

Shape constrained kernel density estimation

Melanie Birke

Ruhr-Universität Bochum

Fakultät für Mathematik

44780 Bochum, Germany

e-mail: melanie.birke@ruhr-uni-bochum.de

March 27, 2008

Abstract

In this paper, a method for estimating monotone, convex and log-concave densities is proposed. The estimation procedure consists of an unconstrained kernel estimator which is modified in a second step with respect to the desired shape constraint by using monotone rearrangements. It is shown that the resulting estimate is a density itself and shares the asymptotic properties of the unconstrained estimate. A short simulation study shows the finite sample behavior.

Key words: Convexity, log-concavity, monotone rearrangements, monotonicity, nonparametric density estimation

Introduction

A typical setting for nonparametric density estimation is the following. Assume that the observations X_1, \dots, X_n are independent identically distributed on $A \subseteq \mathbb{R}$ with density f . Usually, this density is completely unknown and it can be estimated by several nonparametric estimators,

e.g. kernel density estimators. To this end let K be a kernel with support $\Lambda \subseteq \mathbb{R}$ and h the corresponding bandwidth. Then the kernel density estimator is given by

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right). \quad (1)$$

In some context the density f is unknown but there is information like monotonicity, convexity or log-concavity about its shape. In general, the unconstrained estimator defined in (1) will not necessarily fulfill those shape constraints but if there is such information, it is desirable to have estimators which meet these constraints. In the literature, there are several proposals of such estimators in density as well as regression estimation but most methods rely on maximum likelihood estimation or constrained least squares where the estimator is obtained by minimizing the L_2 -distance to the data over the set of all monotone, convex or log-concave density functions. A monotone maximum likelihood estimator is first discussed by Grenander (1956) and Groeneboom (1985). Van der Vaart and van der Laan (2003) propose a smooth monotone estimator based on least squares estimation where the data is first smoothed by an unconstrained kernel density estimator. Recently, Balabdaoui and Wellner (2007) propose a maximum likelihood estimator for a k -monotone density which includes for $k = 1$ the case of a monotone density discussed by Groeneboom (1985). There may occur problems in monotone maximum likelihood estimation in selection biased samples or at points of discontinuity. Methods for those cases are given in Woodroffe and Sun (1993), El Barmi and Nelson (1999), El Barmi and Nelson (2000) and Anevski and Hössjer (2002), respectively. Work on convex density estimation using least squares or maximum likelihood estimators has been done by Groeneboom, Jongbloed and Wellner (2001) and Dümbgen, Rufibach and Wellner (2007) but there has not much work been done on log-concave density estimation which has recently been considered by Dümbgen and Rufibach (2007).

It is the purpose of the present paper to define such shape constrained estimators by only using kernel methods. In a first step, the density is estimated by an unconstrained kernel estimate and

then, in a second step, it is modified with respect to the shape constraints by using monotone, convex or concave rearrangements which have been introduced in nonparametric regression by Dette, Neumeyer and Pilz (2006), Birke and Dette (2005) and Birke and Dette (2007). Those methods produce shape constrained estimators which have the same asymptotic distribution as the unconstrained estimator one starts with. For example, in the case of kernel density estimation the asymptotic distribution is a normal distribution which is appealing for practitioners.

The remaining article is organized as follows. In section 1 monotone, convex and concave rearrangements are shortly introduced. The application to unconstrained estimators is then straight forward and the resulting estimates are presented in section 2 together with some asymptotic properties. It turns out that the asymptotic result for a convex density estimate still holds if the true density is only two times continuously differentiable. This also improves the result in Birke and Dette (2007) where the estimated function is assumed to be three times continuously differentiable. Finally, in section 3 the finite sample behavior is investigated in a small simulation study.

1 Rearrangements

The basic tool for all rearrangements used in the sequel are monotone rearrangements which are first introduced in an analytical context in Hardy, Littlewood and Pólya (1952). Properties of monotone rearrangements are discussed in Bennett and Sharply (1988) or Lieb and Loss (1997). But those rearrangements also have a statistical motivation which is presented here. In the following this motivation is given and several rearrangements are defined. Properties are only discussed as far as they are needed for the discussion of the estimation methods in the following section.

If U is uniformly distributed on $A = [a_l, a_u]$ and $g : A \rightarrow \mathbb{R}$ is a strictly increasing and differentiable function then $g(U)$ has density

$$(g^{-1})'(u)I_{[g(0),g(1)]}(u)$$

and distribution function

$$g^{-1}(t) = \int_{-\infty}^t (g^{-1})'(u) I_{g(A)}(u) du + a_l = \int_{a_l}^{a_u} I\{g(v) \leq t\} dv + a_l, \quad t \in g(A).$$

Now, if g is not strictly increasing, the distribution function can still be calculated as the right integral in the above equation, that is

$$\phi(g)(t) = \int_{a_l}^{a_u} I\{g(v) \leq t\} dv + a_l, \quad t \in g([a_l, a_u]) \quad (2)$$

and the increasing rearrangement is given as the generalized inverse of the distribution function,

$$\phi(g)^{-1}(x) = \inf\{t \in \mathbb{R} | \phi(g)(t) \geq x\}. \quad (3)$$

A differentiable increasing rearrangement is obtained by smoothing the distribution function by a kernel smoother. To this end, K_d is a positive kernel of order 2 and h_d is a smoothing parameter called bandwidth. Then there is

$$\phi_{h_d}(g)(t) = \frac{1}{h_d} \int_{a_l}^{a_u} \int_{-\infty}^t K_d\left(\frac{g(v) - u}{h_d}\right) du dv + a_l \quad (4)$$

and the smoothed increasing rearrangement is given by

$$\phi_{h_d}(g)^{-1}(x) = \inf\{t \in \mathbb{R} | \phi_{h_d}(g)(t) \geq x\}. \quad (5)$$

In an analogous way decreasing rearrangements can be defined via the distribution functions

$$\varphi(g)(t) = \int_{a_l}^{a_u} I\{g(v) \geq t\} dv + a_l, \quad t \in g([a_l, a_u]) \quad (6)$$

and

$$\varphi_{h_d}(g)(t) = \frac{1}{h_d} \int_{a_l}^{a_u} \int_t^\infty K_d\left(\frac{g(v) - u}{h_d}\right) dudv + a_l \quad (7)$$

as

$$\varphi(g)^{-1}(x) = \sup\{t \in \mathbb{R} | \varphi(g)(t) \leq x\} \quad (8)$$

and

$$\varphi_{h_d}(g)^{-1}(x) = \sup\{t \in \mathbb{R} | \varphi_{h_d}(g)(t) \leq x\}, \quad (9)$$

respectively. One property of monotone rearrangements is that they have the same L_p -norm for $p \geq 1$ as the original function, that is

$$\int_A |\phi(g)^{-1}(x)|^p dx = \int_A |g(x)|^p dx$$

(see e.g. Bennett and Sharpley, 1988). Another property is that they reduce the L^p -estimation error for all $p \geq 1$ if applied to an unconstrained approximation (estimate) of the function, that means

$$\left(\int_A |\phi(\hat{g})^{-1}(x) - g(x)|^p \right)^{1/p} \leq \left(\int_A |\hat{g}(x) - g(x)|^p \right)^{1/p}$$

(see Chernozhukov, Fernández-Val and Galichon, 2007).

For defining convex estimators note that a differentiable function is convex if and only if its first derivative is increasing. Therefore one possibility to define a convex rearrangement is to apply the increasing rearrangement to the derivative of the function. Depending on the increasing rearrangement that is used there is the convex rearrangement

$$\rho(f)(x, a) = \int_a^x \phi(f')^{-1}(u) du + f(a) \quad (10)$$

or the smoothed convex rearrangement

$$\rho_{h_d}(f)(x, a) = \int_a^x \phi_{h_d}(f')^{-1}(u) du + f(a). \quad (11)$$

A disadvantage of the last two rearrangements is that they strongly depend on the lower integration bound a . A way out is to use the L_2 -optimal rearrangement

$$\rho^*(f)(x) = \frac{1}{a_u - a_l} \int_{a_l}^{a_u} \rho(f)(x, a) da \quad (12)$$

or

$$\rho_{h_d}^*(f)(x) = \frac{1}{a_u - a_l} \int_{a_l}^{a_u} \rho_{h_d}(f)(x, a) da. \quad (13)$$

Concave rearrangements of a function f can be defined analogously by using decreasing instead of increasing rearrangements for the first derivative. The corresponding concave rearrangements are denoted as $\varrho(f)(x, a)$, $\varrho_{h_d}(f)(x, a)$, $\varrho^*(f)(x)$ and $\varrho_{h_d}^*(f)(x)$.

2 Shape constrained estimators

In this section the different rearrangements are applied to an unconstrained kernel density estimator. This results in monotone, convex or log-concave density estimators and for each the asymptotic behavior is discussed. It turned out that for the asymptotics of the convex estimator it suffices to consider a two times continuously differentiable function. Although that is very helpful in density estimation (see also Remark 2) this result is not restricted to that special setting and therefore also improves the results for regression estimation considered in Birke and Dette (2007).

2.1 Monotone density estimators

Let f be defined on a compact interval $A = [a_l, a_u] \subsetneq \mathbb{R}$ and increasing. Then, an increasing estimator of f on A is given by

$$\hat{f}_I(x) = \phi_{h_d}(\hat{f})^{-1}(x) \quad (14)$$

where the smoothed increasing rearrangement in (5) is applied to the unconstrained density estimator defined in (1). If f is decreasing on A , a decreasing estimator is defined analogly by applying the decreasing rearrangement in (9),

$$\hat{f}_D(x) = \varphi_{h_d}(\hat{f})^{-1}(x).$$

Because of the similar structure of increasing and decreasing estimates the following results are only stated for the increasing case but they also hold true for decreasing estimators. Suppose, that \mathcal{D} is the set of all densities on A . Then for any $n \in \mathbb{N}$ and any bandwidth h it is clear that $\hat{f} \in \mathcal{D}$. Furthermore it can be shown, that for a sufficiently small bandwidth h_d , the increasing estimator defined in (14) is still a density. This result is stated in the following Lemma.

Lemma 1 *For any unconstrained uniformly continuous density estimator $\hat{f} \in \mathcal{D}$, $n \in \mathbb{N}$ and h fixed, the increasing estimator converges for $h_d \rightarrow 0$ pointwise to an element in \mathcal{D} , that is*

$$\hat{f}_I \xrightarrow{h_d \rightarrow 0} \phi^{-1}(\hat{f}) \in \mathcal{D}$$

where

$$\phi(\hat{f})(t) = \int_A I\{\hat{f}(v) \leq t\} dv + a_l$$

Proof. In a first step it is shown, that the inverse $\phi_{h_d}(\hat{f})(t)$ converges for $h_d \rightarrow 0$ to $\phi(\hat{f})(t)$. To simplify notation A is chosen as the interval $[0, 1]$. It is

$$\begin{aligned}
\phi_{h_d}(\hat{f})(t) &= \frac{1}{h_d} \int_0^1 \int_{-\infty}^t K_d\left(\frac{\hat{f}(v) - u}{h_d}\right) dudv \\
&= \frac{1}{h_d} \int_0^1 I\{\hat{f}(v) \leq t + h_d\} \int_{\hat{f}(v) - h_d}^{\hat{f}(v)} K_d\left(\frac{\hat{f}(v) - u}{h_d}\right) dudv \\
&= \int_0^1 I\{\hat{f}(v) \leq \hat{t} + h_d\} \int_{(\hat{f}(v) - t)/h_d} K_d(u) dudv \\
&= \int_0^1 I\{\hat{f}(v) \leq t + h_d\} I\{\hat{f}(v) \leq t - h_d\} dv \\
&\quad + \int_0^1 I\{t - h_d \leq \hat{f}(v) \leq t + h_d\} \int_{(\hat{f}(v) - t)/h_d}^1 K_d(u) dudv \\
&= \int_0^1 I\{\hat{f}(v) \leq \hat{m}(x) - h_d\} dv \\
&\quad + \int_0^1 I\{t - h_d \leq \hat{f}(v) \leq t + h_d\} \int_{(\hat{f}(v) - t)/h_d}^1 K_d(u) dudv
\end{aligned}$$

and hence

$$\begin{aligned}
|\phi_{h_d}(\hat{f})(t) - \phi(\hat{f})(t)| &= \left| \int_0^1 I\{t - h_d \leq \hat{f}(v) \leq t\} dv \right. \\
&\quad \left. - \int_0^1 I\{t - h_d \leq \hat{f}(v) \leq t + h_d\} \int_{(\hat{f}(v) - t)/h_d}^1 K_d(u) dudv \right| \\
&\leq \int_0^1 I\{t - h_d \leq \hat{f}(v) \leq t\} dv \\
&\quad + \int_0^1 I\{t - h_d \leq \hat{f}(v) \leq t + h_d\} dv = o(1)
\end{aligned}$$

using the uniform continuity of \hat{f} . Because of the continuity of the functional which maps a function onto its inverse in a fixed point x , this yields

$$\hat{f}_I(x) \xrightarrow{h_d \rightarrow 0} \phi^{-1}(\hat{f})(x).$$

It remains to show, that $\phi^{-1}(\hat{f})$ is a density. But that is clear from the theory of monotone rearrangements. Since $\hat{f} \geq 0$ and ϕ^{-1} only rearranges the function values in an increasing order, also $\phi^{-1}(\hat{f}) \geq 0$. Monotone rearrangements preserve the L_p norm of the original function (see Bennett and Sharpley (1988) for details). Therefore

$$\int_A \phi(\hat{f})(x)dx = \int_A |\phi(\hat{f})(x)|dx = \int_A \hat{f}(x)dx = 1$$

which proves Lemma 1.

From an asymptotic point of view, if the true density is increasing and two times continuously differentiable, the unconstrained and the isotone density estimator show the same behavior. That means, \hat{f}_I is pointwise consistent if \hat{f} is pointwise consistent and both have the same asymptotic normal distribution. The consistency of \hat{f}_I follows from results stated in Neumeyer (2007). For the sake of simplicity the asymptotic considerations are stated for $A = [0, 1]$ but they hold true for any compact interval A

Theorem 1 *Let the density $f \in \mathcal{C}^2([0, 1])$ be strictly increasing and K be a kernel of order 2. If the bandwidths fulfill*

$$h_d \rightarrow 0, h \rightarrow 0, nh^5 = O(1), \frac{h_d}{h} \rightarrow 0 \text{ and } \frac{\log h^{-1}}{nhh_d^3} \rightarrow 0$$

for $n \rightarrow \infty$, then for any $x \in (0, 1)$ with $f'(x) > 0$

$$\sqrt{nh} \left(\hat{f}_I(x) - f(x) - h^2 b_n^{(2)}(x) \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, s^2(x))$$

where

$$b_n^{(2)}(x) = \frac{f''(x)}{2} \int_{-1}^1 u^2 K(u) du,$$

$$s^2(x) = f(x) \int_{-1}^1 K^2(u) du$$

are the essential bias and asymptotic variance of the unconstrained density estimator \hat{f} .

The proof of this theorem is similar to that for an increasing regression function and therefore omitted. Details for the regression case can be found in Dette, Neumeyer and Pilz (2006).

Remark 1 (Densities with unbounded support) For the sake of simplicity the monotone rearrangements above were only introduced for densities with bounded support. Dette and Volgushev (2007) define monotone rearrangements for densities with unbounded support. In the present case of monotone densities it is interesting to consider supports of the type $A = (-\infty, a_u]$ (increasing case) and $A = [a_l, \infty)$ (decreasing case). To this end let $G : A \rightarrow [0, 1]$ be a strictly increasing known function. Then

$$f \circ G^{-1} : [0, 1] \rightarrow f(A)$$

has again bounded support and the distribution function and monotone rearrangement can be calculated as usual. The increasing rearrangements of f are then obtained by

$$\phi(f \circ G^{-1})^{-1}(G(x)) \text{ and } \phi_{h_d}(f \circ G^{-1})^{-1}(G(x))$$

while the decreasing rearrangements are

$$\varphi(f \circ G^{-1})^{-1}(G(x)) \text{ and } \varphi_{h_d}(f \circ G^{-1})^{-1}(G(x)).$$

Substituting f by an estimator \hat{f} then yields the increasing and decreasing estimates, respectively. In finite samples the rearrangement depends on the choice of G and a proper choice is discussed in Dette and Volgushev (2007) but it is also shown there that the choice of G does not influence the asymptotic behaviour of the rearrangement.

2.2 Convex density estimators

Let again $A = [a_l, a_u] \subsetneq \mathbb{R}$. If the convex rearrangement defined in (11) is applied to the derivative of an unconstrained density estimator, we get a convex estimator of the density. With

$$\hat{f}'(x) = \frac{1}{nh^2} \sum_{i=1}^n K' \left(\frac{x - X_i}{h} \right),$$

for any fixed $x \in A$ there is

$$\hat{f}_C(x, a) = \rho_{h_d}(\hat{f})(x, a) = \int_a^x \phi_{h_d}(\hat{f}')^{-1}(z) dz + \hat{f}(a) \quad (15)$$

$$(16)$$

a convex estimator of f . Applying the L_2 -optimal convex rearrangement (13) yields

$$\hat{f}_C^*(x) = \rho_{h_d}^*(\hat{f})(x) = \frac{1}{a_u - a_l} \int_A \rho_{h_d}(\hat{f})(x, a) da \quad (17)$$

as convex estimator of f independent of the choice of a . If f is concave, an application of the corresponding concave rearrangements results in the concave estimators

$$\hat{f}_{Conc}(x, a) = \varrho_{h_d}(\hat{f})(x, a) = \int_a^x \varphi_{h_d}(\hat{f}')^{-1}(z) dz + \hat{f}(a) \quad (18)$$

$$(19)$$

and

$$\hat{f}_{Conc}^*(x) = \varrho_{h_d}^*(\hat{f})(x) = \frac{1}{a_u - a_l} \int_A \varrho_{h_d}(\hat{f})(x, a) da. \quad (20)$$

The following results are only stated for the convex case but also hold true for concave estimators.

Lemma 2 *For any unconstrained density estimator $\hat{f} \in \mathcal{D}$, $n \in \mathbb{N}$ and h fixed, the increasing*

estimator converges for $h_d \rightarrow 0$ to an element in \mathcal{D} , that is

$$\hat{f}_C^* \xrightarrow{h_d \rightarrow 0} \rho^*(\hat{f}) \in \mathcal{D}$$

where

$$\rho^*(\hat{f})(x) = \frac{1}{a_u - a_l} \int_A \left(\int_a^x \phi(\hat{f}')(v)dv + \hat{f}(a) \right) da.$$

The integration to 1 only holds for the L_2 -optimal convex or concave estimators. It is not true for any convex or concave estimator with arbitrary lower bound $a \in A$.

Proof of Lemma 2. To simplify the notation, the interval A is again chosen as $[0, 1]$. If Φ denotes the primitive of $\phi(\hat{f})^{-1}$, then

$$\begin{aligned} \int_0^1 \rho^*(\hat{f})(x)dx &= \int_0^1 \left(\int_0^1 \left(\int_a^x \phi(\hat{f}')(v)dv + \hat{f}(a) \right) da \right) dx \\ &= \int_0^1 \int_0^1 \left(\Phi(x) - \Phi(a) + \hat{f}(a) \right) dadx \\ &= \int_0^1 \Phi(x)dx - \int_0^1 \Phi(a)da + \int_0^1 \hat{f}(a)da = 1 \end{aligned}$$

The convex estimator also has the same asymptotic behavior as the unconstrained one.

Theorem 2 *Let the density $f \in \mathcal{C}^p([0, 1])$ be strictly convex and K be a kernel of order $p \geq 2$ which is at least two times continuously differentiable. If the bandwidths fulfill*

$$h_d \rightarrow 0, h \rightarrow 0, nh^{5+\delta} = O(1), \frac{h_d}{h^{1+\delta/4}} \rightarrow 0 \text{ and } \frac{(\log h^{-1})^{1/2}}{nh^3 h_d^2} \rightarrow 0 \quad (21)$$

for $n \rightarrow \infty$, then for any $x \in (0, 1)$ with $f''(x) > 0$ and any $a \in (0, 1)$

$$\hat{f}_C(x, a) - f(x) = \hat{f}(x) - f(x) + o_p\left(\frac{1}{\sqrt{nh}}\right)$$

This also yields that the convex estimator is consistent for all $x \in (0, 1)$. The proof of Theorem 2 is similar to that for convex estimators of regression functions in Birke and Dette (2007) and to that of Lemma 3 in Section 2.3 and therefore omitted.

Corollary 1 *If the assumptions of Theorem 2 hold and f is $p \geq 2$ times continuously differentiable, then*

$$\sqrt{nh} \left(\hat{f}_C(x, a) - f(x) - h^p b_n^{(p)}(x) \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, s^2(x)) \quad (22)$$

for any $x, a \in (0, 1)$, where the asymptotic bias and variance are given by

$$b_n^{(p)}(x) = \frac{f^{(p)}(x)}{p!} \int_{-1}^1 u^p K(u) du,$$

$$s^2(x) = f(x) \int_{-1}^1 K^2(u) du$$

respectively.

Remark 2 Theorem 2 and Corollary 1 show that we obtain asymptotic normality of the convex density estimate if f is two times continuously differentiable. But in this case the bandwidth has to be chosen slightly smaller than the optimal bandwidth which is proportional to $n^{-1/5}$. The consequence is, that for $p = 2$ and $0 < \delta < 2$ the bias of the estimates is not of order $O(1/\sqrt{nh})$. The situation changes if f is known to be three times continuously differentiable. Then the optimal bandwidth is proportional to $n^{-1/7}$ which corresponds to $nh^{5+\delta} = O(1)$ with $\delta = 2$ and the convex estimator has the optimal convergence rate $n^{3/7}$. Nevertheless, assuming a three times continuously differentiable density evokes some problems in the choice of the kernel. Like in the regression framework it makes sense to choose a kernel K of order 3 if the density is three times continuously differentiable. But such a kernel is no longer symmetric or positive. Even symmetric kernels of order $p \geq 4$ are no longer positive. This has the effect, that, in finite samples, the density estimator with a higher order kernel is not necessarily positive and therefore no density. Therefore several other bias correction techniques have been considered in literature.

For example, Jones, Signorini and Hjort (1999) consider a modified density estimator which has a bias of order $O(h^4)$ as long as the true density is four times continuously differentiable while the kernel is of order 2 and therefore positive. Ruppert and Cline (1994) use transformation-kernel density estimators to reduce the bias. This results in a bias of order $O(h^3)$ and $h = O(n^{-1/7})$ if the density is three times continuously differentiable. Further bias reduction techniques are given in Abramson (1982), Silverman (1986), Gajek (1986), Hall and Marron (1988) or Hall (1990). If one of those unconstrained estimates is used for convex estimation, the convex density estimator inherits the bias behavior of that estimator.

2.3 Log-concave density estimators

A strictly positive density f is said to be log-concave if $\log f$ is concave. Therefore, to obtain a log-concave density estimator of f , the concave rearrangement defined in section 1 is applied to $\log \hat{f}$ and in a second step transformed back. Log-concave densities have to be strictly positive on their support and, if the procedure described above shall work, so has to be the estimate \hat{f} . To this end, let K be a positive and symmetric kernel with support \mathbb{R} and h the corresponding bandwidth. Then the kernel density estimator

$$\frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)$$

is strictly positive on the set A . A log-concave density estimator can then be defined as

$$\hat{f}_{lc}(x, a) = \exp(\varrho_{h_d}(\log \hat{f})(x, a))$$

and the L_2 -optimal one as

$$\hat{f}_{lc}(x) = \exp(\varrho_{h_d}(\log \hat{f})(x, a^{lc}))$$

with

$$a^{lc} = \arg \min_{a \in A} \int_A (\hat{f}_{lc}(x, a) - \hat{f}(x))^2 dx.$$

Another possibility is to use the L_2 -optimal concave rearrangement for $\log \hat{f}$, that is

$$\hat{f}_{lc}(x) = \exp(\varrho_{h_d}^*(\log \hat{f})(x)).$$

In practice, the two ways to choose a produce only negligibly differences in the resulting estimators.

The following result holds for any choice of a .

Theorem 3 *Let the density f be log-concave and $p \geq 2$ times continuously differentiable on $A \subseteq \mathbb{R}$. If the bandwidths fulfill*

$$h_d \rightarrow 0, h \rightarrow 0, nh^{5+\delta} = O(1), \frac{h_d}{h^{1+\delta/4}} \rightarrow 0 \text{ and } \frac{(\log h^{-1})^{1/2}}{nh^3 h_d^2} \rightarrow 0, \quad (23)$$

then for any fixed $x \in A$ with $f''(x)f(x) - f'^2(x) > 0$ and any $a \in A^\circ$

$$\sqrt{nh}(\hat{f}_{lc}(x) - f(x) - h^p b_n^{(p)}(x)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, s^2(x))$$

with

$$b_n^{(p)}(x) = \frac{f^{(p)}(x)}{p!} \int_0^1 u^p K(u) du$$

$$s^2(x) = f(x) \int_{-1}^1 K^2(u) du.$$

This statement follows from the next Lemma by applying the Delta method.

Lemma 3 *Under the assumptions of Theorem 3 we have for any $a \in A^\circ$*

$$\sqrt{nh} \left(\varrho_{h_d}(\log \hat{f})(x, a) - \log f(x) - h^p \frac{b^{(p)}(x)}{f(x)} \right) \xrightarrow{\mathcal{D}} \mathcal{N} \left(0, \frac{s^2(x)}{f^2(x)} \right)$$

Proof of Lemma 3. To simplify the notation set $\log f = g$ and $\log \hat{f} = \hat{g}$ and let $A = [0, 1]$. For considering the asymptotic theory of $\varrho_{h_d}(\hat{g})(x)$ a representation of the decreasing rearrangement of $(\log \hat{f})' = \hat{g}'$ is necessary. The mean value theorem yields for $\xi_n(t)$ with $|\xi_n(t) - g'(t)| \leq |\hat{g}'(t) - g'(t)|$.

$$\begin{aligned}\varrho_{h_d}(\hat{g})(x, a) - g(x) &= - \int_a^x \frac{\varphi_{h_d}(\hat{g}')(\hat{g}'(t)) - g'(g'(t))}{(\varphi_{h_d}(\hat{g}'))'(\xi_n(t))} dt \\ &= A_n(x)(1 + o_P(1)) + \hat{g}(a) - g(a)\end{aligned}$$

with

$$A_n(x) = - \int_a^x \frac{\varphi_{h_d}(\hat{g}') - g'}{(g'^{-1})'}(g'(t)) dt.$$

This decomposition is true because

$$\begin{aligned}|\hat{g}'(z) - g'(z)| &= \left| \frac{\hat{f}'(z)}{\hat{f}(z)} - \frac{f'(z)}{f(z)} \right| \\ &\leq \frac{|\hat{f}'(z) - f'(z)|f(z) + |\hat{f}(z) - f(z)||f'(z)|}{\hat{f}(z)f(z)} = O_P\left(\frac{\log h^{-1}}{nh^3}\right)^{1/2},\end{aligned}$$

so that with a similar calculation as in Birke and Dette (2005).

$$\begin{aligned}\sup |\varphi_{h_d}(\hat{g}')(t) - g'^{-1}(t)| &= O_P\left(\frac{\log h^{-1}}{nh^3}\right)^{1/2} + O(h_d) = o_P(1) \\ \sup |\varphi_{h_d}(\hat{g}')'(t) - (g'^{-1})'(t)| &= O_P\left(\frac{\log h^{-1}}{nh^3 h_d^2}\right)^{1/2} + o(1) = o_P(1)\end{aligned}$$

which finally yields that $\xi_n(t)$ converges uniformly to $g'(t)$ and hence, that $\varphi_{h_d}(\hat{g}')'(\xi_n(t))$ converges uniformly to $(g'^{-1})'(g'(t))$ which justifies the decomposition above. The leading term $A_n(x)$ can

be decomposed into a stochastic and a deterministic part.

$$A_n(x) = A_{n,1}(x) + A_{n,2}(x)$$

with

$$\begin{aligned} A_{n,1}(x) &= - \int_a^x \frac{\varphi_{h_d}(\hat{g}') - \varphi_{h_d}(g')}{(g'^{-1})'} \circ g'(t) dt \\ A_{n,2}(x) &= - \int_a^x \frac{\varphi_{h_d}(g') - g'}{(g'^{-1})'} \circ g'(t) dt. \end{aligned}$$

The definition of $\hat{\varphi}_{h_d}$ and a Taylor expansion yields for $A_{n,1}$

$$\begin{aligned} A_{n,1}(x) &= - \frac{1}{h_d} \int_{g'(x)}^{g'(a)} \int_0^1 K_d\left(\frac{g'(v) - t}{h_d}\right) (\hat{g}'(v) - g'(v)) dv dt \\ &\quad - \frac{1}{2h_d^2} \int_{g'(x)}^{g'(a)} \int_0^1 K'_d\left(\frac{g'(v) - t}{h_d}\right) (\hat{g}'(v) - g'(v))^2 dv dt (1 + o_P(1)) \\ &= \Delta_n^{(1)}(x) + \frac{1}{2} \Delta_n^{(2)}(x) (1 + o_P(1)) \end{aligned}$$

where the notations $\Delta_n^{(1)}$ and $\Delta_n^{(2)}$ become clear from the previous equality. The quantity $\Delta_n^{(1)}$ can be decomposed into

$$\Delta_n^{(1)}(x) = \Delta_n^{(1.1)}(x) + \Delta_n^{(1.2)}(x) + \Delta_n^{(1.3)}(x) \tag{24}$$

with

$$\begin{aligned} \Delta_n^{(1.1)}(x) &= \int_{g'^{-1}(g'(a)-h_d)}^{g'^{-1}(g'(x)+h_d)} \int_{-1}^1 K_d(t) dt (\hat{g}(v) - g(v))' dv \\ \Delta_n^{(1.2)}(x) &= \int_0^{g'^{-1}(g'(a)-h_d)} \int_{\frac{g'(v)-g'(a)}{h_d}}^1 K_d(t) dt (\hat{g}(v) - g(v))' dv \\ \Delta_n^{(1.3)}(x) &= \int_{g'^{-1}(g'(x)+h_d)}^1 \int_{-1}^{\frac{g'(v)-g'(x)}{h_d}} K_d(t) dt (\hat{g}(v) - g(v))' dv. \end{aligned}$$

A further Taylor expansion yields for the first term

$$\begin{aligned}\Delta_n^{(1.1)}(x) &= \int_{g'^{-1}(g'(a)-h_d)}^{g'^{-1}(g'(x)+h_d)} (\hat{g}(v) - g(v))' dv \\ &= (\hat{g} - g)(g'^{-1}(g'(x) + h_d)) - (\hat{g} - g)(g'^{-1}(g'(a) - h_d))\end{aligned}\quad (25)$$

with $|\zeta_z - g'(z)| \leq h_d$ for $z = x, a$. After substituting $v = g'^{-1}(g'(a) + h_d w)$ and $v = g'^{-1}(g'(x) + h_d w)$, respectively and again using a Taylor expansion, the two remaining terms are

$$\begin{aligned}\Delta_n^{(1.2)}(x) &= \int_0^{g'^{-1}(g'(a)-h_d)} \int_{\frac{g'(v)-g'(a)}{h_d}}^1 K_d(t) dt (\hat{g} - g)'(v) dv \\ &= h_d \int_{-1}^1 K_d(t) \int_{-1}^w (g'^{-1})'(g'(a) + h_d v) (\hat{g} - g)'(g'^{-1}(g'(a) + h_d v)) dv dt \\ &= \int_{-1}^1 K_d(t) [(\hat{g} - g)(g'^{-1}(g'(a) + h_d t)) - (\hat{g} - g)(g'^{-1}(g'(a) - h_d))] dt \\ &= -(\hat{g} - g)(a) - h_d \int_{-1}^1 t K_d(t) (g'^{-1})'(\zeta_{a,t}) (\hat{g} - g)'(g'^{-1}(\zeta_{a,t})) dt - (\hat{g} - g)(g'^{-1}(g'(a) - h_d))\end{aligned}\quad (26)$$

and in the same way

$$\Delta_n^{(1.3)}(x) = \int_{g'^{-1}(g'(x)+h_d)}^1 \int_{-1}^{\frac{g'(v)-g'(x)}{h_d}} K_d(t) dt (\hat{g} - g)'(v) dv \quad (27)$$

$$= h_d \int_{-1}^1 t K_d(t) (g'^{-1})'(\zeta_{x,t}) (\hat{g} - g)'(g'^{-1}(\zeta_{x,t})) dt - h_d (g'^{-1})'(\zeta_x) (\hat{g} - g)'(g'^{-1}(\zeta_x)) \quad (28)$$

with $|\zeta_{z,t} - g'(z)| \leq h_d |t| \leq h_d$ for $z = a, x$. Bearing in mind that

$$\sup_{u \in J} |\hat{g}'(u) - g'(u)| = O_P(1) \sup_{u \in J} |\hat{f}'(u) - f'(u)| = O_P\left(\frac{\log h^{-1}}{nh^3}\right)^{1/2} \quad (29)$$

(see Silverman, 1978), equations (24) - (27) combine to

$$\Delta_n^{(1)}(x) = \hat{g}(x) - g(x) - \hat{g}(a) - g(a) \quad (30)$$

$$\begin{aligned}
& +h_d \int_{-1}^1 tK_d(t)(g'^{-1})'(\zeta_{x,t})(\hat{g} - g)'(g'^{-1}(\zeta_{x,t}))dt \\
& -h_d \int_{-1}^1 tK_d(t)(g'^{-1})'(\zeta_{a,t})(\hat{g} - g)'(g'^{-1}(\zeta_{a,t}))dt \\
& = \hat{g}(x) - g(x) - \hat{g}(a) - g(a) + O_P\left(\frac{h_d^2 \log h^{-1}}{nh^3}\right)^{1/2} \\
& = \hat{g}(x) - g(x) - \hat{g}(a) - g(a) + o_P\left(\frac{1}{\sqrt{nh}}\right).
\end{aligned}$$

The second term is negligible which can be seen by the following estimation

$$\begin{aligned}
\Delta_n^{(2)}(x) & = - \int_{-1}^1 K_d(v)(g'^{-1})'(g'(x) + h_d v)(\hat{g}' - g')^2(g'^{-1}(g'(x) + h_d v))dv \\
& \quad + \int_{-1}^1 K_d(v)(g'^{-1})'(g'(a) + h_d v)(\hat{g}' - g')^2(g'^{-1}(g'(a) + h_d v))dv.
\end{aligned}$$

The uniform convergence rates in (29) and conditions (23) now provide that

$$\Delta_n^{(2)}(x) = O_P\left(\frac{\log h^{-1}}{nh^3}\right) = o_P\left(\frac{1}{\sqrt{nh}}\right).$$

The deterministic part can be similarly calculated as in Dette, Neumeyer and Pilz (2006) where now it has to be regarded that g' is strictly decreasing. If G denotes the primitive of g'^{-1} , integration by parts and a Taylor expansion yield

$$\begin{aligned}
A_{n,2}(x) & = \int_{g'(x)}^{g'(a)} \left\{ \frac{1}{h_d} \int_0^1 \int_t^\infty K_d\left(\frac{g'(v) - u}{h_d}\right) dudv - g'^{-1}(t) \right\} dt \\
& = \frac{1}{h_d} \int_{g'(x)}^{g'(a)} \left\{ \int_0^{g'^{-1}(t-h_d)} \int_t^{g'(v)+h_d} K_d\left(\frac{g'(v) - u}{h_d}\right) dudv - g'^{-1}(t) \right\} dt \\
& = \int_{g'(x)}^{g'(a)} \left\{ g'^{-1}(t + h_d) + \int_{g'^{-1}(t+h_d)}^{g'^{-1}(t-h_d)} \int_{-1}^{(g'(v)-t)/h_d} K_d(u) dudv - g'^{-1}(t) \right\} dt \\
& = \int_{g'(x)}^{g'(a)} \left\{ g'^{-1}(t + h_d) - h_d \int_{-1}^1 (g'^{-1})'(t + h_d v) \int_{-1}^v K_d(u) dudv - g'^{-1}(t) \right\} dt \\
& = \int_{g'(x)}^{g'(a)} \left\{ g'^{-1}(t + h_d) - \left(g'^{-1}(t + h_d v) \int_{-1}^v K_d(u) du \right) \Big|_{-1}^1 + \int_{-1}^1 g'^{-1}(t + h_d v) K_d(v) dv - g'^{-1}(t) \right\} dt
\end{aligned}$$

$$\begin{aligned}
&= \int_{g'(x)}^{g'(a)} \left\{ \int_{-1}^1 g'^{-1}(t + h_d v) K_d(v) dv - g'^{-1}(t) \right\} dt \\
&= \int_{-1}^1 K_d(v) \left\{ G(g'(a + h_d v)) - G(g'(a)) - (G(g'(x) + h_d v) - G(g'(x))) \right\} dv \\
&= h_d^2 \kappa_2(K_d) \left\{ \frac{1}{g''(a)} - \frac{1}{g''(x)} \right\} + o(h_d^2) = o\left(\frac{1}{\sqrt{nh}}\right).
\end{aligned}$$

Therefore it is shown that

$$\varrho_{h_d}(\log \hat{f})(x, a) - \log f(x) = \log \hat{f}(x) - \log f(x) + o_P\left(\frac{1}{\sqrt{nh}}\right).$$

It is well known that

$$\sqrt{nh}(\hat{f}(x) - f(x) - b_n^{(p)}(x)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, s^2(x))$$

and hence an application of the delta method yields the assertion of Lemma 3. \square

3 Simulations

In this section the performance of the antitone, convex and log-concave estimators are demonstrated by first showing some typical estimates and after that comparing their simulated mean squared error (MSE) and integrated mean squared error (IMSE) to those of the unconstrained estimator. The simulations are performed for the densities

$$\begin{aligned}
f_1(x) &= 2 \exp(-2x) I_{[0, \infty)(x)} \\
f_2(x) &= 12 \left(x - \frac{1}{2}\right)^2 I_{[0, 1](x)} \\
f_3(x) &= \frac{15}{16} (1 - x^2)^2 I_{(-1, 1)(x)} \\
f_4(x) &= \frac{1}{2\sqrt{2\pi}} \exp\left(-\frac{x^2}{8}\right).
\end{aligned}$$

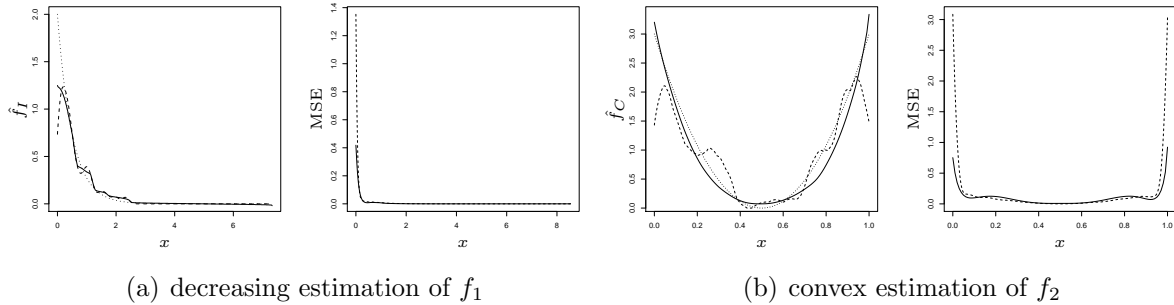
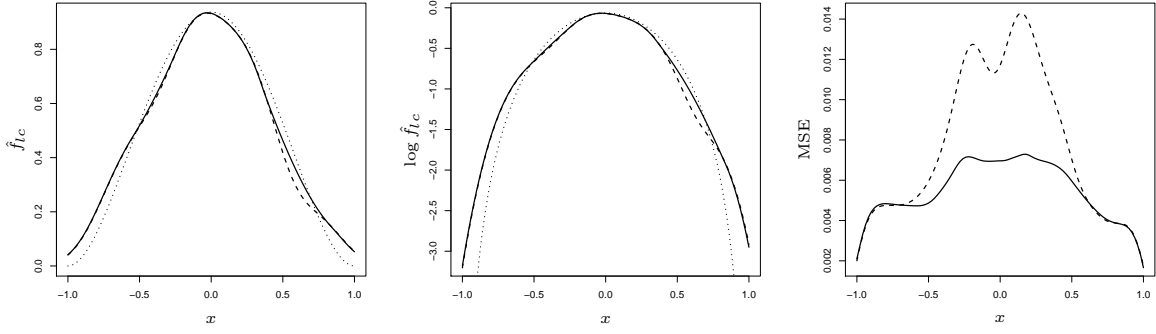
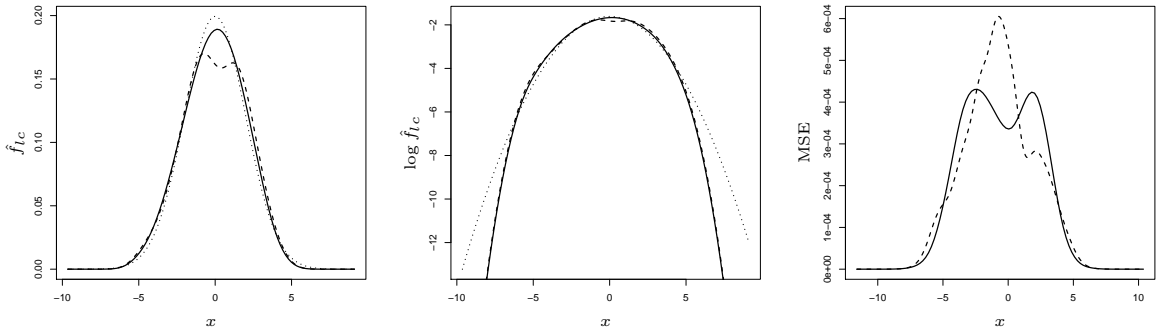


Figure 1: (a) Left part: Typical decreasing (solid line) and unconstrained (dashed line) estimates of the true density f_1 (dotted line), right part: MSE of the decreasing (solid line) and the unconstrained (dashed line) estimates. (b) Left part: Typical convex (solid line) and unconstrained (dashed line) estimates of the true density f_2 (dotted line), right part: MSE of the convex (solid line) and the unconstrained (dashed line) estimates

To this end, $n = 100$ observations are simulated according to the densities $f_1 - f_4$, respectively. From this data the antitone, convex and log-concave estimators are calculated. The unconstrained estimates of f_1 , f_3 and f_4 are based on gaussian kernels while for f_2 the Epanechnikov kernel is used. The bandwidth h is chosen by the plug-in method `dpik` implemented in the `KernSmooth` package in R while h_d is chosen as h^2 . Figure 1(a) presents a typical decreasing estimator (solid line) of f_1 compared to the underlying unconstrained one (dashed line) and the true density f_1 (dotted line) and the corresponding mean squared errors of both estimates. Figure 1(b) shows the same for the convex estimate of f_2 . A first inspection of the representative estimators in the left pictures of (a) and (b) shows that the restricted estimators correct the shape by preserving the quality of the unconstrained estimator. The unconstrained estimators show in particular some problems in estimating f_1 or f_2 correctly near the boundary. This is quite well corrected by the constrained estimators without using boundary corrected kernels. After that first impression the estimators are compared by using the MSE criterion. The pointwise mean squared errors of the constrained and unconstrained estimates in the right parts of (a) and (b) are calculated from 200 simulation runs. Both, the unconstrained as well as the constrained estimators behave very similar in the interior of the domain but there are clear advantages for the constrained estimators at



(a) log-concave estimation of f_3



(b) log-concave estimation of f_4

Figure 2: Left: Typical unconstrained (dashed line) and log-concave (solid line) estimates with the true density (dotted line); middle: logarithm of the unconstrained (dashed line) and the log-concave (solid line) estimates and of the density (dotted line); right: MSE of the unconstrained (dashed line) and the log-concave (solid line) estimates.

the boundary. Seen over the whole interval, the integrated mean squared error is an appropriate measure for the quality of an estimator. The IMSE of the unconstrained estimator of f_1 is 0.01217 which is larger than the IMSE 0.00589 of the decreasing estimate. This, to some extent, confirms the theoretical results of Chernozhukov, Fernández-Val and Galichon (2007) for monotone function estimators and indicates that for f_1 and a fixed sample size the decreasing estimator performs better than the unconstrained one. For f_2 the IMSE of the unconstrained estimator is 0.16429 and that of the convex estimator gives 0.10142. This yields an advantage of the convex estimator over the unconstrained one.

Figure 2 shows typical log-concave estimates (solid line) with the underlying unconstrained estimates (dashed line) and the true density (dotted line) in the left part. Figure 2(a) corresponds to f_3 and Figure 2(b) to f_4 . In the middle there are the log transforms of \hat{f}_{lc} (solid line), \hat{f} (dashed line) and $f_i, i = 3, 4$ (dotted line). Based on the left and middle part of Figure 2 it can be observed that the log-concave transform corrects slight non-log-concavities in the unconstrained estimate while both behave very similar where log-concavity is already fulfilled. Finally, the right part of Figure 2 shows the MSE of the unconstrained (dashed line) and the log-concave (solid line) density estimate. In (a) there are clear advantages for the log-concave estimate over nearly the whole interval. This is also confirmed by the IMSE whose value is 0.00771 for the unconstrained and 0.00548 for the log-concave estimate. In the right part of Figure 2(b), that is for the normal density f_4 , there cannot be observed such clear advantages over the whole interval. Looking at the typical estimates in the left and middle part it seems as if the log-concave estimate better estimates the peak of the density. This first impression is confirmed by the simulated MSE which is considerable smaller for the log-concave estimate around the maximum of the true density. But there are also regions where the unconstrained estimate has a smaller MSE. The IMSE of both estimates is very similar, that is 0.00016 for the unconstrained and 0.00015 for the log-concave one. This means that, seen over the whole interval both estimates are qualitatively comparable but if the peak should be modeled correctly, the log-concave estimate should be preferred.

Acknowledgements. The author would like to thank Holger Dette, Lutz Dümbgen, Nicolai Bissantz and Stanislav Volgushev for helpful discussions. The work of the author was supported by the Sonderforschungsbereich 475, Komplexitätsreduktion in multivariaten Datenstrukturen.

References

- Anevski, D. and Hössjer, O. (2002). Monotone regression and density function estimation at a point of discontinuity. *J. Nonparametr. Stat.* **14**, 279-294.
- Balabdaoui, F. (2007). Consistent estimation of a convex density at the origin. *Math. Methods Statist.* **16**, 77-95
- Bennett, C. and Sharpley, R. (1988). *Interpolation of Operators*. Academic Press, N.Y.
- Birke, M. and Dette, H. (2005). A note on estimating a monotone regression function by combining kernel and density estimates. *Technical report, Department of Mathematics*.
- Birke, M. and Dette, H. (2007). Estimating a convex function in nonparametric regression. *Scand. J. Statist.* **34**, 384 - 404.
- Chernozhukov, V., Fernández-Val, I., Galichon, A. (2007). Improving estimates of monotone functions by rearrangement. *Preprint*, arXiv:0704.368v1
- Dette, H., Neumeyer, N. and Pilz, K.F. (2006). A simple nonparametric estimator of a strictly monotone regression function.
- Dette, H., Volgushev, S. (2007). Non-crossing nonparametric estimates of quantile curves. *J. R. Stat. Soc.*, to appear.
- Dümbgen, L., Rufibach, K. and Wellner, J.A. (2007). Marshall's lemma for convex density estimation. *IMS Lecture Notes-Monograph Series Asymptotics: Particles, Processes and Inverse Problems* **55**, 101-107.
- Dümbgen, L., Rufibach, K. (2007). Maximum likelihood estimation of a log-concave density and its distribution function: basic properties and uniform consistency. *Preprint*, arXiv:0709.0334v2

- El Barmi, H. and Nelson, P.I. (2000). Three monotone density estimators from selection biased samples. *J. Statist. comput. Simul.* **67**, 203-217.
- El Barmi, H. and Nelson, P.I. (2002). A note on estimating a non-increasing density in the presence of selection bias. *J. Stat. Plann. Inference* **107**, 353-364.
- Grenander, U. (1956). On the theory of mortality measurements, Part II, *Skand. Akt.* **39**, 125-153.
- Groeneboom, P. (1985). Estimating a monotone density. *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, Vol. II, Lucien M. LeCam and Richard A. Olshen eds. Wadsworth, New York. 529-555.
- Groeneboom, P., Jongbloed, G. and Wellner, J.A. (2001). Estimation of a convex function: Characterizations and asymptotic theory. *Ann. Statist.* **29**, 1653-1698.
- Hardy, G.H., Littlewood, J.E. and Pólya, G. (1952). *Inequalities*. 2nd ed., Cambridge University Press.
- Lieb, E.H., Loss, M. (1997). *Analysis*. Providence, RI: AMS.
- Neumeyer, N. (2007). A note on uniform consistency of monotone function estimators. *Statist. Probab. Lett.* **77**, 693-703.
- Silverman, B.W. (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *Ann. Statist.* **6**, 177 - 184.
- Van der Vaart, A.W. and van der Laan, M.J. (2003). Smooth estimation of a monotone density. *Statistics* **37**, 189-203.
- Volgushev, S. (2007). Nichtparametrische Quantilsregression. Diploma thesis, Ruhr-Universität Bochum (in German).

Woodroffe, M. and Sun, J. (1993). A penalized maximum likelihood estimate of $f(0^+)$ when f is non-increasing. *Statistica Sinica* **3**, 501-515.