A note on uniform consistency of monotone function estimators

Natalie Neumeyer[i]

Abstract

Recently, Dette, Neumeyer and Pilz (2005a) proposed a new monotone estimator for strictly increasing nonparametric regression functions and proved asymptotic normality. We explain two modifications of their method that can be used to obtain monotone versions of any nonparametric function estimators, for instance estimators of densities, variance functions or hazard rates. The method is appealing to practitioners because they can use their favorite method of function estimation (kernel smoothing, wavelets, orthogonal series,…) and obtain a monotone estimator that inherits desirable properties of the original estimator. In particular, we show that both monotone estimators share the same rates of uniform convergence (almost sure or in probability) as the original estimator.

MSC 2000: 62G05
KEYWORDS: function estimator, kernel method, monotonicity, uniform convergence

1 Introduction

During the last decades much effort has been devoted to the problem of estimating monotone functions. Estimating a monotone density function was considered by Grenander (1956), Groeneboom (1985), Groeneboom and Lopuhaä (1993), Datta (1995), Cheng, Gasser und Hall (1999), and van der Vaart and van der Laan (2003), among others. Even more literature can be found about estimating increasing regression functions, starting with Brunk (1958), Barlow, Bartholomew, Bremmer and Brunk (1972), Mukerjee (1988), Mammen (1991), Ramsay (1988), and Hall and Huang (2001), among many others; see Gijbels (2003) for a good and recent review. Uniform consistency of Brunk's estimator was shown by Wright (1979) and Smythe (1980). For censored data Huang and Zhang (1994) and Huang and Wellner (1995) consider estimators for a monotone density and monotone hazard rate. For monotone estimators of a hazard rate see also Mukerjee and Wang (1993) and Hall, Huang, Gifford and Gijbels (2001).

Appealing to users of common kernel methods is a new method proposed by Dette, Neumeyer and Pilz (2005a) for nonparametric regression functions and by Dette and Pilz (2004) for variance functions in nonparametric regression models. The considered estimator is easy to

[i]Ruhr-Universität Bochum, Fakultät für Mathematik, 44780 Bochum, Germany, e-mail: natalie.neumeyer@rub.de. The financial support of the Deutsche Forschungsgemeinschaft (Research grant and SFB 475) is gratefully acknowledged.

implement, is based on kernel estimators, and, in contrast to many other procedures, does not require any optimization over function spaces. To obtain a monotone estimator for a strictly increasing function g (here $g : [0, 1] \to \mathbb{R}$ denotes the regression or variance function), the method consists of first monotonicitly estimating the distribution function of $g(U)$, i. e. $h(t) = P(g(U) \leq t)$, by a kernel method, where U is uniformly distributed in $[0, 1]$. The first step uses a (not necessarily increasing) kernel estimator $\hat{g}$ for g. More precisely, the estimator for h is an integrated kernel density estimator,

$$\hat{h}(t) \;=\; \int_{-\infty}^{t} \frac{1}{N} \sum_{i=1}^{N} \frac{1}{a} k\Big(\frac{x - \hat{g}(\frac{i}{N})}{a}\Big)\, dx, \tag{1.1}$$

where k denotes a density function, $a = a_N = o(1)$ a sequence of bandwidths and N converges to infinity. Noting that $h(t) = g^{-1}(t)$, an increasing estimator for g is then obtained by inversion of $\hat{h}$. Asymptotic normality of the constrained estimator is shown in Dette et al. (2005a) and Dette and Pilz (2004). A further application of the method can be found in Dette, Neumeyer and Pilz (2005b) where an increasing estimator for the dose response curve in binomial regression is proposed.

An alternative method to obtain the estimator for g^{-1} is mentioned but not further developed in the aforementioned references, namely using

$$\hat{h}(t) \;=\; \int_{0}^{1} I\{\hat{g}(x) \leq t\}\, dx \tag{1.2}$$

(where I denotes the indicator function) as an estimator for $\int_0^1 I\{g(x) \leq t\}\, dx = g^{-1}(t)$ (where g is increasing). Note that Dette et al.'s (2005a) proof for the asymptotic distribution of $\hat{h}$ defined in (1.1) and its inverse is not easily generalized to obtain asymptotic results about the estimator based on (1.2). The approach to use the inverse $\hat{h}^{-1}$ as an estimator for g, where $\hat{h}$ is defined in (1.2) is related to nondecreasing rearrangements of data considered by Ryff (1965,1970), and is in principle similar to Polonik's (1995,1998) work, who constructs estimators for a density f from the identity

$$f(x) \;=\; \int_{0}^{\infty} I\{f(x) \geq t\}\, dt.$$

The density contour clusters $\{x : f(x) \geq t\}$ are estimated by the so-called excess mass approach. By choosing the class of sets appropriately, for example, monotone density estimators are obtained. In this case the estimator coincides with Grenander's (1956) estimator. In a more general context, Polonik (1995) shows L^1-consistency of the obtained estimators. The approach is related to the estimation of density level sets, see Tsybakov (1997), among others.

In the paper at hand properties of the two methods [using the inverse of (1.1) or (1.2), respectively, as a monotone estimator of g] will be compared. Both methods are not restricted to monotone estimation of regression or variance functions, neither to the case of kernel or local linear estimators used in the first step. These restricted cases were considered in Dette et al. (2005a,b) to prove asymptotic normality of the new estimators and first order equivalence to the unconstrained estimator. In these references it was also crucial to assume the function g to be strictly increasing with positive derivative. Here we consider the general case to modify any function estimator (using kernels, local polynomials, nearest-neighbors, wavelets, splines, orthogonal series, ...) for any function (density, regression function, variance function, hazard function, ...) with compact support (or support bounded on one side) to obtain a monotone (either nondecreasing or strictly increasing) estimator. The estimators do not need to be based on an independent and identically distributed sample but can be based on dependent observations such as time series, or on censored observations. Also the original estimators are not supposed to be nonparametric but can be either non-, semi- or parametric. We only assume knowledge about uniform consistency of the original estimator used in the first step.

Both procedures [based on (1.1) and (1.2)] to obtain monotone versions of any function estimator are explained in detail in Section 2. We will show that the monotone modifications of the estimator share the same rates of uniform convergence (almost sure or in probability) as the original unconstrained estimator, see Section 3. Some examples of applications are also given in Section 3 and the details of the proofs are deferred to Section 4.

2 Monotone modifications of function estimators

We explain in the following the method to obtain a monotone modification of any function estimator $\hat{g}$ of an unknown function g, where g is (not necessarily strictly) increasing. We restrict ourselves first to the case of a compact support of the target function g. Only for the ease of presentation this support is assumed to be $[0,1]$. Changes in the methods for noncompact supports will be discussed at the end of Section 3.

For any Lebesgue–measurable function $f : [a,b] \to \mathbb{R}$ we define a function $\Phi(f) : \mathbb{R} \to \mathbb{R}$ by

$$\Phi(f)(z) \;=\; \int_a^b I\{f(x) \leq z\}\, dx + a, \quad z \in \mathbb{R}.$$

For a strictly increasing function f, the function $\Phi(f) I_{[f(a),f(b)]}$ is just the inverse f^{-1}. Is f increasing, but not strictly, then $\Phi(f) I_{[f(a),f(b)]}$ is the generalized inverse $f^{-1}(t) = \inf\{u \mid f(u) > t\}$ that may have jump points when f has constant parts. Whether f is increasing or not, $\Phi(f)$ is always increasing. Also, $\Phi(f)$ is Lebesgue–measurable. Now for a Lebesgue–measurable function $h : [0,1] \to \mathbb{R}$ we define an increasing modification $h_I : [0,1] \to \mathbb{R}$ by

$$h_I \;=\; \Phi\Big(\Phi(h) I_{[h(0),h(1)]}\Big) I_{[0,1]}.$$

Then, for any (not necessarily strictly) increasing function $g : [0,1] \to \mathbb{R}$, we have $g_I = g$

and for an estimator $\hat{g} : [0,1] \to \mathbb{R}$ for g, we call

$$\hat{g}_I \;=\; \Phi\Big(\Phi(\hat{g}) I_{[\hat{g}(0),\hat{g}(1)]}\Big) I_{[0,1]}$$

an isotone modification of $\hat{g}$. We will show in Section 3 that the monotone estimator $\hat{g}_I$ shares the same rates of uniform convergence to g as $\hat{g}$.
A modification of the presented method uses a smooth approximation of the indicator function. To this end, let k denote a density function, $K(y) = \int_{-\infty}^{y} k(u)\,du$ the primitive of k, and a_n a sequence of positive bandwidths converging to zero for increasing sample size. For any estimator $\hat{g} : [0,1] \to \mathbb{R}$ for g we define

$$\Psi(\hat{g})(y) \;=\; \int_0^1 K\Big(\frac{y - \hat{g}(x)}{a_n}\Big)\,dx$$

and an increasing modification $\hat{g}_{SI}$ of $\hat{g}$ by

$$\hat{g}_{SI} \;=\; \Phi\Big(\Psi(\hat{g}) I_{[\hat{g}(0),\hat{g}(1)]}\Big) I_{[0,1]}.$$

This estimator will be strictly increasing (except for small areas at the boundaries) whenever K is strictly increasing.
Both methods are appealing because every practitioner can use his or her favorite method of function estimation like wavelets or orthogonal series and will obtain an increasing estimator that shares the same rate of uniform consistency and also shares a lot of desirable properties of the original estimator because the new estimator will coincide with the original estimator on every intervall where the unconstrained estimator already is nondecreasing and the endpoints are singletons (compare Figure 2). Which of the two methods to apply depends on the requirements one has for the estimator. When using the first method there is no need for the choice of a bandwidth. Also, flat parts of g are better reflected (we obtain a nondecreasing, not a strictly increasing estimator). But the estimator $\hat{g}_I$ may be not differentiable in some points. With the smooth modification of the method we can obtain strictly increasing and smooth estimators $\hat{g}_{SI}$.
The following figures show the monotone modifications h_I and h_{SI} for a monotone (Figure 1) and a not everywhere monotone function h (Figure 2).
We will also give asymptotic results for discrete versions, $\hat{g}_{I,d} = \Phi(\tilde{\Phi}(\hat{g}) I_{[\hat{g}(0),\hat{g}(1)]}) I_{[0,1]}$ and $\hat{g}_{SI,d} = \Phi(\tilde{\Psi}(\hat{g}) I_{[\hat{g}(0),\hat{g}(1)]}) I_{[0,1]}$ where the integrals in the definitions of $\hat{g}_I$ and $\hat{g}_{IS}$ are approximated by Riemann sums, i. e.

$$\tilde{\Phi}(g)(y) \;=\; \frac{1}{N}\sum_{i=1}^{N} I\{\hat{g}(\frac{i}{N}) \leq y\}$$

$$\tilde{\Psi}(g)(y) \;=\; \frac{1}{N}\sum_{i=1}^{N} K\Big(\frac{y - \hat{g}(\frac{i}{N})}{a}\Big).$$

For the estimator $\hat{g}_{SI,d}$ of a strictly increasing regression function Dette et al. (2005a) showed asymptotic normality under some regularity assumptions. One could also consider estimators

$\tilde{\Phi}(\tilde{\Phi}(\hat{g})I_{[\hat{g}(0),\hat{g}(1)]})I_{[0,1]}$ and $\tilde{\Phi}(\tilde{\Psi}(\hat{g})I_{[\hat{g}(0),\hat{g}(1)]})I_{[0,1]}$ but for the second "inversion" a discretization is not expedient as $\tilde{\Phi}(\hat{g})$ and $\tilde{\Psi}(\hat{g})$ are already monotone and for a monotone function h we have just $\Phi(h) = \inf\{u \mid h(u) > \cdot\}$.

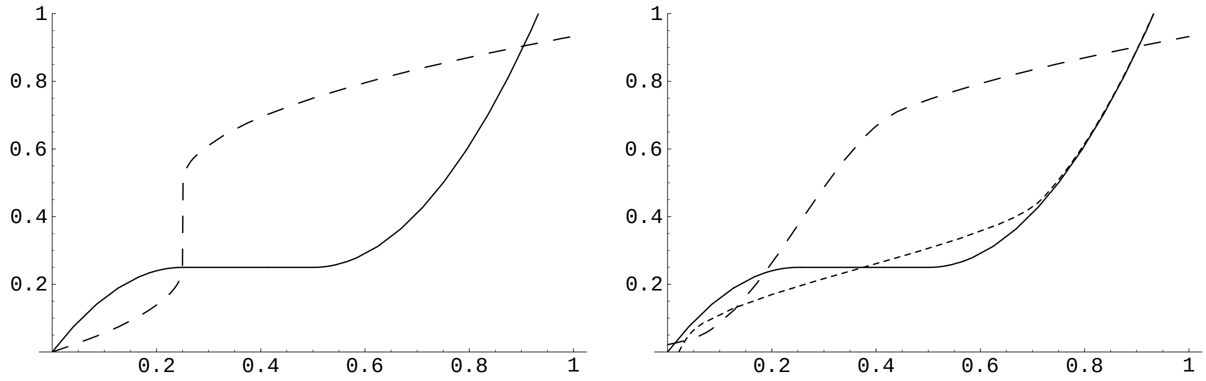

Figure 1: The two graphics show isotone modifications of the nondecreasing function $h(x) = 0.25 - 4(x - 0.25)^2 I\{0 \leq x \leq 0.25\} + 4(x - 0.5)^2 I\{0.5 \leq x \leq 1\}$ (solid line). h_I in the left panel is identical to h, the dotted line in the right panel is h_{SI} for the Epanechnikov–kernel $k(x) = 0.75(1 - x^2)I\{-1 \leq x \leq 1\}$ and bandwidth $a = 0.2$. The dashed curves are $\Phi(h)$ in the left panel and $\Psi(h)$ in the right panel.

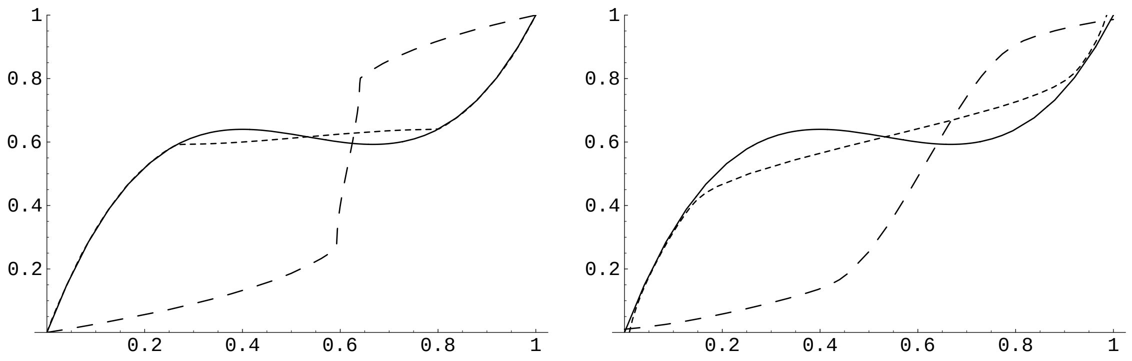

Figure 2: The two graphics show monotone modifications of the not monotone function $h(x) = 5x^3 + 4x - 8x^2$ (solid line). The dotted lines are h_I in the left panel and h_{SI} in the right panel. For the calculation of h_{SI} we used the Epanechnikov–kernel $k(x) = 0.75(1 - x^2)I\{-1 \leq x \leq 1\}$ and bandwidth $a = 0.2$. The dashed curves are $\Phi(h)$ in the left panel and $\Psi(h)$ in the right panel.

3 Main results and applications

In this section we give conditions under which the increasing versions of function estimators share the same rate of uniform convergence as the original estimator. Let in the following $||h||_\infty = \sup_{z\in[0,1]} |h(z)|$ denote the supremum norm of a function $h : [0,1] \to \mathbb{R}$.

Theorem 3.1 **(a)** *Let $g : [0,1] \to \mathbb{R}$ be an increasing function and $\hat{g} : [0,1] \to \mathbb{R}$ an estimator for g. Then there exists a constant c such that for the isotone modification $\hat{g}_I$ of $\hat{g}$ it holds that*

$$||\hat{g}_I - g||_\infty \;\leq\; c||\hat{g} - g||_\infty.$$

(b) *Let $g : [0,1] \to \mathbb{R}$ be a strictly increasing twice differentiable function with bounded second derivative such that the first derivative is bounded away from zero. Let $\hat{g} : [0,1] \to \mathbb{R}$ be an estimator for g. Let k denote a symmetric density function with compact support and two bounded derivatives. Let $a_n = o(1)$ denote a sequence of positive bandwidths. Then there exists a constant c such that for the strictly increasing modification $\hat{g}_{SI}$ of $\hat{g}$ it holds that*

$$||\hat{g}_{SI} - g||_\infty \;\leq\; c\Big(||\hat{g} - g||_\infty + \frac{1}{a_n}||\hat{g} - g||_\infty^2 + \frac{1}{a_n^3}||\hat{g} - g||_\infty^3 + a_n^2\Big).$$

The constant c in Theorem 3.1 obtained in the proof is not claimed to be the best possible. In special cases (for example estimating a regression function by kernel methods) it might be possible to obtain sharper bounds, but our results are valid very general and the given proof is uncomplicated. In the situation of Theorem 3.1 (a) we obtain uniform consistency of the estimator $\hat{g}_I$ whenever $||\hat{g} - g||_\infty = o(1)$ is known. Also, when rates of convergence are known for the original estimator, i. e. $||\hat{g} - g||_\infty = O(c_n)$ for $n \to \infty$ a.s. (in probability), then the same holds for $\hat{g}_I$, i. e.

$$||\hat{g}_I - g||_\infty = O(c_n) \text{ for } n \to \infty \text{ a.s. (in probability).}$$

The estimator $\hat{g}_I$ based on the indicator method works better to estimate constant functions resp. nondecreasing functions with flat parts. Moreover, there is no need for choosing a bandwidth a_n when using this estimator.

In contrast, in the situation of Theorem 3.1 (b) uniform consistency of the estimator $\hat{g}_{SI}$ can only be obtained from rates of the uniform convergence of $\hat{g}$ and by choosing the bandwidth a_n accordingly. When it is known that $||\hat{g} - g||_\infty = O(c_n)$ for $n \to \infty$ a.s. (in probability), then it holds that

$$||\hat{g}_{SI} - g||_\infty = O\Big(c_n + \frac{c_n^2}{a_n} + \frac{c_n^3}{a_n^3} + a_n^2\Big) \text{ for } n \to \infty \text{ a.s. (in probability).}$$

When a sequence of bandwidths a_n is chosen that satisfies $c_n = O(a_n^{5/3})$ and $a_n = O(c_n^{1/2})$ we obtain the same rate $O(c_n)$ for the uniform convergence of the strictly increasing version. For the discrete versions we have the following asymptotic results.

Theorem 3.2 **(a)** *Let $g : [0,1] \to \mathbb{R}$ be a strictly increasing differentiable function such that the first derivative is bounded away from zero and let $\hat{g} : [0,1] \to \mathbb{R}$ be an estimator for g. Then there exists a constant c such that for the isotone modification $\hat{g}_{I,d}$ of $\hat{g}$ it holds that*

$$||\hat{g}_{I,d} - g||_\infty \; \leq \; c\Big(||\hat{g} - g||_\infty + \frac{1}{N} \Big).$$

(b) *Let $g : [0,1] \to \mathbb{R}$ be an strictly increasing twice differentiable function with bounded second derivative such that the first derivative is bounded away from zero. Let $\hat{g} : [0,1] \to \mathbb{R}$ an estimator for g. Let k denote a symmetric density function with compact support and two bounded derivatives. Let $a_n = o(1)$ denote a sequence of positive bandwidths. Then there exists a constant c such that for the strictly increasing modification $\hat{g}_{SI,d}$ of $\hat{g}$ it holds that*

$$||\hat{g}_{SI,d} - g||_\infty \; \leq \; c\Big(||\hat{g} - g||_\infty (1 + \frac{1}{Na_n}) + \frac{1}{a_n} ||\hat{g} - g||_\infty^2 (1 + \frac{1}{Na_n^2}) + \frac{1}{a_n^3} ||\hat{g} - g||_\infty^3 + \frac{1}{N} + \frac{1}{N^2 a_n} + \frac{1}{N^3 a_n^3} + a_n^2 \Big).$$

There are plenty of applications and we only mention a few. Whenever we have knowledge about uniform consistency of a function estimate and a monotone uniformly consistent estimator is desired it is sensible to use one of the above methods. For example, uniform almost sure consistency of kernel density estimators was shown by Silverman (1978), Devroye and Wagner (1978) and Stute (1982), among others. For kernel regression estimators corresponding results can be found in Mack and Silverman (1982), see also Einmahl and Mason (2000). Rates of uniform almost sure convergence for variance function estimators in nonparametric regression models are a by-product of Akritas and Van Keilegom (2001). Further, there is a vast literature about uniform consistency of wavelet estimators for densities and regression functions based on iid or time series or censored data, respectively, see, e. g., Masry (1997), Massiani (2003), Zhang, Sha and Cheng (1999) or Xue (2002). Corresponding results about orthogonal series estimators can be found in publications by Chen (1981), Györfi and Walk (1996), Newey (1997) and de Jong (2002). Moreover, strong uniform consistency of k-nearest neighbor estimators for regression and density functions based on iid or dependent data is considered by Devroye and Wagner (1977), Mack (1983), and Qin and Cheng (1994). Uniform consistency for different estimators of hazard rates is shown by Zhang (1996) and Collomb, Hassani, Sarda and Vieu (1985). For each of the proposed estimators our method yields a monotone version that shares the same rate of uniform convergence.

For example, let m denote the isotone regression function in a nonparametric regression model

$$Y_i \; = \; m(X_i) + \varepsilon_i, \quad i = 1, \ldots, n$$

with independent observations and univariate covariates $X_i \in [0,1]$. Let $\hat{m}$ denote the common kernel regression estimator (Nadaraya, 1964; Watson, 1964),

$$\hat{m}(x) \; = \; \frac{\sum_{i=1}^n \kappa(\frac{x - X_i}{h_n}) Y_i}{\sum_{i=1}^n \kappa(\frac{x - X_i}{h_n})},$$

where κ denotes a kernel function and h_n a sequence of positive bandwidths converging to zero. Under common regularity assumptions (see Mack and Silverman, 1982) it holds that

$$||\hat{m} - m||_\infty = O(c_n) \text{ for } n \to \infty \text{ a.s., where } c_n = \sqrt{\frac{\log h_n^{-1}}{nh_n}}.$$

We obtain an isotone modification of the kernel estimator $\hat{m}$, namely $\hat{m}_I$. This estimator fulfills $||\hat{m}_I - m||_\infty = O(c_n)$ for $n \to \infty$ a.s. For the smooth version $\hat{m}_{SI}$ we obtain $||\hat{m}_{SI} - m||_\infty = O(c_n)$ for $n \to \infty$ a.s. when a sequence of bandwidths a_n is chosen that fulfills $nh_n a_n^4 / \log h_n^{-1} = O(1)$ and $\log h_n^{-1}/(nh_n a_n^{10/3}) = O(1)$. For the common choice $h_n = Cn^{-1/5}$, for instance, $a_n = h_n$ is a possible choice. Note that Birke and Dette (2005) show a rate for uniform convergence of $\hat{m}_{SI}^{-1}$ as a by-product.

Masry (1997) considers d–dimensional wavelet density estimators $\hat{f}$ on compact sets D for strongly mixing stationary processes and densities f in certain Besov spaces B_{spq}. For simplicity we assume $D = [0, 1]$ and consider the one-dimensional case $d = 1$. For example, under certain assumptions in Corollary 1, Masry (1997) obtains the uniform rate of convergence

$$||\hat{f} - f||_\infty \;=\; O\Big(\Big(\frac{\log n}{n}\Big)^{\frac{s}{1+2s}}\Big) \text{ for } n \to \infty \text{ a.s.}$$

for $f \in B_{s\infty\infty}$. The wavelet estimator $\hat{f}$ can be modified to obtain increasing (or, analogously, decreasing) estimators $\hat{f}_I$ and $\hat{f}_{SI}$ such that

$$||\hat{f}_I - f||_\infty \;=\; O\Big(\Big(\frac{\log n}{n}\Big)^{\frac{s}{1+2s}}\Big) \text{ and } ||\hat{f}_{SI} - f||_\infty \;=\; O\Big(\Big(\frac{\log n}{n}\Big)^{\frac{s}{1+2s}}\Big) \text{ for } n \to \infty \text{ a.s.}$$

where for $\hat{f}_{SI}$ a bandwidth a_n is used such that $na_n^{4+2/s}/\log n = O(1)$ and $\log n/(na_n^{2/3+5/(3s)}) = O(1)$. For example, $a_n = Cn^{-1/5}$ is a possible choice for $s = 2$.

Finally, we consider how the assumption of the compact support of the target function can be weakened. For instance, often densities are assumed to be increasing on $(-\infty, 0]$ (respectively decreasing on $[0, \infty)$) and also hazard rates are often defined on $[0, \infty)$. We will describe in the following how the proposed methods are applicable when an increasing function $h : (-\infty, 0] \to \mathbb{R}$ has to be estimated. Assume there is an estimator $\hat{h} : (-\infty, 0] \to \mathbb{R}$ available such that

$$\sup_{z \in (-\infty, 0]} |\hat{h}(z) - h(z)| = O(c_n).$$

Because $\log : (0, 1] \to (-\infty, 0]$ is continuous we have for $g = h \circ \log$, $\hat{g} = \hat{h} \circ \log$ that $||\hat{g} - g||_\infty = O(c_n)$ and from the results of Sections 2 and 3 we obtain a monotone version of $\hat{g}$, i.e. $\hat{g}_I$, such that $||\hat{g}_I - g||_\infty = O(c_n)$. A monotone estimator for h is defined by $\hat{h}_I = \hat{g}_I \circ \exp : (-\infty, 0] \to \mathbb{R}$ and it holds that

$$\sup_{z \in (-\infty, 0]} |\hat{h}_I(z) - h(z)| = ||\hat{g}_I - g||_\infty = O(c_n).$$

4 Proofs

4.1 Proof of Theorem 3.1 (a).

For increasing g we have $g = g_I$ and, hence,

$$\begin{aligned}
||\hat{g}_I - g||_\infty \;&=\; \sup_{z\in[0,1]} \Big| \int_{\hat{g}(0)}^{\hat{g}(1)} I\Big\{ \textstyle\int_0^1 I\{\hat{g}(t) \le x\}\, dt \le z \Big\}\, dx + \hat{g}(0) \\
&\qquad - \int_{g(0)}^{g(1)} I\Big\{ \textstyle\int_0^1 I\{g(t) \le x\}\, dt \le z \Big\}\, dx - g(0) \Big| \\
&\le\; 2|\hat{g}(0) - g(0)| + |\hat{g}(1) - g(1)| + r_n
\end{aligned}$$

where

$$\begin{aligned}
r_n \;&=\; \sup_{z\in[0,1]} \Big| \int_{g(0)}^{g(1)} \Big(I\Big\{ \textstyle\int_0^1 I\{\hat{g}(t) \le x\}\, dt \le z \Big\} - I\Big\{ \textstyle\int_0^1 I\{g(t) \le x\}\, dt \le z \Big\} \Big)\, dx \Big| \\
&\le\; \sup_{z\in[0,1]} \int_{g(0)}^{g(1)} I\Big\{ \textstyle\int_0^1 I\{\hat{g}(t) \le x\}\, dt \le z \text{ and } \int_0^1 I\{g(t) \le x\}\, dt > z \Big\}\, dx \\
&\quad + \sup_{z\in[0,1]} \int_{g(0)}^{g(1)} I\Big\{ \textstyle\int_0^1 I\{\hat{g}(t) \le x\}\, dt > z \text{ and } \int_0^1 I\{g(t) \le x\}\, dt \le z \Big\}\, dx.
\end{aligned}$$

Both summands are bounded in the very same way and we therefore restrict to the first one in the following, i. e.

$$\begin{aligned}
&\sup_{z\in[0,1]} \int_{g(0)}^{g(1)} I\Big\{ \textstyle\int_0^1 I\{g(t) \le x - (\hat{g}(t) - g(t))\}\, dt \le z < \int_0^1 I\{g(t) \le x\}\, dt \Big\}\, dx \\
\le\;& \sup_{z\in[0,1]} \int_{g(0)}^{g(1)} I\Big\{ \textstyle\int_0^1 I\{g(t) \le x - ||\hat{g} - g||_\infty\}\, dt \le z < \int_0^1 I\{g(t) \le x\}\, dt \Big\}\, dx \\
=\;& \sup_{z\in[0,1]} \int_{g(0)}^{g(1)} I\Big\{ g^{-1}(x - ||\hat{g} - g||_\infty) \le z < g^{-1}(x) \Big\}\, dx \\
\le\;& \sup_{z\in[0,1]} \int_{g(0)}^{g(1)} I\Big\{ g(z) \le x \le g(z) + ||\hat{g} - g||_\infty \Big\}\, dx \\
=\;& \sup_{z\in[0,1]} \Big(I\Big\{ g(z) + ||\hat{g} - g||_\infty \le g(1) \Big\} ||\hat{g} - g||_\infty + I\Big\{ g(z) + ||\hat{g} - g||_\infty > g(1) \Big\} (g(1) - g(z)) \Big) \\
\le\;& ||\hat{g} - g||_\infty.
\end{aligned}$$

Altogether we obtain

$$||\hat{g}_I - g||_\infty \;\le\; 2|\hat{g}(0) - g(0)| + |\hat{g}(1) - g(1)| + 2||\hat{g} - g||_\infty \;\le\; 5||\hat{g} - g||_\infty.$$

$\square$

4.2 Proof of Theorem 3.1 (b).

For the proof of Theorem 3.1 (b) we first show that the following Proposition is valid.

Proposition 4.1 *Let $g : [0,1] \to \mathbb{R}$ be a strictly increasing twice differentiable function with bounded second derivative such that the first derivative is bounded away from zero. Let $\hat{g} : [0,1] \to \mathbb{R}$ be an estimator for g. Let k denote a symmetric density function with compact support and two bounded derivatives and let $a_n = o(1)$ denote a sequence of positive bandwidths. Then there exists a constant C such that*

$$\sup_{y\in(g(0),g(1))} |\Psi(\hat{g})(y) - g^{-1}(y)| \ \leq\ C\Big(||\hat{g}-g||_\infty + \frac{1}{a_n}||\hat{g}-g||_\infty^2 + \frac{1}{a_n^3}||\hat{g}-g||_\infty^3 + a_n^2\Big)$$

Proof of Proposition 4.1. During the proof we assume the support of k to be $[-1,1]$. Note that then $K(z) = 0$ for $z \leq -1$ and $K(z) = 1$ for $z \geq 1$. For every fixed $y \in (g(0), g(1))$ we have

$$\begin{aligned}
|\Psi(\hat{g})(y) - g^{-1}(y)| \ \leq\ & \Big|\int_0^1 \Big[K\Big(\frac{y-\hat{g}(x)}{a_n}\Big) - K\Big(\frac{y-g(x)}{a_n}\Big)\Big]\,dx\Big| \\
& + \Big|\int_0^1 K\Big(\frac{y-g(x)}{a_n}\Big)\,dx - g^{-1}(y)\Big|.
\end{aligned} \tag{4.1}$$

The first term on the right hand side of (4.1) is estimated by a Taylor expansion,

$$\begin{aligned}
&\Big|\int_0^1 \Big[K\Big(\frac{y-\hat{g}(x)}{a_n}\Big) - K\Big(\frac{y-g(x)}{a_n}\Big)\Big]\,dx\Big| \\
\leq\ & \Big|\int_0^1 \frac{1}{a_n}k\Big(\frac{y-g(x)}{a_n}\Big)(\hat{g}(x)-g(x))\,dx\Big| \\
& + \Big|\int_0^1 \frac{1}{a_n^2}k'\Big(\frac{y-g(x)}{a_n}\Big)\,(\hat{g}(x)-g(x))^2\,dx\Big| + \sup_{u\in\mathbb{R}}|k''(u)|\frac{1}{a_n^3}||\hat{g}-g||_\infty^3 \\
\leq\ & C_1||\hat{g}-g||_\infty + C_2\frac{1}{a_n}||\hat{g}-g||_\infty^2 + C_3\frac{1}{a_n^3}||\hat{g}-g||_\infty^3
\end{aligned}$$

for some constants C_1, C_2, C_3, where the last line follows by a replacement of variables, $z = (y - g(x))/a_n$, in the integrals. By a change of the variable and integration by parts we obtain that the second term on the right hand side of (4.1) is bounded by

$$\begin{aligned}
&\Big|\int_0^{g^{-1}(y-a_n)} K\Big(\frac{y-g(x)}{a_n}\Big)\,dx + \int_{g^{-1}(y-a_n)}^{g^{-1}(y+a_n)} K\Big(\frac{y-g(x)}{a_n}\Big)\,dx - g^{-1}(y)\Big| \\
\leq\ & \Big|g^{-1}(y-a_n) - \int_{-1}^1 K(z)\frac{\partial}{\partial z}g^{-1}(y-a_nz)\,dz - g^{-1}(y)\Big| \\
=\ & \Big|g^{-1}(y-a_n) - K(z)g^{-1}(y-a_nz)\Big|_{z=-1}^{z=1} + \int_{-1}^1 k(z)g^{-1}(y-a_nz)\,dz - g^{-1}(y)\Big| \\
\leq\ & a_n^2\sup_t|(g^{-1})''(t)|\int k(z)z^2\,dz \ \leq C_4a_n^2
\end{aligned}$$

for some constant C_4. Collecting all bounds together the assertion follows. □

Proof of Theorem 3.1 (b). Denote $D_n = C(||\hat{g}-g||_\infty + \frac{1}{a_n}||\hat{g}-g||_\infty^2 + \frac{1}{a_n^3}||\hat{g}-g||_\infty^3 + a_n^2)$ such that $\sup_{y\in(g(0),g(1))} |\Psi(\hat{g})(y) - g^{-1}(y)| \le D_n$ from Proposition 4.1. Then from $g = \Phi(g^{-1}I_{[g(0),g(1)]})I_{[0,1]}$ it follows that

$$\begin{aligned}
||\hat{g}_{SI} - g||_\infty &= \sup_{z\in[0,1]} \Big| \int_{\hat{g}(0)}^{\hat{g}(1)} I\Big\{\Psi(\hat{g})(x) \le z\Big\}\, dx + \hat{g}(0) - \int_{g(0)}^{g(1)} I\Big\{g^{-1}(x) \le z\Big\}\, dx - g(0)\Big| \\
&\le 2|\hat{g}(0) - g(0)| + |\hat{g}(1) - g(1)| + r_n
\end{aligned}$$

where

$$r_n \le \sup_{z\in[0,1]} \int_{g(0)}^{g(1)} I\Big\{\Psi(\hat{g})(x) \le z < g^{-1}(x)\Big\}\, dx + \sup_{z\in[0,1]} \int_{g(0)}^{g(1)} I\Big\{g^{-1}(x) \le z < \Psi(\hat{g})(x)\Big\}\, dx.$$

Both summands are estimated in the very same way and by Proposition 4.1 the first one is bounded by

$$\sup_{z\in[0,1]} \int_{g(0)}^{g(1)} I\Big\{g^{-1}(x) - D_n \le z < g^{-1}(x)\Big\}\, dx \le \sup_{z\in[0,1]} |g(z+D_n) - g(z)| \le ||g'||_\infty D_n.$$

The assertion follows collecting all bounds together. □

4.3 Proof of Theorem 3.2 (a).

For the proof of Theorem 3.2 (a) we first show that the following Proposition is valid.

Proposition 4.2 *Let $g : [0,1] \to \mathbb{R}$ be a strictly increasing differentiable function such that the first derivative is bounded away from zero, and $\hat{g} : [0,1] \to \mathbb{R}$ an estimator for g. Then there exists a constant C such that*

$$\sup_{y\in(g(0),g(1))} |\tilde{\Phi}(\hat{g})(y) - g^{-1}(y)| \le C\Big(||\hat{g} - g||_\infty + \frac{1}{N}\Big).$$

Proof of Proposition 4.2. We consider the following decomposition,

$$\begin{aligned}
\tilde{\Phi}(g)(y) - g^{-1}(y) &= \frac{1}{N}\sum_{i=1}^N \Big[I\{\hat{g}(\frac{i}{N}) \le y\} - I\{g(\frac{i}{N}) \le y\}\Big] \\
&\quad + \sum_{i=1}^N \int_{\frac{i-1}{N}}^{\frac{i}{N}} \Big[I\{g(\frac{i}{N}) \le y\} - I\{g(x) \le y\}\, dx\Big] \\
&\quad + \int_0^1 I\{g(x) \le y\}\, dx - g^{-1}(y).
\end{aligned} \tag{4.2}$$

Because g is increasing, the last line vanishes. The absolute value of the second term on the right hand side of (4.2) can be bounded, for all $y \in (g(0), g(1))$, by

$$\begin{aligned}
\sum_{i=1}^N \int_{\frac{i-1}{N}}^{\frac{i}{N}} I\{g(x) \le y < g(\frac{i}{N})\}\, dx &\le \sum_{i=1}^N \int_{\frac{i-1}{N}}^{\frac{i}{N}} I\{g(\frac{i-1}{N}) \le y \le g(\frac{i}{N})\}\, dx \\
&\le \sum_{i=1}^N \frac{1}{N} I\{g^{-1}(y) \le \frac{i}{N} \le g^{-1}(y) + \frac{1}{N}\}\, dx \le \frac{2}{N}.
\end{aligned}$$

A bound for the absolute value of the first term on the right hand side of (4.2) is given by

$$\frac{1}{N}\sum_{i=1}^{N} I\{\hat{g}(\frac{i}{N}) \le y \le g(\frac{i}{N})\} + \frac{1}{N}\sum_{i=1}^{N} I\{g(\frac{i}{N}) \le y \le \hat{g}(\frac{i}{N})\}$$

and we only consider the second term in the following. It is bounded by

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N} I\{g(\frac{i}{N}) \le y \le g(\frac{i}{N}) + ||\hat{g}-g||_\infty\} \;&\le\; \frac{1}{N}\sum_{i=1}^{N} I\{g^{-1}(y-||\hat{g}-g||_\infty) \le \frac{i}{N} \le g^{-1}(y)\} \\ &\le\; 2\Big(g^{-1}(y) - g^{-1}(y-||\hat{g}-g||_\infty)\Big) \;\le\; 2||\frac{1}{g'}||_\infty ||\hat{g}-g||_\infty\end{aligned}$$

for all y such that $y - ||\hat{g}-g||_\infty \ge g(0)$. Otherwise we estimate

$$\begin{aligned}&\sup_{y\in[g(0),g(0)+||\hat{g}-g||_\infty]} \frac{1}{N}\sum_{i=1}^{N} I\{g(\frac{i}{N}) \le y \le g(\frac{i}{N}) + ||\hat{g}-g||_\infty\} \\ &\le\; \frac{1}{N}\sharp\{i \mid g(\frac{i}{N}) \le g(0) + ||\hat{g}-g||_\infty\} \\ &\le\; 2g^{-1}(g(0)+||\hat{g}-g||_\infty) \;\le\; 2||\frac{1}{g'}||_\infty ||\hat{g}-g||_\infty\end{aligned}$$

and the assertion of the Proposition follows. □

Proof of Theorem 3.2 (a). Theorem 3.2 (a) follows from Proposition 4.2 in the same way as Theorem 3.1 (b) is deduced from Proposition 4.1. □

4.4 Proof of Theorem 3.2 (b).

For the proof of Theorem 3.2 (b) we first show that the following Proposition is valid.

Proposition 4.3 *Let $g : [0,1] \to \mathbb{R}$ be a strictly increasing twice differentiable function with bounded second derivative such that the first derivative is bounded away from zero. Let $\hat{g} : [0,1] \to \mathbb{R}$ be an estimator for g. Let k denote a symmetric density function with compact support and two bounded derivatives and let $a_n = o(1)$ denote a sequence of positive bandwidths. Then there exists a constant C such that*

$$\begin{aligned}\sup_{y\in(g(0),g(1))} |\tilde{\Psi}(\hat{g})(y) - g^{-1}(y)| \;\le\; C\Big(&||\hat{g}-g||_\infty(1+\frac{1}{Na_n}) + \frac{1}{a_n}||\hat{g}-g||_\infty^2(1+\frac{1}{Na_n^2}) \\ &+ \frac{1}{a_n^3}||\hat{g}-g||_\infty^3 + \frac{1}{N} + \frac{1}{N^2a_n} + \frac{1}{N^3a_n^3} + a_n^2\Big)\end{aligned}$$

Proof of Proposition 4.3. We have

$$\begin{aligned}(4.3)\qquad |\tilde{\Psi}(\hat{g})(y) - \Psi(\hat{g})(y)| \;=\; \Big|&\frac{1}{N}\sum_{i=1}^{N}\Big[K\Big(\frac{y-\hat{g}(\frac{i}{N})}{a_n}\Big) - K\Big(\frac{y-g(\frac{i}{N})}{a_n}\Big)\Big] \\ &+ \sum_{i=1}^{N}\int_{\frac{i-1}{N}}^{\frac{i}{N}}\Big[K\Big(\frac{y-g(\frac{i}{N})}{a_n}\Big) - K\Big(\frac{y-g(x)}{a_n}\Big)\Big]\,dx\Big|\end{aligned}$$

and by a Taylor expansion the first term on the right hand side of (4.3) is bounded by

$$
\begin{aligned}
&\Big|\frac{1}{N}\sum_{i=1}^{N}\frac{1}{a_n}k\Big(\frac{y-g(\frac{i}{N})}{a_n}\Big)(g(\frac{i}{N})-\hat{g}(\frac{i}{N}))\Big| \\
&\quad+\Big|\frac{1}{N}\sum_{i=1}^{N}\frac{1}{a_n^2}k'\Big(\frac{y-g(\frac{i}{N})}{a_n}\Big)(g(\frac{i}{N})-\hat{g}(\frac{i}{N}))^2\Big|+\sup_{u\in\mathbb{R}}|k''(u)|\frac{1}{a_n^3}||\hat{g}-g||_\infty^3 \\
&\le ||\hat{g}-g||_\infty\int\frac{1}{a_n}k\Big(\frac{y-g(x)}{a_n}\Big)\,dx+\frac{1}{a_n}||\hat{g}-g||_\infty^2\int\frac{1}{a_n}\Big|k'\Big(\frac{y-g(x)}{a_n}\Big)\Big|\,dx \\
&\quad+||\hat{g}-g||_\infty\sum_{i=1}^{N}\int_{(i-1)/N}^{i/N}\frac{1}{a_n}\Big|k\Big(\frac{y-g(\frac{i}{N})}{a_n}\Big)-k\Big(\frac{y-g(x)}{a_n}\Big)\Big|\,dx \\
&\quad+\frac{1}{a_n^2}||\hat{g}-g||_\infty^2\sum_{i=1}^{N}\int_{(i-1)/N}^{i/N}\Big|k'\Big(\frac{y-g(\frac{i}{N})}{a_n}\Big)-k'\Big(\frac{y-g(x)}{a_n}\Big)\Big|\,dx \\
&\quad+\sup_{u\in\mathbb{R}}|k''(u)|\frac{1}{a_n^3}||\hat{g}-g||_\infty^3 \\
&\le C_1\Big(||\hat{g}-g||_\infty(1+\frac{1}{Na_n}+\frac{1}{N^2a_n^3})+\frac{1}{a_n}||\hat{g}-g||_\infty^2(1+\frac{1}{Na_n^2})+\frac{1}{a_n^3}||\hat{g}-g||_\infty^3\Big)
\end{aligned}
$$

for some constant C_1, where the last inequality follows by similar calculations as in the argumentation for the second term on the right hand side of (4.3). This one is bounded by

$$
\begin{aligned}
&\Big|\sum_{i=1}^{N}\int_{\frac{i-1}{N}}^{\frac{i}{N}}\frac{1}{a_n}k\Big(\frac{y-g(x)}{a_n}\Big)(g(\frac{i}{N})-g(x))\,dx\Big| \\
&\quad+\Big|\sum_{i=1}^{N}\int_{\frac{i-1}{N}}^{\frac{i}{N}}\frac{1}{a_n^2}k'\Big(\frac{y-g(x)}{a_n}\Big)(g(\frac{i}{N})-g(x))^2\,dx\Big| \\
&\quad+\frac{1}{a_n^3}\sup_{u\in\mathbb{R}}|k''(u)|\sum_{i=1}^{N}\int_{\frac{i-1}{N}}^{\frac{i}{N}}|g(\frac{i}{N})-g(x)|^3\,dx \\
&\le ||g'||_\infty\frac{1}{N}\int_0^1\frac{1}{a_n}k\Big(\frac{y-g(x)}{a_n}\Big)\,dx+(||g'||_\infty\frac{1}{N})^2\int_0^1\frac{1}{a_n^2}\Big|k'\Big(\frac{y-g(x)}{a_n}\Big)\Big|\,dx \\
&\quad+\sup_{u\in\mathbb{R}}|k''(u)|(||g'||_\infty\frac{1}{a_nN})^3 \\
&\le C_2\Big(\frac{1}{N}+\frac{1}{N^2a_n}+\frac{1}{N^3a_n^3}\Big)
\end{aligned}
$$

for some constant C_2 uniformly with respect to y, where the last line follows from a change of variable $z=(y-g(x))/a_n$ in the integrals and because g^{-1} is bounded. The assertion now follows by Proposition 4.1. □

Proof of Theorem 3.2 (b). Theorem 3.2 (b) follows from Proposition 4.3 in the same way as Theorem 3.1 (b) is deduced from Proposition 4.1. □

Acknowledgements. This work was done while the author visited the Australian National University in Canberra and she would like to thank the members of the Mathematical Science Institute, in particular Peter Hall, for their hospitality.

5 References

M. Akritas and I. Van Keilegom (2001). *Nonparametric estimation of the residual distribution.* Scand. J. Statist. 28, 549–567.

R. E. Barlow, D. J. Bartholomew, J. M. Bremmer and H. D. Brunk (1972). *Statistical Inference under order restrictions.* Wiley, New York.

M. Birke and H. Dette (2005). *A note on estimating a monotone regression by combining kernel and density estimates.* preprint, http://www.rub.de/mathematik3/preprint.htm

H. D. Brunk (1958). *On the estimation of parameters restricted by inequalities.* Annals of Mathematical Statistics 29, 437–454.

T. W. Chen (1981). *On the strong consistency of density estimation by orthogonal series methods.* Tamkang J. Math. 12, 265–268.

M-Y. Cheng, T. Gasser and P. Hall (1999). *Nonparametric density estimation under unimodality and monotonicity constraints.* Journal of Computational and Graphical Statistics 8, 1–21.

G. Collomb, S. Hassani, P. Sarda, P. Vieu (1985). *Convergence uniforme d'estimateurs de la fonction de hasard pour des observations dépendantes: méthodes du noyau et des k-points les plus proches. (French) [Strong uniform consistency of kernel and k-nearest neighbor estimates of the hazard rate from dependent observations].* C. R. Acad. Sci. Paris Sér. I Math. 301, 653–656.

S. Datta (1995). *A minimax optimal estimator for continuous monotone densities.* Journal of Statistical Planning and Inference 46, 183–193.

R. M. de Jong (2002). *A note on: "Convergence rates and asymptotic normality for series estimators" [J. Econometrics 79 (1997), no. 1, 147–168] by W. K. Newey: uniform convergence rates.* J. Econometrics 111, 1–9.

H. Dette, N. Neumeyer and K. Pilz (2005a). *A simple nonparametric estimator of a strictly increasing regression function.* forthcoming in Bernoulli. http://www.rub.de/mathematik3/preprint.htm

H. Dette, N. Neumeyer and K. Pilz (2005b). *A note on nonparametric estimation of the effective dose in quantal bioassay.* forthcoming in Journal of the American Statistical Association. http://www.rub.de/mathematik3/preprint.htm

H. Dette, K. Pilz (2004). *On the estimation of a monotone conditional variance in nonparametric regression.* technical report Ruhr-Universität Bochum. http://www.ruhr-uni-bochum.de/mathematik3/preprint.htm

L. P. Devroye, T. J. Wagner (1977). *The strong uniform consistency of nearest neighbor density estimates.* Ann. Statist. 5, 536–540.

L. P. Devroye, T. J. Wagner (1978). *The strong uniform consistency of kernel density estimates.* Multivariate analysis, V (Proc. Fifth Internat. Sympos., Univ. Pittsburgh, Pittsburgh, Pa., 1978), pp. 59–77, North-Holland, Amsterdam-New York, 1980.

U. Einmahl, D. M. Mason (2000). *An empirical process approach to the uniform consistency of kernel-type function estimators.* J. Theoretical Probab. 13, 1–37.

I. Gijbels (2003). *Monotone regression.* Discussion paper 0334, Institute de Statistique, Université Catholique de Louvain. http://www.stat.ucl.ac.be/ISpub/ISdp.html

U. Grenander (1956). *On the theory of mortality measurement II.* Skand. Aktuarietidskr. 39, 125–153.

P. Groeneboom (1985) *Estimating a monotone density.* Proceedings of the Berkeley conference in honor of Jerzy Neyman and Jack Kiefer, Vol. II (Berkeley, Calif., 1983), 539–555, Wadsworth Statistics/Probability Series.

P. Groeneboom and H. P. Lopuhaä (1993). *Isotonic estimators of monotone densities and distribution functions: basic facts.* Statistica Neerlandica 47, 175–183.

L. Györfi, H. Walk (1996). *On the strong universal consistency of a series type regression estimate.* Math. Methods Statist. 5, 332–342.

P. Hall and L-S. Huang (2001). *Nonparametric kernel regression subject to monotonicity constraints.* Annals of Statistics 29, 624–647.

P. Hall, L-S. Huang, J. A. Gifford and I. Gijbels (2001). *Nonparametric estimation of hazard rate under the constraint of monotonicity.* Journal of Computational and Graphical Statistics 10, 592–614.

J. Huang and J. A. Wellner (1995). *Estimation of a monotone density or monotone hazard under random censoring.* Scandinavian Journal of Statistics 22, 3–33.

Y. Huang and C.–H. Zhang (1994). *Estimating a monotone density from censored observations.* Annals of Statistics 22, 1256–1274.

Y. P. Mack (1983). *Rate of strong uniform convergence of k-NN density estimates.* J. Statist. Plann. Inference 8, 185–192.

Y. P. Mack, B. W. Silverman (1982). *Weak and strong uniform consistency of kernel regression estimates.* Z. Wahrsch. Verw. Gebiete 61, 405–415.

E. Mammen (1991). *Estimating a smooth monotone regression function.* Annals of Statistics 19, 724–740.

A. Massiani (2003). *Vitesse de convergence uniforme presque sûre de l'estimateur linéaire par méthode d'ondelettes. (French) [Rate of almost sure uniform convergence of the linear wavelet density estimator].* C. R. Math. Acad. Sci. Paris 337, 67–70.

E. Masry (1997). *Multivariate probability density estimation by wavelet methods: strong consistency and rates for stationary time series.* Stoch. Process. Appl. 67, 177–193.

H. Mukerjee (1988). *Monotone nonparametric regression.* Ann. Statist. 16, 741–750.

H. Mukerjee and J-L. Wang (1993). *Nonparametric maximum likelihood estimation of an increasing hazard rate for uncertain cause-of-death data.* Scandinavian Journal of Statistics 20, 17–33.

É. A. Nadaraya (1964). *On non–parametric estimates of density functions and regression curves.* J. Probab. Appl. 10, 186–190.

W. Newey (1997). *Convergence rates and asymptotic normality for series estimators.* J. Econometrics 79, 147–168.

W. Polonik (1995). *Density estimation under qualitative assumptions in higher dimensions.* J. Multivariate Anal. 55, 61–81.

W. Polonik (1998). *The silhouette, concentration functions and ML-density estimation under order restrictions.* Ann. Statist. 26, 1857–1877.

G. S. Qin, P. Cheng (1994). *Strong uniform consistency of k-nearest neighbor regression function estimators.* Sci. China Ser. A 37, 1032–1040.

J. O. Ramsay (1988). *Monotone regression splines in action (with discussion).* Statistical Science, 3, 425–441.

J. V. Ryff (1965). *Orbits of L^1-functions under doubly stochastic transformations.* Trans. Americ. Math. Soc. 117, 92-100.

J. V. Ryff (1970). *Measure preserving transformations and rearrangements.* J. Math. Anal. Appl. 31, 449-458.

B. W. Silverman (1978). *Weak and strong uniform consistency of the kernel estimate of a density and its derivatives.* Ann. Statist. 6, 177–184.

R. T. Smythe (1980). *Maxima of partial sums and a monotone regression estimator.* Ann. Probab. 8, 630–635.

W. Stute (1982). *A law of the logarithm for kernel density estimators.* Ann. Probab. 10, 414–422.

A. B. Tsybakov (1997). *On nonparametric estimation of density level sets.* Ann. Statist. 25, 948–969.

A. W. van der Vaart and M. J. van der Laan (2003). *Smooth estimation of a monotone density.* Statistics 37, 189–203.

G. S. Watson (1964). *Smooth Regression Analysis.* Sankhya A 26, 359–372.

F. T. Wright (1979) *A strong law for variables indexed by a partially ordered set with applications to isotone regression.* Ann. Probab. 7, 109–127.

L. G. Xue (2002). *Strong uniform convergence rates of wavelet estimates of regression function under complete and censored data.* Acta Math. Appl. Sin. 25, 430–438.

B. Zhang (1996). *A note on strong uniform consistency of kernel estimators of hazard functions under random censorship. Lifetime data: models in reliability and survival analysis.* (Cambridge, MA, 1994), 395–399, Kluwer Acad. Publ., Dordrecht, 1996.

S. L. Zhang, Q. Y. Sha, M. Y. Cheng (1999). *Uniform strong consistency of nonlinear wavelet estimates of regression functions.* Chinese J. Appl. Probab. Statist. 15, 375–380.