehat{\sigma}^{(1)}-\sigma|>2\varepsilon\big) &= \mathbb{P}\big(\widehat{\sigma}^{(1)}-\sigma>2\varepsilon\big)+\mathbb{P}\big(\sigma-\widehat{\sigma}^{(1)}>2\varepsilon\big)\\
&\leqslant \mathbb{P}\big(\widehat{\sigma}^{(1)}-\sigma>2\varepsilon\big)+\mathbb{P}\big(\sigma-\widehat{\sigma}^{(1)}>2R\big)+\mathbb{P}(R>\varepsilon)\\
&=: T_1+T_2+T_3.
\end{aligned}$$

We will consider all three terms separately. For T_1 we write on the event Λ

$$\widehat{\sigma}^{(1)} = -2(\min_j Y_j - X_{(1)}) - 2X_{(1)} + 2\overline{X}_{\widehat{(l)}} \leqslant \sigma - 2X_{(1)} + 2\overline{X}_{(3l)}.$$

Since conditional on $|\mathcal{X}| = n_x$ it holds $X_{(l)} \sim \mathrm{Beta}(l, n_x + 1 - l)$, we obtain

$$\begin{aligned} T_1 &= \mathbb{P}\big(\widehat{\sigma}^{(1)} - \sigma > 2\varepsilon\big) \leqslant \mathbb{P}\big(- 2X_{(1)} + 2\overline{X}_{(3l)} > 2C\frac{l}{n}\big) + \mathbb{P}(\Lambda^c) \\ &\leqslant \frac{n}{Cl}\mathbb{E}\big[\overline{X}_{(3l)}\big] + o(1) = \frac{n}{Cl}\mathbb{E}\Big[\frac{1}{3l}\sum_{i=1}^{3l}\frac{i}{|\mathcal{X}|+1}\Big] + o(1) \leqslant \frac{n}{Cl}\mathbb{E}\Big[\frac{1+3l}{|\mathcal{X}|+1}\Big] + o(1). \end{aligned}$$

Due to $|\mathcal{X}| \sim \mathrm{Poiss}(\lambda n)$, T_1 is arbitrarily small for sufficiently large C.

To bound T_2, we note that R is a random variable depending only on $\mathcal{X}$. Conditional on $\mathcal{X}$, the offspring trait Y_1 has the distribution function $F_{\sigma|\mathcal{X}}(z) := \frac{1}{|\mathcal{X}|}\sum_i F_\sigma(z - X_i)$ where $F_\sigma = F(\cdot/\sigma)$ is the c.d.f. corresponding to the scaled dispersal density f_σ. Therefore, we write T_2 as

$$\begin{aligned} T_2 = \mathbb{P}(\overline{\sigma}^{(1)} < \sigma - 2R) &\leqslant \mathbb{P}\big(\min_j Y_j > R - \frac{\sigma}{2} + \overline{X}_{(l)}\big) + \mathbb{P}(\Lambda^c) \\ &= \mathbb{P}\big(\forall j : Y_j > X_{(l+1)} - \frac{\sigma}{2}\big) + o(1) \\ &= \mathbb{E}\Big[\mathbb{P}\big(\forall j : Y_j > X_{(l+1)} - \frac{\sigma}{2}|\mathcal{X}, |\mathcal{Y}|\big)\Big] + o(1) \\ &= \mathbb{E}\Big[\mathbb{P}\big(Y_1 > X_{(l+1)} - \frac{\sigma}{2}|\mathcal{X}\big)^{|\mathcal{Y}|}\Big] + o(1) \\ &= \mathbb{E}\Big[\big(1 - F_{\sigma|\mathcal{X}}\big(X_{(l+1)} - \frac{\sigma}{2}\big)\big)^{|\mathcal{Y}|}\Big] + o(1). \end{aligned}$$

The boundary assumption on F yields

$$\begin{aligned} F_{\sigma|\mathcal{X}}\big(X_{(l+1)} - \frac{\sigma}{2}\big) &= \frac{1}{|\mathcal{X}|}\sum_i F_\sigma\big(X_{(l+1)} - \frac{\sigma}{2} - X_i\big) \\ &\geqslant \frac{\gamma}{\sigma|\mathcal{X}|}\sum_i \big(0 \vee \big(X_{(l+1)} - X_i\big)\big) \\ &= \frac{\gamma}{\sigma|\mathcal{X}|}\sum_{i:X_i<X_{(l+1)}} \big(X_{(l+1)} - X_i\big) \\ &= \frac{\gamma}{\sigma|\mathcal{X}|}\sum_{i=1}^{l}\big(X_{(l+1)} - X_{(i)}\big) = \frac{\gamma l R}{\sigma|\mathcal{X}|}. \end{aligned}$$

We conclude for some $C_1 > 0$

$$\begin{aligned} T_2 &\leqslant \mathbb{E}\Big[\big(1 - \frac{\gamma l R}{\sigma|\mathcal{X}|}\big)^{|\mathcal{Y}|}\Big] \\ &\leqslant \mathbb{E}\Big[\big(1 - \frac{\gamma C_1}{|\mathcal{X}|}\big)^{|\mathcal{Y}|}\Big] + \mathbb{P}(lR < C_1\sigma) \\ &\leqslant \exp(-\gamma C_1^{1/3}) + \mathbb{P}(|\mathcal{Y}| < C_1^{-1/3}n) + \mathbb{P}(|\mathcal{X}| > C_1^{1/3}n) + \mathbb{P}(lR < C_1\sigma) \end{aligned}$$

using $(1-\kappa)^{n_y} \leqslant \exp(n_y \log(1-\kappa)) \leqslant \exp(-n_y\kappa)$ for $\kappa = \gamma C^{2/3}/n \to 0$. Since $|\mathcal{X}| \sim \mathrm{Poiss}(\lambda n)$ and conditional on $|\mathcal{X}|$ we have $|\mathcal{Y}| \sim \mathrm{Poiss}(\mu|\mathcal{X}|)$, we conclude for an arbitrary small δ and sufficiently large $C_1 = C_1(\delta, \lambda, \mu, \gamma)$ that

$$T_2 \leqslant \delta + \mathbb{P}(lR < C_1\sigma).$$

We bound $T_3 + \mathbb{P}(lR < C_1\sigma)$ in the same line of arguments, having

$$\begin{aligned} T_3 + \mathbb{P}(lR < C_1\sigma) &= \mathbb{P}(R > \varepsilon) + \mathbb{P}(R < C_1\frac{\sigma}{l}). \\ &\leqslant \varepsilon^{-1}\mathbb{E}[R] + \mathbb{P}(R - \mathbb{E}[R|\,|\mathcal{X}|] < C_1\frac{\sigma}{l} - \mathbb{E}[R|\,|\mathcal{X}|]) \\ &\leqslant \varepsilon^{-1}\mathbb{E}[R] + \mathbb{E}\Big[\frac{\mathrm{Var}(R|\,|\mathcal{X}|)}{(\mathbb{E}[R|\,|\mathcal{X}|] - C_1\frac{\sigma}{l})^2}\Big]. \end{aligned}$$

Note that $\mathbb{P}(|\mathcal{X}| = 0) \to 0$ as $n \to \infty$. For $n_x \geqslant 1$ we calculate

$$\begin{aligned} \mathbb{E}[R|\,|\mathcal{X}| = n_x] &= \mathbb{E}\Big[X_{(l+1)} - \overline{X}_{(l)}|\,|\mathcal{X}| = n_x\Big] \\ &= \frac{l+1}{n_x+2} - \frac{1}{l}\sum_{j=1}^{l}\frac{j}{n_x+1} \\ &= \frac{l+1}{n_x+2} - \frac{(l+1)}{2(n_x+1)} \\ &= \frac{n_x(l+1)}{2(n_x+1)(n_x+2)} \in (\frac{l}{12n_x}, \frac{l}{n_x}). \end{aligned}$$

The properties of order statistics under the uniform distribution yield $\mathrm{Cov}(X_{(j)}, X_{(k)}|\,|\mathcal{X}| = n_x) = \frac{j(n_x-k+1)}{(n_x+1)^2(n_x+2)} \leqslant \frac{l+1}{n_x^2}$ for $1 \leqslant j \leqslant k \leqslant l+1$. We infer

$$\begin{aligned} \mathrm{Var}(R|\,|\mathcal{X}| = n_x) = \mathrm{Var}(X_{(l+1)}|\,|\mathcal{X}| = n_x) &- \frac{2}{l}\sum_{j=1}^{l}\mathrm{Cov}(X_{(j)}, X_{(l+1)}|\,|\mathcal{X}| = n_x) \\ &+ \frac{1}{l^2}\sum_{j_1,j_2=1}^{l}\mathrm{Cov}(X_{(j_1)}, X_{(j_2)}|\,|\mathcal{X}| = n_x) \lesssim \frac{l+1}{n_x^2}. \end{aligned}$$

Therefore,

$$\begin{aligned} T_3 + \mathbb{P}(lR < C_1\sigma) &\lesssim \frac{l}{\varepsilon n} + \frac{l}{n^2}\Big(\frac{l}{24n} - C_1\frac{\sigma}{l}\Big)^{-2} \\ &= \frac{1}{C} + \frac{1}{l}\Big(\frac{1}{24} - C_1\frac{n\sigma}{l^2}\Big)^{-2} \lesssim \frac{1}{C} + O\Big(\frac{1}{\kappa_n\sqrt{\sigma n}}\Big). \end{aligned}$$

This upper bound is arbitrary small for sufficiently large C and n. □

Proof of Proposition 10. We decompose for some $c \in (0,1)$

$$\begin{aligned} \mathbb{P}(|\frac{\widehat{\sigma}^{(2)}}{\sigma} - 1| \geqslant \varepsilon) &= \mathbb{P}(|\widehat{\sigma}^{(2)} - \sigma| \geqslant \sigma\varepsilon) \\ &\leqslant \mathbb{P}(\widehat{\sigma}^{(2)} - \sigma \geqslant \sigma\varepsilon) + \mathbb{P}(\sigma - \widehat{\sigma}^{(2)} \geqslant \sigma\varepsilon) \end{aligned}$$

$$\leqslant \mathbb{P}\big(\widehat{\sigma}^{(2)} \geqslant \sigma(1+\varepsilon)\big) + \mathbb{P}\big(\sigma(1-\varepsilon) \geqslant \widehat{\sigma}^{(2)} \geqslant c\sigma\big) + \mathbb{P}\big(\widehat{\sigma}^{(2)} < c\sigma\big)$$
$$=: P_1 + P_2 + P_3.$$

In the following we will prove that all three probabilities tend to zero. To this end, note that we can write $\widehat{\sigma}^{(2)}$ as

$$\widehat{\sigma}^{(2)} = \min\Big\{h > 0 : \mathbb{P}(\sigma|D_1| \leqslant \frac{h}{2}) + \xi(h) \geqslant 1 - \sqrt{nh^2 + n^{-1}}\kappa_n\Big\}.$$

For the first term P_1, we set $h^\circ = \sigma$ such that the support of D_1 implies $\mathbb{P}(\sigma|D_1| \leqslant \frac{h^\circ}{2}) = \mathbb{P}(|D_1| \leqslant \frac{1}{2}) = 1$. Therefore,

$$\begin{aligned} P_2 \leqslant \mathbb{P}(\widehat{\sigma}^{(2)} > h^\circ) &\leqslant \mathbb{P}\big(\xi(h^\circ) + \mathbb{P}(\sigma|D_1| \leqslant \frac{h^\circ}{2}) < 1 - \sqrt{n(h^\circ)^2 + n^{-1}}\kappa_n\big) \\ &= \mathbb{P}\big(- \xi(h^\circ) > \sqrt{n(h^\circ)^2 + n^{-1}}\kappa_n\big) \\ &\leqslant \mathbb{P}(\Xi^c) \to 0 \end{aligned}$$

with the good event Ξ from Lemma 24.

To bound the second probability P_2, we set $h^* = \sigma(1-\varepsilon) \in (0, 1/2)$ and estimate on $\Xi \cap \{h^* \geqslant \widehat{\sigma}^{(2)} > c\sigma\}$

$$\begin{aligned} \mathbb{P}(\sigma|D_1| \leqslant \frac{h^*}{2}) + \sqrt{n(h^*)^2 + n^{-1}}\kappa_n &\geqslant \mathbb{P}(\sigma|D_1| \leqslant \frac{\widehat{\sigma}^{(2)}}{2}) + \sqrt{n(\widehat{\sigma}^{(2)})^2 + n^{-1})}\kappa_n \\ &\geqslant \mathbb{P}(\sigma|D_1| \leqslant \frac{\widehat{\sigma}^{(2)}}{2}) + \xi(\widehat{\sigma}^{(2)}) \\ &\geqslant 1 - \sqrt{n(\widehat{\sigma}^{(2)})^2 + n^{-1})}\kappa_n \\ &\geqslant 1 - \sqrt{n(h^*)^2 + n^{-1}}\kappa_n. \end{aligned}$$

Since $\operatorname{Supp} f \subseteq [-1/2, 1/2]$ and f is bounded from below, we conclude on $\Xi \cap \{h^* \geqslant \widehat{\sigma}^{(2)} \geqslant c\sigma\}$,

$$2\sqrt{n(h^*)^2 + n^{-1}}\kappa_n \geqslant \mathbb{P}(\sigma|D_1| > \frac{h^*}{2}) = \mathbb{P}(|D_1| > \frac{1-\varepsilon}{2}) \geqslant \varepsilon \min_{|x| \leqslant 1/2} f(x).$$

Hence, for $\varepsilon = 3(\min_{|x|\leqslant 1/2} f(x))^{-1}\sqrt{n\sigma^2 + n^{-1}}\kappa_n > 2(\min_{|x|\leqslant 1/2} f(x))^{-1}\sqrt{n(h^*)^2 + n^{-1}}\kappa_n$, we have

$$P_2 \leqslant \mathbb{P}(\Xi^c) \to 0.$$

It remains to prove $P_3 \to 0$. Since $\psi^\dagger$ is a triangle kernel, the function $h \mapsto \psi^\dagger(x/h)$ is non decreasing, and so is $h \mapsto \frac{1}{\mu\lambda n}\sum_{i,j}\psi^\dagger((Y_j - X_i)/h)$. Therefore, on the event $\{\overline{h} \geqslant \widehat{\sigma}^{(2)} > \underline{h}\}$ for any $0 \leqslant \underline{h} < \overline{h} < \sigma$ that

$$\begin{aligned} \frac{1}{\mu\lambda n}\sum_{i,j}\psi^\dagger\big((Y_j - X_i)/\overline{h}\big) &\geqslant \frac{1}{\mu\lambda n}\sum_{i,j}\psi^\dagger\big((Y_j - X_i)/\widehat{\sigma}^{(2)}\big) \\ &\geqslant n\lambda\widehat{\sigma}^{(2)} + 1 - \sqrt{n(\widehat{\sigma}^{(2)})^2 + n^{-1}}\kappa_n \\ &\geqslant n\lambda\underline{h} + 1 - \sqrt{n\sigma^2 + n^{-1}}\kappa_n. \end{aligned}$$

Therefore, assuming κ_n is such that $\sqrt{n\sigma^2+n^{-1}}\kappa_n = o(1)$, a choice which is always possible and under the condition and $n\lambda(\overline{h}-\underline{h}) < \mathbb{P}(|D_1| \geqslant \frac{\overline{h}}{2\sigma})$, Markov's inequality yields

$$\begin{aligned}
\mathbb{P}\big(\overline{h} \geqslant \widehat{\sigma}^{(2)} > \underline{h}\big) &\leqslant \mathbb{P}\Big(\frac{1}{\mu\lambda n}\sum_{i,j}\psi^\dagger((Y_j-X_i)/\overline{h}) \geqslant n\lambda\underline{h} + 1 - o(1)\Big) \\
&= \mathbb{P}\Big(\xi(\overline{h}) \geqslant n\lambda(\underline{h}-\overline{h}) + 1 - \mathbb{P}\big(\sigma|D_1| \leqslant \frac{\overline{h}}{2}\big) - o(1)\Big) \\
&\leqslant \mathbb{P}\Big(\xi(\overline{h}) \geqslant \mathbb{P}\big(|D_1| > \frac{\overline{h}}{2\sigma}\big) - n\lambda(\overline{h}-\underline{h}) - o(1)\Big) \\
&\leqslant \frac{\operatorname{Var}\left(\xi(\overline{h})\right))}{\big(\mathbb{P}(|D_1| > \overline{h}/(2\sigma)) - n\lambda(\overline{h}-\underline{h}) - \mathbb{E}[\xi(\overline{h})] - o(1)\big)^2} \\
&= \frac{n\overline{h}^2 + n^{-1}}{\big(\mathbb{P}(|D_1| > \overline{h}/(2\sigma)) - n\lambda(\overline{h}-\underline{h}) + o(1)\big)^2}, \qquad (43)
\end{aligned}$$

where we used (21) for the last estimate. In the case $n\sigma \leqslant c_1 := \mathbb{P}(|D_1| > 1/4)/\lambda$, we can choose $\underline{h} = 0, \overline{h} = \sigma/2$ and conclude

$$\mathbb{P}\Big(\widehat{\sigma}^{(2)} \leqslant \frac{\sigma}{2}\Big) \lesssim \frac{n\sigma^2+n^{-1}}{\big(\mathbb{P}(|D_1|>1/4) - n\sigma\lambda/2 + o(1)\big)^2} \to 0.$$

If $n\sigma > c_1$, we first note that (43) with $\underline{h} = 0, \overline{h} = 1/(2n)$ yields

$$\begin{aligned}
\mathbb{P}\Big(\widehat{\sigma}^{(2)} \leqslant \frac{1}{2n}\Big) &\lesssim \frac{n^{-1}}{\big(\mathbb{P}(|D_1| > 1/(4n\sigma)) - \sigma\lambda/2 + o(1)\big)^2} \\
&\lesssim \frac{n^{-1}}{\big(\mathbb{P}(|D_1| > 1/(4c_1)) + o(1)\big)^2} \to 0.
\end{aligned}$$

To improve this bound in the case $\sigma > \frac{1}{n}$, we choose $h_i := \frac{1}{2n} + \frac{c_1}{2}\frac{i}{n} \in [\frac{1}{2n}, \frac{1}{2n} + \frac{\sigma}{2}]$ for $0 \leqslant i \leqslant I := \lceil \frac{n}{c_1}(\sigma - \frac{1}{n})\rceil = O(\sigma n)$ and estimate using (43)

$$\begin{aligned}
\mathbb{P}\Big(\widehat{\sigma}^{(2)} \leqslant \frac{\sigma}{2}\Big) &\leqslant \mathbb{P}(\widehat{\sigma}^{(2)} \leqslant \frac{1}{2n}) + \sum_{i=1}^{I}\mathbb{P}(h_{i-1} < \widehat{\sigma}^{(2)} \leqslant h_i) \\
&\leqslant \sum_{i=1}^{I}\frac{nh_i^2+n^{-1}}{\big(\mathbb{P}(|D_1| > h_i/(2\sigma)) - n\lambda(\overline{h}-\underline{h}) + o(1)\big)^2} + o(1) \\
&\leqslant \sum_{i=1}^{I}\frac{nh_i^2+n^{-1}}{\big(\mathbb{P}(|D_1| > 1/(4\sigma n) + \frac{1}{4}) - c_1\lambda/2 + o(1)\big)^2} + o(1) \\
&\leqslant \sum_{i=1}^{I}\frac{nh_i^2+n^{-1}}{\big(\mathbb{P}(|D_1| > \frac{1}{4})/2 + o(1)\big)^2} + o(1) \\
&\lesssim \frac{I}{n} + \frac{1}{n}\sum_{i=1}^{I} i^2 + o(1) \\
&\lesssim \frac{I}{n} + \frac{I^3}{n} + o(1) \lesssim \sigma + n^2\sigma^3. \qquad \square
\end{aligned}$$

LEMMA 24. *Let $c \in (0,1)$ and $\sqrt{n}\sigma = O(1)$. There is for all $\varepsilon > 0$ some $\kappa > 0$ such that*

$$\mathbb{P}(\Xi^c) \leqslant \varepsilon \qquad \text{for} \qquad \Xi := \Big\{ \sup_{h\in[c\sigma,\sigma]} \frac{|\xi(h)|}{\sqrt{nh^2+n^{-1}}} \leqslant \kappa \Big\}. \tag{44}$$

PROOF. *Step 1: We first bound $\mathbb{E}[\xi(h)]$.* For $\psi^\dagger = \mathbb{1}_{[-\frac{1}{2},\frac{1}{2}]} * \mathbb{1}_{[-\frac{1}{2},\frac{1}{2}]}$ and $D \sim f$, we have $Y \sim f_\sigma * p$

$$\begin{aligned}
\mathbb{E}\Big[\frac{1}{\mu\lambda n}\sum_{i,j}\psi^\dagger((Y_j - X_i)/h)\Big] &= \int \psi^\dagger(z/h) f_\sigma(z)\mathrm{d}z + n\lambda \int_0^1 \int \psi^\dagger(\frac{y+x}{h})(f_\sigma * p)(y)\mathrm{d}y\mathrm{d}x \\
&= \mathbb{E}[\psi^\dagger(\sigma D/h)] + n\lambda h \int_0^1 \mathbb{E}[\psi_h^\dagger(Y+x)]\mathrm{d}x \\
&= \mathbb{E}[\psi^\dagger(\sigma D/h)] + n\lambda h - n\lambda h \int_{\mathbb{R}\setminus[0,1]} \mathbb{E}[\psi_h^\dagger(Y+x)]\mathrm{d}x,
\end{aligned}$$

using $\int_{\mathbb{R}} \mathbb{E}[\psi_h^\dagger(Y+x)]\mathrm{d}x = \int \psi_h^\dagger(x)\mathrm{d}x = 1$. Since $\|f_\sigma * p\|_\infty \leqslant 1$ we have

$$\begin{aligned}
\int_{\mathbb{R}\setminus[0,1]} \mathbb{E}[\psi_h^\dagger(Y+x)]\mathrm{d}x &= \mathbb{E}\Big[\int_{(-\infty,Y-1]\cup[Y,\infty)} \psi_h^\dagger(x)\mathrm{d}x\Big] \\
&\leqslant h^{-1}\|\psi^\dagger\|_\infty \mathbb{E}\big[\mathbb{1}_{\{-h<Y-1\}}\big(h \wedge (Y-1) + h\big) + \mathbb{1}_{\{Y<h\}}\big(h - (-h \vee Y)\big)\big] \\
&\leqslant 2\|\psi^\dagger\|_\infty \mathbb{P}(\{Y > 1-h\} \cup \{Y < h\}) \\
&\leqslant 2\|\psi^\dagger\|_\infty (2h + \sigma).
\end{aligned}$$

With the notation $K = \mathbb{1}_{[-\frac{1}{2},\frac{1}{2}]}$ we have $\psi^\dagger(z/h) = \mathbb{1}_{[-h/2,h/2]} * K_h(z)$ and since f has a bounded density, we obtain

$$\begin{aligned}
\mathbb{E}[\psi^\dagger(\sigma D/h)] &= \mathbb{P}(\sigma|D_1| \leqslant \frac{h}{2}) + (F * K_h - F)(\frac{h}{2\sigma}) - (F * K_h - F)(-\frac{h}{2\sigma}) \\
&= \mathbb{P}(\sigma|D_1| \leqslant \frac{h}{2}) + O(h).
\end{aligned}$$

Therefore,

$$\mathbb{E}[\xi(h)] = \mathbb{E}\Big[\frac{1}{\mu\lambda n}\sum_{i,j}\mathbb{1}^{*2}_{[-h/2,h/2]}(Y_j - X_i)\Big] - \mathbb{P}(\sigma|D_1| \leqslant \frac{h}{2}) - n\lambda h = O(nh(h+\sigma) + h).$$

In particular, since $\sigma \leq n\sigma^2 + n^{-1}$ always, we have for all $h \in (0,\sigma)$

$$\frac{|\mathbb{E}[\xi(h)]|}{\sqrt{nh^2+n^{-1}}} \leqslant \frac{nh(\sigma+h)}{\sqrt{n}h} + \sqrt{n}h \leqslant 3\sqrt{n}\sigma$$

which is uniformly bounded.

Step 2: It remains to prove tightness of

$$\sup_{h\in[c\sigma,\sigma]} \frac{|\xi(h) - \mathbb{E}[\xi(h)]|}{\sqrt{nh^2+n^{-1}}}.$$

To this end, we apply the Kolmogorov-Chentsov criterion to the process

$$V_t := \frac{\xi(h_t) - \mathbb{E}[\xi(h_t)]}{\sqrt{nh_t^2+n^{-1}}}, \qquad h_t := t\sigma, t \in (c,1].$$

Due to (21), we have $V_c = O_{\mathbb{P}}(1)$. For $0 < s < t \leqslant 1$ we decompose increments into

$$V_t - V_s = \frac{1}{n\lambda\mu}\sum_{i,j}\big(\Delta_{s,t}(Y_j - X_i) - \mathbb{E}[\Delta_{s,t}(Y_j - X_i)]\big),$$
$$\Delta_{s,t}(z) := \frac{1}{\sqrt{nh_t^2 + n^{-1}}}\psi^\dagger(z/h_t) - \frac{1}{\sqrt{nh_s^2 + n^{-1}}}\psi^\dagger(z/h_s).$$

Proposition 20 with $\psi_2 = \Delta_{s,t}(\sigma\cdot)$ and $\psi_1 = \mathbb{1}_{[-2,2]}$ yields

$$\begin{aligned}\mathrm{Var}\Big(\frac{1}{n\lambda\mu}\sum_{i,j}\big(\Delta_{s,t}(Y_j - X_i)\big)\Big) &\lesssim \frac{1}{n}\big((n\sigma + n^2\sigma^2)\sigma^{-2}\|\Delta_{s,t}\|_{L^1}^2 + (n\sigma+1)\sigma^{-1}\|\Delta_{s,t}\|_{L^2}^2\big)\\ &= (\sigma^{-1} + n)\|\Delta_{s,t}\|_{L^1}^2 + (1 + (n\sigma)^{-1})\|\Delta_{s,t}\|_{L^2}^2. \qquad (45)\end{aligned}$$

We can bound the above L^1-norm by

$$\begin{aligned}\|\Delta_{s,t}\|_{L^1}^2 &\leqslant 2\Big(\frac{1}{\sqrt{nh_t^2 + n^{-1}}} - \frac{1}{\sqrt{nh_s^2 + n^{-1}}}\Big)^2\Big(\int|\psi^\dagger(z/h_t)|\mathrm{d}z\Big)^2\\ &\quad + \frac{2}{nh_s^2 + n^{-1}}\Big(\int|\psi^\dagger(z/h_t) - \psi(z/h_s)|\mathrm{d}z\Big)^2 = 2T_1^2 + 2T_2^2.\end{aligned}$$

Using $\frac{1}{\sqrt{a}} - \frac{1}{\sqrt{b}} = \frac{b-a}{\sqrt{ab}(\sqrt{a}+\sqrt{b})} \leqslant \frac{b-a}{\sqrt{ab}}$, $a + b \geqslant 2\sqrt{ab}$ and $s \leqslant t$, we estimate

$$\begin{aligned}(\sigma^{-1} + n)T_1^2 &\leqslant (\sigma^{-1} + n)\frac{n^2h_t^2(h_t^2 - h_s^2)^2}{(nh_t^2 + n^{-1})^2(nh_s^2 + n^{-1})}\Big(\int|\psi^\dagger(z)|\mathrm{d}z\Big)^2\\ &\lesssim (\sigma^{-1} + n)\frac{n^2\sigma^6(t^2 - s^2)^2}{(n\sigma^2t^2 + n^{-1})^2(n\sigma^2s^2 + n^{-1})}\\ &\lesssim \frac{n^2\sigma^5(t-s)^2(t+s)^2}{n^2\sigma^5t^3s^2} + \frac{n^3\sigma^6(t-s)^2(t+s)^2}{n^3\sigma^6t^4s^2} \lesssim c^{-4}(t-s)^2\end{aligned}$$

For T_2 the mean value theorem yields

$$\begin{aligned}(\sigma^{-1} + n)T_2^2 &\leqslant \frac{\sigma^{-1} + n}{nh_s^2 + n^{-1}}\sup_{r\in[s,t]}\Big(\int\Big|\frac{z}{h_t} - \frac{z}{h_s}\Big||(\psi^\dagger)'(\frac{z}{h_r})|\mathrm{d}z\Big)^2\\ &= \frac{\sigma^{-1} + n}{nh_s^2 + n^{-1}}\sup_{r\in[s,t]}h_r^2\Big(\frac{h_r}{h_s} - \frac{h_r}{h_t}\Big)^2\Big(\int|z||(\psi^\dagger)'(z)|\mathrm{d}z\Big)^2\\ &\lesssim \frac{\sigma + n\sigma^2}{n\sigma^2s^2 + n^{-1}}\sup_{r\in[s,t]}r^2\Big(\frac{r}{s} - \frac{r}{t}\Big)^2\\ &\lesssim \frac{t^4}{s^2}\Big(\frac{t-s}{ts}\Big)^2 \leqslant c^{-4}(t-s)^2.\end{aligned}$$

Similarly we proceed with the L^2-norm in (45):

$$\begin{aligned}\|\Delta_{s,t}\|_{L^2}^2 &\leqslant 2\Big(\frac{1}{\sqrt{nh_t^2 + n^{-1}}} - \frac{1}{\sqrt{nh_s^2 + n^{-1}}}\Big)^2\int|\psi^\dagger(z/h_t)|^2\mathrm{d}z\\ &\quad + \frac{2}{nh_s^2 + n^{-1}}\int|\psi^\dagger(z/h_t) - \psi^\dagger(z/h_s)|^2\mathrm{d}z =: 2S_1^2 + 2S_2^2.\end{aligned}$$

with

$$\begin{aligned}(1+(n\sigma)^{-1})S_1^2 &\lesssim \Big(1+\frac{1}{n\sigma}\Big)h_t\Big(\frac{1}{\sqrt{nh_t^2+n^{-1}}}-\frac{1}{\sqrt{nh_s^2+n^{-1}}}\Big)^2\\ &\leqslant \Big(1+\frac{1}{n\sigma}\Big)\frac{n^2h_t(h_t^2-h_s^2)^2}{(nh_t^2+n^{-1})^2(nh_s^2+n^{-1})}\\ &\lesssim \Big(1+\frac{1}{n\sigma}\Big)\frac{n^2\sigma^4(t+s)^2(t-s)^2}{(n\sigma^2t^2+n^{-1})(n\sigma^2s^2+n^{-1})}\\ &\lesssim \frac{n^2\sigma^4(t+s)^2}{n^2\sigma^4t^2s^2}(t-s)^2+\frac{n\sigma^3(t+s)^2}{n\sigma^3t^2s^2}(t-s)^2\lesssim c^{-2}(t-s)^2\end{aligned}$$

and

$$\begin{aligned}(1+(n\sigma)^{-1})S_2^2 &\lesssim \frac{1+(n\sigma)^{-1}}{nh_s^2+n^{-1}}\sup_{r\in[s,t]}\int\Big|\frac{z}{h_t}-\frac{z}{h_s}\Big|^2|(\psi^\dagger)'(\frac{z}{h_r})|^2\mathrm{d}z\\ &\lesssim \frac{1+(n\sigma)^{-1}}{nh_s^2+n^{-1}}\sup_{r\in[s,t]}h_r\Big(\frac{h_r}{h_s}-\frac{h_r}{h_t}\Big)^2\lesssim\frac{\sigma+n^{-1}}{n\sigma^2s^2+n^{-1}}\frac{t^3}{t^2s^2}(t-s)^2\lesssim c^{-3}(t-s)^2.\end{aligned}$$

These calculations verify

$$\mathbb{E}\big[(V_t-V_s)^2\big]=\mathrm{Var}\Big(\frac{1}{n\lambda\mu}\sum_{i,j}\big(\Delta_{s,t}(Y_j-X_i)\big)\Big)\lesssim(t-s)^2.$$

Hence, (V_t) has an α-Hölder regular modification for any $\alpha\in(0,1/2)$ implying tightness. □

6.5. *Proofs for the plug-in estimators.*

Proof of Theorem 12. *(i)* We analyse the deconvolution estimator in four steps.
Step 1: Prefactor. Defining

$$\overline{f}^{(1)}_{\widehat{\sigma}}(x_0):=\frac{1}{\widehat{\sigma}\widehat{h}_1^2\lambda\mu n}\sum_j K'\Big(\frac{z_0}{\widehat{h}_1}-\frac{Y_j}{\widehat{\sigma}\widehat{h}_1}\Big)\Big(\frac{1}{n\lambda}\sum_i K\Big(\frac{\widehat{\sigma}z_0}{9}-\frac{Y_j-X_i}{9}\Big)\Big),$$

we have

$$\widetilde{f}^{(1)}_{\widehat{\sigma}}(z_0)-f(z_0)=\frac{\lambda\mu n}{|\mathcal{Y}|}\frac{\lambda n}{|\mathcal{X}|}(\overline{f}_{\widehat{\sigma}}(z_0)-f(z_0))+\Big(\frac{\lambda\mu n}{|\mathcal{Y}|}\frac{\lambda n}{|\mathcal{X}|}-1\Big)f(z_0).$$

For $\tau_n\to\infty$ the event

$$\Lambda:=\Big\{\Big|\frac{|\mathcal{Y}|}{\lambda\mu n}-1\Big|\leqslant\frac{\tau_n}{\sqrt{n}}\Big\}\cup\Big\{\Big|\frac{|\mathcal{X}|}{\lambda n}-1\Big|\leqslant\frac{\tau_n}{\sqrt{n}}\Big\}$$

satisfies

$$\begin{aligned}\mathbb{P}(\Lambda^c)&\leqslant\mathbb{P}\Big(\Big|\frac{|\mathcal{Y}|}{\lambda\mu n}-1\Big|>\frac{\tau_n}{\sqrt{n}}\Big)+\mathbb{P}\Big(\Big|\frac{|\mathcal{X}|}{\lambda n}-1\Big|>\frac{\tau_n}{\sqrt{n}}\Big)\\ &\leqslant\frac{n}{\tau_n^2(\lambda\mu n)^2}\mathbb{E}\big[\,\mathrm{Var}\left(|\mathcal{Y}|\,\big|\,|\mathcal{X}|\right)\big]+\frac{n}{\tau_n^2(\lambda n)^2}\mathrm{Var}(|\mathcal{X}|)\\ &=\frac{1}{\tau_n^2(\lambda\mu)^2n}\mathbb{E}[\mu|\mathcal{X}|]+\frac{1}{\tau_n^2\lambda}\end{aligned}$$

$$= \frac{1}{\tau_n^2\lambda\mu} + \frac{1}{\tau_n^2\lambda} \to 0,$$

due to $|\mathcal{Y}||\mathcal{X}| \sim \mathrm{Poiss}(\mu|\mathcal{X}|)$ and $|\mathcal{X}| \sim \mathrm{Poiss}(\lambda n)$. On Λ we have for $\tau_n/\sqrt{n} \leqslant 1/2$ that

$$\begin{aligned}\frac{\lambda\mu n}{|\mathcal{Y}|}\frac{\lambda n}{|\mathcal{X}|} - 1 &= \Big(\frac{\lambda\mu n}{|\mathcal{Y}|} - 1\Big)\frac{\lambda n}{|\mathcal{X}|} + \frac{\lambda n}{|\mathcal{X}|} - 1 \\ &= \Big(\frac{1 - |\mathcal{Y}|/(\lambda\mu n)}{1 - (1 - |\mathcal{Y}|/\lambda\mu n)}\Big)\frac{\lambda n}{|\mathcal{X}|} + \frac{1 - |\mathcal{X}|/(\lambda n)}{1 - (1 - |\mathcal{X}|/\lambda n)} \leqslant 6\tau_n/\sqrt{n}.\end{aligned}$$

Since r_n is always slower than $n^{-1/2}$, we conclude

$$r_n^{-1}|\widetilde{f}^{(1)}_{\widehat{\sigma}}(z_0) - f(z_0)| = O_{\mathbb{P}}\Big(r_n^{-1}|\overline{f}^{(1)}_{\widehat{\sigma}}(z_0) - f(z_0)|\Big) + o_{\mathbb{P}}(1).$$

Step 2: From $\widehat{\sigma}$ to σ. Consider the event

$$\Sigma := \{\widehat{\sigma} \in [\sigma(1-\varepsilon_n), \sigma(1+\varepsilon_n)]\}, \qquad \varepsilon_n = \frac{\log n}{\sqrt{\sigma n}}$$

satisfying

$$\mathbb{P}(\Sigma^c) = \mathbb{P}(|\widehat{\sigma} - \sigma| > \varepsilon_n\sigma) = \mathbb{P}(|\frac{\widehat{\sigma}}{\sigma} - 1| > \frac{\log n}{\sqrt{\sigma n}}) \to 0$$

due to Theorem 11. Writing $\overline{h}_1 = (n\overline{\sigma})^{-1/(2s+3)}$ for any $\overline{\sigma} > 0$, we have on Σ

$$\begin{aligned} r_n^{-1}|\overline{f}^{(1)}_{\widehat{\sigma}}(z_0) - f(z_0)| &\leqslant \sup_{\overline{\sigma}:|\overline{\sigma}-\sigma|\leqslant\varepsilon_n\sigma} r_n^{-1}\Big|\overline{f}^{(1)}_{\overline{\sigma}}(z_0) - \mathbb{E}[\overline{f}^{(1)}_{\overline{\sigma}}(z_0)] - \big(f(z_0) - \mathbb{E}[\overline{f}^{(1)}_{\overline{\sigma}}(z_0)]\big)\Big| \\ &\leqslant \sup_{\overline{\sigma}:|\overline{\sigma}-\sigma|\leqslant\varepsilon_n\sigma} \frac{\sqrt{\overline{\sigma}n\overline{h}_1^3}}{r_n\sqrt{\overline{\sigma}n\overline{h}_1^3}}\Big|\overline{f}^{(1)}_{\overline{\sigma}}(z_0) - \mathbb{E}[\overline{f}^{(1)}_{\overline{\sigma}}(z_0)]\Big| + \sup_{\overline{\sigma}:|\overline{\sigma}-\sigma|\leqslant\varepsilon_n\sigma} \frac{\overline{h}_1^{-s}}{r_n\overline{h}_1^{-s}}\Big|\big(f(z_0) - \mathbb{E}[\overline{f}^{(1)}_{\overline{\sigma}}(z_0)]\big)\Big| \\ &\lesssim \sup_{\overline{\sigma}:|\overline{\sigma}-\sigma|\leqslant\varepsilon_n\sigma} \sqrt{\overline{\sigma}n\overline{h}_1^3}\Big|\overline{f}^{(1)}_{\overline{\sigma}}(z_0) - \mathbb{E}[\overline{f}^{(1)}_{\overline{\sigma}}(z_0)]\Big| + \sup_{\overline{\sigma}:|\overline{\sigma}-\sigma|\leqslant\varepsilon_n\sigma} \overline{h}_1^{-s}\Big|\big(f(z_0) - \mathbb{E}[\overline{f}^{(1)}_{\overline{\sigma}}(z_0)]\big)\Big| \end{aligned} \tag{46}$$

using in the last step that the minimax rate satisfies $r_n = (h_1)^s = \big(n\sigma(h_1)^3\big)^{-1/2}$ and thus

$$\begin{aligned}\frac{1}{r_n^2\overline{\sigma}n\overline{h}_1^3} &= \frac{\sigma n(h_1)^3}{\overline{\sigma}n\overline{h}_1^3} = \Big(\frac{\sigma}{\overline{\sigma}}\Big)^{2s/(2s+3)} \leqslant (1-\varepsilon_n)^{-2s/(2s+3)}, \\ \frac{1}{r_n\overline{h}_1^{-s}} &= \Big(\frac{\overline{h}_1}{h_1}\Big)^s = \Big(\frac{\overline{\sigma}}{\sigma}\Big)^{s/(2s+3)} \leqslant (1+\varepsilon_n)^{s/(2s+3)}.\end{aligned}$$

Subsequently, we will bound both terms in (46) separately. To this end, we proceed similarly to the proof of Proposition 5(i). To incorporate $\overline{\sigma}$ we set

$$\overline{\psi}_1 := \frac{1}{\overline{\sigma}\overline{h}_1^2}K'\Big(\frac{z_0}{\overline{h}_1} - \frac{\cdot}{\overline{\sigma}\overline{h}_1}\Big) \qquad \text{and} \qquad \overline{\psi}_2 := \frac{1}{\overline{h}_2}K\Big(\frac{z_0}{\overline{h}_2} - \frac{\cdot}{\overline{h}_2}\Big). \tag{47}$$

where $\overline{h}_2 := 9/\overline{\sigma}$.

Step 3: Bias. The analog to decomposition (7) leads to

$$\overline{\mathcal{U}}_\sigma(f*p) = \int_{\mathbb{R}} \overline{\psi}_1(y)(\overline{\psi}_2 * \mathbb{1}_{[0,1/\overline{\sigma}]})(y/\overline{\sigma})(f_\sigma * p)(y)\mathrm{d}y, \quad \overline{\mathcal{V}}_\sigma(f) = \int_{\mathbb{R}} (\overline{\psi}_1 * \mathbb{1}_{[-1,0]})(\sigma z)\overline{\psi}_2(z)f(z)\mathrm{d}z$$

where σ is the true data-generating parameter. Along the lines of the proof of Propositions 18(i) and 19(ii) we obtain

$$\mathbb{E}[\overline{\mathcal{U}}_\sigma(f*p)] = \frac{1}{\overline{\sigma}\overline{h}_2} f_{\sigma/\overline{\sigma}}(z_0) + O(\frac{h_1^s}{\sigma h_2}) \qquad \text{and} \qquad \mathbb{E}[\overline{\mathcal{V}}_\sigma(f)] \lesssim \overline{h}_1^{-1}.$$

Note that

$$f_{\sigma/\overline{\sigma}}(z_0) - f(z_0) = \frac{\overline{\sigma}}{\sigma}(f(\overline{\sigma}z_0/\sigma) - f(z_0)) + (\frac{\overline{\sigma}}{\sigma} - 1)f(z_0) \lesssim \Big(\frac{\log(n\sigma)}{n\sigma}\Big)^{(1\wedge s)/2}.$$

Therefore, we obtain the following modification of (34):

$$\begin{aligned}\mathbb{E}\Big[\overline{f}_{\overline{\sigma}}^{(1)}(z_0)\Big] &= \overline{\sigma}\overline{h}_2\,\overline{\mathcal{U}}_{\overline{\sigma}}(f*p) + \frac{\overline{h}_2}{n\lambda}\overline{\mathcal{V}}_{\overline{\sigma}}(f) \\ &= f(z_0) + O\Big(h_1^s + \Big(\frac{\log(n\sigma)}{n\sigma}\Big)^{(1\wedge s)/2} + \frac{h_2}{nh_1}\Big).\end{aligned}$$

We conclude

$$\sup_{\overline{\sigma}:|\overline{\sigma}-\sigma|\leqslant\varepsilon_n\sigma} \overline{h}_1^{-s}\Big|(f(z_0) - \mathbb{E}[\overline{f}_{\overline{\sigma}}^{(1)}(z_0)])\Big| \lesssim 1 + \Big(\frac{\log(\sigma n)}{\sigma n}\Big)^{(1\wedge s)/2} r_n^{-1} + \frac{r_n^{-1}}{n\sigma h_1} \lesssim 1.$$

Step 4: Stochastic error term. Define

$$\sigma_t := \sigma(1 - \varepsilon_n + 2\varepsilon_n t), \qquad h_t := (n\sigma_t)^{-1/(2s+3)}, \qquad t \in [0,1],$$

as well as

$$V_t := \sqrt{\sigma_t n h_t^3}(\overline{f}_{\sigma_t}^{(1)}(z_0) - \mathbb{E}[\overline{f}_{\sigma_t}^{(1)}(z_0)])$$

We thus have to prove tightness of the process $(V_t)_{t\in[0,1]}$ that is $\sup_{t\in[0,1]}|V_t| = O_{\mathbb{P}}(1)$. As in the proof of Proposition 10, we apply the Kolmogorov-Chentsov criterion. Writing

$$\begin{aligned}V_t - V_s =& \frac{1}{(\lambda\mu\sqrt{n})(\lambda n)} \sum_{i,j} (\Delta_{s,t}^{(1)}(X_i, Y_j) - \mathbb{E}[\Delta_{s,t}^{(1)}(X_i, Y_j)]) \\ &+ \frac{1}{(\lambda\mu\sqrt{n})(\lambda n)} \sum_{i,j} (\Delta_{s,t}^{(2)}(X_i, Y_j) - \mathbb{E}[\Delta_{s,t}^{(2)}(X_i, Y_j)])\end{aligned}$$

with

$$\Delta_{s,t}^{(1)}(x,y) = \underbrace{\Big(\frac{1}{\sqrt{\sigma_t h_t}} K'\Big(\frac{z_0}{h_t} - \frac{y}{\sigma_t h_t}\Big) - \frac{1}{\sqrt{\sigma_s h_s}} K'\Big(\frac{z_0}{h_s} - \frac{y}{\sigma_s h_s}\Big)\Big)}_{=:\Delta_{s,t}^{(1,1)}(y)} \underbrace{K\Big(\frac{\sigma_t z_0}{9} - \frac{y-x}{9}\Big)}_{=:\Delta_t^{(1,2)}((y-x)/\sigma)},$$

$$\Delta^{(2)}_{s,t}(x,y)=\underbrace{\frac{1}{\sqrt{\sigma_s h_s}}K'\Big(\frac{z_0}{h_s}-\frac{y}{\sigma_s h_s}\Big)}_{=:\Delta^{(2,1)}_s(y)}\Big(\underbrace{K\Big(\frac{\sigma_t z_0}{9}-\frac{y-x}{9}\Big)-K\Big(\frac{\sigma_s z_0}{9}-\frac{y-x}{9}\Big)}_{=:\Delta^{(2,2)}_{s,t}((y-x)/\sigma)}\Big).$$

Proposition 20 yields

$$\begin{aligned}\mathbb{E}[(V_t-V_s)^2]\lesssim&\Big((\frac{\sigma}{n}+\sigma^2)\|\Delta^{(1,2)}_t\|^2_{L^1}+(\frac{\sigma}{n}+\frac{1}{n^2})\|\Delta^{(1,2)}_t\|^2_{L^2}\Big)\|\Delta^{(1,1)}_{s,t}\|^2_{L^2}\\&+\Big((\frac{\sigma}{n}+\sigma^2)\|\Delta^{(2,2)}_{s,t}\|^2_{L^1}+(\frac{\sigma}{n}+\frac{1}{n^2})\|\Delta^{(2,2)}_{s,t}\|^2_{L^2}\Big)\|\Delta^{(2,1)}_s\|^2_{L^2}\\\lesssim&\Big(\frac{1}{n\sigma}+1+\frac{1}{n}+\frac{1}{n^2\sigma}\Big)\|\Delta^{(1,1)}_{s,t}\|^2_{L^2}\\&+\Big((\frac{\sigma}{n}+\sigma^2)\|\Delta^{(2,2)}_{s,t}\|^2_{L^1}+(\frac{\sigma}{n}+\frac{1}{n^2})\|\Delta^{(2,2)}_{s,t}\|^2_{L^2}\Big)\end{aligned}\tag{48}$$

We have to bound the norms in the previous line. We have

$$\begin{aligned}\|\Delta^{(1,1)}_{s,t}\|^2_{L^2}\leqslant&\,3\Big(\frac{1}{\sqrt{\sigma_t h_t}}-\frac{1}{\sqrt{\sigma_s h_s}}\Big)^2\int K'\Big(\frac{z_0}{h_t}-\frac{y}{\sigma_t h_t}\Big)\Big)^2\mathrm{d}y\\&+\frac{3}{\sigma_s h_s}\int\Big(K'\Big(\frac{z_0}{h_t}-\frac{y}{\sigma_t h_t}\Big)-K'\Big(\frac{z_0}{h_s}-\frac{y}{\sigma_t h_s}\Big)\Big)^2\mathrm{d}y\\&+\frac{3}{\sigma_s h_s}\int\Big(K'\Big(\frac{z_0}{h_s}-\frac{y}{\sigma_t h_s}\Big)-K'\Big(\frac{z_0}{h_s}-\frac{y}{\sigma_s h_s}\Big)\Big)^2\mathrm{d}y\\=:&\,3T_1+3T_2+3T_3.\end{aligned}$$

Using $K'\in L^2$, we have

$$T_1\lesssim\Big(1-\frac{\sqrt{\sigma_t h_t}}{\sqrt{\sigma_s h_s}}\Big)^2=\Big(1-\Big(\frac{\sigma_t}{\sigma_s}\Big)^{(s+1)/(2s+3)}\Big)^2\lesssim(\frac{\sigma_s-\sigma_t}{\sigma_s})^2\lesssim(t-s)^2.$$

Moreover, for some intermediate point $r\in[s,t]$ we have

$$\begin{aligned}T_2=\frac{1}{h_s}\int\Big(K'\Big(\frac{y}{h_t}\Big)-K'\Big(\frac{y}{h_s}\Big)\Big)^2\mathrm{d}y&=\frac{1}{h_s}\int\Big(\frac{y}{h_t}-\frac{y}{h_s}\Big)^2K''\Big(\frac{y}{h_r}\Big)^2\mathrm{d}y\\&\leqslant\sup_{r\in[s,t]}\frac{h_r}{h_s}\Big(\frac{h_r}{h_t}-\frac{h_r}{h_s}\Big)^2\int y^2K''(y)^2\mathrm{d}y\\&\lesssim(t-s)^2,\end{aligned}$$

and

$$\begin{aligned}T_3=\frac{1}{\sigma_s}\int\Big(K'\Big(\frac{z_0}{h_s}-\frac{y}{\sigma_t}\Big)-K'\Big(\frac{z_0}{h_s}-\frac{y}{\sigma_s}\Big)\Big)^2\mathrm{d}y&=\frac{1}{\sigma_s}\int\Big(\frac{y}{\sigma_t}-\frac{y}{\sigma_s}\Big)^2K''\Big(\frac{z_0}{h_s}-\frac{y}{\sigma_r}\Big)^2\mathrm{d}y\\&\leqslant\sup_{r\in[s,t]}\frac{\sigma_r}{\sigma_s}\Big(\frac{\sigma_r}{\sigma_t}-\frac{\sigma_r}{\sigma_s}\Big)^2\int y^2K''(\frac{z_0}{h_s}-y)^2\mathrm{d}y\\&\lesssim\frac{(\sigma_t-\sigma_s)^2}{\sigma_s^2h_s^2}\\&\lesssim\frac{\varepsilon_n^2(t-s)^2}{h_s^2}\lesssim(t-s)^2\end{aligned}$$

because $\varepsilon_n^2 h_s^{-2} \to 0$. For $\Delta_{s,t}^{(2,2)}$ we have similarly

$$\begin{aligned}\|\Delta_{s,t}^{(2,2)}\|_{L^1} &= \int \Big|K\Big(\frac{\sigma_t z_0}{9} - \frac{\sigma z}{9}\Big) - K\Big(\frac{\sigma_s z_0}{9} - \frac{\sigma z}{9}\Big)\Big|\mathrm{d}z \\ &= \frac{1}{81}\int |\sigma_t z_0 - \sigma_s z_0|\Big|K\Big(\frac{\sigma_r z_0}{9} - \frac{\sigma z}{9}\Big)\Big|\mathrm{d}z \lesssim \frac{|\sigma_t - \sigma_s|}{\sigma} \lesssim |t-s|,\end{aligned}$$

and

$$\begin{aligned}\|\Delta_{s,t}^{(2,2)}\|_{L^2}^2 &= \int \Big(K\Big(\frac{\sigma_t z_0}{9} - \frac{\sigma z}{9}\Big) - K\Big(\frac{\sigma_s z_0}{9} - \frac{\sigma z}{9}\Big)\Big)^2\mathrm{d}z \\ &= \frac{1}{81}\int \big(\sigma_t z_0 - \sigma_s z_0\big)^2 K\Big(\frac{\sigma_r z_0}{9} - \frac{\sigma z}{9}\Big)\Big)\Big)^2\mathrm{d}z. \lesssim \frac{(\sigma_t - \sigma_s)^2}{\sigma} \lesssim \sigma(t-s)^2\end{aligned}$$

It follows that $\mathbb{E}\big[(V_t - V_s)^2\big] \lesssim (t-s)^2$ and (V_t) has an α-Hölder regular modification for any $\alpha \in (0, 1/2)$ implying tightness. We have shown (i).

(ii) We verify the convergence rate for the direct estimator similiarly to (i).
Step 1: Reduction. We define

$$\overline{f}_{\widehat{\sigma}}^{(2)}(z_0) = \frac{1}{\widehat{h}_2 \lambda\mu n}\sum_{i,j} 2K'\big(2(\widehat{\sigma}z_0 - Y_j)\big)K\Big(\frac{z_0}{\widehat{h}_2} - \frac{Y_j - X_i}{\widehat{\sigma}\widehat{h}_2}\Big) - \widehat{\sigma}|\mathcal{X}|.$$

Exactly as in Step 1 of the proof of (i), we see that

$$r_n^{-1}|\widetilde{f}_{\widehat{\sigma}}^{(2)}(z_0) - f(z_0)| = O_{\mathbb{P}}\Big(r_n^{-1}|\overline{f}_{\widehat{\sigma}}^{(2)}(z_0) - f(z_0)|\Big) + o_{\mathbb{P}}(1).$$

For $\sigma = o(n^{-2/3})$ we have $\mathbb{P}(\Sigma^c) \to 0$ for $\Sigma := \{\widehat{\sigma} \in [\sigma(1-\varepsilon_n), \sigma(1+\varepsilon_n)]\}$ with

$$\varepsilon_n = \kappa_n\Big(\sqrt{n\sigma^2 + n^{-1}} \wedge \frac{1}{\sqrt{n\sigma}}\Big), \qquad \kappa_n = \log n$$

due to Theorem 11. Writing $\overline{h}_2 = (n \wedge \overline{\sigma}^{-1})^{-1/(2s+1)}$ for $\overline{\sigma} \in [\sigma(1-\varepsilon_n), \sigma(1+\varepsilon_n)]$ we note that

$$\begin{aligned} r_n\overline{h}_2^{-s} \gtrsim \big((n \wedge \sigma^{-1})^{-s(2s+1)} + \sqrt{n\sigma^2}\big)h_2^{-s} \geqslant 1 \qquad &\text{and} \\ r_n\big((n \wedge \sigma^{-1})^{-1/2}\overline{h}_2^{-1/2} + \sqrt{n\sigma^2}\big)^{-1} \gtrsim 1.&\end{aligned}$$

Hence, as in Step 2 of the proof of (i) we have on Σ

$$\begin{aligned} r_n^{-1}|\overline{f}_{\widehat{\sigma}}^{(2)}(z_0) - f(z_0)| \lesssim &\sup_{\overline{\sigma}:|\overline{\sigma}-\sigma|\leqslant\varepsilon_n\sigma} \sqrt{\min(n\overline{h}_2, \sigma^{-1}\overline{h}_2, (n\sigma^2)^{-1})}\Big|\overline{f}_{\overline{\sigma}}^{(2)}(z_0) - \mathbb{E}[\overline{f}_{\overline{\sigma}}^{(2)}(z_0)]\Big| \\ &+ \sup_{\overline{\sigma}:|\overline{\sigma}-\sigma|\leqslant\varepsilon_n\sigma} \overline{h}_2^{-s}\Big|\big(f(z_0) - \mathbb{E}[\overline{f}_{\overline{\sigma}}^{(1)}(z_0)]\big)\Big| \end{aligned} \tag{49}$$

Step 2: Bias. We will use the notation $\overline{\psi}_1$ and $\overline{\psi}_2$ from (47) and the resulting $\overline{\mathcal{U}}_\sigma(f * p)$ and $\overline{\mathcal{V}}_\sigma(f)$. Note that $\overline{h}_1 = 1/(2\overline{\sigma})$ in this case. With minor modifications in the proofs of Propositions 18(ii) and 19(i) we obtain

$$\mathbb{E}[\overline{\mathcal{U}}_\sigma(f * p)] = \frac{1}{\overline{h}_1} \qquad \text{and} \qquad \mathbb{E}[\overline{\mathcal{V}}_\sigma(f)] = \overline{h}_1^{-1}f(z_0) + O(h_1^{-1}h_2^s).$$

Therefore, we obtain the following modification of (35):

$$\mathbb{E}\Big[\overline{f}^{(2)}_{\overline{\sigma}}(z_0)\Big] = \overline{h}_1\overline{\mathcal{V}}_\sigma(f) + \overline{\sigma}n\lambda h_1\,\overline{\mathcal{U}}_\sigma(f*p) - \overline{\sigma}\mathbb{E}[|\mathcal{X}|] = f(z_0) + O(h_2^s).$$

We conclude

$$\sup_{\overline{\sigma}:|\overline{\sigma}-\sigma|\leqslant\varepsilon_n\sigma} \overline{h}_2^{-s}\Big|\big(f(z_0) - \mathbb{E}[\overline{f}^{(2)}_{\overline{\sigma}}(z_0)]\big)\Big| \lesssim 1.$$

Step 3: Stochastic error term. First note that due to the bias correction we have the additional stochastic error term:

$$\sup_{\overline{\sigma}:|\overline{\sigma}-\sigma|\leqslant\varepsilon_n\sigma} \sqrt{\min(n\overline{h}_2, \sigma^{-1}\overline{h}_2, (n\sigma^2)^{-1})}\Big|\overline{\sigma}|\mathcal{X}| - \overline{\sigma}n\lambda\Big| \leqslant \frac{1}{\sigma n^{1/2}} 2\sigma\big||\mathcal{X}| - n\lambda\big| = O_{\mathbb{P}}(1),$$

where we used $|\mathcal{X}| \sim \mathrm{Poiss}(n\lambda)$. To bound the stochastic error due to the terms involving $\overline{\psi}_1$ and $\overline{\psi}_2$, we use again the Kolmogorov-Chentsov criterion for the process

$$V_t := \varpi_t\Big(\overline{f}^{(1)}_{\sigma_t}(z_0) - \mathbb{E}[\overline{f}^{(1)}_{\sigma_t}(z_0)] + \sigma_t(|\mathcal{X}| - n\lambda)\Big)$$

with

$$\varpi_t := \sqrt{\min(nh_t, \sigma^{-1}h_t, (n\sigma^2)^{-1})},$$
$$\sigma_t := \sigma(1 - \varepsilon_n + 2\varepsilon_n t), \qquad h_t := (n \wedge \sigma_t^{-1})^{-1/(2s+1)}, \qquad t \in [0,1].$$

We decompose

$$\begin{aligned} V_t - V_s =& \frac{1}{\lambda\mu\sqrt{n}} \sum_{i,j} \big(\Delta^{(1)}_{s,t}(X_i, Y_j) - \mathbb{E}[\Delta^{(1)}_{s,t}(X_i, Y_j)]\big) \\ &+ \frac{1}{\lambda\mu\sqrt{n}} \sum_{i,j} \big(\Delta^{(2)}_{s,t}(X_i, Y_j) - \mathbb{E}[\Delta^{(2)}_{s,t}(X_i, Y_j)]\big) \end{aligned}$$

with

$$\Delta^{(1)}_{s,t}(x,y) = \Big(\underbrace{2K'\big(2(\sigma_t z_0 - y)\big) - 2K'\big(2(\sigma_s z_0 - y)\big)}_{=:\Delta^{(1,1)}_{s,t}(y)}\Big)\underbrace{\frac{\varpi_t}{\sqrt{n}h_t} K\Big(\frac{z_0}{h_t} - \frac{y-x}{\sigma_t h_t}\Big)}_{=:\Delta^{(1,2)}_t((y-x)/\sigma)},$$

$$\Delta^{(2)}_{s,t}(x,y) = \underbrace{2K'\big(2(\sigma_s z_0 - y)\big)}_{=:\Delta^{(2,1)}_s(y)}\Big(\underbrace{\frac{\varpi_t}{\sqrt{n}h_t} K\Big(\frac{z_0}{h_t} - \frac{y-x}{\sigma_t h_t}\Big) - \frac{\varpi_s}{\sqrt{n}h_s} K\Big(\frac{z_0}{h_s} - \frac{y-x}{\sigma_s h_s}\Big)}_{=:\Delta^{(2,2)}_{s,t}((y-x)/\sigma)}\Big).$$

With these definitions, the bound (48) remains valid up to a factor n^2 (coming from the missing factor $\frac{1}{n}$ in $V_t - V_s$) and we obtain

$$\begin{aligned} \mathbb{E}[(V_t - V_s)^2] &\lesssim \varpi_t\Big(\sigma + n\sigma^2 + \frac{\sigma}{h_t} + \frac{1}{nh_t}\Big)\|\Delta^{(1,1)}_{s,t}\|^2_{L^2} \\ &\quad + (\sigma n + \sigma^2 n^2)\|\Delta^{(2,2)}_{s,t}\|^2_{L^1} + (\sigma n + 1)\|\Delta^{(2,2)}_{s,t}\|^2_{L^2} \\ &\lesssim \|\Delta^{(1,1)}_{s,t}\|^2_{L^2} + (\sigma n + (\sigma n)^2)\|\Delta^{(2,2)}_{s,t}\|^2_{L^1} + (\sigma n + 1)\|\Delta^{(2,2)}_{s,t}\|^2_{L^2}. \end{aligned}$$

Next, we have

$$\begin{aligned}\|\Delta_{s,t}^{(1,1)}\|_{L^2}^2 &= 4\int \Big(K'(2(\sigma_t z_0 - y)) - 2K'(2(\sigma_s z_0 - y))\Big)^2 \mathrm{d}y \\ &= 8(\sigma_t-\sigma_s)^2 z_0 \int K''(2(\xi-y))^2 \mathrm{d}y \lesssim \sigma(t-s)^2 \lesssim (t-s)^2.\end{aligned}$$

Moreover, the term $\|\Delta_{s,t}^{(2,2)}\|_{L^1}$ is bounded above by

$$\begin{aligned}&\Big|\frac{\varpi_t}{\sqrt{n}h_t} - \frac{\varpi_s}{\sqrt{n}h_s}\Big| \int \Big|K\Big(\frac{z_0}{h_t} - \frac{\sigma z}{\sigma_t h_t}\Big)\Big| \mathrm{d}z \\ &+ \frac{\varpi_s}{\sqrt{n}h_s}\int \Big|K\Big(\frac{z_0}{h_t} - \frac{\sigma z}{\sigma_t h_t}\Big) - K\Big(\frac{z_0}{h_s} - \frac{\sigma z}{\sigma_t h_s}\Big)\Big| \mathrm{d}z \\ &+ \frac{\varpi_s}{\sqrt{n}h_s}\int \Big|K\Big(\frac{z_0}{h_s} - \frac{\sigma z}{\sigma_t h_s}\Big) - K\Big(\frac{z_0}{h_s} - \frac{\sigma z}{\sigma_s h_s}\Big)\Big| \mathrm{d}z \\ &\leqslant \Big|\frac{\varpi_t}{\sqrt{n}h_t} - \frac{\varpi_s}{\sqrt{n}h_s}\Big| h_t + \frac{\varpi_s}{\sqrt{n}h_s}\sup_{r\in[s,t]}\frac{\sigma_t^2 h_r^2}{\sigma^2}\Big|\frac{\sigma}{\sigma_t h_t} - \frac{\sigma}{\sigma_t h_s}\Big| + \frac{\varpi_s}{\sqrt{n}h_s}\sup_{r\in[s,t]}\frac{\sigma_r^2 h_s}{\sigma^2}\Big|\frac{\sigma}{\sigma_t h_s} - \frac{\sigma}{\sigma_s h_s}\Big| \\ &\lesssim \Big|\frac{\varpi_t}{\sqrt{n}h_t} - \frac{\varpi_s}{\sqrt{n}h_s}\Big| h_t + \frac{\varpi_s}{\sqrt{n}h_s}(h_s|t-s| + \frac{|\sigma_t-\sigma_s|}{\sigma}).\end{aligned}$$

Since h_s and h_s are of the same order in terms of n and σ both minima in the above difference are obtained at the same argument. Separate upper bounds in all three cases yield

$$\begin{aligned}(\sigma n + \sigma^2 n^2)\|\Delta_{s,t}^{(2,2)}\|_{L^1}^2 &\lesssim \sigma n h_t^2 (h_s^{-1/2} - h_t^{-1/2})^2 \mathbb{1}_{\sigma < n^{-(2s+1)/(2s+2)}} + h_t^2 (h_t^{-1} - h_s^{-1})^2 \\ &\quad + \Big(\frac{\sigma n}{h_s}\mathbb{1}_{\sigma < n^{-(2s+1)/(2s+2)}} + \frac{1}{h_s^2}\mathbb{1}_{\sigma \geqslant n^{-(2s+1)/(2s+2)}}\Big)(h_s^2 + \varepsilon_n^2)|t-s|^2 \\ &\lesssim |t-s|^2,\end{aligned}$$

noting that $n\sigma h_t \leqslant 1$ as well as $\frac{\sigma n \varepsilon_n^2}{h_s} < n^{(2-2s)/2s+2} \leqslant 1$ for $\sigma < n^{-(2s+1)/(2s+2)}$ and $s \geqslant 1$ while $\frac{\varepsilon_n}{h_s} = \kappa_n \min\Big(n^{1/2}\sigma^{\frac{2s}{2s+1}}, n^{-1/2}\sigma^{-\frac{2s+3}{4s+2}}\Big) \lesssim \kappa_n n^{(3-2s)/(12s+6)} \lesssim 1$ for $\sigma \geqslant n^{-(2s+1)/(2s+2)}$ and $s > 3/2$. Similarly, $\|\Delta_{s,t}^{(2,2)}\|_{L^2}^2$ is less than

$$\begin{aligned}&\Big|\frac{\varpi_t}{\sqrt{n}h_t} - \frac{\varpi_s}{\sqrt{n}h_s}\Big|^2 \int \Big|K\Big(\frac{z_0}{h_t} - \frac{\sigma z}{\sigma_t h_t}\Big)\Big|^2 \mathrm{d}z \\ &\leq \frac{\varpi_s}{n h_s^2}\int \Big|K\Big(\frac{z_0}{h_t} - \frac{\sigma z}{\sigma_t h_t}\Big) - K\Big(\frac{z_0}{h_s} - \frac{\sigma z}{\sigma_t h_s}\Big)\Big|^2 \mathrm{d}z + \frac{\varpi_s}{n h_s^2}\int \Big|K\Big(\frac{z_0}{h_s} - \frac{\sigma z}{\sigma_t h_s}\Big) - K\Big(\frac{z_0}{h_s} - \frac{\sigma z}{\sigma_s h_s}\Big)\Big|^2 \mathrm{d}z \\ &\leqslant \Big|\frac{\varpi_t}{\sqrt{n}h_t} - \frac{\varpi_s}{\sqrt{n}h_s}\Big|^2 h_t + \frac{\varpi_s^2}{n h_s^2}\sup_{r\in[s,t]}\frac{\sigma_t^3 h_r^3}{\sigma^3}\Big|\frac{\sigma}{\sigma_t h_t} - \frac{\sigma}{\sigma_t h_s}\Big|^2 + \frac{\varpi_s}{\sqrt{n}h_s}\sup_{r\in[s,t]}\frac{\sigma_r^3 h_s}{\sigma^3}\Big|\frac{\sigma}{\sigma_t h_s} - \frac{\sigma}{\sigma_s h_s}\Big|^2 \\ &\lesssim \Big|\frac{\varpi_t}{\sqrt{n}h_t} - \frac{\varpi_s}{\sqrt{n}h_s}\Big|^2 h_t + \frac{\varpi_s}{n h_s^2}\Big(h_s|t-s|^2 + \frac{|\sigma_t-\sigma_s|^2}{\sigma^2 h_s}\Big)\end{aligned}$$

and thus we conclude

$$\begin{aligned}(\sigma n + 1)\|\Delta_{s,t}^{(2,2)}\|_{L^2}^2 &\lesssim h_t(h_s^{-1/2} - h_t^{-1/2})^2 + \frac{h_t}{\sigma n}(h_t^{-1} - h_s^{-1})^2 \mathbb{1}_{\sigma \geqslant n^{-(2s+1)/(2s+2)}} \\ &\quad + (\sigma n + 1)\frac{\min(n h_s, \sigma^{-1}h_s, (n\sigma^2)^{-1})}{n h_s^2}(h_s + \frac{\varepsilon_n^2}{h_s})|t-s|^2 \lesssim |t-s|^2\end{aligned}$$

by distinguishing the three different cases where the minima can be attained. In particular, we have for the last term:

$$(\sigma n+1)\frac{\min(nh_s,\frac{h_s}{\sigma},\frac{1}{n\sigma^2})}{nh_s^3}\varepsilon_n^2 \lesssim \begin{cases} \frac{\kappa_n^2}{n^2h_s^2}=\kappa_n^2 n^{-4s/(2s+1)}, & \sigma\leqslant n^{-1},\\ \frac{\kappa_n^2\sigma^2 n}{h_s^2}\leqslant \kappa_n^2 n^{-(s-1)/(s+1)}, & n^{-1}<\sigma<n^{-\frac{2s+1}{2s+2}},\\ \frac{\kappa_n^2\sigma}{h_s^3}=\kappa_n^2\sigma^{(2s-2)/(2s+1)}, & \text{otherwise,}\end{cases}$$

which is uniformly bounded if $s>1$. □

APPENDIX A: REMAINING PROOFS

A.1. Proof of the covariance structure of (M,N).

Proof of Lemma 16. (i) to (iii): For $A_1,A_2\subseteq[-1,1]$ and $B_1,B_2\subseteq\mathbb{R}$ with $A_1\cap A_2=\emptyset$ and $B_1\cap B_2=\emptyset$ we write in view of Proposition 1:

$$\begin{aligned}\Psi(\eta_1,\eta_2,\xi_1,\xi_2) :=& \mathbb{E}\Big[e^{\eta_1 M(A_1)+\eta_2 M(A_2)+\xi_1 N(B_1)+\xi_2 N(B_2)}\Big]\\ =&\exp\Big(n\lambda(e^{\eta_1}-1)|A_1|+n\lambda(e^{\eta_2}-1)|A_2|+n\lambda\int_0^1(e^{\psi_1(\xi_1,x)+\psi_2(\xi_2,x)}-1)\mathrm{d}x\\ &+n\lambda\sum_{i=1}^2(e^{\eta_i}-1)\int_{A_i}(e^{\psi_1(\xi_1,x)+\psi_2(\xi_2,x)}-1)\mathrm{d}x\Big).\end{aligned}$$

where

$$\psi_j(\xi,x):=\mu(e^{\xi}-1)\int_{B_j}f_\sigma(y-x)\mathrm{d}y.$$

We moreover abbreviate

$$h(x):=(e^{\psi_1(\xi_1,x)+\psi_2(\xi_2,x)}-1),\qquad h_j'(x):=\partial_{\xi_j}h(x)=e^{\psi_1(\xi_1,x)+\psi_2(\xi_2,x)}\partial_{\xi_j}\psi_j(\xi_j,x).$$
$$h''(x):=\partial_{\xi_1}\partial_{\xi_2}h(x)=e^{\psi_1(\xi_1,x)+\psi_2(\xi_2,x)}\partial_{\xi_1}\psi_1(\xi_1,x)\partial_{\xi_2}\psi_2(\xi_2,x).$$

Then the first order partial derivatives are given by:

$$\partial_{\eta_1}\Psi(\eta_1,\eta_2,\xi_1,\xi_2)=\Psi(\eta_1,\eta_2,\xi_1,\xi_2)\,n\lambda e^{\eta_1}\Big(|A_1|+\int_{A_1}h(x)\mathrm{d}x\Big),$$
$$\partial_{\xi_1}\Psi(\eta_1,\eta_2,\xi_1,\xi_2)=\Psi(\eta_1,\eta_2,\xi_1,\xi_2)\,n\lambda\Big(\int_0^1 h_1'(x)\mathrm{d}x+\sum_{i=1}^2(e^{\eta_i}-1)\int_{A_i}h_1'(x)\mathrm{d}x\Big).$$

We moreover need the second order derivatives

$$\partial_{\eta_1}\partial_{\eta_2}\Psi(\eta_1,\eta_2,\xi_1,\xi_2)=\Psi(\eta_1,\eta_2,\xi_1,\xi_2)n^2\lambda^2e^{\eta_1+\eta_2}\Big(|A_1|+\int_{A_1}h(x)\mathrm{d}x\Big)\Big(|A_2|+\int_{A_2}h(x)\mathrm{d}x\Big),$$
$$\begin{aligned}\partial_{\xi_1}\partial_{\xi_2}\Psi(\eta_1,\eta_2,\xi_1,\xi_2)=&\,\Psi(\eta_1,\eta_2,\xi_1,\xi_2)n^2\lambda^2\Big(\int_0^1h_1'(x)\mathrm{d}x+\sum_{i=1}^2(e^{\eta_i}-1)\int_{A_i}h_1'(x)\mathrm{d}x\Big)\\ &\times\Big(\int_0^1h_2'(x)\mathrm{d}x+\sum_{i=1}^2(e^{\eta_i}-1)\int_{A_i}h_2'(x)\mathrm{d}x\Big)\\ &+\Psi(\eta_1,\eta_2,\xi_1,\xi_2)n\lambda\Big(\int_0^1h''(x)\mathrm{d}x+\sum_{i=1}^2(e^{\eta_i}-1)\int_{A_i}h''(x)\mathrm{d}x\Big).\end{aligned}\tag{50}$$

Therefore,

$$\begin{aligned}\frac{\partial_{\eta_1}\partial_{\eta_2}\partial_{\xi_1}\Psi(\eta_1,\eta_2,\xi_1,\xi_2)}{\Psi(\eta_1,\eta_2,\xi_1,\xi_2)} =& n^3\lambda^3 e^{\eta_1+\eta_2}\Big(|A_1|+\int_{A_1}h(x)\mathrm{d}x\Big)\Big(|A_2|+\int_{A_2}h(x)\mathrm{d}x\Big)\\ &\times\Big(\int_0^1 h_1'(x)\mathrm{d}x+\sum_{i=1}^{2}(e^{\eta_i}-1)\int_{A_i}h_1'(x)\mathrm{d}x\Big)\\ &+n^2\lambda^2 e^{\eta_1+\eta_2}\Big(\int_{A_1}h_1'(x)\mathrm{d}x\Big)\Big(|A_2|+\int_{A_2}h(x)\mathrm{d}x\Big)\\ &+n^2\lambda^2 e^{\eta_1+\eta_2}\Big(|A_1|+\int_{A_1}h(x)\mathrm{d}x\Big)\Big(\int_{A_2}h_1'(x)\mathrm{d}x\Big) \qquad (51)\end{aligned}$$

and

$$\begin{aligned}&\frac{\partial_{\eta_1}\partial_{\xi_1}\partial_{\xi_2}\Psi(\eta_1,\eta_2,\xi_1,\xi_2)}{\Psi(\eta_1,\eta_2,\xi_1,\xi_2)}\\ &=n^3\lambda^3 e^{\eta_1}\Big(|A_1|+\int_{A_1}h(x)\mathrm{d}x\Big)\Big(\int_0^1 h_1'(x)\mathrm{d}x+\sum_{i=1}^{2}(e^{\eta_i}-1)\int_{A_i}h_1'(x)\mathrm{d}x\Big)\\ &\quad\times\Big(\int_0^1 h_2'(x)\mathrm{d}x+\sum_{i=1}^{2}(e^{\eta_i}-1)\int_{A_i}h_2'(x)\mathrm{d}x\Big)\\ &\quad+n^2\lambda^2 e^{\eta_1}\int_{A_1}h_1'(x)\mathrm{d}x\Big(\int_0^1 h_2'(x)\mathrm{d}x+\sum_{i=1}^{2}(e^{\eta_i}-1)\int_{A_i}h_2'(x)\mathrm{d}x\Big)\\ &\quad+n^2\lambda^2 e^{\eta_1}\Big(\int_0^1 h_1'(x)\mathrm{d}x+\sum_{i=1}^{2}(e^{\eta_i}-1)\int_{A_i}h_1'(x)\mathrm{d}x\Big)\int_{A_1}h_2'(x)\mathrm{d}x\\ &\quad+n^2\lambda^2 e^{\eta_1}\Big(|A_1|+\int_{A_1}h(x)\mathrm{d}x\Big)\Big(\int_0^1 h''(x)\mathrm{d}x+\sum_{i=1}^{2}(e^{\eta_i}-1)\int_{A_i}h''(x)\mathrm{d}x\Big)\\ &\quad+n\lambda e^{\eta_1}\int_{A_1}h''(x)\mathrm{d}x. \qquad (52)\end{aligned}$$

Evaluating (50), (51) and (52) at $\eta_1=\eta_2=\xi_1=\xi_2=0$ yields (i), (ii) and (iii), respectively.

(*iv*) It remains to calculate $\partial_{\eta_1}\partial_{\eta_2}\partial_{\xi_1}\partial_{\xi_2}\Psi(\eta_1,\eta_2,\xi_1,\xi_2)$ which can be deduced straightfroward from the previous formulas:

$$\begin{aligned}&\frac{\partial_{\eta_1}\partial_{\eta_2}\partial_{\xi_1}\partial_{\xi_2}\Psi(\eta_1,\eta_2,\xi_1,\xi_2)}{\Psi(\eta_1,\eta_2,\xi_1,\xi_2)}\\ &=n^4\lambda^4 e^{\eta_1+\eta_2}\Big(|A_1|+\int_{A_1}h(x)\mathrm{d}x\Big)\Big(|A_2|+\int_{A_2}h(x)\mathrm{d}x\Big)\\ &\quad\times\Big(\int_0^1 h_1'(x)\mathrm{d}x+\sum_{i=1}^{2}(e^{\eta_i}-1)\int_{A_i}h_1'(x)\mathrm{d}x\Big)\Big(\int_0^1 h_2'(x)\mathrm{d}x+\sum_{i=1}^{2}(e^{\eta_i}-1)\int_{A_i}h_2'(x)\mathrm{d}x\Big)\\ &\quad+n^3\lambda^3 e^{\eta_1+\eta_2}\Big(|A_1|+\int_{A_1}h(x)\mathrm{d}x\Big)\int_{A_2}h_1'(x)\mathrm{d}x\Big(\int_0^1 h_2'(x)\mathrm{d}x+\sum_{i=1}^{2}(e^{\eta_i}-1)\int_{A_i}h_2'(x)\mathrm{d}x\Big)\\ &\quad+n^3\lambda^3 e^{\eta_1+\eta_2}\Big(|A_1|+\int_{A_1}h(x)\mathrm{d}x\Big)\Big(\int_0^1 h_1'(x)\mathrm{d}x+\sum_{i=1}^{2}(e^{\eta_i}-1)\int_{A_i}h_1'(x)\mathrm{d}x\Big)\int_{A_2}h_2'(x)\mathrm{d}x\end{aligned}$$

$$
\begin{aligned}
&+ n^3\lambda^3 e^{\eta_1+\eta_2}\Big(|A_2| + \int_{A_2} h(x)\mathrm{d}x\Big)\int_{A_1} h_1'(x)\mathrm{d}x\Big(\int_0^1 h_2'(x)\mathrm{d}x + \sum_{i=1}^2 (e^{\eta_i}-1)\int_{A_i} h_2'(x)\mathrm{d}x\Big)\\
&+ n^2\lambda^2 e^{\eta_1+\eta_2}\int_{A_1} h_1'(x)\mathrm{d}x\int_{A_2} h_2'(x)\mathrm{d}x\\
&+ n^3\lambda^3 e^{\eta_1+\eta_2}\Big(|A_2| + \int_{A_2} h(x)\mathrm{d}x\Big)\Big(\int_0^1 h_1'(x)\mathrm{d}x + \sum_{i=1}^2 (e^{\eta_i}-1)\int_{A_i} h_1'(x)\mathrm{d}x\Big)\int_{A_1} h_2'(x)\mathrm{d}x\\
&+ n^2\lambda^2 e^{\eta_1+\eta_2}\int_{A_2} h_1'(x)\mathrm{d}x\int_{A_1} h_2'(x)\mathrm{d}x\\
&+ n^3\lambda^3 e^{\eta_1+\eta_2}\Big(|A_1| + \int_{A_1} h(x)\mathrm{d}x\Big)\Big(|A_2| + \int_{A_2} h(x)\mathrm{d}x\Big)\\
&\quad\times\Big(\int_0^1 h''(x)\mathrm{d}x + \sum_{i=1}^2 (e^{\eta_i}-1)\int_{A_i} h''(x)\mathrm{d}x\Big)\\
&+ n^2\lambda^2 e^{\eta_1+\eta_2}\Big(\Big(|A_1| + \int_{A_1} h(x)\mathrm{d}x\Big)\int_{A_2} h''(x)\mathrm{d}x + \Big(|A_2| + \int_{A_2} h(x)\mathrm{d}x\Big)\int_{A_1} h''(x)\mathrm{d}x\Big).
\end{aligned}
$$

Evaluating this partial derivative at 0 yields

$$
\begin{aligned}
&\mathbb{E}[M(A_1)M(A_2)N(B_1)N(B_2)]\\
&= n^4\lambda^4\mu^2|A_1||A_2|Q_\sigma([0,1],B_1)Q_\sigma([0,1],B_2)\\
&\quad + n^3\lambda^3\mu^2\big(|A_1|Q_\sigma(A_2,B_1)Q_\sigma([0,1],B_2) + |A_1|Q_\sigma([0,1],B_1)Q_\sigma(A_2,B_1)\big)\\
&\quad + n^3\lambda^3\mu^2\big(|A_2|Q_\sigma(A_1,B_1)Q_\sigma([0,1],B_2) + |A_2|Q_\sigma([0,1],B_1)Q_\sigma(A_1,B_2)\big)\\
&\quad + n^3\lambda^3\mu^2|A_1||A_2|Q_\sigma^2([0,1])\\
&\quad + n^2\lambda^2\mu^2\big(Q_\sigma(A_1,B_1)Q_\sigma(A_2,B_2) + Q_\sigma(A_2,B_1)Q_\sigma(A_1,B_2)\big)\\
&\quad + n^2\lambda^2\mu^2\big(|A_1|Q_\sigma^2(A_2) + |A_2|Q_\sigma^2(A_1)\big).
\end{aligned}
$$

Combining this formula with Corollary 2 yields the assertion. □

REFERENCES

Arnaud, J.-F., Viard, F., Delescluse, M., and Cuguen, J. (2003). Evidence for gene flow via seed dispersal from crop to wild relatives in Beta vulgaris (Chenopodiaceae): consequences for the release of genetically modified crop species with weedy lineages. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 270(1524):1565–1571.

Baccelli, F., Kauffmann, B., and Veitch, D. (2009). Inverse problems in queueing theory and Internet probing. *Queueing Systems*, 63(1):59–107.

Belomestny, D. and Goldenshluger, A. (2021). Density deconvolution under general assumptions on the distribution of measurement errors. *The Annals of Statistics*, 49(2):615–649.

Berenfeld, C. and Hoffmann, M. (2021). Density estimation on an unknown submanifold. *Electronic Journal of Statistics*, 15(1):2179–2223.

Bingham, N. H. and Dunham, B. (1997). Estimating diffusion coefficients from count data: Einstein-smoluchowski theory revisited. *Annals of the Institute of Statistical Mathematics*, 49(4):667–679.

Blanghaps, N., Nov, Y., and Weiss, G. (2013). Sojourn time estimation in an M/G/∞ queue with partial information. *Journal of Applied Probability*, 50(4):1044–1056.

Bonnet, A., Lacour, C., Picard, F., and Rivoirard, V. (2022). Uniform deconvolution for poisson point processes. *Journal of Machine Learning Research*, 23(194):1–36.

Brown, M. (1970). An M/G/∞ estimation problem. *The Annals of Mathematical Statistics*, pages 651–654.

Carstensen, L., Sandelin, A., Winther, O., and Hansen, N. R. (2010). Multivariate Hawkes process models of the occurrence of regulatory elements. *BMC bioinformatics*, 11(1):1–19.

Chorowski, J. (2018). Nonparametric volatility estimation in scalar diffusions: Optimality across observation frequencies. *Bernoulli*, 24(4A):2934–2990.

Comte, F. and Lacour, C. (2011). Data-driven density estimation in the presence of additive noise with unknown distribution. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, 73(4):601–627.

Dattner, I., Reiß, M., and Trabs, M. (2016). Adaptive quantile estimation in deconvolution with unknown error distribution. *Bernoulli*, 22(1):143–192.

Delaigle, A. and Meister, A. (2011). Nonparametric function estimation under Fourier-oscillating noise. *Statist. Sinica*, 21(3):1065–1092.

Duval, C. and Hoffmann, M. (2011). Statistical inference across time scales. *Electronic Journal of Statistics*, 5:2004–2030.

Giné, E. and Nickl, R. (2016). *Mathematical foundations of infinite-dimensional statistical models*. Cambridge Series in Statistical and Probabilistic Mathematics, [40]. Cambridge University Press, New York.

Goldenshluger, A. (2016). Nonparametric estimation of service time distribution in the M/G/∞ queue. *Advances in Applied Probability*, 48:1117–1138.

Goldenshluger, A. (2018). The $M/G/\infty$ estimation problem revisited. *Bernoulli*, 24:2531–2568.

Goldenshluger, A. and Jacobovic, R. (2021). Smoluchowski processes and nonparametric estimation of functionals of particle displacement distributions from count data. *arXiv preprint arXiv:2108.06954*.

Goldenshluger, A. and Koops, D. T. (2019). Nonparametric estimation of service time characteristics in infinite-server queues with nonstationary Poisson input. *Stochastic Systems*, 9(3):183–207.

Groeneboom, P. and Jongbloed, G. (2003). Density estimation in the uniform deconvolution model. *Statistica Neerlandica*, 57(1):136–157.

Gusto, G. and Schbath, S. (2005). FADO: a statistical method to detect favored or avoided distances between occurrences of motifs using the Hawkes models. *Statistical Applications in Genetics and Molecular Biology*, 4(1).

Hall, P. and Park, J. (2004). Nonparametric inference about service time distribution from indirect measurements. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(4):861–875.

Heuertz, M., Vekemans, X., Hausman, J.-F., Palada, M., and Hardy, O. J. (2003). Estimating seed vs. pollen dispersal from spatial genetic structure in the common ash. *Molecular Ecology*, 12(9):2483–2495.

Hunt, X. J., Reynaud-Bouret, P., Rivoirard, V., Sansonnet, L., and Willett, R. (2018). A data-dependent weighted LASSO under Poisson noise. *IEEE Transactions on Information Theory*, 65(3):1589–1613.

Isagi, Y., Kanazashi, T., Suzuki, W., Tanaka, H., and Abe, T. (2000). Microsatellite analysis of the regeneration process of Magnolia obovata Thunb. *Heredity*, 84(2):143–151.

Lavorel, S., Gardner, R. H., and O'Neill, R. V. (1995). Dispersal of annual plants in hierarchically structured landscapes. *Landscape Ecology*, 10(5):277–289.

Le Cam, L. (2012). *Asymptotic methods in statistical decision theory*. Springer Science & Business Media.

Meister, A. (2007). Deconvolution from Fourier-oscillating error densities under decay and smoothness restrictions. *Inverse Problems*, 24(1):015003.

Neumann, M. H. (2007). Deconvolution from panel data with unknown error distribution. *Journal of Multivariate Analysis*, 98(10):1955–1968.

Nickl, R., Reiß, M., Söhl, J., and Trabs, M. (2016). High-frequency Donsker theorems for Lévy measures. *Probability Theory and Related Fields*, 164(1-2):61–108.

Randol, B. (1969). On the asymptotic behavior of the Fourier transform of the indicator function of a convex set. *Transactions of the American Mathematical Society*, 139:279–285.

Tsybakov, A. B. (2009). *Introduction to nonparametric estimation*. Springer Series in Statistics. Springer, New York. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats.

van der Vaart, A. (2002). The statistical work of Lucien Le Cam. *The Annals of Statistics*, 30(3):631–682.

Marc Hoffmann,
Université Paris-Dauphine PSL,
Place du Maréchal De Lattre de Tassigny,
75016 Paris, France,
E-mail: hoffmann@ceremade.dauphine.fr

Mathias Trabs,
Karlsruhe Institute of Technology,
Institut für Stochastik
Englerstr. 2,
76131 Karlsruhe, Germany
E-mail: trabs@kit.edu