Some asymptotics for local least-squares regression with regularization

Jürgen Franke and Joseph Tadjuidje Kamgaing
Department of Mathematics, University of Kaiserslautern

Stefan Didas and Joachim Weickert
Department of Mathematics, University of Saarbrücken

Abstract: We derive some asymptotics for a new approach to curve estimation proposed by Mrázek et al. [3] which combines localization and regularization. This methodology has been considered as the basis of a unified framework covering various different smoothing methods in the analogous two-dimensional problem of image denoising. As a first step for understanding this approach theoretically, we restrict our discussion here to the least-squares distance where we have explicit formulas for the function estimates and where we can derive a rather complete asymptotic theory from known results for the Priestley-Chao curve estimate. In this paper, we consider only the case where the bias dominates the mean-square error. Other situations are dealt with in subsequent papers.

This paper is a corrected and extended version of a previous preliminary report.

1 Introduction

We consider data generated from the nonparametric regression model

$$f_j = \mu(x_j) + \varepsilon_j, \; j = 1, \ldots, N, \tag{1}$$

where $\varepsilon_1, \ldots, \varepsilon_N$ are independent and identically distributed with mean 0 and finite variance σ^2 and $x_j = \frac{j}{N}$, $j = 1, \ldots, N$, form an equidistant grid on the unit interval $[0, 1]$. Mrázek et al. [3] have described a general approach for image denoising which, by combining localization and regularization, includes most of the known image denoising algorithms like local smoothing or nonlinear diffusion filtering. Estimating the function $\mu(x)$ at the grid points $x_1, \ldots, x_N$ from the data $f_1, \ldots, f_N$ generated by (1) is the one-dimensional analogon of the image denoising problem. Mrázek et al. [3] propose to solve this problem by minimizing a target function like

$$Q(\mathbf{u}) = \sum_{i,j=1}^{N} \Psi_D(|u_i - f_j|^2) w_D(|x_i - x_j|^2) + \frac{\lambda}{2} \sum_{i,j=1}^{N} \Psi_S(|u_i - u_j|^2) w_S(|x_i - x_j|^2)$$

w.r.t. $\mathbf{u} = (u_1, \ldots, u_N)^T$.

Let $u(x)$ denote the function estimate of $\mu(x)$ which we first consider only at $x_1, \ldots, x_N$. Ψ_D, Ψ_S are penalizing functions measuring the fit of $u(x)$ to the observations $f_1, \ldots, f_N$ and the smoothness of $u(x)$ resp. The spatial weighting functions w_D, w_S guarantee a localization effect, and $\lambda \geq 0$ balances between smoothness and the quality of the fit. To get some first insight into the asymptotic properties of the resulting estimates $u_1, \ldots, u_N$ for $\mu(x_1), \ldots, \mu(x_N)$, we investigate the following special case:

$$\Psi_D(s^2) = s^2, \; \Psi_S(s^2) = s^2, w_D(x^2) = K_h(x) = \frac{1}{h} K(\frac{x}{h}), \; w_S(x^2) = L_g(x) = \frac{1}{g} L(\frac{x}{g})$$

where the kernels K and L are nonnegative, symmetric functions on $\mathbb{R}$ and the bandwidths $h, g > 0$ can be chosen to control the smoothness of the function estimate together with the balancing factor λ. Therefore, we consider minimizing

$$Q(\mathbf{u}) = \sum_{i,j=1}^{N} (u_i - f_j)^2 K_h(x_i - x_j) + \frac{\lambda}{2} \sum_{i,j=1}^{N} (u_i - u_j)^2 L_g(x_i - x_j). \tag{2}$$

We call the resulting estimate $\mathbf{u}$ a *regularized local least-squares estimate.*

For $\lambda = 0$, we immediately get from $\frac{\partial}{\partial u_i} Q(\mathbf{u}) = 0$, $i = 1, \ldots, N$, that $u_i = \tilde{\mu}(x_i, h)$, $i = 1, \ldots, N$, where $\tilde{\mu}(x, h)$ is the classical Nadaraya-Watson kernel regression estimate (compare, e.g., Härdle [2])

$$\begin{aligned} \tilde{\mu}(x, h) &= \frac{\hat{\mu}(x, h)}{\hat{p}_K(x, h)} \qquad \text{with} \qquad\qquad (3) \\ \hat{\mu}(x, h) &= \frac{1}{N} \sum_{j=1}^{N} f_j K_h(x - x_j), \quad \hat{p}_K(x, h) = \frac{1}{N} \sum_{j=1}^{N} K_h(x - x_j). \end{aligned}$$

As $\hat{p}_K(x, h)$ converges to 1 for the equidistant x_i, the Nadaraya-Watson estimate $\tilde{\mu}(x, h)$ and the Priestley-Chao estimate $\hat{\mu}(x, h)$ are asymptotically equivalent.

2 Asymptotic expansion

The main goal of this section is to show that the regularized local least-squares estimate, which we get by minimizing (2), is closely related asymptotically for $N \to \infty$ to the simple Priestley-Chao estimate (Priestley and Chao, [4]) with bandwidth h where, however, the regularization parameter λ provides an additional tool for fine tuning the properties of the estimate. As a first step, we show that the solution of (2) has an explicit representation in terms of the Priestley-Chao estimate. For convenience, we use the following notation for the values of this estimate at the grid points $x_i, i = 1, \ldots, N$:

$$\hat{\mu} = (\hat{\mu}_1, \ldots, \hat{\mu}_N)^T \quad \text{with } \hat{\mu}_i = \hat{\mu}(x_i, h),\ i = 1, \ldots, N.$$

Proposition 1 *Let $\hat{p}_L(x,g)$ be defined analogously to $\hat{p}_K(x,h)$ with L, g replacing K, h, and let $\hat{p}_\lambda(x,h,g) = \hat{p}_K(x,h) + \lambda \hat{p}_L(x,g)$. Let Λ denote the $N \times N$-matrix with entries $\Lambda_{i,j} = \frac{1}{N} L_g(x_i - x_j)$, and let $\hat{P}$ denote the $N \times N$-diagonal matrix with entries $\hat{P}_{ii} = \hat{p}_\lambda(x_i, h, g)$. Then, if $\hat{P} - \lambda\,\Lambda$ is invertible, the solution of (2) is given as*

$$\mathbf{u} = \{\hat{P} - \lambda\Lambda\}^{-1}\hat{\mu}. \tag{4}$$

Proof: Setting the partial derivatives of $Q(\mathbf{u})$ w.r.t. $u_k,\ k = 1, \ldots, N$, to 0, we get

$$0 = 2\sum_{j=1}^{N}(u_k - f_j)K_h(x_k - x_j) + 2\lambda \sum_{j=1}^{N}(u_k - u_j)L_g(x_k - x_j),\ k = 1, \ldots, N,$$

where we have used the symmetry of L_g. Therefore, we have for $k = 1, \ldots, N$,

$$u_k \left\{ \sum_{j=1}^{N} K_h(x_k - x_j) + \lambda \sum_{j=1}^{N} L_g(x_k - x_j) \right\}$$

$$= \lambda \sum_{j=1}^{N} u_j L_g(x_k - x_j) + \sum_{j=1}^{N} f_j K_h(x_k - x_j)$$

which, using the definition of $\hat{p}_K, \hat{p}_L$ and $\hat{\mu}$, implies the assertion. ∎

Now, we want to prove that the estimate $\mathbf{u}$, given by (4), is consistent in a certain sense if $N \to \infty$. First, we investigate the asymptotic behaviour of $\hat{p}_K(x,h)$. If the x_j would be i.i.d. random variables, then $\hat{p}_K$ would be the well-known Rosenblatt-Parzen estimate of their common probability density (compare, e.g., Silverman, [6]). In our case, $x_1, \ldots, x_N$ are equidistant and behave similar to uniform random variables, i.e. in particular $\hat{p}_K(x,h) \to 1$ under appropriate assumptions. We consider throughout the paper only kernels satisfying

(A1) a) K is a nonnegative, symmetric kernel function with compact support $[-1, +1]$.

b) $\int K(y)dy = 1$

c) K is Lipschitz continuous with Lipschitz constant C_K.

In the following, we use the abbreviations

$$V_K = \int z^2 K(z)dz, \quad Q_K = \int K^2(z)dz$$

for the second moment and the L^2-norm of a kernel K which both are finite under (A1) a), c). We could relax the assumptions of symmetry and compactness of the support of K, but we want to keep our arguments simple in this paper. Due to the same reason, we mainly neglect boundary effects, which could be dealt with as in section 4.4 of Härdle [2], by restricting our attention to $x \in [h, 1-h]$. Asymptotically, this will have no effect as we shall have $h \to 0$ for $N \to \infty$ anyhow.

Sometimes, we need to extend assumption (A1) with the following conditions

d) $K(\pm 1) = 0$.

e) K is decreasing in $[0, 1]$.

f) $K \in C^2(-1, +1)$ with bounded second derivative K''.

g) K'' is Lipschitz continuous, and $K'(\pm 1) = 0$.

In the following, we will frequently approximate a Riemann sum by the corresponding integral. For reference, we, therefore, state the following Lemma:

Lemma 1 *a) Let $g(y)$ be Lipschitz continuous on $[0, 1]$ with Lipschitz constant C. Then,*

$$\left| \int_0^1 g(y)dy - \frac{1}{N} \sum_{j=1}^N g(x_j) \right| \leq \frac{C}{N}.$$

b) Let $g \in C^2(a, b)$. Then, for $x'_j = a + j(b-a)/M, i = 0, \ldots, M$,

$$\left| \frac{1}{b-a} \int_a^b g(y)dy - \frac{1}{M} \sum_{j=1}^M g(x'_j) \right| \leq \frac{(b-a)^2}{12M^2} \sup_{a<z<b} |g''(z)| + \frac{1}{2M} |g(b) - g(a)|$$

Proof: a) By the mean-value theorem of integration there are $y_j \in [x_{j-1}, x_j],\ j = 1, \ldots, N$, where $x_0 = 0$, such that, using $x_j - x_{j-1} = \frac{1}{N}$ for all j,

$$\begin{aligned} \left| \int_0^1 g(y)dy - \frac{1}{N} \sum_{j=1}^N g(x_j) \right| &= \left| \sum_{j=1}^N \{ \int_{x_{j-1}}^{x_j} g(y)dy - \frac{1}{N} g(x_j) \} \right| \\ &= \left| \sum_{j=1}^N \{ g(y_j) - g(x_j) \} \frac{1}{N} \right| \\ &\leq \frac{1}{N} \sum_{j=1}^N C |y_j - x_j| \leq \frac{C}{N} \end{aligned}$$

b) The assertion is a version of the Euler-Maclaurin formula (compare, e.g., [7], ch. 3.2). Let $B_1(x) = x - \frac{1}{2}, B_2(x) = x^2 - x + \frac{1}{6}$ be the first two Bernoulli polynomials. We first consider the case $a = 0, b = 1$. Using integration by parts twice, we get

$$\begin{aligned} \int_0^1 g(y)dy &= B_1(y)g(y)\big|_0^1 - \frac{1}{2} B_2(y)g'(y)\big|_0^1 + \frac{1}{2} \int_0^1 B_2(y)g''(y)dy \\ &= \frac{1}{2}\{g(1) + g(0)\} - \frac{1}{12}\{g'(1) - g'(0)\} + \frac{1}{2} \int_0^1 B_2(y)g''(y)dy \\ &= \frac{1}{2}\{g(1) + g(0)\} + \frac{1}{2} \int_0^1 \{x^2 - x\} g''(y)dy \\ &= \frac{1}{2}\{g(1) + g(0)\} - \frac{1}{12} g''(\xi) \end{aligned}$$

for some $\xi \in (0,1)$, using $x^2 - x \le 0$ in $[0,1]$ and the mean-value theorem of integration. We now consider $a = 0, b = M$, and we apply this argument repeatedly for each subinterval of length 1 to get

$$\begin{aligned}\int_0^M g(y)dy &= \sum_{i=1}^M \int_{i-1}^i g(y)dy = \frac{1}{2}\sum_{i=1}^M \{g(i) + g(i-1)\} - \frac{1}{12}\sum_{i=1}^M g''(\xi_i)\\ &= \sum_{i=1}^M g(i) - \frac{1}{2}\{g(M) - g(0)\} - \frac{1}{12}\sum_{i=1}^M g''(\xi_i)\end{aligned}$$

for some $\xi_i \in (i-1, i), i = 1, \dots, M$. For general a, b, the assertion now follows from substituting $y \mapsto a + y(b-a)/M$. ∎

As an immediate consequence, we have

Corollary 1 *a) Assuming (A1), a)-c), for the kernel K, we have*

$$|1 - \hat{p}_K(x,h)| \le C_K \frac{1}{Nh^2} \quad \textit{for all } x \in [h, 1-h].$$

b) Assuming additionally (A1), d), f) for the kernel K, we have

$$|1 - \hat{p}_K(x,h)| = O\left(\frac{1}{N^2h^2}\right) \quad \textit{uniformly in } x \in [h, 1-h].$$

Proof: a) follows immediately from Lemma 1 a) with $g(y) = K_h(x-y)$ as the Lipschitz constant of K_h is C_K/h^2.

For b), we set $n_0 = \min\{k; x_k \ge x - h\}, n_M = \max\{k; x_k \le x + h\}$ such that $M = n_M - n_0 \le 2Nh$. We apply Lemma 1 b) to $a = n_0/N, b = n_M/N, b - a = M/N$ such that $x'_j = a + j(b-a)/M = a + j/N = x_{n_0+j}$ and get

$$\begin{aligned}&\left|\int_a^b K_h(x-y)dy - \frac{1}{M}\sum_{j=1}^M K_h(x - x'_j)\right|\\ \le\ &\frac{M}{12N^3}\sup_{a<z<b}|\frac{1}{h^3}K''(\frac{x-z}{h})| + \frac{1}{2N}|K_h(x-b) - K_h(x-a)|\end{aligned}$$

Using that the support of $K_h(x - \cdot)$ is $[x-h, x+h]$, that a, b differ from $x - h$ resp. $x + h$ by at most $1/N$ and using that K is Lipschitz continuous and $K_h(x - (x \pm h)) = 0$ we get

$$\left|\int_{-1}^{+1} K(z)dz - \frac{1}{N}\sum_{j=1}^N K_h(x - x_j)\right| \le \frac{M}{12N^3h^3}\sup_{|z|<1}|K''(z)| + O\big(\frac{1}{N^2h^2}\big).$$

The assertion follows from $M \le 2Nh$. ∎

As a next step, we investigate the asymptotic behaviour of $\hat{\mu}$ as an estimate of $(\mu(x_1), \dots, \mu(x_N))^T$. We assume

(A2) a) μ is twice continuously differentiable

b) $\mu''(x)$ is Hoelder continuous on $[0,1]$ with exponent β, i.e. for some $\beta > 0, H < \infty$ $|\mu''(x) - \mu''(y)| \leq H|x-y|^\beta$ for all $x, y \in [0,1]$

The following asymptotic expansions for bias and variance of the Priestley-Chao estimate is well known. We only use assumption (A2b) to get a slightly more precise assertion about the remainder of the bias which will turn out to be useful later on, and we add a more detailed statement about the asymptotic covariance of estimates at different locations.

Proposition 2 *Assuming (A1), a)-f), and (A2), we have for the Priestley-Chao estimate $\hat{\mu}(x,h)$, based on (1), for $N \to \infty,\ h \to 0$ such that $Nh^3 \to \infty$:*

i) bias $\hat{\mu}_i = \mathrm{E}\ \hat{\mu}_i - \mu(x_i) = \frac{h^2}{2}\mu''(x_i)V_K + O(h^{2+\beta})$ *uniformly in $x_i \in [h, 1-h]$, where the main remainder term is*

$$r(x,h) = \frac{h^2}{2}\int K(z)z^2\{\mu''(x-\vartheta hz) - \mu''(x)\}dz = O(h^{2+\beta})$$

uniformly in x.

ii) var $\hat{\mu}_i = \mathrm{E}(\hat{\mu}_i - \mathrm{E}\ \hat{\mu}_i)^2 = \frac{\sigma^2}{Nh}Q_K + O(\frac{1}{N^3h^3})$ *uniformly in $x_i \in [h, 1-h]$.*

iii) mse $\hat{\mu}_i = \mathrm{E}(\hat{\mu}_i - \mu(x_i))^2 = \frac{\sigma^2}{Nh}Q_K + \frac{h^4}{4}\{\mu''(x_i)\}^2V_K^2 + O(h^{4+\beta})$ *uniformly in $x_i \in [h, 1-h]$. In particular,*

$$\hat{\mu}_i - \mu(x_i) \to 0 \ \text{ in probability.}$$

iv) cov $(\hat{\mu}_i, \hat{\mu}_k) = 0$ *if $|x_i - x_k| > 2h$, and*
cov $(\hat{\mu}_i, \hat{\mu}_k) = \frac{\sigma^2}{Nh}\int K(z)K(z + \frac{x_i - x_k}{h})dz + O(\frac{1}{N^3h^3})$ *uniformly in $x_i, x_k \in [h, 1-h]$, else.*

If μ does not satisfy the smoothness condition (A2) everywhere, then Proposition 1 still holds in every subinterval $[a,b] \subset [0,1]$ where (A2) is satisfied, as is obvious from the proof. So, if μ jumps in x^*, but otherwise is smooth enough, the assertions of the Proposition hold uniformly in $x_i, x_k \in [h, x^* - h] \cup [x^* + h, 1-h]$.

Proof: a) We use the common decomposition of mean-squared error into variance and squared bias

$$\text{mse}\ \hat{\mu}_i = \mathrm{E}(\hat{\mu}(x_i,h) - \mu(x_i))^2 = \text{var}\ \hat{\mu}(x_i,h) + \{\text{bias}\ \hat{\mu}(x_i,h)\}^2.$$

We have uniformly in $x \in [h, 1-h]$

$$\begin{aligned}
\text{var}\ \hat{\mu}(x,h) &= \frac{1}{N^2}\sum_{j=1}^N K_h^2(x - x_j)\ \text{var}\ f_j \\
&= \frac{\sigma^2}{N}\int K_h^2(x-y)dy + O(\frac{1}{N^3h^3}) \\
&= \frac{\sigma^2}{Nh}\int K^2(z)dz + O(\frac{1}{N^3h^3})
\end{aligned}$$

using the same argument as in the proof of Corollary 1 b) with $K_h^2(x-y)$ instead of $K_h(x-y)$. Moreover, by the same argument, now for $K_h(x-y)\mu(y)$,

$$\begin{aligned}
\text{bias } \hat{\mu}(x,h) &= \mathrm{E}\ \hat{\mu}(x,h) - \mu(x) \\
&= \frac{1}{N}\sum_{j=1}^{N} K_h(x-x_j)\mu(x_j) - \mu(x) \\
&= \int K(z)\{\mu(x-hz) - \mu(x)\}dz + O(\frac{1}{N^2h^2}) \\
&= \int K(z)(-hz)dz\,\mu'(x) + \int K(z)\frac{h^2z^2}{2}\mu''(x-\theta hz)dz\ + O(\frac{1}{N^2h^2}) \\
&= \frac{h^2}{2}V_K\mu''(x) + \frac{h^2}{2}\int K(z)z^2\{\mu''(x-\theta hz) - \mu''(x)\}dz + O(\frac{1}{N^2h^2})
\end{aligned}$$

with $\theta = \theta(z) \in [0,1]$, again uniformly in $x \in [h, 1-h]$. We have used Assumption (A2) together with a Taylor expansion of μ and symmetry of K. By Hoelder continuity of μ, the middle term on the right-hand side is of order $O(h^{2+\beta})$, which is asymptotically larger than the last term on the right-hand side as $Nh^3 \to \infty$.

Combining the bias and variance expansion, we get

$$\text{mse } \hat{\mu}_i = \frac{\sigma^2}{Nh}Q_K + \frac{h^4}{4}(\mu''(x_i))^2V_K^2 + O(h^{4+\beta}),$$

as, again due to assuming $Nh^3 \to \infty$, the remainder in the variance expansion is of smaller order than $O(h^{4+\beta})$.

b) Analogously to a), we conclude using the independence of the f_j.

$$\begin{aligned}
\text{cov}(\hat{\mu}(x,h), (\hat{\mu}(x',h)) &= \frac{\sigma^2}{N^2}\sum_{j=1}^{N} K_h(x-x_j)K_h(x'-x_j) \\
&= \frac{\sigma^2}{Nh}\int K(z)K(z+\frac{x-x'}{h})dz + O(\frac{1}{N^3h^3}).
\end{aligned}$$

By compactness of the support of K, we have for $|x-x'| > 2h$ that $K_h(x-x_j)K_h(x'-x_j) = 0$ and, therefore, cov $(\hat{\mu}(x,h), \hat{\mu}(x',h)) = 0$. ∎

We need the following generalization of Corollary 1 a) which takes care of the boundary effects:

Corollary 2 *Assuming (A1), a)-e), we have*

i) $|1 - \hat{p}_K(x,h)| \leq C_K\frac{1}{Nh^2}$ *for* $h \leq x \leq 1-h$

ii) $\min\{\frac{1}{2}, 1 - \frac{1}{2}C_K(1-\frac{x}{h})^2\} - C_K\frac{1}{Nh^2} \leq \hat{p}_K(x,h) \leq 1 + C_K \cdot \frac{1}{Nh^2}$ *for* $0 \leq x \leq h$

iii) $\min\{\frac{1}{2}, 1 - \frac{1}{2}C_K(1-\frac{1-x}{h})^2\} - C_K\frac{1}{Nh^2} \leq \hat{p}_K(x,h) \leq 1 + C_K \cdot \frac{1}{Nh^2}$ *for* $1-h \leq x \leq 1$

Proof: i) follows from Corollary 1 a). By Lemma 1 a), we have for $0 \le x \le h$:

$$|\int_0^1 K_h(x-y)dy - \frac{1}{N}\sum_{j=1}^N K_h(x-x_j)| \le C_K \cdot \frac{1}{Nh^2}.$$

Now, $\int_0^1 K_h(x-y)dy = \int_{-\frac{x}{h}}^{\frac{1-x}{h}} K(z)dz = 1 - \int_{-1}^{-\frac{x}{h}} K(z)dz = 1 - \int_{\frac{x}{h}}^1 K(z)dz$ using the symmetry of K. By symmetry and nonnegativity of K, the right-hand side is in $[\frac{1}{2}, 1]$. As $K(1) = 0$ and K is Lipschitz continuous, $K(z) \le C_K(1-z)$ for $0 \le z \le 1$, i.e.

$$\int_0^1 K_h(x-y)dy \ge 1 - C_K \int_{\frac{x}{h}}^1 (1-z)dz = 1 - C_K \frac{1}{2}(1-\frac{x}{h})^2$$

ii) follows. iii) can be shown analogously. ■

Proposition 2 describes the asymptotic behaviour of $\hat{\mu}$ which is related to the final estimate $\mathbf{u}$ by (4). To get the asymptotic properties of $\mathbf{u}$, we have to investigate the matrix factor of (4). First, we show that Λ is uniformly of order $1/\sqrt{g}$ for all $N \ge 1$ w.r.t. the common matrix norm $||\Lambda|| = \sup_{||\mathbf{z}||=1} ||\Lambda \mathbf{z}||$. As an immediate consequence, Λ^t is of order $g^{-t/2}$ for all integer $t \ge 1$. Additionally, we have that $\hat{P}^{-1}$ is approximately $I/(1+\lambda)$, where I denotes the $N \times N$ identity matrix, if it is applied to vectors which are 0 in certain coordinates corresponding to the boundary. We again use the abbreviation $Q_L = \int L^2(u)du$.

Lemma 2 *Let K, L satisfy (A1), a)-e). Let $N \to \infty, h, g \to 0, Nh^2, Ng^2 \to \infty$*

i) $\quad ||\Lambda||^2 \le \frac{1}{g}\int L^2(u)du - \int |u| L^2(u)du + O\Big(\frac{1}{Ng^3}\Big) = \frac{Q_L}{g} + O(1) + O\Big(\frac{1}{Ng^3}\Big)$

ii) $\quad ||\Lambda^t||^2 \le \Big(\frac{Q_L}{g}\Big)^t + O\Big(\frac{1}{g^{t-1}}\Big) + O\Big(\frac{1}{Ng^{2+t}}\Big)$

iii) $\quad ||\hat{P}^{-1}||$ *is uniformly bounded for all N large enough.*

iv) $\quad$ *Let K, L satisfy additionally (A1), f). Then, for $\mathbf{z} \in \mathbb{R}^N$ with $z_j = 0$ for all $j < N(h+g), j > N - N(h+g)$, we have*

$\quad ||\hat{P}^{-1}\mathbf{z} - \frac{1}{1+\lambda}\mathbf{z}|| \le \delta_N\, ||\mathbf{z}|| \qquad$ *with* $\delta_N = O(\frac{1}{N^2h^2}) + O(\frac{\lambda}{N^2g^2})$.

Proof: i) Note that Λ is a symmetric matrix, and let $\eta, \mathbf{e}$ be the largest (in absolute value) eigenvalue and a corresponding unit eigenvector of Λ. Then,

$$\begin{aligned}
||\Lambda||^2 &= \eta^2 = ||\Lambda \mathbf{e}||^2 = \sum_{j=1}^N (\Lambda \mathbf{e})_j^2 = \sum_{j=1}^N \big(\sum_{k=1}^N \Lambda_{jk} e_k\big)^2 \\
&\le \sum_{j=1}^N \big(\sum_{k=1}^N \Lambda_{jk}^2\big)\big(\sum_{k=1}^N e_k^2\big) = \sum_{j,k=1}^N \Lambda_{jk}^2 = \frac{1}{N^2}\sum_{j,k=1}^N L_g^2(x_j - x_k) \\
&= \int_0^1\int_0^1 \frac{1}{g^2} L^2\big(\frac{v-w}{g}\big)dvdw + O\big(\frac{1}{Ng^3}\big) = \int_0^{1/g}\int_0^{1/g} L^2(y-z)dydz + O\big(\frac{1}{Ng^3}\big) \\
&= \int L^2(u) \int 1_{[0,g^{-1}]}(z) 1_{[-u, g^{-1}-u]}(z) dzdu + O\big(\frac{1}{Ng^3}\big) \\
&= \int L^2(u)\left(\frac{1}{g} - |u|\right)du + O\big(\frac{1}{Ng^3}\big),
\end{aligned}$$

where we have used the Cauchy-Schwarz inequality for getting line 2 and Lemma 1 a) for getting line 3.

ii) Using the notation of i), $\eta^t, \mathbf{e}$ are the largest eigenvalue of the symmetric matrix Λ^t and a corresponding unit eigenvector. Therefore, for $Ng^2 \to \infty$,

$$||\Lambda^t||^2 = ||\Lambda||^{2t} \leq \frac{1}{g^t}\Big(Q_L + O(g) + O(\frac{1}{Ng^2})\Big)^t = \Big(\frac{Q_L}{g}\Big)^t + O(\frac{1}{g^{t-1}}) + O(\frac{1}{Ng^{2+t}})$$

iii) By Corollary 2, for some constant c we have $\hat{p}_\lambda(x_j, h, g) = \hat{P}_{jj} \geq c > 0$ uniformly in $j = 1, \ldots, N$, for all large enough N. Therefore, for any unit vector $\mathbf{z}$

$$||\hat{P}^{-1}\mathbf{z}||^2 = \sum_{j=1}^N \frac{z_j^2}{\hat{p}_\lambda(x_j, h, g)^2} \leq \frac{1}{c^2}\ ||\mathbf{z}||^2 = \frac{1}{c^2}.$$

iv) If $\max(h, g) \leq x \leq 1 - \max(h, g)$, we have by Corollary 1 b)

$$|1 - \hat{p}_K(x, h)| = O(\frac{1}{N^2h^2}),\ |1 - \hat{p}_L(x, g)| = O(\frac{1}{N^2g^2}).$$

As, by Corollary 2, $\hat{p}_K$ and $\hat{p}_L$ are uniformly bounded from below by a positive constant for N large enough, we have uniformly for all those x

$$\left|\frac{1}{\hat{p}_\lambda(x, h, g)} - \frac{1}{1+\lambda}\right| \leq\ \delta_N.$$

Therefore, as all x_j which correspond to nonvanishing z_j satisfy this condition

$$||\hat{P}^{-1}\mathbf{z} - \frac{1}{1+\lambda}\ \mathbf{z}||^2 = \sum_{j=1}^N \left|\frac{1}{\hat{p}_\lambda(x_j, h, g)} - \frac{1}{1+\lambda}\right|^2 z_j^2 \leq \delta_N^2\ ||\mathbf{z}||^2$$

■

As $\hat{P}^{-1}$ is bounded and $\lambda\Lambda$ is asymptotically negligible if $\lambda/\sqrt{g} \to 0$, we can expand the factor of $\hat{\mu}$ in (4) for some given integer $t \geq 1$ to

$$\begin{aligned}
\{\hat{P} - \lambda\Lambda\}^{-1} &= \{I - \lambda\ \hat{P}^{-1}\Lambda\}^{-1}\hat{P}^{-1} \\
&= \Big\{\sum_{n=0}^t \lambda^n \big(\hat{P}^{-1}\Lambda\big)^n\Big\}\hat{P}^{-1} + O\big(\lambda^{t+1}||(\hat{P}^{-1}\Lambda)^{t+1}\hat{P}^{-1}||\big).
\end{aligned}$$

where the remainder term is of order $(\lambda/\sqrt{g})^{t+1}$ by Lemma 2 ii) and iii). Therefore, we have

$$\mathbf{u} = \Big\{\sum_{n=0}^t \lambda^n\big(\hat{P}^{-1}\Lambda\big)^n\hat{P}^{-1} + O\Big(\frac{\lambda^{t+1}}{\sqrt{g^{t+1}}}\Big)\Big\}\hat{\mu} = \Big\{\hat{P}^{-1}\sum_{n=0}^t \lambda^n\big(\Lambda\hat{P}^{-1}\big)^n + O\Big(\frac{\lambda^{t+1}}{\sqrt{g^{t+1}}}\Big)\Big\}\hat{\mu}$$

or, coordinatewise, as $\hat{\mu}_i$ is stochastically bounded by Proposition 2 and Chebyshev's inequality uniformly in $\{i; x_i \in [h, 1-h]\}$

$$\begin{aligned}
u_i &= \frac{1}{\hat{p}_\lambda(x_i, h, g)}\hat{\mu}(x_i, h) \\
&\quad + \frac{1}{\hat{p}_\lambda(x_i, h, g)}\sum_{n=1}^t \frac{\lambda^n}{N^n}\sum_{j_1,\cdots,j_n} \frac{L_g(x_i - x_{j_1})}{\hat{p}_\lambda(x_{j_1}, h, g)}\frac{L_g(x_{j_1} - x_{j_2})}{\hat{p}_\lambda(x_{j_2}, h, g)}\cdots\frac{L_g(x_{j_{n-1}} - x_{j_n})}{\hat{p}_\lambda(x_{j_n}, h, g)}\hat{\mu}(x_{j_n}, h) \\
&\quad + O_p\Big(\frac{\lambda^{t+1}}{\sqrt{g^{t+1}}}\Big). \qquad (5)
\end{aligned}$$

We neglect the boundary and consider only x_i with

$$\max(h,g) + tg \le x_i \le 1 - \max(h,g) - tg. \tag{6}$$

Then, as L_g has support $[-g,g]$, we have $\max(h,g) \le x_{j_n} \le 1 - \max(h,g)$ for all $j_n, n = 1,\dots,t$, corresponding to the nonvanishing terms in the n-fold sum of (5). Therefore, we may apply Lemma 2 iv) to replace $\hat{p}_\lambda(x_{j_n},h,g)$ by $1+\lambda$ in (5), and we get

$$\begin{aligned} u_i &= \frac{1}{1+\lambda}\left\{\hat{\mu}(x_i,h) + \sum_{n=1}^{t} \frac{\lambda^n}{(1+\lambda)^n}\hat{\mu}_{n+1}(x_i,h,g)\right\} + R^*_{N,i} \qquad (7) \\ &= \frac{1}{1+\lambda}\sum_{n=0}^{t}\frac{\lambda^n}{(1+\lambda)^n}\hat{\mu}_{n+1}(x_i,h,g) + R^*_{N,i} \end{aligned}$$

where the remainder term $R^*_{N,i}$ is $O_p(\lambda^{t+1}/\sqrt{g^{t+1}}) + O_p(\delta_N)$ uniformly in all i satisfying (6). Here, we define $\hat{\mu}_1(x,h,g) = \hat{\mu}(x,h)$ and, for $n \ge 1$,

$$\begin{aligned} \hat{\mu}_{n+1}(x,h,g) &= \frac{1}{N^n}\sum_{j_1,\dots,j_n} L_g(x-x_{j_1})L_g(x_{j_1}-x_{j_2})\dots L_g(x_{j_{n-1}}-x_{j_n})\hat{\mu}(x_{j_n},h) \\ &= \frac{1}{N^{n+1}}\sum_{j_1,\dots,j_n,\ell} L_g(x-x_{j_1})L_g(x_{j_1}-x_{j_2})\dots L_g(x_{j_{n-1}}-x_{j_n})K_h(x_{j_n}-x_\ell)f_\ell \\ &= \sum_\ell \gamma_\ell^{(n)}(x,h,g)f_\ell + \hat{\mu}_n(x,h,g) = \hat{\nu}_{n+1}(x,h,g) + \hat{\mu}_n(x,h,g) \end{aligned}$$

where, for $n \ge 1$,

$$\begin{aligned} \gamma_\ell^{(n)}(x,h,g) &= \frac{1}{N^{n+1}}\sum_{j_1,\dots,j_n} L_g(x-x_{j_1})L_g(x_{j_1}-x_{j_2})\dots L_g(x_{j_{n-1}}-x_{j_n})K_h(x_{j_n}-x_\ell) \\ &\quad - \frac{1}{N^n}\sum_{j_1,\dots,j_{n-1}} L_g(x-x_{j_1})L_g(x_{j_1}-x_{j_2})\dots L_g(x_{j_{n-2}}-x_{j_{n-1}})K_h(x_{j_{n-1}}-x_\ell). \end{aligned}$$

We can calculate the weights $\gamma_\ell^{(n)}$ also recursively from

$$\begin{aligned} \gamma_\ell^{(1)}(x,h,g) &= \frac{1}{N^2}\sum_j L_g(x-x_j)K_h(x_j-x_\ell) - \frac{1}{N}K_h(x-x_\ell), \\ \gamma_\ell^{(n)}(x,h,g) &= \frac{1}{N}\sum_j L_g(x-x_j)\gamma_\ell^{(n-1)}(x_j,h,g), \quad n \ge 2. \qquad (8) \end{aligned}$$

Writing as abbreviation $\vartheta = \lambda/(1+\lambda)$ such that $1-\vartheta = 1/(1+\lambda)$, we get from (7) and $\hat{\nu}_n(x,h,g) = \hat{\mu}_n(x,h,g) - \hat{\mu}_{n-1}(x,h,g), n \ge 2$, and setting for convenience $\hat{\nu}_1(x,h,g) =$

$\hat{\mu}_1(x,h,g) = \hat{\mu}(x,h)$

$$
\begin{aligned}
u_i &= (1-\vartheta)\sum_{n=0}^{t}\vartheta^n \hat{\mu}_n(x_i,h,g) + R^*_{N,i} \\
&= (1-\vartheta)\sum_{n=1}^{t}\vartheta^n \sum_{k=1}^{n+1}\hat{\nu}_k(x_i,h,g) + R^*_{N,i} \\
&= (1-\vartheta)\sum_{k=0}^{t}\hat{\nu}_{k+1}(x_i,h,g)\sum_{n=k}^{t}\vartheta^n + R^*_{N,i} \\
&= \sum_{k=0}^{t}\vartheta^k(1-\vartheta^{t+1-k})\hat{\nu}_{k+1}(x_i,h,g) + R^*_{N,i}.
\end{aligned}
\tag{9}
$$

As we already know the asymptotic behaviour of $\hat{\nu}_1 = \hat{\mu}$, we have to investigate that of $\hat{\nu}_n, n \geq 2$. The first step is to look at the weights $\gamma_\ell^{(n)}$.

Lemma 3 *For $n \geq 1$*

$$
\begin{aligned}
&\gamma_\ell^{(n)}(x,h,g) = O(\frac{g}{Nh^2}) && \textit{uniformly in} \quad h+ng \leq x \leq 1-(h+ng), \\
&\gamma_\ell^{(n)}(x,h,g) = 0 && \textit{if } |x-x_\ell| > h+ng.
\end{aligned}
$$

Proof: For $n=1$, the assertion follows analogously to Lemma A2 of Franke and Härdle [1]. For $n \geq 2$, we get it straightforwardly by induction, applying the recursion (8). ∎

Proposition 3 *Assume that K and L satisfy (A1), a)-g), and that μ satisfies (A2). Then, if $h, g \to 0, Ng^4, Nh^4 \to \infty$ for $N \to \infty$ we have for all $n \geq 1$ uniformly in $h+ng \leq x \leq 1-(h+ng)$*

$$
\begin{aligned}
&i) \qquad \text{var } \hat{\nu}_{n+1}(x,h,g) = O(\frac{g^2}{Nh^3}) + O(\frac{g^3}{Nh^4}). \\
&ii) \qquad E\hat{\nu}_{n+1}(x,h,g) = \sum_\ell \gamma_\ell^{(n)}(x,h,g)\mu(x_\ell) = \text{bias } \hat{\mu}_L(x,g) + o(g^2),
\end{aligned}
$$

where $\hat{\mu}_L$ denotes the Priestley-Chao estimate with kernel L instead of K.

Proof: a) As in the proof of (A7) of Franke and Härdle [1], we get, using Lemma 3,

$$
Nh \text{ var}(\sum_l \gamma_l^{(n)}(x,h,g)f_l) = Nh\ E(\sum_l \gamma_l^{(n)}(x,h,g)\varepsilon_l)^2 \leq c\ \frac{(h+ng)g^2}{h^3}
$$

for some constant $c > 0$. i) follows.

b) We first show ii) for $n=1$ similar to the proof of Theorem 1, part d) of Franke and Härdle [1]. In this part of the proof, we use the abbreviation $\gamma_\ell = \gamma_\ell^{(1)}$. We have

$$
\begin{aligned}
E\hat{\nu}_2(x,h,g) &= \sum_\ell \gamma_\ell(x,h,g)\mu(x_\ell) \\
&= \frac{1}{N^2}\sum_{\ell,j} L_g(x-x_j)K_h(x_j-x_\ell)\mu(x_\ell) - \frac{1}{N}\sum_\ell K_h(x-x_\ell)\mu(x_\ell) \\
&= \frac{1}{N}\sum_j L_g(x-x_j)\Big\{\frac{1}{N}\sum_\ell K_h(x_j-x_\ell)\mu(x_\ell)-\mu(x_j)\Big\} \\
&\quad +\frac{1}{N}\sum_j L_g(x-x_j)\mu(x_j)-\mu(x)-\Big\{\frac{1}{N}\sum_\ell K_h(x-x_\ell)\mu(x_\ell)-\mu(x)\Big\} \\
&= b_L(x,g)+\frac{1}{N}\sum_j L_g(x-x_j)b(x_j,h)-b(x,h) \qquad (10)
\end{aligned}
$$

with $b(x,h) = \text{bias } \hat{\mu}(x,h), b_L(x,g) = \text{bias } \hat{\mu}_L(x,g)$. Using (A1) f), (A2) we have that $b(x,h)$ is twice continuously differentiable with derivative

$$
b''(x,h) = \frac{1}{Nh^3}\sum_\ell K''(\frac{x-x_\ell}{h})\mu(x_\ell) - \mu''(x).
$$

Using Taylor's formula with remainder in Lagrangian form we get, recalling that $L_g(x-x_j) = 0$ for $|x-x_j| > g$,

$$
\begin{aligned}
&\left|\frac{1}{N}\sum_j L_g(x-x_j)b(x_j,h)-b(x,h)\right| \\
\leq &\left|\frac{1}{N}\sum_j L_g(x-x_j)\{b(x_j,h)-b(x,h)\}\right| + |b(x,h)|O(\frac{1}{N^2g^2}) \\
\leq &\ |b'(x,h)|\left|\frac{1}{N}\sum_j L_g(x-x_j)\{x_j-x\}\right| \\
&+\sup_{|z-x|\leq g}|b''(z,h)|\ \frac{g^2}{2}\ \frac{1}{N}\sum_j L_g(x-x_j) + |b(x,h)|O(\frac{1}{N^2g^2}),
\end{aligned}
$$

where we have used Corollary 1 b) and (A1), b) for the kernel L in getting the first inequality. Applying this argument again to the last line and an analogous result for the kernel $-uL(u)$ for the second line we get

$$
\begin{aligned}
&\left|\frac{1}{N}\sum_j L_g(x-x_j)b(x_j,h)-b(x,h)\right| \\
\leq &\ |b'(x,h)|\left| -g\int L(u)\ u\ du + O(\frac{1}{N^2g^2})\right| \\
&+\sup_{|z-x|\leq g}|b''(z,h)|\ \frac{g^2}{2}\ \left(1+O(\frac{1}{N^2g^2})\right) + |b(x,h)|O(\frac{1}{N^2g^2}) \\
= &\sup_{|z-x|\leq g}|b''(z,h)|O(g^2) + (|b'(x,h)|+|b(x,h)|)O(\frac{1}{N^2g^2}) \qquad (11)
\end{aligned}
$$

by symmetry of L. From Proposition 2, we know $b(x,h) = O(h^2)$. For the first derivative of the bias, we have, applying Lemma 1 a) to $g(z) = K'(\frac{x-z}{h})\mu(z)$,

$$\begin{aligned}
|b'(x,h)| &= \left|\frac{1}{Nh^2}\sum_\ell K'(\frac{x-x_\ell}{h})\mu(x_\ell) - \mu'(x)\right| \\
&= \left|\frac{1}{h^2}\int K'(\frac{x-u}{h})\mu(u)du - \mu'(x)\right| + O\big(\frac{1}{Nh^3}\big) \\
&= \left|\int K(z)\{\mu'(x-hz) - \mu'(x)\}dz\right| + O\big(\frac{1}{Nh^3}\big) \\
&\leq O(h^{1+\beta})\int K(z)|z|dz \ \sup_z |\mu''(z)| + O\big(\frac{1}{Nh^3}\big) = O(h^{1+\beta}) + O\big(\frac{1}{Nh^3}\big)
\end{aligned}$$

where we have used integration by parts, (A1) d) and substitution for getting the third line and Taylor expansion of μ' as well as symmetry of K and (A2), b) for getting the last line. Analogously, we get, using (A1) g)

$$|b''(x,h)| = \left|\int K(z)\{\mu''(x-hz) - \mu''(x)\}dz\right| + O\big(\frac{1}{Nh^4}\big) = O(h^\beta) + O\big(\frac{1}{Nh^4}\big),$$

using again (A2), b), i.e. Hoelder continuity of μ''. Finally, we get from (11)

$$\left|\frac{1}{N}\sum_j L_g(x-x_j)b(x_j,h) - b(x,h)\right| = o(g^2) \tag{12}$$

which, together with (10), implies ii) for $n = 1$.

c) For $n \geq 2$, we get ii) by induction. We assume that ii) holds for $n-1$, and we get for $h + ng \leq x \leq 1 - (h + ng)$

$$\sum_\ell \gamma_\ell^{(n)}(x,h,g)\mu(x_\ell) = \frac{1}{N}\sum_j L_g(x-x_j)E\hat{\nu}_n(x_j,h,g) = \frac{1}{N}\sum_j L_g(x-x_j)\big\{b_L(x_j,g) + o(g^2)\big\}$$

by (8), as for all x_j with nonvanishing $L_g(x-x_j)$ we have $h + (n-1)g \leq x \leq 1 - (h + (n-1)g)$. The remainder on the right-hand side is, again by Lemma 1 b), of order $\{1 + O(1/(N^2g^2)\}o(g^2) = o(g^2)$, and the first term is

$$\left\{\frac{1}{N}\sum_j L_g(x-x_j)b_L(x_j,g) - b_L(x,g)\right\} + b_L(x,g) = o(g^2) + b_L(x,g)$$

by the same arguments as in proving (12) with $b_L(\cdot,g)$ replacing $b(\cdot,h)$. ∎

By Proposition 3, we have uniformly for all $h + tg \leq x \leq 1 - (h + tg), 1 \leq n \leq t$,

$$\begin{aligned}
\hat{\nu}_{n+1}(x,h,g) &= E\ \hat{\nu}_{n+1}(x,h,g) + \hat{\nu}_{n+1}(x,h,g) - E\ \hat{\nu}_{n+1}(x,h,g) \\
&= \text{bias } \hat{\mu}_L(x,g) + o(g^2) + \left(1 + \sqrt{\frac{g}{h}}\right)O_p\left(\frac{g}{h}\frac{1}{\sqrt{Nh}}\right).
\end{aligned}$$

In this paper, we consider only the situation where the last part is neglible to the first one which, by Proposition 2 a), is of order $O(g^2)$. For that purpose, we either have to assume

$g = O(h), Nh^3g^2 \to \infty$ or $h = o(g), Nh^4g \to \infty$. Then, we get from (9) with remainder term $R^{**}_{N,i} = R^*_{N,i} + o(\vartheta g^2)$

$$\begin{aligned} u_i &= (1-\vartheta^{t+1})\hat{\mu}(x_i,h) + \sum_{k=1}^{t} \vartheta^k (1-\vartheta^{t+1-k})\hat{\nu}_{k+1}(x_i,h,g) + R^*_{N,i} \\ &= (1-\vartheta^{t+1})\hat{\mu}(x_i,h) + \sum_{k=1}^{t} \vartheta^k (1-\vartheta^{t+1-k}) \text{bias } \hat{\mu}_L(x_i,g) + R^{**}_{N,i} \\ &= (1-\vartheta^{t+1})\hat{\mu}(x_i,h) + \{\vartheta \frac{1-\vartheta^t}{1-\vartheta} + t\vartheta^{t+1}\} \text{bias } \hat{\mu}_L(x_i,g) + R^{**}_{N,i} \\ &= \hat{\mu}(x_i,h) + \frac{\vartheta}{1-\vartheta} \text{bias } \hat{\mu}_L(x_i,g) + R_{N,i} \end{aligned} \tag{13}$$

setting $R_{N,i} = R^{**}_{N,i} + O_p(\vartheta^{t+1})$. We summarize this to the main result of this section:

Theorem 2.1 *Let K, L satisfy (A1), a)-g), and let (A2) hold. Then, for $N \to \infty,\ h, g, \lambda \to 0$, such that $\lambda = O(g^s)$ for some $s > 1/2$, $Nh^4 \to \infty, Ng^4 \to \infty$, and either $g = O(h), Nh^3g^2 \to \infty$ or $h = o(g), Nh^4g \to \infty$.*

Then, with $t = \max\{[5/(2s-1)] + 1, 4\}$ the smallest integer satisfying $s - \frac{1}{2} - \frac{5}{2t} > 0$ and $t \geq 4$, we have uniformly for all i satisfying $h + tg \leq x_i \leq 1 - (h + tg)$

i) $u_i = \hat{\mu}(x_i,h) + \lambda \text{ bias } \hat{\mu}_L(x_i,g) + o_p(\lambda g^2) + O_p\big(\frac{1}{N^2h^2}\big)$.

ii) bias $u_i = E\ u_i - \mu(x_i) = \frac{1}{2}\{h^2V_K + \lambda\ g^2V_L\}\mu''(x_i) + o(\lambda g^2) + O(h^{2+\beta}) + O\big(\frac{1}{N^2h^2}\big)$.

iii) var $u_i = \mathrm{E}(u_i - \mathrm{E}u_i)^2 = \frac{\sigma^2}{Nh}Q_K + o\big(\frac{\lambda g^2}{\sqrt{Nh}}\big) + o(\lambda^2 g^4) + O\big(\{\frac{1}{Nh}\}^{5/2}\big)$

iv) mse $u_i = E\ (u_i - \mu(x_i))^2 = \frac{1}{4}\{h^2V_K + \lambda\ g^2V_L\}^2\{\mu''(x_i)\}^2 + \frac{\sigma^2}{Nh}\ Q_K + O(h^{4+\beta})$
$+\ o(\lambda g^2) \max\{\frac{1}{\sqrt{Nh}}, h^2 + \lambda g^2\}$

Proof: i) follows immediately from replacing ϑ by $\lambda/(1+\lambda)$ in (13) and looking at the remainder term

$$R_{N,i} = O_p\big(\frac{\lambda^{t+1}}{\sqrt{g^{t+1}}}\big) + O_p\big(\frac{1}{N^2h^2}\big) + O_p\big(\frac{\lambda}{N^2g^2}\big) + o(\lambda g^2) + O_p(\lambda^{t+1}).$$

By the choice of t and the assumption on λ, the first and the last term both are $o_p(\lambda g^2)$. For the third term, this also holds as $Ng^2 \to \infty$.

Using a dominated convergence argument, we get from i)

$$\begin{aligned} E\ u_i - \mu(x_i) &= \text{bias}\,\hat{\mu}(x_i,h) + \lambda \text{ bias}\,\hat{\mu}_L(x_i,g) + o(\lambda g^2) + O\big(\frac{1}{N^2h^2}\big) \\ &= \frac{h^2}{2}\mu''(x_i)V_K + O(h^{2+\beta}) + \lambda\ \frac{g^2}{2}\mu''(x_i)V_L + \lambda o(g^2) + O\big(\frac{1}{N^2h^2}\big) \end{aligned}$$

by Proposition 2 i).

As the bias of $\hat{\mu}_L(x_i, g)$ is not random and using

$$\hat{\mu}(x_i, h) - \mathrm{E}\hat{\mu}(x_i, h) = O_p(\frac{1}{\sqrt{Nh}})$$

by Proposition 2, we have

$$\text{var } u_i = \text{var } \hat{\mu}(x_i, h) + O(\frac{1}{\sqrt{Nh}})\big(o(\lambda g^2) + O_p\big(\frac{1}{N^2 h^2}\big)\big) + o(\lambda^2 g^4)$$

by Proposition 2 ii) and using $Nh \to \infty$.

Finally, iv) follows from mse u_i = var u_i + (bias $u_i)^2$ by applying ii) and iii) and neglecting remainder terms which are of smaller order under our rate assumptions. ■

3 Applications

Once we have the asymptotic equivalence of the regularized least-squares estimate and the shifted Priestley-Chao estimate given by Theorem 2.1 i), we can exploit it to answer various questions showing up in the practice of curve estimation or, analogously, in the two-dimensional case of image denoising. As an illustration, we derive here the asymptotic distribution of the function estimate which allows to quantify the reliability of the estimate or to test if the observed data are a noisy version of a given curve resp. image μ.

If we are not only interested in point estimates u_i for the function values $\mu(x_i)$, but also in confidence intervals which provide information about the reliability of the estimates u_i, we need an asymptotically valid approximation of the probability distribution of u_i. For that purpose, we prove asymptotic normality of our function estimates which may also be used for constructing hypothesis tests for comparing functions. As the grid points x_i, where we observe data and where we calculate estimates u_i, depend on N, we have to extend our method to estimates $u(x)$ of the function of interest $\mu(x)$ at an arbitrary fixed location x. We may interpolate the estimates u_i at $x_i, i = 1, \ldots, N$, smoothly by, e.g., splines, but let us first consider the case where we just define

$$u(x) = u_i \quad \text{for } x_{i-1} < x \leq x_i, i = 1, \ldots, N,$$

with $x_0 = 0$.

Theorem 3.1 *Under the assumptions of Theorem 2.1 we have*

i)

$$\text{bias } u(x) = \mathrm{E}\ u(x) - \mu(x) = \frac{1}{2}\big\{h^2 V_K + \lambda\ g^2 V_L\big\}\mu''(x) + R_N$$

with remainder $R_N = o(\lambda g^2) + O\big(\frac{h^2}{N^\beta}\big) + O(h^{2+\beta}) + O\big(\frac{1}{N^2 h^2}\big) + O(\frac{1}{N})$ *uniformly in* $h + tg \leq x \leq 1 - (h + tg) - \frac{1}{N}$.

ii) If, additionally, $\lambda^2 h = O(\frac{1}{Ng^4})$

$$\sqrt{Nh}\left(u(x) - \mathrm{E}\ u(x)\right) \to_d \mathcal{N}(0, \sigma^2 Q_K) \quad \textit{for } N \to \infty.$$

Proof: For given x, we choose the sequence $i(N)$ depending on N such that $x_{i(N)-1} < x \leq x_{i(N)}$ for all N. Then, $u(x) = u_{i(N)}$. We remark that $x_{i(N)}$ satisfies the uniformity condition on x_i in Theorem 2.1 if $h + tg \leq x \leq 1 - (h + tg) - \frac{1}{N}$.

i) By assumption (A2), we have $\mu(x_{i(N)}) - \mu(x) = O(\frac{1}{N})$, $\mu''(x_{i(N)}) - \mu''(x) = O(\frac{1}{N^\beta})$ uniformly in $x \in (0, 1)$. Therefore, by Theorem 2.1 ii)

$$\begin{aligned}
\text{bias } u(x) &= \text{bias } u_{i(N)} + \mu(x_{i(N)}) - \mu(x) \\
&= \text{bias } u_{i(N)} + O(\frac{1}{N}) \\
&= \frac{1}{2}\big\{h^2 V_K + \lambda\ g^2 V_L\big\}\mu''(x_{i(N)}) + o(\lambda g^2) + O(h^{2+\beta}) + O\big(\frac{1}{N^2 h^2}\big) + O(\frac{1}{N}) \\
&= \frac{1}{2}\big\{h^2 V_K + \lambda\ g^2 V_L\big\}\mu''(x) + O\big(\frac{1}{N^\beta}\max(h^2, \lambda g^2)\big) \\
&\quad + o(\lambda g^2) + O(h^{2+\beta}) + O\big(\frac{1}{N^2 h^2}\big) + O(\frac{1}{N}),
\end{aligned}$$

and i) follows.

ii) Using (1), the centered Priestley-Chao estimate $\hat{\mu}(x_{i(N)}, h) - E\ \hat{\mu}(x_{i(N)}, h)$ is the sample mean of the zero mean random variables $Y_{Nj} = \varepsilon_j K_h(x_{i(N)} - x_j), j = 1, \ldots, N$. Using assumption (A1), a)-c) on the kernel K, it is easy to check that these random variables satisfy the Lindeberg condition, and we get by the Lindeberg-Feller central limit theorem for triangular arrays of random variables (compare section 1.9.3 of Serfling [5]) that

$$\hat{\mu}(x_{i(N)}, h) - E\ \hat{\mu}(x_{i(N)}, h) = \frac{1}{N}\sum_{j=1}^{N} Y_{Nj} \tag{14}$$

is asymptotically normal with mean 0 and variance

$$B_N^2 = \text{var}\big(\ \hat{\mu}(x_{i(N)}, h)\ \big) = \frac{\sigma^2}{Nh} Q_K + O(\frac{1}{N^3 h^3})$$

uniformly in $x \in [h+tg, 1-(h+tg)-\frac{1}{N}]$ by part a) of the proof of Proposition 2. Multiplying the left hand-side of (14) with $\sqrt{Nh}$ and using Slutsky's Theorem (compare section 1.5.4 of Serfling [5]), we get a sequence of random variables with the non-degenerate limit distribution $\mathcal{N}(0, \sigma^2 Q_K)$. By Theorem 2.1 i), using, again, Slutsky's Theorem, we get

$$\begin{aligned}
&\sqrt{Nh}\,\big(u(x) - E\ u(x)\big) = \sqrt{Nh}\,\big(u_{i(N)} - E\ u_{i(N)}\big) \\
=\ & \sqrt{Nh}\big(\hat{\mu}(x_{i(N)}, h) - E\ \hat{\mu}(x_{i(N)}, h)\big) + o_p\big(\lambda g^2 \sqrt{Nh}\big) + O\big(\frac{1}{\sqrt{N^3 h^3}}\big) \\
\to_d\ & \mathcal{N}(0, \sigma^2 Q_K),
\end{aligned}$$

■

Finally, we consider linear interpolation of the estimates u_i at $x_i, i = 1, \ldots, N$, as an alternative. Analogously, we could as well use, e.g., higher-order splines for an even more smooth interpolation. Here, we define for $x \geq x_1$

$$u(x_i) = u_i,\ i = 1, \ldots, N, \qquad u(x) = (1-\theta)u_{i-1} + \theta u_i \quad \text{for } x_{i-1} \leq x \leq x_i,\ i = 1, \ldots, N,$$

with $0 \leq \theta = N(x - x_{i-1}) \leq 1$ depending on x and N. Then, we get the following result analogous to Theorem 3.1:

Theorem 3.2 *Under the assumptions of Theorem 2.1 we have*

i)

$$\text{bias } u(x) = \text{E } u(x) - \mu(x) = \frac{1}{2}\{h^2 V_K + \lambda \, g^2 V_L\}\mu''(x) + R_N$$

with remainder R_N as in Theorem 3.1 uniformly in $h + tg \leq x \leq 1 - (h + tg) - \frac{1}{N}$.

ii) If, additionally, $\lambda^2 h = O(\frac{1}{Ng^4})$

$$\sqrt{Nh}\left(u(x) - \text{E } u(x)\right) \to_d \mathcal{N}(0, \sigma^2 Q_K) \quad \textit{for } N \to \infty.$$

Proof: For given x, we choose the sequence $i(N)$ depending on N such that $x_{i(N)-1} < x \leq x_{i(N)}$ for all N. Then, $u(x) = (1-\theta)u_{i(N)-1} + \theta u_{i(N)}$. We remark that $x_{i(N)-1}, x_{i(N)}$ satisfy the uniformity condition on x_i in Theorem 2.1 if $h + tg \leq x \leq 1 - (h + tg) - \frac{1}{N}$.

a) Using (1), the centered Priestley-Chao estimate $\hat{\mu}(x_{i(N)}, h) - E \; \hat{\mu}(x_{i(N)}, h)$ is the sample mean of the zero mean random variables $Y_{Nj} = \varepsilon_j K_h(x_{i(N)} - x_j), j = 1, \ldots, N$. Using assumption (A1), a)-c), on the kernel K, it is easy to check that these random variables satisfy the Lindeberg condition, and we get by the Lindeberg-Feller central limit theorem for triangular arrays of random variables (compare section 1.9.3 of Serfling [5]) that

$$\hat{\mu}(x_{i(N)}, h) - E \; \hat{\mu}(x_{i(N)}, h) = \frac{1}{N}\sum_{j=1}^{N} Y_{Nj} \tag{15}$$

is asymptotically normal with mean 0 and variance

$$B_N^2 = \text{var}\big(\; \hat{\mu}(x_{i(N)}, h) \; \big) = \frac{\sigma^2}{Nh} Q_K + O(\frac{1}{N^3 h^3})$$

uniformly in $x \in [h + tg, 1 - (h + tg) - \frac{1}{N}]$ by part a) of the proof of Proposition 2. Multiplying the left hand-side of (15) with $\sqrt{Nh}$ and using Slutsky's Theorem (compare section 1.5.4 of Serfling [5]), we get a sequence of random variables with the non-degenerate limit distribution $\mathcal{N}(0, \sigma^2 Q_K)$. The same result, of course, holds for $x_{i(N)-1}$ instead of $x_{i(N)}$.

b) By Theorem 2.1 i), using, again, Slutsky's Theorem, we get

$$\begin{aligned}
& \sqrt{Nh}\left(u(x) - E \; u(x)\right) \\
= \; & \sqrt{Nh}\left((1-\theta)(u_{i(N)-1} - E \; u_{i(N)-1}) + \theta(u_{i(N)} - E \; u_{i(N)})\right) \\
= \; & \sqrt{Nh}\big((1-\theta)[\hat{\mu}(x_{i(N)-1}, h) - E \; \hat{\mu}(x_{i(N)-1}, h)] + \theta[\hat{\mu}(x_{i(N)}, h) - E \; \hat{\mu}(x_{i(N)}, h)]\big) \\
& + o_p\big(\lambda g^2 \sqrt{Nh}\big) + O\big(\frac{1}{\sqrt{N^3 h^3}}\big) \\
\to_d \; & \mathcal{N}(0, v),
\end{aligned}$$

as the sum of two asymptotically normal random variables is again asymptotically normal. We only have to calculate the asymptotic variance v, i.e. the limit of

$$\begin{aligned} & Nh \,\mathrm{var}\big((1-\theta)\hat\mu(x_{i(N)-1},h)+\theta\hat\mu(x_{i(N)},h)\big) = Nh(1-\theta)^2 \mathrm{var}\ \hat\mu(x_{i(N)-1},h) \\ + \quad & Nh\ \theta^2 \mathrm{var}\ \hat\mu(x_{i(N)},h) + 2Nh\ \theta(1-\theta)\ \mathrm{cov}(\hat\mu(x_{i(N)-1},h),\hat\mu(x_{i(N)},h)) \end{aligned}$$

The first two terms coincide asymptotically with $(1-\theta)^2\sigma^2 Q_K + \theta^2\sigma^2 Q_K$ by a). For the third term, we use that by Proposition 2 iv) and as $x_{i(N)} - x_{i(N)-1} = \frac{1}{N}$

$$\frac{\sigma^2}{Nh}\int K(z)K(z+\frac{1}{Nh})dz + O(\frac{1}{N^3h^3}) = \frac{\sigma^2}{Nh}\ Q_K + O(\frac{1}{N^2h^2}) + O(\frac{1}{N^3h^3})$$

by assumption (A1), a)-f). We conclude

$$v = (1-\theta)^2\sigma^2 Q_K + \theta^2\sigma^2 Q_K + 2(1-\theta)\theta\sigma^2 Q_K = \sigma^2 Q_K,$$

and ii) follows.

c) By assumption (A2), we have $\mu(x_{i(N)}) - \mu(x) = O(\frac{1}{N}), \mu''(x_{i(N)}) - \mu''(x) = O(\frac{1}{N^\beta})$ and correspondingly for $x_{i(N)-1}$ uniformly in $x \in (0,1)$. Therefore,

$$\begin{aligned} \mathrm{bias}\ u(x) &= (1-\theta)\ \mathrm{bias}\ u_{i(N)-1} + \theta\ \mathrm{bias}\ u_{i(N)} + (1-\theta)\ \mu(x_{i(N)-1}) + \theta\ \mu(x_{i(N)}) - \mu(x) \\ &= (1-\theta)\ \mathrm{bias}\ u_{i(N)-1} + \theta\ \mathrm{bias}\ u_{i(N)} + O(\frac{1}{N}) \end{aligned}$$

Moreover, by Theorem 2.1 ii),

$$\begin{aligned} \mathrm{bias}\ u_{i(N)} &= \frac{1}{2}\big\{h^2 V_K + \lambda\ g^2 V_L\big\}\mu''(x_{i(N)}) + o(\lambda g^2) + O(h^{2+\beta}) + O\big(\frac{1}{N^2h^2}\big) \\ &= \frac{1}{2}\big\{h^2 V_K + \lambda\ g^2 V_L\big\}\mu''(x) +\ O(\frac{h^2}{N^\beta}) + o(\lambda g^2) + O(h^{2+\beta}) + O\big(\frac{1}{N^2h^2}\big), \end{aligned}$$

and analogously for $x_{i(N)-1}$. We conclude i). ■

Acknowledgements
The work was supported by the Deutsche Forschungsgemeinschaft (DFG) as part of the priority research program 1114 *Mathematical Methods of Time Series and Digital Image Analysis* as well as by the center of excellence *Dependable Adaptive Systems and Mathematical Modelling* funded by the state of Rhineland-Palatinate.

References

[1] J. Franke and W. Härdle. On bootstrapping kernel spectrum estimates. *Ann. Statist.* 20: 121–145, 1992.

[2] W. Härdle. *Applied Nonparametric Regression.* Cambridge University Press, Cambridge, 1990.

[3] P. Mrázek, J. Weickert and A. Bruhn. On robust estimation and smoothing with spatial and tonal kernels. In R. Klette, R. Kozera, L. Noakes, J. Weickert eds., *Geometric Properties from Incomplete Data.* Springer, Berlin-Heidelberg-New York, 2006.

[4] M. B. Priestley and M. T. Chao. Nonparametric function fitting. *J. Royal Stat. Soc., Ser. B,* 34: 385–392, 1972.

[5] R. J. Serfling. *Approximation Theorems of Mathematical Statistics.* Wiley, New York, 1980.

[6] B. W. Silverman. *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, London, 1986.

[7] J. Stoer. *Einführung in die Numerische Mathematik I* (4th ed.). Springer, Berlin-Heidelberg-New York, 1983.

[8] C. J. Stone. Optimal global rates of convergence for nonparametric estimators. *Ann. Statist.* 8: 1348–1360, 1980.

[9] C. J. Stone. Optimal rates of convergence for nonparametric regression. *Ann. Statist.* 10: 1040–1053, 1982.