

# Estimation of a Convex Function: Characterizations and Asymptotic Theory

Piet Groeneboom, Geurt Jongbloed, and Jon A. Wellner<sup>1</sup>

*Delft University of Technology, Vrije Universiteit Amsterdam,  
and University of Washington*

*July 2, 2001*

## Abstract

**Abstract:** We study nonparametric estimation of convex regression and density functions by methods of least squares (in the regression and density cases) and maximum likelihood (in the density estimation case). We provide characterizations of these estimators, prove that they are consistent, and establish their asymptotic distributions at a fixed point of positive curvature of the functions estimated. The asymptotic distribution theory relies on the existence of a “invelope function” for integrated two-sided Brownian motion  $+t^4$  which is established in the companion paper GROENEBOOM, JONGBLOED AND WELLNER (2001A).

---

<sup>1</sup> Research supported in part by National Science Foundation grant DMS-95-32039 and NIAID grant 2R01 AI291968-04

*AMS 2000 subject classifications.* Primary: 62G05; secondary 62G07, 62G08, 62E20.

*Key words and phrases.* convex, density estimation, regression function, maximum likelihood, least squares, integrated Brownian motion

# 1 Introduction

Estimation of functions restricted by monotonicity or other inequality constraints has received much attention. Estimation of monotone regression and density functions goes back to work by BRUNK (1958), VAN EEDEN (1956), VAN EEDEN (1957), and GRENANDER (1956). Asymptotic distribution theory for monotone regression estimators was established by BRUNK (1970), and for monotone density estimators by PRAKASA RAO (1969). The asymptotic theory for monotone regression function estimators was re-examined by WRIGHT (1981), and the asymptotic theory for monotone density estimators was re-examined by GROENEBOOM (1985). The “universal component” of the limit distribution in these problems is the distribution of the location of the maximum of two-sided Brownian motion minus a parabola. GROENEBOOM (1988) examined this distribution and other aspects of the limit Gaussian problem with canonical monotone function  $f_0(t) = 2t$  in great detail. GROENEBOOM (1985) provided an algorithm for computing this distribution, and this algorithm has recently been implemented by GROENEBOOM AND WELLNER (2000). See BARLOW, BARTHOLOMEW, BREMNER AND BRUNK (1972) and ROBERTSON, WRIGHT AND DYKSTRA (1988) for a summary of the earlier parts of this work.

In the case of estimation of a concave regression function, HILDRETH (1954) first proposed least squares estimators, and these were proved to be consistent by HANSON AND PLEDGER (1976). MAMMEN (1991) established rates of convergence for a least squares convex or concave regression function estimator and the slope thereof at a fixed point  $x_0$ . In the case of estimating a convex density function the first work seems to be that of ANEVSKI (1994), who was motivated by some problems involving the migration of birds discussed by HAMPEL (1987) and LAVÉE, SAFRIE AND MEILIJSON (1991). JONGBLOED (1995) established lower bounds for minimax rates of convergence, and established rates of convergence for a “sieved maximum likelihood estimator”.

Until now, the limiting distributions of these convex function estimators at a fixed point  $x_0$  have not been available. We establish these limiting distributions in Section 5 of this paper. In Sections 2-4 we lay the groundwork for these limit distributions by introducing the estimators to be studied, giving careful characterizations thereof, and proving the needed consistency and rates of convergence, or giving references to the earlier literature when consistency or rates of convergence have already been established. Our proofs of the limit distributions in Section 5 here rely strongly on the characterization of the solution of a corresponding continuous Gaussian problem for the canonical convex function  $f_0(t) = 12t^2$  given in GROENEBOOM, JONGBLOED AND WELLNER (2001A). This solution is given by a (random) piecewise cubic function  $H$  which lies above  $Y$ , two-sided integrated Brownian motion plus the drift function  $t^4$  (note that  $12t^2$  is the second derivative of  $t^4$ ), with the property that  $H''$  is piecewise linear and convex. Thus we call  $H$  an *invelope* of the process  $Y$ . The key universal component of the limiting distribution of a convex function estimator and its derivative is given by the joint distribution of  $(H''(0), H'''(0))$ . Although no analytic expressions are currently available for this joint distribution, it is in principle possible to get Monte-Carlo evidence for it, using the characterization as an invelope of integrated Brownian motion.

One previous attempt at finding these limiting distributions is due to WANG (1994), who examined the convex regression function problem studied by MAMMEN (1991). See GROENEBOOM,

JONGBLOED AND WELLNER (2001A) for a discussion of some of the difficulties in Wang’s arguments.

Here is an outline of this paper: Section 2 gives definitions and characterizations of the estimators to be considered. Consistency of each of the estimators is proved in Section 3, and rates of convergence of the estimators are established in Section 4. Section 5, based on parts of Chapter 6 of JONGBLOED (1995), gives a brief discussion of local asymptotic minimax lower bounds for estimation of a convex density function and its derivative at a fixed point  $x_0$ . Finally, Section 6 contains our results concerning the asymptotic distributions of the estimators at a fixed point  $x_0$ . This section relies strongly on GROENEBOOM, JONGBLOED AND WELLNER (2001A).

Because of the length of the current manuscript we will examine computational methods and issues in GROENEBOOM, JONGBLOED AND WELLNER (2001B). For computational methods for the canonical limit Gaussian problem, see Section 3 of GROENEBOOM, JONGBLOED AND WELLNER (2001A). For some work on computation of the estimators studied here, see MAMMEN (1991), JONGBLOED (1998), and MEYER (1997).

## 2 Estimators of a convex density or regression function

In this section we will study two different estimators of a convex density function  $f_0$ , a least squares estimator and the nonparametric maximum likelihood estimator (MLE), and the least squares estimator of a convex regression function  $r_0$ . We begin with the least squares estimator for a convex and decreasing density. First, in Lemma 2.1, existence and uniqueness of the least squares estimator  $\tilde{f}$  will be established. Moreover, it will be shown that the estimator is piecewise linear, having at most one change of slope between successive observations. In Lemma 2.2 necessary and sufficient conditions will be derived for a convex decreasing density to be the least squares estimator. These conditions can be rephrased and interpreted geometrically, saying that the second integral of  $\tilde{f}$  is an ‘envelope’ of the integral of the empirical distribution function based on the data. Then we will proceed to the MLE. In Lemma 2.3, existence and uniqueness of the MLE is established. This estimator will also turn out to be piecewise linear. In Lemma 2.4, the MLE is characterized geometrically in terms of a certain convex envelope of the function  $\frac{1}{2}t^2$ .

It is interesting that the least squares estimator and the MLE are really different in general. This differs from the situation for monotone densities. In the related problem of estimating a monotone density, the least squares estimator and the MLE coincide: the least squares estimator is identical to the MLE found by GRENDER (1956).

### 2.1 The Least Squares estimator of a convex decreasing density

The least squares (LS) estimator  $\tilde{f}_n$  of a convex decreasing density function  $f_0$  is defined as minimizer of the criterion function

$$Q_n(f) = \frac{1}{2} \int f(x)^2 dx - \int f(x) d\mathbb{F}_n(x),$$

over  $\mathcal{K}$ , the class of convex and decreasing nonnegative functions on  $[0, \infty)$ ; here  $\mathbb{F}_n$  is the empirical distribution function of the sample. The definition of  $Q_n$  is motivated by the fact that if  $\mathbb{F}_n$  had

density  $f_n$  with respect to Lebesgue measure, then the least squares criterion would be

$$\begin{aligned} \frac{1}{2} \int (f(x) - f_n(x))^2 dx &= \frac{1}{2} \int f(x)^2 dx - \int f(x) f_n(x) dx + \int f_n(x)^2 dx \\ &= \frac{1}{2} \int f(x)^2 dx - \int f(x) d\mathbb{F}_n(x) + \int f_n(x)^2 dx \end{aligned}$$

where the last (really undefined) term does not depend on the unknown  $f$  which we seek to minimize the criterion with respect to. Note that  $\mathcal{C}$ , the class of convex and decreasing density functions on  $[0, \infty)$ , is the subclass of  $\mathcal{K}$  consisting of functions with integral 1. In Corollary 2.1 we will see that the minimizer of  $Q_n$  over  $\mathcal{K}$  belongs to this smaller set  $\mathcal{C}$ , implying that the estimate is a genuine convex and decreasing *density*.

**Lemma 2.1** *There exists a unique  $\tilde{f}_n \in \mathcal{K}$  that minimizes  $Q_n$  over  $\mathcal{K}$ . This solution is piecewise linear, and has at most one change of slope between two successive observations  $X_{(i)}$  and  $X_{(i+1)}$  and no changes of slope at observation points. The first change of slope is to the right of the first order statistic and the last change of slope, which is also the right endpoint of the support of  $\tilde{f}_n$ , is to the right of the largest order statistic.*

**Proof:** Existence follows from a compactness argument. We will show that there is a bounded convex decreasing function  $\bar{g}$  with bounded support such that the minimization can be restricted to the compact subset

$$\{g \in \mathcal{K} : g \leq \bar{g}\} \tag{2.1}$$

of  $\mathcal{K}$ .

First note that there is a  $c_1 > 0$  such that any candidate to be the minimizer of  $Q_n$  should have a left derivative at  $X_{(1)}$  bounded above in absolute value by  $c_1 = c_1(\omega)$ . Indeed, if  $g$  is a function in  $\mathcal{K}$ , then

$$g(x) \geq g(X_{(1)}) + g'(X_{(1)-})(x - X_{(1)}) \quad \text{for } x \in [0, X_{(1)}],$$

and

$$\begin{aligned} Q_n(g) &\geq \frac{1}{2} \int_0^{X_{(1)}} g(x)^2 dx - g(X_{(1)}) \\ &\geq \frac{1}{2} \int_0^{X_{(1)}} (g(X_{(1)}) + g'(X_{(1)-})(x - X_{(1)}))^2 dx - g(X_{(1)}) \\ &\geq \frac{1}{2} X_{(1)} g(X_{(1)})^2 + \frac{1}{6} X_{(1)}^3 g'(X_{(1)-})^2 - g(X_{(1)}) \\ &\geq -(2X_{(1)})^{-1} + \frac{1}{6} X_{(1)}^3 g'(X_{(1)-})^2, \end{aligned}$$

showing that  $Q_n(g)$  tends to infinity as the left derivative of  $g$  at  $X_{(1)}$  tends to minus infinity. In the last inequality we use that  $u \mapsto \frac{1}{2} X_{(1)} u^2 - u$  attains its minimum at  $u = 1/X_{(1)}$ . This same argument can be used to show that the right derivative at  $X_{(n)}$  of any solution candidate  $g$  is bounded below in absolute value by some  $c_2 = c_2(\omega)$ , whenever  $g(X_{(n)}) > 0$ .

Additionally, it is clear that  $g(X_{(1)})$  is bounded by some constant  $c_3 = c_3(\omega)$ . This follows from the fact that

$$Q_n(g) \geq \frac{1}{2}g(X_{(1)})^2 X_{(1)} - g(X_{(1)}),$$

which tends to infinity as  $g(X_{(1)})$  tends to infinity.

To conclude the existence argument, observe that we may restrict attention to functions in  $\mathcal{K}$  that are linear on the interval  $[0, X_{(1)}]$ . Indeed, any element  $g$  of  $\mathcal{K}$  can be modified to a  $\tilde{g} \in \mathcal{K}$  which is linear on  $[0, X_{(1)}]$  as follows:

$$\tilde{g}(x) = \begin{cases} g(X_{(1)}) + g'(X_{(1)+})(x - X_{(1)}) & \text{for } x \in [0, X_{(1)}] \\ g(x) & \text{for } x > X_{(1)}, \end{cases}$$

and if  $g \neq \tilde{g}$ ,  $Q_n(g) > Q_n(\tilde{g})$  (only first term is influenced by going from  $g$  to  $\tilde{g}$ ). For the same reason, attention can be restricted to functions that behave linearly between the point  $X_{(n)}$  and the point where it hits zero, by extending a function using its left derivative at the point  $X_{(n)}$ . In fact, this argument can be adapted to show that a solution of the minimization problem has at most 1 change of slope between successive observations. Indeed, let  $g$  be a given convex decreasing function, and fix its values at the observation points. Then one can construct a piecewise linear function which lies entirely below  $g$ , and has the same values at the observation points. This shows that  $Q_n$  is decreased when going from  $g$  to this piecewise linear version, since the first term of  $Q_n$  decreases and the second term stays the same.

Hence, defining the function

$$\bar{g}(x) = \begin{cases} c_3 + c_1(X_{(1)} - x) & \text{for } x \in [0, X_{(1)}] \\ (c_3 - c_2(x - X_{(1)})) \vee 0 & \text{for } x > X_{(1)}, \end{cases}$$

we see that the minimization of  $Q_n$  over  $\mathcal{K}$  may be restricted to the compact set (2.1). Uniqueness of the solution follows from the strict convexity of  $Q_n$  on  $\mathcal{K}$ .  $\square$

**Lemma 2.2** *Let  $Y_n$  be defined by*

$$Y_n(x) = \int_0^x \mathbb{F}_n(t) dt, \quad x \geq 0.$$

*Then the piecewise linear function  $\tilde{f}_n \in \mathcal{K}$  minimizes  $Q_n$  over  $\mathcal{K}$  if and only if the following conditions are satisfied for  $\tilde{f}_n$  and its second integral  $\tilde{H}_n(x) = \int_{0 < t < u < x} \tilde{f}_n(t) dt du$ :*

$$\tilde{H}_n(x) \begin{cases} \geq Y_n(x), & \text{if } x \geq 0, \\ = Y_n(x) & \text{if } \tilde{f}'_n(x+) > \tilde{f}'_n(x-). \end{cases} \quad (2.2)$$

**Proof.** Let  $\tilde{f}_n \in \mathcal{K}$  satisfy (2.2), and note that this implies

$$\int_{(0, \infty)} \left\{ \tilde{H}_n(x) - Y_n(x) \right\} d\tilde{f}'_n(x) = 0. \quad (2.3)$$

Choose  $g \in \mathcal{K}$  arbitrary. Then we get, using integration by parts,

$$Q_n(g) - Q_n(\tilde{f}_n) \geq \int_{(0,\infty)} \left\{ \tilde{H}_n(x) - Y_n(x) \right\} d(g' - \tilde{f}'_n)(x).$$

But using (2.3) and (2.2), we get

$$\int_{(0,\infty)} \left\{ \tilde{H}_n(x) - Y_n(x) \right\} d(g' - \tilde{f}'_n)(x) = \int_{(0,\infty)} \left\{ \tilde{H}_n(x) - Y_n(x) \right\} dg'(x) \geq 0.$$

Hence  $\tilde{f}_n$  minimizes  $Q_n$  over  $\mathcal{K}$ .

Conversely, suppose that  $\tilde{f}_n$  minimizes  $Q_n(g)$  over  $\mathcal{K}$ . Consider, for  $x > 0$ , the function  $g_x \in \mathcal{K}$ , defined by

$$g_x(t) = (x - t)_+, \quad t \geq 0. \quad (2.4)$$

Then we must have:

$$\lim_{\epsilon \downarrow 0} \frac{Q_n(\tilde{f}_n + \epsilon g_x) - Q_n(\tilde{f}_n)}{\epsilon} = \tilde{H}_n(x) - Y_n(x) \geq 0.$$

This yields the inequality part of (2.2). We must also have

$$\lim_{\epsilon \rightarrow 0} \frac{Q_n((1 + \epsilon)\tilde{f}_n) - Q_n(\tilde{f}_n)}{\epsilon} = \int_{(0,\infty)} \left\{ \tilde{H}_n(x) - Y_n(x) \right\} d\tilde{f}'_n(x) = 0,$$

which is (2.3). This can, however, only hold if the equality part of (2.2) also holds.  $\square$

Lemma 2.2 characterizes the LS estimator  $\tilde{f}_n$  as the second derivative of a very special ‘invelope’ of the integrated empirical distribution function. The term ‘invelope’ is coined for this paper, in contrast to the term “envelope” that will be encountered in the characterization of the MLE.

Figure 1 shows a picture of  $Y_n$  and the “invelope”  $\tilde{H}_n$  for a sample of size 20, generated by the density

$$f_0(x) = 3(1 - x)^2 1_{[0,1]}(x), \quad x \geq 0. \quad (2.5)$$

We take such a small sample, because otherwise the difference between  $Y_n$  and  $\tilde{H}_n$  is not visible. The algorithm used works equally well for big sample sizes (like 5000 or 10,000). The algorithm that was used in producing these pictures (and likewise the algorithm that produced the pictures of the MLE in the sequel) will be discussed in GROENEBOOM, JONGBLOED AND WELLNER (2001B).

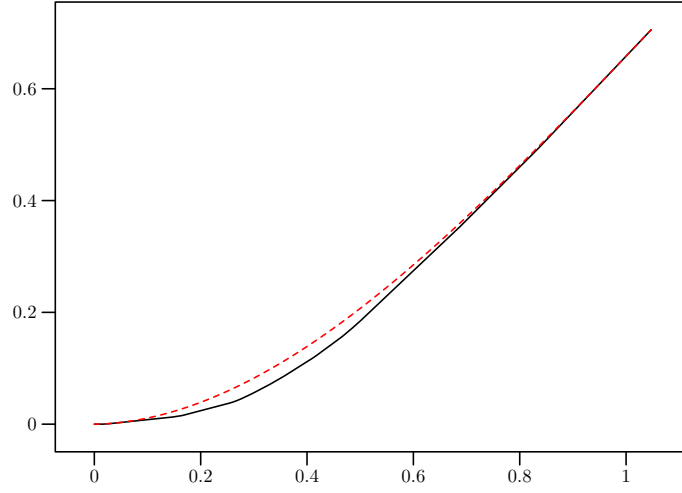


Figure 1: Solid:  $Y_n$  (black), dashed:  $\tilde{H}_n$  (red).

Figure 2 shows a picture of  $\mathbb{F}_n$  and  $\tilde{H}'_n$  for the same sample.

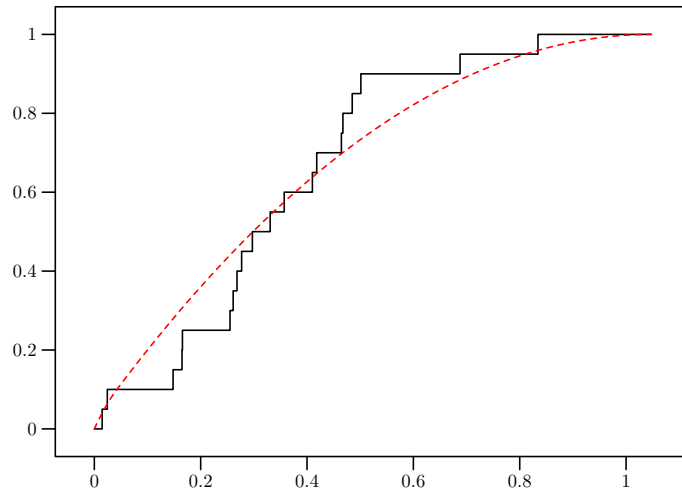


Figure 2: Solid:  $\mathbb{F}_n$  (black), dashed:  $\tilde{H}'_n$  (red).

**Corollary 2.1** Let  $\tilde{H}_n$  satisfy condition (2.2) of Lemma 2.2 and let  $\tilde{f}_n = \tilde{H}_n''$ . Then:

- (i)  $\tilde{F}_n(x) = \mathbb{F}_n(x)$  for each  $x$  such that  $\tilde{f}'_n(x-) < \tilde{f}'_n(x+)$ , where  $\tilde{F}_n(x) = \int_0^x \tilde{f}_n(t) dt$ .
- (ii)  $\tilde{f}_n(X_{(n)}) > 0$ , where  $X_{(n)}$  is the largest order statistic of the sample.
- (iii)  $\tilde{f}_n \in \mathcal{C}$ , i.e.,  $\int \tilde{f}_n(x) dx = 1$ .

(iv) Let  $0 < t_1 < \dots < t_m$  be the points of change of slope of  $\tilde{H}_n''$  and let  $t_0 = 0$ . Then  $\tilde{f}_n$  and  $\tilde{H}_n$  have the following “midpoint properties”:

$$\tilde{f}_n(\bar{t}_k) = \frac{1}{2} \left\{ \tilde{f}_n(t_{k-1}) + \tilde{f}_n(t_k) \right\} = \frac{\mathbb{F}_n(t_k) - \mathbb{F}_n(t_{k-1})}{t_k - t_{k-1}}, \quad (2.6)$$

and

$$\tilde{H}_n(\bar{t}_k) = \frac{1}{2} \{Y_n(t_{k-1}) + Y_n(t_k)\} - \frac{1}{8} \{\mathbb{F}_n(t_k) - \mathbb{F}_n(t_{k-1})\} (t_k - t_{k-1}), \quad (2.7)$$

for  $k = 1, \dots, m$ , where  $\bar{t}_k = (t_{k-1} + t_k)/2$ .

**Proof.** For proving (i), note that at each point  $x$  such that  $\tilde{f}'_n(x-) < \tilde{f}'_n(x+)$  (note that such a point cannot be an observation point by Lemma 2.1) we have by (2.2) that  $Y_n(x) = \tilde{H}_n(x)$ . Since  $\tilde{H}_n \geq Y_n$  throughout and both  $Y_n$  and  $\tilde{H}_n$  are differentiable at  $x$ , we have that  $\tilde{F}_n(x) = \mathbb{F}_n(x)$ .

For (ii), we will prove that the upper support point of the piecewise linear density  $\tilde{f}_n$ ,  $x(\tilde{f}_n)$ , satisfies  $x(\tilde{f}_n) > X_{(n)}$ . From Lemma 2.1 we already know that  $x(\tilde{f}_n) \neq X_{(n)}$ . Now suppose that  $x(\tilde{f}_n) < X_{(n)}$ . Then for all  $x > X_{(n)}$

$$\tilde{H}'_n(x) = \tilde{F}'_n(x) = \tilde{F}'_n(x(\tilde{f}_n)) \stackrel{(i)}{=} \mathbb{F}'_n(x(\tilde{f}_n)) < 1.$$

However, since  $Y'_n(x) = \mathbb{F}'_n(x) = 1$  for all  $x > X_{(n)}$ , inevitably the inequality part of (2.2) would be violated eventually. Hence  $x(\tilde{f}_n) > X_{(n)}$  and (ii) follows.

For (iii), combine (i) and (ii) to get

$$\int \tilde{f}_n(x) dx = \tilde{F}_n(x(\tilde{f}_n)) = \mathbb{F}_n(x(\tilde{f}_n)) = 1.$$

The first part of (iv) is an easy consequence of the fact that  $\tilde{F}_n(t_i) = \mathbb{F}_n(t_i)$ ,  $i = 0, \dots, m$  (part (i)), combined with the property that  $\tilde{f}_n$  is linear on the intervals  $[t_{i-1}, t_i]$ . Again by the fact that  $\tilde{f}_n$  is linear on  $[t_{k-1}, t_k]$ , we get that  $\tilde{H}_n$  is a cubic polynomial on  $[t_{k-1}, t_k]$ , determined by

$$\tilde{H}_n(t_{k-1}) = Y_n(t_{k-1}), \tilde{H}_n(t_k) = Y_n(t_k), \tilde{H}'_n(t_{k-1}) = \mathbb{F}'_n(t_{k-1}), \tilde{H}'_n(t_k) = \mathbb{F}'_n(t_k),$$

using that  $\tilde{H}_n$  is tangent to  $Y_n$  at  $t_{k-1}$  and  $t_k$ . This implies (2.7).  $\square$

**Remark:** We know from Lemma 2.1 and Corollary 2.1 that for the case  $n = 1$ , the LS estimator is a function on  $[0, \infty)$  which only changes slope at the endpoint of its support. Denoting this point by  $\theta$  and the observation by  $X_1$ , we see, in view of Corollary 2.1 (iii), that

$$\tilde{f}_1(x) = f_\theta(x) = \frac{2}{\theta^2}(\theta - x)_+. \quad (2.8)$$

Consequently, we have that

$$Q_n(f_\theta) = \frac{1}{2} \int_0^\theta f_\theta^2(x) dx - f_\theta(x_1) = \begin{cases} 2x_1/\theta^2 - 4/(3\theta) & \text{if } \theta > X_1 \\ 2/(3\theta) & \text{if } \theta \leq X_1 \end{cases}$$

and the least squares estimator corresponds to  $\theta = 3X_1$ . Note that this least squares estimator can also be obtained directly via the characterization of the estimator given in Lemma 2.2.



## 2.2 The Nonparametric Maximum Likelihood Estimator of a convex decreasing density

For  $g \in \mathcal{C}$ , the convex subset of  $\mathcal{K}$  corresponding to convex and decreasing densities on  $[0, \infty)$ , define ‘minus the loglikelihood function’ by

$$-\int \log g(x) d\mathbb{F}_n(x), \quad g \in \mathcal{C}$$

and the nonparametric maximum likelihood estimator (MLE) as minimizer of this function over  $\mathcal{C}$ . In order to relax the constraint  $\int g(x)dx = 1$  and get a criterion function to minimize over all of  $\mathcal{K}$ , we define

$$\psi_n(g) = -\int \log g(x) d\mathbb{F}_n(x) + \int g(x)dx, \quad g \in \mathcal{K}.$$

Lemma 2.3 shows that the minimizer of  $\psi_n$  over  $\mathcal{K}$  is a function  $\hat{f}_n \in \mathcal{C}$ , and hence  $\hat{f}_n$  is the MLE.

**Lemma 2.3** *The MLE  $\hat{f}_n$  exists and is unique. It is a piecewise linear function and has at most one change of slope in each interval between successive observations. It is also the unique minimizer of  $\psi_n$  over  $\mathcal{K}$*

**Proof:** Fix an arbitrary  $g \in \mathcal{C}$ . We show that there exists a  $\bar{g} \in \mathcal{C}$  which is piecewise linear with at most one change of slope between successive observations and for which  $\psi_n(\bar{g}) \leq \psi_n(g)$ . It is easily seen that if we define  $h$  by requiring that  $h(X_{(i)}) = g(X_{(i)})$  for all  $i = 1, \dots, n$ ,  $h'(X_{(i)}) = \frac{1}{2}(g'(X_{(i)}-) + g'(X_{(i)}+))$  and that  $h$  is piecewise linear with at most one change of slope between successive observations,  $\bar{g} = h/\int h$  has  $\psi_n(\bar{g}) < \psi_n(g)$  whenever  $\bar{g} \neq g$ . Thus minimizers of  $\psi_n$  over  $\mathcal{C}$  must be of the form of  $\bar{g}$ .

We will show that the minimizer of  $\psi_n$  exists by showing that the minimization of  $\psi_n$  may be restricted to a compact subset  $\mathcal{C}_M$  of  $\mathcal{C}$  given by

$$\mathcal{C}_M = \{g \in \mathcal{C} : g(0) \leq M, g(M) = 0\}$$

for some fixed  $M > 0$  (depending on the data). Indeed, since  $g$  satisfies  $\int g(x)dx = 1$ , any element of  $\mathcal{C}$  which is piecewise linear with at most one change of slope between successive observations satisfies  $g(0) \leq 2/X_{(1)}$ . Moreover, if for some  $c > X_{(n)}$ ,  $g(c) > 0$ , this automatically implies that  $g(X_{(n)}) \leq 2/(c - X_{(n)})$ , which tends to zero as  $c \rightarrow \infty$ . However, this again implies  $\psi_n(g_c) \rightarrow \infty$ .

Now for the uniqueness. Suppose  $g_1$  and  $g_2$  are both piecewise linear with at most one change of slope between successive observations and with  $\psi_n(g_1) = \psi_n(g_2)$ , minimal. Then the first claim is that  $g_1(X_{(i)}) = g_2(X_{(i)})$  for all  $i = 1, \dots, n$ . This follows from strict concavity of  $u \rightarrow \log u$  on  $(0, \infty)$ , implying that  $\psi_n((g_1 + g_2)/2) < \psi_n(g_1)$  whenever inequality at some observation holds, contradicting the fact that  $\psi_n(g_1)$  is minimal. The second claim is that  $g_1$  and  $g_2$  have the same endpoints of their support. This has to be the case since otherwise the function  $\bar{g} = (g_1 + g_2)/2$  would minimize  $\psi_n$ , whereas it would have two changes of slope in the interval  $(X_{(n)}, \infty)$ , contradicting the fact that any solution can only have one change of slope. Consequently, since  $g_1(X_{(n)}) = g_2(X_{(n)})$ ,  $g_1'(X_{(n)}) = g_2'(X_{(n)})$  necessarily. Now observe that between  $X_{(n-1)}$  and  $X_{(n)}$  in principle three

things can happen:

- (i)  $g_1$  and  $g_2$  have a change of slope at a (common) point between  $X_{(n-1)}$  and  $X_{(n)}$ .
- (ii)  $g_1$  and  $g_2$  both have a change of slope between  $X_{(n-1)}$  and  $X_{(n)}$ , but at different points.
- (iii) Only one of  $g_1$  and  $g_2$  has a change of slope.

Note that (i) implies (using  $g_1(X_{(n-1)}) = g_2(X_{(n-1)})$ ), that  $g_1'(X_{(n-1)}) = g_2'(X_{(n-1)})$ . Also note that (ii) and (iii) cannot happen. Indeed, (iii) is impossible since it contradicts the fact that  $g_1(X_{(n-1)}) = g_2(X_{(n-1)})$ , and (ii) is impossible by the same argument used to show that  $g_1$  and  $g_2$  have the same support. This same argument can be used recursively for the intervals between successive observations, and uniqueness follows.

Finally, we show that  $\hat{f}_n$  actually minimizes  $\psi_n$  over  $\mathcal{K}$ . To this end choose  $g \in \mathcal{K}$  with  $\int_0^\infty g(x) dx = c \in (0, \infty)$  and observe that, since  $g/c \in \mathcal{C}$ ,

$$\begin{aligned} \psi_n(g) - \psi_n(\hat{f}_n) &= - \int \log \left( \frac{g(x)}{c} \right) d\mathbb{F}_n(x) - \log c + 1 - 1 + c + \int \log \hat{f}(x) d\mathbb{F}_n(x) - 1 \\ &= \psi_n(g/c) - \psi_n(\hat{f}_n) - \log c - 1 + c \geq -\log c - 1 + c \geq 0 \end{aligned}$$

with strict inequality if  $g \neq \hat{f}_n$ . □

**Remark:** From Lemma 2.3 we see that for the case  $n = 1$ , the MLE is a function on  $[0, \infty)$  which only changes slope at the endpoint of its support. Denoting this point by  $\theta$ , the observation by  $X_1$ , and the resulting form of the estimator by  $f_\theta$  as in (2.8), it follows that

$$\psi_n(f_\theta) = -\log f_\theta(X_1) + 1 = \begin{cases} 2 \log \theta - \log 2 + 1 - \log(\theta - X_1) & \text{if } \theta > X_1 \\ \infty & \text{if } \theta \leq X_1 \end{cases},$$

and the maximum likelihood estimator corresponds to  $\theta = 2X_1$ , which differs from the LS estimator we encountered in the remark following Corollary 2.1 for each  $X_1 > 0$ . Note that the MLE can also be determined from the characterization that is given in Lemma 2.4 below.

Now, for a characterization of the MLE  $\hat{f}_n$ , let  $G_n : \mathbb{R}^+ \times \mathcal{K} \rightarrow \mathbb{R} \cup \{\infty\}$  be defined by

$$G_n(t, f) = \int_0^t f(u)^{-1} d\mathbb{F}_n(u). \quad (2.9)$$

Then define  $H_n : \mathbb{R}^+ \times \mathcal{K} \rightarrow \mathbb{R} \cup \{\infty\}$  by

$$H_n(t, f) = \int_0^t G_n(u, f) du = \int_0^t \frac{t-u}{f(u)} d\mathbb{F}_n(u). \quad (2.10)$$

**Lemma 2.4** (i) *The piecewise linear function  $\hat{f}_n \in \mathcal{K}$  minimizes  $\psi_n$  over  $\mathcal{K}$  if and only if*

$$\hat{H}_n(t) := H_n(t, \hat{f}_n) \begin{cases} \leq \frac{1}{2}t^2, & x \geq 0 \\ = \frac{1}{2}t^2, & \hat{f}'_n(t-) < \hat{f}'_n(t+) \end{cases}. \quad (2.11)$$

(ii) Let  $t_1 < \dots < t_m$  be the changes of slope of  $\hat{H}_n''$ , where  $\hat{H}_n$  is defined as defined in (i), and let  $t_0 = 0$ . Then  $\hat{f}_n$  and  $\hat{H}_n$  have the following “midpoint properties”:

$$\hat{f}_n(\bar{t}_k) = \frac{1}{2} \left\{ \hat{f}_n(t_{k-1}) + \hat{f}_n(t_k) \right\} = \frac{\mathbb{F}_n(t_k) - \mathbb{F}_n(t_{k-1})}{t_k - t_{k-1}}, \quad (2.12)$$

$$H_n(\bar{t}_k) = \frac{1}{2} \left\{ \int_{[t_{k-1}, \bar{t}_k]} \frac{\bar{t}_k - x}{\hat{f}_n(x)} d\mathbb{F}_n(x) + \int_{[\bar{t}_k, t_k]} \frac{x - \bar{t}_k}{\hat{f}_n(x)} d\mathbb{F}_n(x) + t_k t_{k-1} \right\} \quad (2.13)$$

for  $k = 1, \dots, m$ , where  $\bar{t}_k = (t_{k-1} + t_k)/2$ .

**Proof:** First suppose that  $\hat{f}_n$  minimizes  $\psi_n$  over  $\mathcal{K}$ . Then for any  $g \in \mathcal{K}$  and  $\epsilon > 0$  we have

$$\psi_n(\hat{f}_n + \epsilon g) \geq \psi_n(\hat{f}_n),$$

and hence

$$0 \leq \lim_{\epsilon \downarrow 0} \frac{\psi_n(\hat{f}_n + \epsilon g) - \psi_n(\hat{f}_n)}{\epsilon} = - \int \frac{g(x)}{\hat{f}_n(x)} d\mathbb{F}_n(x) + \int g(x) dx. \quad (2.14)$$

Taking  $g(x) = (t - x)_+$  for fixed  $t > 0$  yields the inequality part of (i). To see the equality part of (2.11), note that for  $g(x) = (t - x)_+$  and  $t$  belonging to the set of changes of slope of  $\hat{f}_n'$ , the function  $\hat{f}_n + \epsilon g \in \mathcal{K}$  for  $\epsilon < 0$  and  $|\epsilon|$  sufficiently small; repeating the argument for these values of  $t$  and  $\epsilon$  yields the equality part of (2.11).

Now suppose that (2.11) is satisfied for  $\hat{f}_n$ . We first show that this implies (ii). Let  $t_1 < \dots < t_m$  be the changes of slope  $\hat{H}_n''$  and let  $t_0 = 0$ . At the points  $t_k$  the equality condition can be written as follows:

$$\int_0^{t_k} \frac{t_k - x}{\hat{f}_n(x)} d\mathbb{F}_n(x) = \frac{1}{2} t_k^2, \quad k = 1, \dots, m.$$

After some algebra, it is seen that this means

$$\int_{t_{k-1}}^{t_k} \frac{t_k - x}{\hat{f}_n(x)} d\mathbb{F}_n(x) = \frac{1}{2} (t_k - t_{k-1})^2, \quad k = 1, \dots, m, \quad (2.15)$$

where  $t_0 = 0$ .

But the equality conditions together with the inequality conditions in (2.11) imply that the function  $\hat{H}_n$  has to be tangent to the function  $t \mapsto \frac{1}{2}t^2$  at the points  $t_i$ ,  $i \geq 1$  and at  $t_0 = 0$ , and this implies that also the following equations hold (at the “derivative level”):

$$\int_{t_{k-1}}^{t_k} \frac{1}{\hat{f}_n(x)} d\mathbb{F}_n(x) = t_k - t_{k-1}, \quad k = 1, \dots, m. \quad (2.16)$$

We can write

$$\begin{aligned} \mathbb{F}_n(t_k) - \mathbb{F}_n(t_{k-1}) &= \int_{t_{k-1}}^{t_k} d\mathbb{F}_n(x) = \int_{t_{k-1}}^{t_k} \frac{\hat{f}_n(x)}{\hat{f}_n(x)} d\mathbb{F}_n(x) \\ &= \int_{t_{k-1}}^{t_k} \frac{\hat{f}_n(\bar{t}_k)}{\hat{f}_n(x)} d\mathbb{F}_n(x) + \hat{f}_n'(\bar{t}_k) \int_{t_{k-1}}^{t_k} \frac{x - \bar{t}_k}{\hat{f}_n(x)} d\mathbb{F}_n(x) \\ &= \hat{f}_n(\bar{t}_k) \{t_k - t_{k-1}\} + \hat{f}_n'(\bar{t}_k) \int_{t_{k-1}}^{t_k} \frac{x - \bar{t}_k}{\hat{f}_n(x)} d\mathbb{F}_n(x), \end{aligned}$$

where we use (2.16) in the last step. But by (2.16) we also get

$$\int_{t_{k-1}}^{t_k} \frac{t_k - x}{\hat{f}_n(x)} d\mathbb{F}_n(x) + \int_{t_{k-1}}^{t_k} \frac{x - t_{k-1}}{\hat{f}_n(x)} d\mathbb{F}_n(x) = \{t_k - t_{k-1}\} \int_{t_{k-1}}^{t_k} \frac{1}{\hat{f}_n(x)} d\mathbb{F}_n(x) = \{t_k - t_{k-1}\}^2,$$

and hence, using (2.15), it is seen that

$$\int_{t_{k-1}}^{t_k} \frac{x - t_{k-1}}{\hat{f}_n(x)} d\mathbb{F}_n(x) = \int_{t_{k-1}}^{t_k} \frac{t_k - x}{\hat{f}_n(x)} d\mathbb{F}_n(x) = \frac{1}{2} \{t_k - t_{k-1}\}^2. \quad (2.17)$$

Hence we obtain the first part of (ii), since

$$\hat{f}'_n(\bar{t}_k) = \frac{\hat{f}_n(t_k) - \hat{f}_n(t_{k-1})}{t_k - t_{k-1}},$$

using the linearity of  $\hat{f}_n$  on the interval  $[t_{k-1}, t_k]$ .

To prove the second part of (ii) we first note that

$$\begin{aligned} \int_{\bar{t}_k}^{t_k} \frac{x - \bar{t}_k}{\hat{f}_n(x)} d\mathbb{F}_n(x) &= \int_0^{t_k} \frac{x - \bar{t}_k}{\hat{f}_n(x)} d\mathbb{F}_n(x) + \int_0^{\bar{t}_k} \frac{\bar{t}_k - x}{\hat{f}_n(x)} d\mathbb{F}_n(x) \\ &= H_n(\bar{t}_k) + \frac{1}{2} (t_k - t_{k-1}) \int_0^{t_k} \frac{1}{\hat{f}_n(x)} d\mathbb{F}_n(x) - \int_0^{t_k} \frac{t_k - x}{\hat{f}_n(x)} d\mathbb{F}_n(x) \\ &= H_n(\bar{t}_k) + \frac{1}{2} (t_k - t_{k-1}) t_k - \frac{1}{2} t_k^2 = H_n(\bar{t}_k) - \frac{1}{2} t_k t_{k-1}. \end{aligned} \quad (2.18)$$

In a similar way, we get

$$\int_{t_{k-1}}^{\bar{t}_k} \frac{\bar{t}_k - x}{\hat{f}_n(x)} d\mathbb{F}_n(x) = H_n(\bar{t}_k) - \frac{1}{2} t_k t_{k-1}. \quad (2.19)$$

Combining (2.18) and (2.19) we get the result.

Part (ii) immediately implies that  $\hat{f}_n$  belongs to  $\mathcal{C}$ , since

$$\int_0^\infty \hat{f}_n(x) dx = \sum_{k=1}^m \hat{f}_n(\bar{t}_k) (t_k - t_{k-1}) = \sum_{k=1}^m \{\mathbb{F}_n(t_k) - \mathbb{F}_n(t_{k-1})\} = 1. \quad (2.20)$$

To show that  $\hat{f}_n$  minimizes  $\psi_n$  over  $\mathcal{K}$ , note that all  $g \in \mathcal{K}$  have the following representation

$$g(x) = \int_0^\infty (t - x)_+ d\nu(t) \quad (2.21)$$

for some finite positive measure  $\nu$ . Then, using  $-\log(u) \geq 1 - u$  and the definition of  $G_n(\cdot, \hat{f}_n)$ , we have

$$\psi_n(g) - \psi_n(\hat{f}_n) = - \int_0^\infty \log\left(\frac{g}{\hat{f}_n}\right) d\mathbb{F}_n + \int_0^\infty (g(x) - \hat{f}_n(x)) dx$$

$$\begin{aligned}
&\stackrel{(2.20)}{\geq} \int_0^\infty \left(1 - \frac{g}{\hat{f}_n}\right) d\mathbb{F}_n + \int_0^\infty g(x)dx - 1 = - \int_0^\infty \frac{g}{\hat{f}_n} d\mathbb{F}_n + \int_0^\infty g(x)dx \\
&\stackrel{(2.21)}{=} - \int_0^\infty \int_0^\infty (t-x)_+ d\nu(t) dG_n(x, \hat{f}_n) + \int_0^\infty \int_0^\infty (t-x)_+ d\nu(t) dx \\
&= \int_0^\infty \left\{ - \int_0^\infty (t-x)_+ dG_n(x, \hat{f}_n) + \int_0^\infty (t-x)_+ dx \right\} d\nu(t) \\
&= \int_0^\infty \left\{ \frac{1}{2}t^2 - H_n(t, \hat{f}_n) \right\} d\nu(t) \geq 0,
\end{aligned}$$

where we use the inequality condition in (2.11) in the last step. Thus  $\hat{f}_n$  minimizes  $\psi_n$  over  $\mathcal{K}$ .  $\square$

Note that the property that the MLE can have at most one change of slope between two observations (and cannot change slope at any of the observations) that was part of the statement of Lemma 2.3, can also be seen from the characterization given in Lemma 2.4. A piecewise linear envelope of the function  $t \mapsto \frac{1}{2}t^2$  cannot touch this function (the location of any such touch coincides with a change of slope of the MLE) at a point where it bends (i.e., an observation point). Moreover, a straight line cannot touch a parabola at two distinct points.

The MLE shares the “midpoint property” with the LS estimator (but clearly for different points  $t_k$ ), see Corollary 2.1, part (iv), and Lemma 2.4, part (ii). So both are a kind of “derivative” of the empirical distribution function, just like the Grenander estimator of a decreasing density. We note in passing that the MLE  $\hat{f}_n$  solves the following weighted least squares problem with “self-induced” weights: minimize  $\tilde{\psi}_n(g)$  over  $g \in \mathcal{K}$  where

$$\tilde{\psi}_n(g) = \frac{1}{2} \int_0^\infty \frac{g(t)^2}{\hat{f}_n(t)} dt - \int_0^\infty \frac{g(t)}{\hat{f}_n(t)} d\mathbb{F}_n(t).$$

Figure 3 shows a picture of  $\hat{H}_n$  and the function  $t \mapsto t^2/2$  for the same sample of size 20, as used for figures 1 and 2; figure 4 shows  $\hat{H}'_n$  and the identity function. Figure 5 gives a comparison of the LS estimator and the MLE for the same sample.

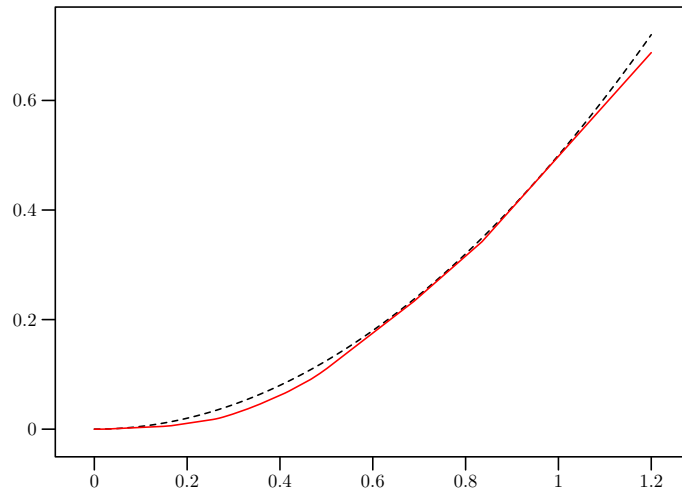


Figure 3: Function  $\hat{H}_n$  of Lemma 2.4. Solid:  $\hat{H}_n$  (red), dashed:  $t \mapsto t^2/2$  (black).

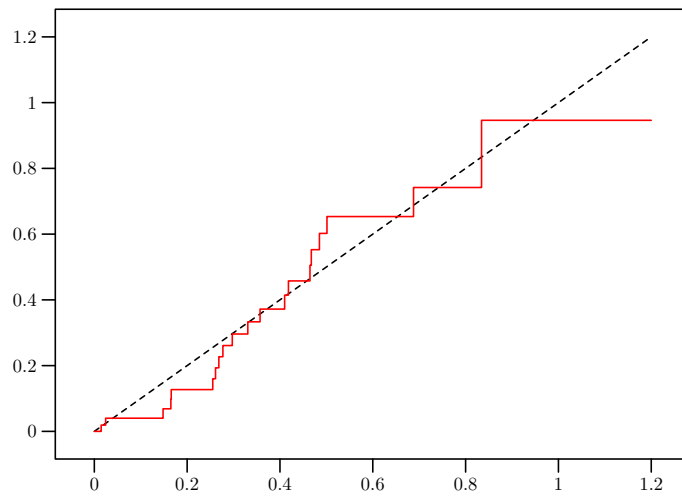


Figure 4: Solid:  $\hat{H}'_n$  (red), dashed:  $t \mapsto t$  (black).

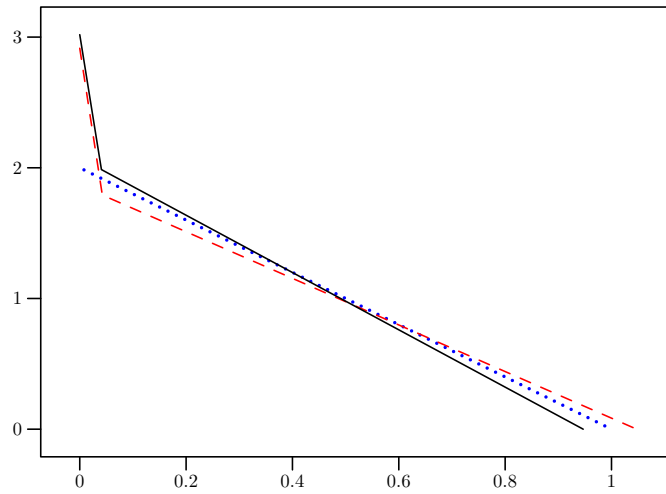


Figure 5: Dotted: real density (blue), solid: MLE (black), dashed: LS estimator (red).

We chose the small sample size because otherwise the difference between  $\tilde{H}_n$  and  $Y_n$  (resp.  $\hat{H}_n$  and  $t^2/2$ ) is hardly visible. For the same reason we chose the “borderline” convex function that is linear on  $[0, 1]$ . Figure 6 shows a comparison of the LS estimator and the MLE for a more “normal” sample size 100 and the strictly convex density function

$$x \mapsto 3(1 - x)^2 1_{[0,1]}(x), x \geq 0.$$

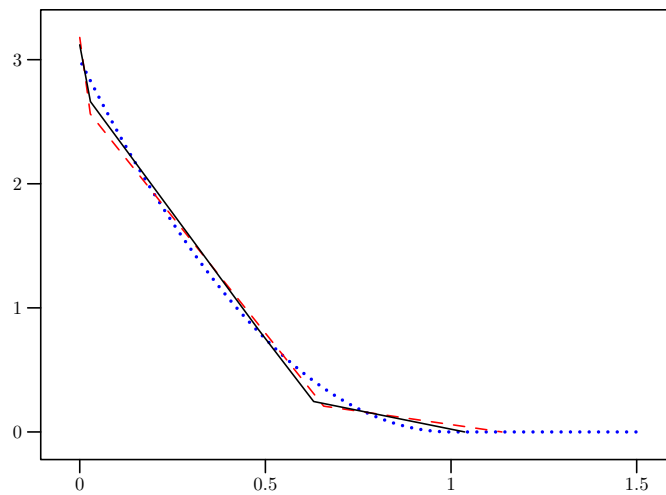


Figure 6: Dotted: real density (blue), solid: MLE (black), dashed: LS estimator (red).

### 2.3 The Least Squares estimator of a convex regression function.

Consider given the following data for  $n = 1, 2, \dots$ :  $\{(x_{n,i}, Y_{n,i}) : i = 1, \dots, n\}$ , where

$$Y_{n,i} = r_0(x_{n,i}) + \epsilon_{n,i} \quad (2.22)$$

for a convex function  $r_0$  on  $\mathbb{R}$ . Here  $\{\epsilon_{n,i} : i = 1, \dots, n, n \geq 1\}$  is a triangular array of i.i.d. random variables satisfying  $Ee^{t\epsilon_{1,1}} < \infty$  for some  $t > 0$ , and the  $x_{n,i}$ 's are ordered as  $x_{n,1} < x_{n,2} < \dots < x_{n,n}$ . Writing  $\mathcal{K}$  for the set of all convex functions on  $\mathbb{R}$ , the first suggestion for a least squares estimate of  $r_0$  is

$$\operatorname{argmin}_{r \in \mathcal{K}} \phi_n(r), \quad \text{where} \quad \phi_n(r) = \frac{1}{2} \sum_{i=1}^n (Y_{n,i} - r(x_{n,i}))^2.$$

It is immediately clear, however, that this definition needs more specification. For instance, any solution to the minimization problem can be extended quite arbitrarily (although convex) outside the range of the  $x_{n,i}$ 's. Also, between the  $x_{n,i}$ 's there is some arbitrariness in the way a function can be chosen. We therefore confine ourselves to minimizing  $\phi_n$  over the subclass  $\mathcal{K}_n$  of  $\mathcal{K}$  consisting of the functions that are linear between successive  $x_{n,i}$ 's, as well as to the left and the right of the range of the  $x_{n,i}$ 's. Hence, we define

$$\hat{r}_n = \operatorname{argmin}_{r \in \mathcal{K}_n} \phi_n(r), \quad \text{where} \quad \phi_n(r) = \frac{1}{2} \sum_{i=1}^n (Y_{n,i} - r(x_{n,i}))^2.$$

Note that  $r \in \mathcal{K}_n$  can be parameterized naturally by  $(r_{n,1}, \dots, r_{n,n}) = (r(x_{n,1}), \dots, r(x_{n,n})) \in \tilde{\mathcal{K}}_n \subset \mathbb{R}^n$  where

$$\tilde{\mathcal{K}}_n = \left\{ r_n \in \mathbb{R}^n : \frac{r_{n,i} - r_{n,i-1}}{x_{n,i} - x_{n,i-1}} \leq \frac{r_{n,i+1} - r_{n,i}}{x_{n,i+1} - x_{n,i}} \text{ for all } i = 2, \dots, n-1 \right\}.$$

The identification  $\mathcal{K}_n = \tilde{\mathcal{K}}_n$  will be made throughout.

As for both density estimators, we have existence and uniqueness of this least squares estimator. For completeness we state the lemma.

**Lemma 2.5** *There is a unique function  $\hat{r}_n \in \mathcal{K}_n$  that minimizes  $\phi_n$  over  $\mathcal{K}_n$ .*

**Proof:** Follows immediately from the strict convexity of  $\phi_n : \mathcal{K}_n \rightarrow \mathbb{R}$  and the fact that  $\phi_n(r) \rightarrow \infty$  as  $\|r\|_2 \rightarrow \infty$ .  $\square$

Next step is to characterize the least squares estimator.

**Lemma 2.6** *Define  $\hat{R}_{n,k} = \sum_{i=1}^k \hat{r}_{n,i}$  and  $S_{n,k} = \sum_{i=1}^k Y_{n,i}$ . Then  $\hat{r}_n = \operatorname{argmin}_{r \in \mathcal{K}_n} \phi_n(r)$  if and only if  $\hat{R}_{n,n} = S_{n,n}$  and*

$$\sum_{k=1}^{j-1} \hat{R}_{n,k}(x_{n,k+1} - x_{n,k}) \begin{cases} \geq \sum_{k=1}^{j-1} S_{n,k}(x_{n,k+1} - x_{n,k}) & j = 2, 3, \dots, n \\ = \sum_{k=1}^{j-1} S_{n,k}(x_{n,k+1} - x_{n,k}) & \text{if } \hat{r}_n \text{ has a kink at } x_{n,j} \text{ or } j = n. \end{cases} \quad (2.23)$$



**Proof:** First note that the convex cone  $\mathcal{K}_n$  is generated by the functions  $\pm 1$ ,  $\pm x$ , and  $(x - x_{n,i})_+$  for  $1 \leq i \leq n - 1$ . Hence, by Corollary 2.1 in GROENEBOOM (1996), we get that  $\hat{r}_n = \operatorname{argmin}_{r \in \mathcal{K}_n} \phi_n(r)$  if and only if

$$\sum_{i=1}^n \hat{r}_{n,i} = \sum_{i=1}^n Y_{n,i}, \quad \sum_{i=1}^n x_{n,i} \hat{r}_{n,i} = \sum_{i=1}^n x_{n,i} Y_{n,i}$$

and

$$\sum_{i=1}^{j-1} (\hat{r}_{n,i} - Y_{n,i})(x_{n,j} - x_{n,i}) \begin{cases} \geq 0 & \text{for all } j = 2, 3, \dots, n \\ = 0 & \text{if } \hat{r}_n \text{ has a kink at } x_{n,j}. \end{cases}$$

The first equality can be restated as  $\hat{R}_{n,n} = S_{n,n}$ . Using this, the second equality can be covered by forcing the final inequality for  $j = n$  to be an equality. Rewriting the sum

$$\sum_{i=1}^{j-1} \hat{r}_{n,i}(x_{n,j} - x_{n,i}) = \sum_{i=1}^{j-1} \hat{r}_{n,i} \sum_{k=i}^{j-1} (x_{n,k+1} - x_{n,k}) = \sum_{k=1}^{j-1} \hat{R}_{n,k}(x_{n,k+1} - x_{n,k})$$

and similarly for  $Y_{n,i}$ , the result follows.  $\square$

### 3 Consistency of the estimators

In this section we will prove consistency of the estimators introduced in Section 2. A useful inequality that holds for all convex decreasing densities  $f$  on  $(0, \infty)$  is the following

$$f(x) \leq \frac{1}{2x} \quad \text{for all } x > 0. \quad (3.1)$$

To see this, fix a convex decreasing density  $f$  on  $(0, \infty)$  and  $x_0 > 0$ . Then there exists an  $\alpha < 0$  (subgradient of  $f$  at  $x_0$ ) such that the function  $l_\alpha(x) = (f(x_0) + \alpha(x - x_0))1_{[0, x_0 - f(x_0)/\alpha]}(x)$  satisfies  $f(x) \geq l_\alpha(x)$  for all  $x \geq 0$ . Hence

$$\begin{aligned} 1 = \int_0^\infty f(x) dx &\geq \int_0^\infty l_\alpha(x) dx = \frac{1}{2}(x_0 - f(x_0)/\alpha)(f(x_0) - \alpha x_0) \\ &= x_0 f(x_0) - \frac{1}{2}(\alpha x_0^2 + f(x_0)^2/\alpha) \geq 2x_0 f(x_0). \end{aligned}$$

The final inequality holds for all  $\alpha < 0$ , with equality if and only if  $\alpha = -f(x_0)/x_0$ .

**Theorem 3.1** (*Consistency of LS density estimator*)

Suppose that  $X_1, X_2, \dots$  are i.i.d. random variables with density  $f_0 \in \mathcal{C}$ . Then the least squares estimator is uniformly consistent on closed intervals bounded away from 0: i.e., for each  $c > 0$ , we have, with probability one,

$$\sup_{c \leq x < \infty} |\tilde{f}_n(x) - f_0(x)| \rightarrow 0. \quad (3.2)$$

**Proof:** The proof is based on the characterization of the estimator given in Lemma 2.2. We let  $\mathcal{T}_n$  denote the set of locations of change of slope of  $\tilde{H}_n''$ , where  $\tilde{H}_n$  is defined as in Lemma 2.2.

First assume that  $f_0(0) < \infty$ . Fix  $\delta > 0$ , such that  $[0, \delta]$  is contained in the interior of the support of  $f_0$ , and let  $\tau_{n,1} \in \mathcal{T}_n$  be the last point of change of slope in  $(0, \delta]$ , or zero if there is no such point. Since, with probability one,

$$\liminf_{n \rightarrow \infty} X_{(n)} > \delta$$

and, by Lemma 2.1, the last point of change of slope is to the right of  $X_{(n)}$ , we may assume that there exists a point of change of slope  $\tau_{n,2}$  strictly to the right of  $\delta$ . Let  $\tau_{n,2}$  be the first point of change of slope that is strictly to the right of  $\delta$ . Then the sequence  $(\tilde{f}_n(\tau_{n,1}))$  is uniformly bounded. This is seen in the following way. Let  $\tau_n = \{\tau_{n,1} + \tau_{n,2}\}/2$ . Then  $\tau_n \geq \delta/2$  and hence, by (3.1),

$$\tilde{f}_n(\tau_n) \leq \tilde{f}_n(\delta/2) \leq 1/\delta.$$

This implies that we have an upper bound for  $\tilde{f}_n(\tau_{n,1})$  that only depends on  $\delta$ . Indeed, if  $\tau_{n,1} > \delta/2$ ,  $\tilde{f}_n(\tau_{n,1}) \leq \tilde{f}_n(\delta/2) \leq 1/\delta$  by (3.1). If  $\tau_{n,1} \leq \delta/2$ , we can use linearity of  $\tilde{f}_n$  on  $[\tau_{n,1}, \delta]$  to get

$$1 \geq \int_{\tau_{n,1}}^{\delta} \tilde{f}_n(x) dx = \frac{1}{2}(\delta - \tau_{n,1})(\tilde{f}_n(\delta) + \tilde{f}_n(\tau_{n,1})) \geq \frac{1}{4}\delta \tilde{f}_n(\tau_{n,1})$$

giving  $\tilde{f}_n(\tau_{n,1}) \leq 4/\delta$ . Moreover, the right derivative of  $\tilde{f}_n$  has a uniform absolute upper bound at  $\tau_{n,1}$ , also only depending on  $\delta$ . This can be verified analogously.

On the interval  $[\tau_{n,1}, \infty)$ , we have:

$$\frac{1}{2} \int_{[\tau_{n,1}, \infty)} \tilde{f}_n(x)^2 dx - \int_{[\tau_{n,1}, \infty)} \tilde{f}_n(x) d\mathbb{F}_n(x) \leq \frac{1}{2} \int_{[\tau_{n,1}, \infty)} f_0(x)^2 dx - \int_{[\tau_{n,1}, \infty)} f_0(x) d\mathbb{F}_n(x).$$

This follows from writing  $f_0^2 - \tilde{f}_n^2 = (f_0 - \tilde{f}_n)^2 + 2\tilde{f}_n(f_0 - \tilde{f}_n)$ , implying, using integration by parts,

$$\begin{aligned} & \frac{1}{2} \int_{[\tau_{n,1}, \infty)} f_0(x)^2 dx - \int_{[\tau_{n,1}, \infty)} f_0(x) d\mathbb{F}_n(x) - \frac{1}{2} \int_{[\tau_{n,1}, \infty)} \tilde{f}_n(x)^2 dx + \int_{[\tau_{n,1}, \infty)} \tilde{f}_n(x) d\mathbb{F}_n(x) \\ & \geq \int_{[\tau_{n,1}, \infty)} \tilde{f}_n(x) \{f_0(x) - \tilde{f}_n(x)\} dx - \int_{[\tau_{n,1}, \infty)} \{f_0(x) - \tilde{f}_n(x)\} d\mathbb{F}_n(x) \\ & = \int_{[\tau_{n,1}, \infty)} \{\tilde{H}_n(x) - Y_n(x)\} d(f_0' - \tilde{f}_n')(x) = \int_{[\tau_{n,1}, \infty)} \{\tilde{H}_n(x) - Y_n(x)\} df_0'(x) \geq 0. \end{aligned}$$

This argument was used in the proof of Lemma 2.2 on the interval  $(0, \infty)$ .

Since  $\tau_{n,1} \in [0, \delta]$ , for each subsequence there must be a further subsequence converging to a point  $\tau_1 \in [0, \delta]$ . Using a Helly argument, there will be a further subsequence  $(n_k)$  so that, for each  $x \in (\tau_1, \infty)$ ,  $\tilde{f}_{n_k}(x) \rightarrow \tilde{f}(x) = \tilde{f}(x, \omega)$ , where  $\tilde{f}$  is a convex function on  $[\tau_1, \infty)$ , satisfying  $\tilde{f}(\tau_1) < \infty$ . The function  $\tilde{f}$  satisfies:

$$\frac{1}{2} \int_{[\tau_1, \infty)} \tilde{f}(x)^2 dx - \int_{[\tau_1, \infty)} \tilde{f}(x) dF_0(x) \leq \frac{1}{2} \int_{[\tau_1, \infty)} f_0(x)^2 dx - \int_{[\tau_1, \infty)} f_0(x) dF_0(x), \quad (3.3)$$

where the integrals on the right side are finite, also if  $\tau_1 = 0$ , since  $f_0(0) < \infty$ . But this implies

$$\int_{[\tau_1, \infty)} \left\{ \tilde{f}(x) - f_0(x) \right\}^2 dx \leq 0, \quad (3.4)$$

and hence  $\tilde{f}(x) = f_0(x)$ , for  $x \geq \tau_1$ . Since  $\delta > 0$  can be chosen arbitrarily small, we get that for any  $c > 0$ , each subsequence  $\tilde{f}_\ell$  has a subsequence that converges to  $f_0$  at each point  $x \geq c$ . By the monotonicity of  $f_0$ , the convergence has to be uniform.

If  $f_0$  is unbounded in a neighborhood of zero, we cannot use (exactly) the same proof, since the integrals on the right side of (3.3) could be infinite, if the limit point  $\tau_1$  would be equal to zero. But we can still follow the same idea of proving a relation of type (3.4), by proving that for any  $\delta > 0$  there exist limit points  $\tau_1$  of this type that are strictly positive. The existence of points of this type will follow from the fact that, for each  $\delta > 0$ , there exist points  $x \in (0, \delta)$  such that in each open neighborhood of  $x$  there exist points  $x_1, x_2$  and  $x_3$ , such that  $0 < x_1 < x_2 < x_3$ , and

$$\frac{f_0(x_3) - f_0(x_2)}{x_3 - x_2} > \frac{f_0(x_2) - f_0(x_1)}{x_2 - x_1}. \quad (3.5)$$

We shall denote these points by *points of strict convexity* of  $f_0$ .

For suppose that  $x > 0$  is such a point of strict convexity of  $f_0$ . Then it is plausible that the points of change of slope  $\tau_n$ , closest to  $x$ , have to converge to  $x$  with probability one. In that case we can let  $x$  play the role of  $\tau_1$  on (3.4), and we would be through.

So, two things remain to be proved in this situation:

- (i) The existence of points of strict convexity  $x$  in each interval  $(0, b]$ ,  $b > 0$ .
- (ii) The a.s. convergence to such a point  $x$  of the closest point of change of slope  $\tau_n$ .

ad (i): If  $(0, b]$ , with  $b > 0$ , would be an interval without points of this type, we could cover  $(0, b]$  by a collection of intervals  $(x - \delta_x, x + \delta_x)$  such that  $f_0$  is linear on each interval  $(0 \vee (x - \delta_x), x + \delta_x)$ . But then  $f_0$  would be linear on  $(0, b]$ , since each interval  $[a, b] \subset (0, b]$  would have a finite subcover, and hence  $f_0$  would be linear on each such interval  $[a, b]$ , contradicting  $f_0(0) = \lim_{x \downarrow 0} f_0(x) = \infty$ .

ad (ii): Let  $x > 0$  be such a point of strict convexity of  $f_0$  and let  $\tau_{n,1}$  and  $\tau_{n,2}$  be the last point of touch  $\leq x$  between  $\tilde{H}_n$  and  $Y_n$  and the first point of touch  $> x$  between  $\tilde{H}_n$  and  $Y_n$ , respectively. Moreover, let  $\bar{\tau}_n$  be the midpoint of the interval  $[\tau_{n,1}, \tau_{n,2}]$ . Since  $x > 0$  can be chosen arbitrarily close to zero, we may assume that  $f_0(x) > 0$ . By part (iv) of Corollary 2.1 we get

$$\tilde{f}_n(\bar{\tau}_n) = \frac{1}{2} \left\{ \tilde{f}_n(\tau_{n,1}) + \tilde{f}_n(\tau_{n,2}) \right\} = \frac{\mathbb{F}_n(\tau_{n,2}) - \mathbb{F}_n(\tau_{n,1})}{\tau_{n,2} - \tau_{n,1}}. \quad (3.6)$$

Now, if  $\tau_{n,2} \rightarrow \infty$ , possibly along a subsequence, we would get:

$$\frac{1}{2} \left\{ \tilde{f}_n(\tau_{n,1}) + \tilde{f}_n(\tau_{n,2}) \right\} \rightarrow 0,$$

and in particular  $\tilde{f}_n(\tau_{n,1}) \rightarrow 0$ . But this would contradict the property

$$\int_{[\tau_{n,1}, t]} (t-y)\tilde{f}_n(y) dy \geq \int_{[\tau_{n,1}, t]} (t-y) d\mathbb{F}_n(y), \quad t \geq \tau_{n,1},$$

for large  $n$ , since, almost surely,

$$\liminf_{n \rightarrow \infty} \int_{[\tau_{n,1}, t]} (t-y) d\mathbb{F}_n(y) \geq \int_x^t (t-y)f_0(y) dy > 0, \quad \text{for } t > x.$$

So we may assume that the sequences  $(\tau_{n,1})$  and  $(\tau_{n,2})$  are bounded and have subsequences converging to finite points  $\tau_1$  and  $\tau_2$ , respectively. For convenience we denote these subsequences again by  $(\tau_{n,1})$  and  $(\tau_{n,2})$ . Suppose that

$$\tau_1 < x < \tau_2. \tag{3.7}$$

Then, by (3.6),  $\tilde{f}_n(\tau_{n,1})$  is uniformly bounded, with a uniformly bounded right derivative at  $\tau_{n,1}$ , so we can extend the function linearly on  $[0, \tau_{n,1}]$  to a convex function on  $[0, \infty)$  such that the sequence thus obtained has a convergent subsequence. So  $(\tilde{f}_n)$  has a subsequence, converging to a convex decreasing function  $\tilde{f}$ , at each point in  $(\tau_1, \infty)$ , where  $\tilde{f}(\tau_1) < \infty$ . Suppose  $\tau_1 = 0$ . Then we need to have:

$$\int_0^t (t-y)\tilde{f}(y) dy \geq \int_0^t (t-y)f_0(y) dy, \quad t \geq 0,$$

which cannot occur since  $f_0$  is unbounded near zero and  $\tilde{f}(0) < \infty$  in this case. If  $\tau_1 > 0$ , we would get

$$\frac{1}{2} \int_{[\tau_1, \infty)} \tilde{f}(y)^2 dx - \int_{[\tau_1, \infty)} \tilde{f}(y) f_0(y) dy \leq -\frac{1}{2} \int_{[\tau_1, \infty)} f_0(y)^2 dx, \tag{3.8}$$

implying  $\tilde{f}(y) = f_0(y)$ ,  $y \geq \tau_1$ . This cannot occur either, since  $\tilde{f}$  is linear on  $[\tau_1, \tau_2]$  and  $f_0$  is not linear on that interval, because  $x$  is a point of strict convexity of  $f_0$ . Since the argument can be repeated for subsequences, we can conclude that, with probability one, the point of change of slope  $\tau_n$ , closest to  $x$ , has to converge to  $x$ .  $\square$

**Remark:** It is well known that the Grenander estimator of a bounded decreasing density on  $[0, \infty)$  is inconsistent at zero. See e.g. WOODROOFE AND SUN (1993). A similar result holds for the LS estimator of a bounded convex decreasing density. Indeed, from its characterization in Lemma 2.2 we have

$$\tilde{H}_n(X_{(2)}) \geq Y_n(X_{(2)}) = (X_{(2)} - X_{(1)})/n.$$

Moreover, we have by monotonicity of  $\tilde{f}_n$  that

$$\tilde{H}_n(X_{(2)}) = \int_0^{X_{(2)}} \int_0^y \tilde{f}_n(x) dx dy \leq \frac{1}{2} \tilde{f}_n(0) X_{(2)}^2$$

Hence,

$$\tilde{f}_n(0) \geq \frac{2(X_{(2)} - X_{(1)})}{nX_{(2)}^2}$$

Using the well known representation of the order statistics as transformed rescaled cumulative sums of an exponential sample  $E_1, \dots, E_{n+1}$  (see e.g. SHORACK AND WELLNER (1986), Proposition 8.2.1, page 335), it follows that

$$\begin{aligned} \liminf_{n \rightarrow \infty} P(\tilde{f}_n(0) \geq 2f_0(0)) &\geq \liminf_{n \rightarrow \infty} P\left(\frac{2(X_{(2)} - X_{(1)})}{nX_{(2)}^2} \geq 2f_0(0)\right) = P\left(\frac{E_2}{(E_1 + E_2)^2} \geq 1\right) \\ &= P(E_1 + E_2 \leq \sqrt{E_2}) \geq P(E_1 \leq 2/9)P(E_2 \in [1/9, 1/4]) \geq p > 0. \end{aligned}$$

**Theorem 3.2** (*Consistency of MLE of density*) Suppose that  $X_1, X_2, \dots$  are i.i.d. random variables with density  $f_0 \in \mathcal{C}$ . Then the MLE is uniformly consistent on closed intervals bounded away from 0: i.e., for each  $c > 0$ , we have

$$\sup_{c \leq x < \infty} |\hat{f}_n(x) - f_0(x)| \rightarrow_{a.s.} 0. \quad (3.9)$$

**Proof:** Taking  $g = f_0$  in (2.14), it follows that

$$\int_0^\infty \frac{f_0(x)}{\hat{f}_n(x)} d\mathbb{F}_n(x) \leq 1. \quad (3.10)$$

Now by Glivenko Cantelli we have  $\Omega_0 \equiv \{\omega \in \Omega : \|\mathbb{F}_n(\cdot, \omega) - F_0\|_\infty \rightarrow 0\}$  has  $P(\Omega_0) = 1$ . Now fix  $\omega \in \Omega_0$ . Let  $\{k\}$  be an arbitrary subsequence of  $\{n\}$ . By (3.1), we can use Helly's diagonalization procedure together with the fact that a convex function is continuous to extract a further subsequence  $(n_k)$  along which  $\hat{f}_{n_k}(x) \rightarrow \hat{f}(x)$  for each  $x > 0$ , where  $\hat{f}$  is a convex decreasing function on  $(0, \infty)$ . Note that  $\hat{f}$  may depend on  $\omega$  and on the particular choices of the subsequences  $\{k\}$  and  $\{l\}$ , and that, by Fatou's lemma

$$\int_0^\infty \hat{f}(x) dx \leq 1. \quad (3.11)$$

Note also that  $\hat{f}_l \rightarrow \hat{f}$  uniformly on intervals of the form  $[c, \infty)$  for  $c > 0$ . This follows from the monotonicity of  $\hat{f}_l$  and  $\hat{f}$  and the continuity of  $\hat{f}$ .

Now define, for  $0 < \alpha < 1$ ,  $\eta_\alpha = F_0^{-1}(1 - \alpha)$ , and fix  $\epsilon > 0$  such that  $\epsilon < \eta_\epsilon$ . From (3.10) it follows that there exists a number  $\tau_\epsilon > 0$  such that for  $k$  sufficiently large  $\hat{f}_l(\eta_\epsilon) \geq \tau_\epsilon$ . Consequently, there exist numbers  $0 < c_\epsilon < C_\epsilon < \infty$ , such that for all  $k$  sufficiently large,  $c_\epsilon \leq f_0(x)/\hat{f}_{n_k}(x) \leq C_\epsilon$  whenever  $x \in [\epsilon, \eta_\epsilon]$ . Therefore, we have that

$$\sup_{x \in [\epsilon, \eta_\epsilon]} \left| \frac{f_0(x)}{\hat{f}_{n_k}(x)} - \frac{f_0(x)}{\hat{f}(x)} \right| \rightarrow 0.$$

This yields, for all  $k$  sufficiently large,

$$\int_\epsilon^{\eta_\epsilon} \frac{f_0(x)}{\hat{f}(x)} d\mathbb{F}_{n_k}(x) \leq \int_\epsilon^{\eta_\epsilon} \left( \frac{f_0(x)}{\hat{f}_{n_k}(x)} + \epsilon \right) d\mathbb{F}_{n_k}(x) \leq 1 + \epsilon,$$

where we also use (3.10). However, since  $\mathbb{F}_{n_k} \rightarrow_d F_0$  for our  $\omega$ , and  $f_0/\hat{f}$  is bounded and continuous on  $[\epsilon, \eta_\epsilon]$ , we may conclude that

$$\int_\epsilon^{\eta_\epsilon} \frac{f_0(x)}{\hat{f}(x)} dF_0(x) \leq 1 + \epsilon.$$

Since  $\epsilon > 0$  was arbitrary (yet small), we can apply the monotone convergence theorem to conclude that

$$\int_0^\infty \frac{f_0(x)^2}{\hat{f}(x)} dx \leq 1. \quad (3.12)$$

On the other hand, we have for each  $\epsilon < 1$  and continuous subdensity  $f$  that

$$0 \leq \int_\epsilon^{1/\epsilon} \frac{(f_0(x) - f(x))^2}{f(x)} dx = \int_\epsilon^{1/\epsilon} \frac{f_0(x)^2}{f(x)} dx - 2 \int_\epsilon^{1/\epsilon} f_0(x) dx + \int_\epsilon^{1/\epsilon} f(x) dx,$$

with equality only if  $f \equiv f_0$  on  $[\epsilon, 1/\epsilon]$ . Using monotone convergence, we see that for each continuous subdensity  $f$ ,

$$\int_0^\infty \frac{f_0(x)^2}{f(x)} dx \geq 1$$

with equality only if  $f \equiv f_0$ . Applying this to the subdensity  $\hat{f}$  (see (3.11)), we get that the inequality in (3.12) is an equality, which again implies that  $\hat{f} \equiv f_0$ .

Therefore, we have proved that for each  $\omega \in \Omega_0$  with  $P(\Omega_0) = 1$ , each subsequence  $\{\hat{f}_{n_k}(\cdot; \omega)\}$  of  $\{\hat{f}_n(\cdot; \omega)\}$  contains a further subsequence  $\{\hat{f}_{n_k}(\cdot; \omega)\}$  such that  $\hat{f}_{n_k}(x, \omega) \rightarrow f_0(x)$  all  $x > 0$ . Continuity of  $f_0$  and the monotonicity of  $f_0$  imply (3.9).  $\square$

**Remark:** Just as the LS estimator, the MLE is inconsistent at zero. Using the characterization of Lemma 2.4 at  $t = X_{(2)}$ , this inconsistency at zero follows analogously to that of the LS estimator.

**Lemma 3.1** *Suppose that  $\bar{f}_n$  is a sequence of functions in  $\mathcal{K}$  satisfying  $\sup_{x \geq c} |\bar{f}_n(x) - f_0(x)| \rightarrow 0$  for each  $c > 0$ . Then*

$$-\infty < f_0'(x-) \leq \liminf_{n \rightarrow \infty} \bar{f}_n'(x-) \leq \limsup_{n \rightarrow \infty} \bar{f}_n'(x+) \leq f_0'(x+) < 0 \quad (3.13)$$

for all  $x > 0$ .

**Proof:** For each  $h > 0$  (sufficiently small) the fact that  $\bar{f}_n \in \mathcal{K}$  implies that

$$\frac{\bar{f}_n(x-h) - \bar{f}_n(x)}{-h} \leq \bar{f}_n'(x-) \leq \bar{f}_n'(x+) \leq \frac{\bar{f}_n(x+h) - \bar{f}_n(x)}{h}.$$

Letting  $n \rightarrow \infty$ , we get

$$\frac{f_0(x-h) - f_0(x)}{-h} \leq \liminf_{n \rightarrow \infty} \bar{f}_n'(x-) \leq \limsup_{n \rightarrow \infty} \bar{f}_n'(x+) \leq \frac{f_0(x+h) - f_0(x)}{h}.$$

Now, letting  $h \downarrow 0$ , we obtain (3.13).  $\square$

**Corollary 3.1** *The derivatives of the MLE and LS estimator are consistent for the derivative of  $f_0$  in the sense that (3.13) holds almost surely.*

**Proof:** Combine Theorem 3.1 and 3.2 with Lemma 3.1.  $\square$

Having derived strong consistency of both density estimators, and their derivatives, we now turn to the regression problem. This problem is studied more extensively in the literature, and consistency was proved under more general conditions in HANSON AND PLEDGER (1976).

**Theorem 3.3** *(Consistency of Least Squares regression estimator) Consider model (2.22) with  $x_i$ 's contained in  $[0, 1]$ . Suppose that  $\epsilon_{n,i}$  are independent, identically and symmetrically distributed with finite exponential moment. Furthermore suppose that for each subinterval  $A$  of  $[0, 1]$  of positive Lebesgue measure,  $\liminf_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n 1_A(x_{n,i}) > 0$  almost surely. Then for each  $\epsilon \in (0, 1/2)$ ,*

$$\sup_{[\epsilon, 1-\epsilon]} |\hat{r}_n(x) - r(x)| \rightarrow 0 \text{ a.s.}$$

and for each  $x \in (0, 1)$ ,

$$-\infty < r'(x-) \leq \liminf_{n \rightarrow \infty} r'_n(x-) \leq \limsup_{n \rightarrow \infty} r'_n(x+) \leq r'(x+) < \infty$$

**Proof:** Follows from the theorem in Section 1 of HANSON AND PLEDGER (1976) and Lemma 3.1.  $\square$

## 4 Rates of convergence

A key step in establishing the rate of convergence is to show that, for the estimators considered in Sections 2.1 and 2.2, the distance between successive changes of slope of the estimator is of order  $\mathcal{O}_p(n^{-1/5})$ . A similar result was established for the estimator considered in Section 2.3 in MAMMEN (1991). The result is given in Lemma 4.2. Using Lemma 4.2, we will prove  $n^{-2/5}$ -tightness of the estimators in Lemma 4.4, and  $n^{-1/5}$ -tightness of their derivatives. This will prove to be crucial in Section 6.

As in the previous section, we denote by  $\mathcal{T}_n$  the set of changes of slope of the estimator under consideration.

**Lemma 4.1** *Let  $x_0$  be an interior point of the support of  $f_0$ . Then:*

(i) *Let, for  $0 < x \leq y$ , the random function  $U_n(x, y)$  be defined by*

$$U_n(x, y) = \int_{[x, y]} \left\{ z - \frac{1}{2}(x + y) \right\} d(\mathbb{F}_n - F_0)(z), \quad y \geq x. \quad (4.1)$$

*Then there exist constants  $\delta > 0$  and  $c_0 > 0$  such that, for each  $\epsilon > 0$  and each  $x$  satisfying  $|x - x_0| < \delta$ :*

$$|U_n(x, y)| \leq \epsilon(y - x)^4 + \mathcal{O}_p\left(n^{-4/5}\right), \quad 0 \leq y - x \leq c_0. \quad (4.2)$$

(ii) Let, for  $0 < x \leq y$ , and  $x$  in a neighborhood of  $x_0$ , the random function  $V_n(x, y)$  be defined by

$$V_n(x, y) = \int_{[x, y]} \frac{z - \frac{1}{2}(x + y)}{\hat{f}_n(z)} d(\mathbb{F}_n - F_0)(z), \quad y \geq x, \quad (4.3)$$

where  $\hat{f}_n$  is the MLE. Then there exist constants  $\delta > 0$  and  $c_0 > 0$  such that, for each  $\epsilon > 0$  and each  $x$  satisfying  $|x - x_0| < \delta$ :

$$V_n(x, y) = \epsilon(y - x)^4 (1 + o_p(1)) + \mathcal{O}_p\left(n^{-4/5}\right), \quad 0 \leq y - x \leq c_0, \quad (4.4)$$

**Proof:** ad (i). We have:

$$\sup_{y: 0 \leq y - x \leq R} |U_n(x, y)| = \sup_{y: 0 \leq y - x \leq R} |(\mathbb{P}_n - P)(f_{x, y})|,$$

where

$$f_{x, y}(z) = (z - x)1_{[x, y]}(z) - \frac{1}{2}(y - x)1_{[x, y]}(z), \quad y \geq x.$$

But the collection of functions

$$\mathcal{F}_{x, R} = \{f_{x, y}(z) : x \leq y \leq x + R\}$$

is a VC-subgraph class of functions with envelope function

$$F_{x, R}(z) = (z - x)1_{[x, x+R]}(z) + \frac{1}{2}R1_{[x, x+R]}(z),$$

so that

$$EF_{x, R}^2(X_1) = \frac{1}{3}R^3 \{f_0(x_0) + O(1)\} + \frac{1}{4}R^2 \{F_0(x + R) - F_0(x)\} = \frac{7}{12}R^3 \{f_0(x_0) + O(1)\}. \quad (4.5)$$

for  $x$  in some appropriate neighborhood  $[x_0 - \delta, x_0 + \delta]$  of  $x_0$ . It now follows from Theorem 2.14.1 in VAN DER VAART AND WELLNER (1996) that

$$E \left\{ \left( \sup_{f_{x, y} \in \mathcal{F}_{x, R}} |(\mathbb{P}_n - P)(f_{x, y})| \right)^2 \right\} \leq \frac{1}{n} KEF_{x, R}^2 = \mathcal{O}(n^{-1}R^3)$$

for small values of  $R$  and a constant  $K > 0$ .

Hence there exists a  $\delta > 0$  such that, for  $\epsilon > 0$ ,  $A > 0$  and  $jn^{-1/5} \leq \delta$ :

$$\begin{aligned} & P \left\{ \exists u \in \left[ (j-1)n^{-1/5}, jn^{-1/5} \right) : n^{4/5} |U_n(x, x+u)| > A + \epsilon(j-1)^4 \right\} \\ & \leq cn^{8/5} E \left\{ \|\mathbb{P}_n - P\|_{\mathcal{F}_{x, jn^{-1/5}}} \right\}^2 / \{A + \epsilon(j-1)^4\}^2 \\ & \leq c'j^3 / \{A + \epsilon(j-1)^4\}^2 \end{aligned} \quad (4.6)$$



for constants  $c, c' > 0$ , independent of  $x \in [x_0 - \delta, x_0 + \delta]$ . The result now easily follows, see, e.g., KIM AND POLLARD (1990), page 201, for an analogous argument in the case of “cube root  $n$ ” instead of “fifth root  $n$ ” asymptotics.

Part (ii) is proved in a similar way, using the fact that we can choose a neighborhood of  $x_0$  such that, for  $x$  in this neighborhood,

$$\hat{f}_n(x) \geq \frac{1}{2}f_0(x_0)(1 + o_p(1)), \quad n \rightarrow \infty.$$

□

The proof that the distance between successive changes of slope of the LS estimator and the MLE is of order  $\mathcal{O}_p(n^{-1/5})$  will be based on the characterizations of these estimators, developed in Section 2.

**Lemma 4.2** *Let  $x_0$  be a point at which  $f_0$  has a continuous and strictly positive second derivative. Let  $\xi_n$  be an arbitrary sequence of numbers converging to  $x_0$  and define  $\tau_n^- = \max\{t \in \mathcal{T}_n : t \leq \xi_n\}$  and  $\tau_n^+ = \min\{t \in \mathcal{T}_n : t > \xi_n\}$  (of course  $\mathcal{T}_n$  for the MLE and LS estimator are different). Then,*

$$\tau_n^+ - \tau_n^- = \mathcal{O}_p(n^{-1/5})$$

for both the LS estimator and MLE.

**Proof:** We first prove the result for the LS estimator. Let  $\tau_n^-$  be the last point of change of slope of  $\tilde{H}_n'' < \xi_n$  and  $\tau_n^+$  the first point of change of slope of  $\tilde{H}_n'' \geq \xi_n$ . Note that, since the number of changes of slope is bounded above by  $n$  by Lemma 2.1, we can only have strict changes of slope. Moreover, let  $\tau_n$  be the midpoint of the interval  $[\tau_n^-, \tau_n^+]$ . Then, by the characterization of Lemma 2.2:

$$\tilde{H}_n(\tau_n) \geq Y_n(\tau_n).$$

Using (2.7), this can be written:

$$\frac{1}{2} \{Y_n(\tau_n^-) + Y_n(\tau_n^+)\} - \frac{1}{8} \{\mathbb{F}_n(\tau_n^+) - \mathbb{F}_n(\tau_n^-)\} (\tau_n^+ - \tau_n^-) \geq Y_n(\tau_n). \quad (4.7)$$

Replacing  $Y_n$  and  $\mathbb{F}_n$  by their deterministic counterparts, and expanding the integrands at  $\tau_n$ , we get for large  $n$ :

$$\begin{aligned} & \int_{\tau_n}^{\tau_n^+} \{\tau_n^+ - x\} f_0(x) dx + \int_{\tau_n^-}^{\tau_n} \{x - \tau_n^-\} f_0(x) dx - \frac{1}{4} (\tau_n^+ - \tau_n^-) \int_{\tau_n^-}^{\tau_n^+} f_0(x) dx \\ &= \int_{[\tau_n^-, \tau_n]} \left\{ \frac{1}{2} (\tau_n^- + \tau_n) - x \right\} f_0(x) dx + \int_{[\tau_n, \tau_n^+]} \left\{ x - \frac{1}{2} (\tau_n + \tau_n^+) \right\} f_0(x) dx \\ &= -\frac{1}{384} f_0''(\tau_n) (\tau_n^+ - \tau_n^-)^4 + o_p(\tau_n^+ - \tau_n^-)^4, \end{aligned}$$

using the consistency of  $\tilde{f}_n$  to ensure that  $\tau_n$  belongs to a sufficiently small neighborhood of  $x_0$  to allow this expansion. But, by Lemma 4.1 and the inequality (4.7), this implies:

$$-\frac{1}{384} f_0''(x_0) (\tau_n^+ - \tau_n^-)^4 + \mathcal{O}_p(n^{-4/5}) + o_p(\tau_n^+ - \tau_n^-)^4 \geq 0.$$

Hence:

$$\tau_n^+ - \tau_n^- = \mathcal{O}_p\left(n^{-1/5}\right).$$

Similarly, for the MLE, let  $\tau_n^-$  be the last point of change of slope  $< \xi_n$  and  $\tau_n^+$  the first point of change of slope  $\geq \xi_n$ . Moreover, let  $\tau_n$  be the midpoint of the interval  $[\tau_n^-, \tau_n^+]$ . Then, by the characterization of Lemma 2.4:

$$H_n(\tau_n) \leq \tau_n^2/2.$$

Using (2.13), this can be written:

$$\begin{aligned} & \int_{[\tau_n^-, \tau_n]} \frac{\tau_n - x}{\hat{f}_n(x)} d\mathbb{F}_n(x) + \int_{[\tau_n, \tau_n^+]} \frac{x - \tau_n}{\hat{f}_n(x)} d\mathbb{F}_n(x) - \frac{1}{4} (\tau_n^+ - \tau_n^-)^2 \\ &= \int_{[\tau_n^-, \tau_n]} \frac{\tau_n - x - \frac{1}{4} (\tau_n^+ - \tau_n^-)}{\hat{f}_n(x)} d\mathbb{F}_n(x) + \int_{[\tau_n, \tau_n^+]} \frac{x - \tau_n - \frac{1}{4} (\tau_n^+ - \tau_n^-)}{\hat{f}_n(x)} d\mathbb{F}_n(x) \\ &= \int_{[\tau_n^-, \tau_n]} \frac{\frac{1}{2} (\tau_n^- + \tau_n) - x}{\hat{f}_n(x)} d\mathbb{F}_n(x) + \int_{[\tau_n, \tau_n^+]} \frac{x - \frac{1}{2} (\tau_n + \tau_n^+)}{\hat{f}_n(x)} d\mathbb{F}_n(x) \\ &\leq 0, \end{aligned}$$

where we used (2.16) to obtain the first equality. But we have:

$$\begin{aligned} & \int_{[\tau_n^-, \tau_n]} \frac{\frac{1}{2} (\tau_n^- + \tau_n) - x}{\hat{f}_n(x)} d\mathbb{F}_n(x) + \int_{[\tau_n, \tau_n^+]} \frac{x - \frac{1}{2} (\tau_n + \tau_n^+)}{\hat{f}_n(x)} d\mathbb{F}_n(x) \\ &= \int_{[\tau_n^-, \tau_n]} \frac{\frac{1}{2} (\tau_n^- + \tau_n) - x}{\hat{f}_n(x)} d(\mathbb{F}_n - F_0)(x) + \int_{[\tau_n, \tau_n^+]} \frac{x - \frac{1}{2} (\tau_n + \tau_n^+)}{\hat{f}_n(x)} d(\mathbb{F}_n - F_0)(x) \\ &\quad + \int_{[\tau_n^-, \tau_n]} \frac{\frac{1}{2} (\tau_n^- + \tau_n) - x}{\hat{f}_n(x)} dF_0(x) + \int_{[\tau_n, \tau_n^+]} \frac{x - \frac{1}{2} (\tau_n + \tau_n^+)}{\hat{f}_n(x)} dF_0(x). \end{aligned}$$

Here we use that  $\tau_n^+ - \tau_n^- = o_p(1)$ , which is implied by the consistency of  $\hat{f}_n$  and the fact that  $f_0''(x_0) > 0$  and  $f_0''$  is continuous at  $x_0$  ( $\hat{f}_n$  cannot be linear on an interval of length bounded away from zero in a neighborhood of  $x_0$ ). Now note that we have:

$$\begin{aligned} & \int_{[\tau_n^-, \tau_n]} \frac{\frac{1}{2} (\tau_n^- + \tau_n) - x}{\hat{f}_n(x)} dF_0(x) + \int_{[\tau_n, \tau_n^+]} \frac{x - \frac{1}{2} (\tau_n + \tau_n^+)}{\hat{f}_n(x)} dF_0(x) \\ &= \int_{[\tau_n^-, \tau_n]} \left\{ \frac{1}{2} (\tau_n^- + \tau_n) - x \right\} \left\{ \frac{1}{\hat{f}_n(x)} - \frac{1}{f_0(x)} \right\} dF_0(x) \\ &\quad + \int_{[\tau_n, \tau_n^+]} \left\{ x - \frac{1}{2} (\tau_n + \tau_n^+) \right\} \left\{ \frac{1}{\hat{f}_n(x)} - \frac{1}{f_0(x)} \right\} dF_0(x) \\ &= \frac{1}{192} f_0''(x_0) (\tau_n^+ - \tau_n^-)^4 + o_p\left((\tau_n^+ - \tau_n^-)^4\right), \end{aligned}$$

expanding the functions  $f_0$  and  $\hat{f}_n$  at  $\tau_n$ , and using the linearity of  $\hat{f}_n$  on  $[\tau_n^-, \tau_n^+]$  and the consistency of  $\hat{f}_n$  and  $\hat{f}'_n$ . Moreover, again using  $\tau_n^+ - \tau_n^- = o_p(1)$ , we have that

$$\inf_{x \in [\tau_n^-, \tau_n^+]} \hat{f}_n(x) > \frac{1}{2} f_0(x_0) + o_p(1),$$

and therefore

$$\begin{aligned} & \int_{[\tau_n^-, \tau_n]} \frac{\frac{1}{2}(\tau_n^- + \tau_n) - x}{\hat{f}_n(x)} d(\mathbb{F}_n - F_0)(x) + \int_{[\tau_n, \tau_n^+]} \frac{x - \frac{1}{2}(\tau_n + \tau_n^+)}{\hat{f}_n(x)} d(\mathbb{F}_n - F_0)(x) \\ &= \mathcal{O}_p(n^{-4/5}) + o_p((\tau_n^+ - \tau_n^-)^4), \end{aligned}$$

using part (ii) of Lemma 4.1. Combining these results we obtain

$$f_0''(x_0) (\tau_n^+ - \tau_n^-)^4 + \mathcal{O}_p(n^{-4/5}) + o_p((\tau_n^+ - \tau_n^-)^4) \leq 0$$

This again implies

$$\tau_n^+ - \tau_n^- = \mathcal{O}_p(n^{-1/5}).$$

□

Having established the order of the difference of successive points of changes of slope of  $\tilde{H}_n''$  and  $H_n''$ , we can turn the consistency result into a rate result saying that there will, with high probability, be a point in an  $\mathcal{O}_p(n^{-1/5})$  neighborhood of  $x_0$  where the difference between the estimator and the estimand will be of order  $n^{-2/5}$ . The lemma below has the exact statement.

**Lemma 4.3** *Suppose  $f_0'(x_0) < 0$ ,  $f_0''(x_0) > 0$ , and  $f_0''$  is continuous in a neighbourhood of  $x_0$ . Let  $\xi_n$  be a sequence converging to  $x_0$ . Then for any  $\epsilon > 0$  there exists an  $M > 1$  and a  $c > 0$  such that the following holds with probability bigger than  $1 - \epsilon$ . There are bend points  $\tau_n^- < \xi_n < \tau_n^+$  of  $\tilde{f}_n$  with  $2n^{-1/5} \leq \tau_n^+ - \tau_n^- \leq 2Mn^{-1/5}$  and for any of such points we have that*

$$\inf_{t \in [\tau_n^-, \tau_n^+]} |f_0(t) - \tilde{f}_n(t)| < cn^{-2/5} \quad \text{for all } n.$$

The same result holds for  $\hat{f}_n$  instead of  $\tilde{f}_n$ .

**Proof.** Fix  $\epsilon > 0$  and observe that Lemma 4.2 applied to the sequences  $\xi_n \pm n^{-1/5}$ , gives that there is an  $M > 0$  such that with probability bigger than  $1 - \epsilon$ , there exist jump-points  $\tau_n^-$  and  $\tau_n^+$  of  $\tilde{f}'_n$  (or  $\hat{f}'_n$ ) satisfying  $\xi_n - Mn^{-1/5} \leq \tau_n^- \leq \xi_n - n^{-1/5} \leq \xi_n + n^{-1/5} \leq \tau_n^+ \leq \xi_n + Mn^{-1/5}$  for all  $n$ .

First consider the LS estimator  $\tilde{f}_n$ . Let  $\tau_n^- < \tau_n^+$  be such points of jump. Fix  $c > 0$  and consider the event

$$\inf_{t \in [\tau_n^-, \tau_n^+]} |f_0(t) - \tilde{f}_n(t)| \geq cn^{-2/5}. \quad (4.8)$$

On this set we have:

$$\left| \int_{\tau_n^-}^{\tau_n^+} (f_0(t) - \tilde{f}_n(t)) (\tau_n^+ - t) dt \right| \geq \frac{1}{2} cn^{-2/5} (\tau_n^+ - \tau_n^-)^2.$$

On the other hand, the equality conditions in (2.2) imply:

$$\begin{aligned} 0 &= \int_{[\tau_n^-, \tau_n^+]} (\tau_n^+ - t) d(\tilde{F}_n - \mathbb{F}_n)(t) \\ &= \int_{\tau_n^-}^{\tau_n^+} \left\{ \tilde{f}_n(t) - f_0(t) \right\} (\tau_n^+ - t) dt - \int_{[\tau_n^-, \tau_n^+]} (\tau_n^+ - t) d(\mathbb{F}_n - F_0)(t). \end{aligned}$$

Therefore, by (4.8),

$$\left| \int_{[\tau_n^-, \tau_n^+]} (\tau_n^+ - t) d(\mathbb{F}_n - F_0)(t) \right| \geq \frac{1}{2} cn^{-2/5} (\tau_n^+ - \tau_n^-)^2 \geq 2cn^{-4/5}. \quad (4.9)$$

But the collection of functions

$$\mathcal{F}_{x,R} = \{f_{x,y}(z) : x \leq y \leq x + R\}$$

where

$$f_{x,y}(z) = (y - z)1_{[x,y]}(z), \quad y \geq x$$

is a VC-subgraph class of functions with envelope function

$$F_{x,R}(z) = R1_{[x,x+R]}(z),$$

so that

$$EF_{x,R}^2(X_1) = R^2\{F_0(x+R) - F_0(x)\} = R^3\{f_0(x_0) + o(1)\}. \quad (4.10)$$

for  $x$  in some appropriate neighborhood  $[x_0 - \delta, x_0 + \delta]$  of  $x_0$ . Therefore, just as in Lemma 4.1, we get

$$\left| \int_{[\tau_n^-, \tau_n^+]} (\tau_n^+ - t) d(\mathbb{F}_n - F_0)(t) \right| = \mathcal{O}_p(n^{-4/5}) + o_p((\tau_n^+ - \tau_n^-)^4) = \mathcal{O}_p(n^{-4/5}).$$

So the probability of (4.8) can be made arbitrarily small by taking  $c$  sufficiently big. This proves the result for  $\tilde{f}_n$ .

Now consider the MLE  $\hat{f}_n$ . We get from (i) of Lemma 2.4 that

$$\begin{aligned} 0 &= \hat{H}_n(\tau_n^+) - \frac{1}{2}\tau_n^{+2} - \hat{H}_n(\tau_n^-) + \frac{1}{2}\tau_n^{-2} - (\hat{H}'_n(\tau_n^-) - \tau_n^-)(\tau_n^+ - \tau_n^-) \\ &= \int_{t=\tau_n^-}^{\tau_n^+} \int_{u=\tau_n^-}^t \frac{d\mathbb{F}_n(u)}{\hat{f}_n(u)} - \frac{1}{2}(\tau_n^+ - \tau_n^-)^2 = \\ &= \int_{t=\tau_n^-}^{\tau_n^+} (\tau_n^+ - t) \frac{f_0(t) - \hat{f}_n(t)}{\hat{f}_n(t)f_0(t)} d\mathbb{F}_n(t) - \int_{t=\tau_n^-}^{\tau_n^+} \frac{\tau_n^+ - t}{f_0(t)} d(\mathbb{F}_n - F_0)(t). \end{aligned}$$

Under (4.8) (with  $\hat{f}_n$  instead of  $\tilde{f}_n$ ), the absolute value of the first term in this decomposition will be bounded below asymptotically by  $2cf_0(x_0)^{-1}n^{-4/5}$ , whereas the second term is  $\mathcal{O}_P(n^{-4/5})$ .  $\square$

Using Lemma 4.3 monotonicity of the derivatives of the estimators and the limit density  $f_0$ , we obtain the local  $n^{-2/5}$ -consistency of the density estimators and  $n^{-1/5}$ -consistency of their derivatives.

**Lemma 4.4** Suppose  $f'_0(x_0) < 0$ ,  $f''_0(x_0) > 0$ , and  $f''_0$  is continuous in a neighborhood of  $x_0$ . Then, for  $\bar{f}_n = \bar{f}_n$  or  $\hat{f}_n$ , the following holds. For each  $M > 0$

$$\sup_{|t| \leq M} |\bar{f}_n(x_0 + n^{-1/5}t) - f_0(x_0) - n^{-1/5}t f'_0(x_0)| = \mathcal{O}_p(n^{-2/5}) \quad (4.11)$$

and, interpreting  $\bar{f}'_n$  as left- or right derivative

$$\sup_{|t| \leq M} |\bar{f}'_n(x_0 + n^{-1/5}t) - f'_0(x_0)| = \mathcal{O}_p(n^{-1/5}). \quad (4.12)$$

**Proof.** We start proving (4.12). Fix  $x_0$ ,  $M > 0$  and  $\epsilon > 0$ . Define  $\sigma_{n,1}$  to be the first point of change of slope after  $x_0 + Mn^{-1/5}$ ,  $\sigma_{n,2}$  the first point of change of slope after  $\sigma_{n,1} + n^{-1/5}$  and  $\sigma_{n,3}$  the first point of change of slope after  $\sigma_{n,2} + n^{-1/5}$ . Define the points  $\sigma_{n,i}$  for  $i = -1, -2, -3$  similarly, but then argued from  $x_0$  to the left. Then, according to Lemma 4.3 there are numbers  $\xi_{n,i} \in (\sigma_{n,i}, \sigma_{n,i+1})$  ( $i = 1, 2$ ) and  $\xi_{n,i} \in (\sigma_{n,i-1}, \sigma_{n,i})$  ( $i = -1, -2$ ) and  $c > 0$ , so that, with probability bigger than  $1 - \epsilon$ ,  $|\bar{f}_n(\xi_{n,i}) - f_0(\xi_{n,i})| \leq cn^{-2/5}$ . Hence, we have for each  $t \in [x_0 - Mn^{-1/5}, x_0 + Mn^{-1/5}]$  with probability bigger than  $1 - \epsilon$  that

$$\bar{f}_n(t-) \leq \bar{f}'_n(t+) \leq \bar{f}'_n(\xi_1) \leq \frac{\bar{f}_n(\xi_2) - \bar{f}_n(\xi_1)}{\xi_2 - \xi_1} \leq \frac{f_0(\xi_2) - f_0(\xi_1) + 2cn^{-2/5}}{\xi_2 - \xi_1} \leq f'_0(\xi_2) + 2cn^{-1/5}.$$

In the final step we use that  $\xi_2 - \xi_1 \geq n^{-1/5}$ . Similarly, we get that for each  $t \in [x_0 - Mn^{-1/5}, x_0 + Mn^{-1/5}]$ , with probability above  $1 - \epsilon$  that

$$\bar{f}'_n(t+) \geq \bar{f}_n(t-) \geq f'_0(\xi_{-2}) - 2cn^{-1/5}.$$

Using that  $\xi_{\pm 2} = x_0 + \mathcal{O}_p(n^{-1/5})$  and smoothness of  $f'_0$ , we obtain (4.12).

Now consider (4.11). Fix  $M > 0$  and  $\epsilon > 0$ . By Lemma 4.2, we can find a  $K > M$  such that there will be at least two points of change of slope at mutual distance at least  $n^{-1/5}$  in both the intervals  $[x_0 - Kn^{-1/5}, x_0 - Mn^{-1/5}]$  and  $[x_0 + Mn^{-1/5}, x_0 + Kn^{-1/5}]$  with probability exceeding  $1 - \epsilon$ . From Lemma 4.3 we know that then there are points  $\xi_{-1} \in [x_0 - Kn^{-1/5}, x_0 - Mn^{-1/5}]$  and  $\xi_1 \in [x_0 + Mn^{-1/5}, x_0 + Kn^{-1/5}]$  such that  $|\bar{f}_n(\xi_{n,i}) - f_0(\xi_{n,i})| \leq cn^{-2/5}$  for  $i = -1, 1$ .

From (4.12) we know that a  $c'$  can be chosen to get the probability of

$$\sup_{t \in [x_0 - Kn^{-1/5}, x_0 + Kn^{-1/5}]} |\bar{f}'_n(t) - f'_0(x_0)| \leq c'n^{-1/5}$$

bigger than  $1 - \epsilon$ . Hence, with probability bigger than  $1 - 3\epsilon$ , we have for any  $t \in [x_0 - Mn^{-1/5}, x_0 + Mn^{-1/5}]$  for  $n$  sufficiently large that

$$\begin{aligned} \bar{f}_n(t) &\geq \bar{f}_n(\xi_1) + \bar{f}'_n(\xi_1)(t - \xi_1) \geq f_0(\xi_1) - cn^{-2/5} + (f'_0(x_0) - c'n^{-1/5})(t - \xi_1) \\ &\geq f_0(x_0) + (\xi_1 - x_0)f'_0(x_0) + f'_0(x_0)(t - \xi_1) - (c + 2Kc')n^{-2/5} \\ &= f_0(x_0) + (t - x_0)f'_0(x_0) - (c + 2Kc')n^{-2/5}. \end{aligned}$$

For the reverse inequality, we use convexity again, but now “from above”. Indeed, for  $t \in [x_0 - Mn^{-1/5}, x_0 + Mn^{-1/5}]$  and  $n$  sufficiently large we have that

$$\begin{aligned}
\bar{f}_n(t) &\leq \bar{f}_n(\xi_{-1}) + \frac{\bar{f}_n(\xi_1) - \bar{f}_n(\xi_{-1})}{\xi_1 - \xi_{-1}}(t - \xi_{-1}) \\
&\leq f_0(\xi_{-1}) + cn^{-2/5} + \frac{f_0(\xi_1) - f_0(\xi_{-1}) + 2cn^{-2/5}}{\xi_1 - \xi_{-1}}(t - \xi_{-1}) \\
&\leq f_0(x_0) + (\xi_{-1} - x_0)f'_0(x_0) + \frac{1}{2}(\xi_{-1} - x_0)^2 f''_0(\nu_{1,n}) \\
&\quad + \frac{t - \xi_{-1}}{\xi_1 - \xi_{-1}} \left( f_0(x_0) + (\xi_1 - x_0)f'_0(x_0) + \frac{1}{2}(\xi_1 - x_0)^2 f''_0(\nu_{2,n}) \right. \\
&\quad \left. - f_0(x_0) - (\xi_{-1} - x_0)f'_0(x_0) - \frac{1}{2}(\xi_{-1} - x_0)^2 f''_0(\nu_{3,n}) \right) + (c + c/M)n^{-2/5} \\
&\leq f_0(x_0) + (t - x_0)f'_0(x_0) + f''_0(x_0)(K^2 + K^3/M)n^{-2/5} + (c + c/M)n^{-2/5}
\end{aligned}$$

and the result follows.  $\square$

In the case of convex regression, MAMMEN (1991) established (a result more general than) the first part of the following lemma. As in Theorem 3.3 we will assume that all the  $x_i$ 's are in  $[0, 1]$ .

**Assumption 4.1.** The design points  $x_i = x_{n,i}$  satisfy

$$\frac{c}{n} \leq x_{n,i+1} - x_{n,i} \leq \frac{C}{n}, \quad i = 1, \dots, n$$

for some constants  $0 < c < C < \infty$ .

**Assumption 4.2.** The  $\epsilon_i$ 's are i.i.d. with  $E \exp(t\epsilon_1^2) < \infty$  for some  $t > 0$ .

**Lemma 4.5** *Suppose  $r'(x_0) < 0$ ,  $r''(x_0) > 0$ ,  $r''$  is continuous in a neighborhood of  $x_0$ , and also assume that Assumptions 4.1 and 4.2 hold. Then the least squares estimator  $\hat{r}_n$  satisfies the following: for each  $M > 0$*

$$\sup_{|t| \leq M} |\hat{r}_n(x_0 + n^{-1/5}t) - r(x_0) - n^{-1/5}tr'(x_0)| = \mathcal{O}_p(n^{-2/5}) \quad (4.13)$$

and, interpreting  $\hat{r}'_n$  as left- or right derivative,

$$\sup_{|t| \leq M} |\hat{r}'_n(x_0 + n^{-1/5}t) - r'(x_0)| = \mathcal{O}_p(n^{-1/5}). \quad (4.14)$$

**Proof.** The first assertion with  $M = 0$  follows from Theorem 4 of MAMMEN (1991), and in fact the result with a supremum over  $|t| \leq M$  follows from his methods. The second assertion follows along the lines of our proofs in the density case.  $\square$

## 5 Asymptotic lower bounds for the minimax risk

In this section we briefly describe local asymptotic minimax lower bounds for the behavior of any estimator of a convex density function at a point  $x_0$  for which the second derivative exists and is positive. A similar treatment is possible for the corresponding regression setting, but we will treat only the density case here. The results of this section are from JONGBLOED (1995). See also JONGBLOED (2000).

Let the class of densities  $\mathcal{C}$  be defined by

$$\mathcal{C} = \left\{ f : [0, \infty) \rightarrow [0, \infty) : \int_0^\infty f(x) dx = 1, f \text{ is convex and decreasing} \right\}.$$

We will derive asymptotic lower bounds for the local minimax risks for estimating the convex and decreasing density  $f$  and its derivative at a fixed point. First some definitions. The  $(L_1-)$  minimax risk for estimating a functional  $T$  of  $f_0$  based on a sample  $X_1, X_2, \dots, X_n$  of size  $n$  from  $f_0$  which is known to be in a suitable subset  $\mathcal{C}_n$  of  $\mathcal{C}$ , is defined by

$$\text{MMR}_1(n, T, \mathcal{C}_n) = \inf_{t_n} \sup_{f \in \mathcal{C}_n} E_f |T_n - Tf|.$$

Here the infimum ranges over all possible measurable functions  $t_n : \mathbb{R}^n \rightarrow \mathbb{R}$ , and  $T_n = t_n(X_1, \dots, X_n)$ . When the subclasses  $\mathcal{C}_n$  are taken to be shrinking to one fixed  $f_0 \in \mathcal{C}$ , the minimax risk is called *local* at  $f_0$ . The shrinking classes (parametrized by  $\tau > 0$ ) used here are Hellinger balls centered at  $f_0$ :

$$\mathcal{C}_{n,\tau} = \left\{ f \in \mathcal{C} : H^2(f, f_0) = \frac{1}{2} \int_0^\infty \left( \sqrt{f(z)} - \sqrt{f_0(z)} \right)^2 dz \leq \tau/n \right\}.$$

The behavior, for  $n \rightarrow \infty$ , of such a local minimax risk  $\text{MMR}_1$ , will depend on  $n$  (rate of convergence to zero) and the density  $f_0$  where the subclasses shrink towards. The following lemma will be the key to the lower bound.

**Lemma 5.1** *Assume that there exists some subset  $\{f_\epsilon : \epsilon > 0\}$  of densities in  $\mathcal{C}$  such that, as  $\epsilon \downarrow 0$ ,*

$$H^2(f_\epsilon, f_0) \leq \epsilon(1 + o(1)) \text{ and } |Tf_\epsilon - Tf_0| \geq (c\epsilon)^r(1 + o(1))$$

*for some  $c > 0$  and  $r > 0$ . Then*

$$\sup_{\tau > 0} \liminf_{n \rightarrow \infty} n^r \text{MMR}_1(n, T, \mathcal{C}_{n,\tau}) \geq \frac{1}{4} \left( \frac{cr}{2e} \right)^r.$$

PROOF. By Lemma 4.1 in GROENEBOOM (1996), we get that for each  $\tau > 0$ ,

$$\text{MMR}_1(n, T, \mathcal{C}_{n,\tau}) \geq \frac{1}{4} |Tf_{\tau/n} - Tf_0| (1 - H^2(f_{\tau/n}, f_0))^{2n},$$

so that

$$\liminf_{n \rightarrow \infty} n^r \text{MMR}_1(n, T, \mathcal{C}_{n,\tau}) \geq \frac{1}{4} (c\tau)^r e^{-2\tau}.$$

Maximizing this lower bound with respect to  $\tau > 0$  gives the desired result.  $\square$

**Remark.** The argument used in the proof of Lemma 5.1, bounding the minimax risk from below by the modulus of continuity of the functional  $T$ , appeared (probably) for the first time in DONOHO AND LIU (1987). We want to thank a referee for pointing this out to us.

The functionals to be considered are, for some  $x_0 > 0$ ,

$$T_1 f = f(x_0) \quad \text{and} \quad T_2 f = f'(x_0). \quad (5.1)$$

Let  $f \in \mathcal{C}$  and  $x_0 > 0$  be fixed such that  $f_0$  is twice continuously differentiable at  $x_0$ . Using one family  $\{f_\epsilon : \epsilon > 0\}$  of densities, we will derive asymptotic lower bounds on the minimax risks for estimating  $T_1$  and  $T_2$  over  $\mathcal{C}$ .

Define, for  $\epsilon > 0$ , the functions  $\tilde{f}_\epsilon$  as follows:

$$\tilde{f}_\epsilon(z) = \begin{cases} f_0(x_0 - c_\epsilon \epsilon) + (z - x_0 + c_\epsilon \epsilon) f_0'(x_0 - c_\epsilon \epsilon), & \text{for } z \in (x_0 - c_\epsilon \epsilon, x_0 - \epsilon) \\ f_0(x_0 + \epsilon) + f_0'(x_0 + \epsilon)(z - x_0 - \epsilon), & \text{for } z \in [x_0 - \epsilon, x_0 + \epsilon) \\ f_0(z), & \text{elsewhere.} \end{cases}$$

Here  $c_\epsilon$  is chosen such that  $\tilde{f}_\epsilon$  is continuous at  $x_0 - \epsilon$ . The function  $f_\epsilon$  is then obtained from  $\tilde{f}_\epsilon$  by adding a linear correction term for the fact that  $\tilde{f}_\epsilon$  does not integrate to one,

$$f_\epsilon(z) = \tilde{f}_\epsilon(z) + \tau_\epsilon(x_0 - \epsilon - z)1_{[0, x_0 - \epsilon]}(z).$$

Obviously, for  $\epsilon \downarrow 0$ ,

$$|T_1(f_\epsilon - f_0)| = \frac{1}{2} f_0''(x_0) \epsilon^2 + o(\epsilon^2) \quad (5.2)$$

and

$$|T_2(f_\epsilon - f_0)| = f_0''(x_0) \epsilon + o(\epsilon). \quad (5.3)$$

Moreover, for the functions  $f_\epsilon$  we have the following lemma.

**Lemma 5.2** For  $\epsilon \downarrow 0$ ,

$$H^2(f_\epsilon, f_0) = \frac{2f_0''(x_0)^2}{5f_0(x_0)} \epsilon^5 + o(\epsilon^5) \equiv \nu_0 \epsilon^5 + o(\epsilon^5).$$

For the proof of this Lemma we refer to JONGBLOED (1995), sections 6.2 and 6.4, pages 110-111 and 121-122. From Lemma 5.2, (5.2) and (5.3), it follows that

$$\left| T_1 f_{(\epsilon/\nu_0)^{1/5}} - T_1 f_0 \right| \geq \left( \frac{5f_0(x_0) \sqrt{f_0''(x_0)} \epsilon}{8\sqrt{2}} \right)^{2/5} (1 + o(1))$$

and

$$\left| T_2 f_{(\epsilon/\nu_0)^{1/5}} - T_2 f_0 \right| \geq \left( \frac{5}{2} f_0(x_0) f_0''(x_0)^3 \epsilon \right)^{1/5} (1 + o(1))$$

as  $\epsilon \downarrow 0$ . An application of Lemma 5.1 finishes the proof of the following theorem.



**Theorem 5.1** For the functionals  $T_1$  and  $T_2$  as defined in (5.1),

$$\sup_{\tau > 0} \liminf_{n \rightarrow \infty} n^{2/5} \text{MMR}_1(n, T_1, \mathcal{C}_{n,\tau}) \geq \frac{1}{4} \left( \frac{f_0(x_0) \sqrt{f_0''(x_0)}}{8e\sqrt{2}} \right)^{2/5}$$

and

$$\sup_{\tau > 0} \liminf_{n \rightarrow \infty} n^{1/5} \text{MMR}_1(n, T_2, \mathcal{C}_{n,\tau}) \geq \frac{1}{4} \left( \frac{1}{4} f_0(x_0) f_0''(x_0)^3 e^{-1} \right)^{1/5}.$$

The constants appearing in these lower bounds will appear again in the asymptotic distributions of the maximum likelihood and least squares estimators in Section 6.

## 6 Asymptotic distribution theory

In this section we will establish the pointwise asymptotic distribution of the estimators introduced in Section 2. We will do this in three steps. The first is to show that for all estimators considered, the characterizations can be localized in an appropriate sense. Some terms in this “local characterization” can be shown to converge to a limiting process involving integrated Brownian motion.

Using the results of Section 4, we will see that the limiting distributions can be expressed in terms of a function related to integrated Brownian motion. This *envelope* function is studied in depth in GROENEBOOM, JONGBLOED AND WELLNER (2001A), from which we use the following result.

**Theorem 6.1** (Theorem 2.1 and Corollary 2.1(ii) in GROENEBOOM, JONGBLOED AND WELLNER (2001A)). Let  $X(t) = W(t) + 4t^3$  where  $W(t)$  is standard two-sided Brownian motion starting from 0, and let  $Y$  be the integral of  $X$ , satisfying  $Y(0) = 0$ . Thus  $Y(t) = \int_0^t W(s) ds + t^4$  for  $t \geq 0$ . Then there exists an almost surely uniquely defined random continuous function  $H$  satisfying the following conditions:

(i) The function  $H$  is everywhere above the function  $Y$ :

$$H(t) \geq Y(t), \text{ for each } t \in \mathbb{R}. \quad (6.4)$$

(ii)  $H$  has a convex second derivative, and, with probability one,  $H$  is three times differentiable at  $t = 0$ .

(iii) The function  $H$  satisfies

$$\int_{\mathbb{R}} \{H(t) - Y(t)\} dH^{(3)}(t) = 0. \quad (6.5)$$

The main results of this section are stated in Theorems 6.2 and 6.3.

**Theorem 6.2** (Asymptotic distributions at a point for convex densities.) Suppose that  $f_0 \in \mathcal{C}$  has  $f_0''(x_0) > 0$  and that  $f_0''$  is continuous in a neighborhood of  $x_0$ . Then the nonparametric maximum

likelihood estimator and least squares estimator studied in Section 2 are asymptotically equivalent in the following sense: if  $\bar{f}_n = \hat{f}_n$  or  $\tilde{f}_n$ , then

$$\begin{pmatrix} n^{2/5}c_1(f_0)(\bar{f}_n(x_0) - f_0(x_0)) \\ n^{1/5}c_2(f_0)(\bar{f}'_n(x_0) - f'_0(x_0)) \end{pmatrix} \rightarrow_d \begin{pmatrix} H''(0) \\ H^{(3)}(0) \end{pmatrix}$$

where  $(H''(0), H^{(3)}(0))$  are the second and third derivatives at 0 of the envelope  $H$  of  $Y$  as described in Theorem 6.1 and

$$c_1(f_0) = \left( \frac{24}{f_0^2(x_0)f_0''(x_0)} \right)^{1/5}, \quad c_2(f_0) = \left( \frac{24^3}{f_0(x_0)f_0''(x_0)^3} \right)^{1/5}. \quad (6.6)$$

The derivatives  $\bar{f}'_n(x_0)$  may be interpreted as left or right derivatives.

**Remark.** Note that the constants  $c_i(f_0)$ ,  $i = 1, 2$  also arise naturally in the asymptotic minimax lower bounds obtained by JONGBLOED (1995), Theorem 6.1, page 111.

For the least squares regression estimator  $\hat{r}$ , we need a stronger version of Assumption 4.1 as follows: for  $0 \leq x \leq 1$ , let  $F_n(x) = n^{-1} \sum_{i=1}^n 1_{[0,x]}(x_{n,i})$ .

**Assumption 6.1.** For some  $\delta > 0$  the functions  $\{F_n\}$  satisfy

$$\sup_{x:|x-x_0|\leq\delta} |F_n(x) - x| = o(n^{-1/5}).$$

**Theorem 6.3** (*Asymptotic distributions at a point for convex regression.*) Suppose that  $r_0 \in \mathcal{C}_r$  has  $r_0''(x_0) > 0$ , that Assumptions 4.1, 4.2, and 6.1 hold, and that  $r_0''$  is continuous in a neighborhood of  $x_0$ . Then for the least squares estimator  $\hat{r}_n$  introduced in Section 2 it follows that

$$\begin{pmatrix} n^{2/5}d_1(r_0)(\hat{r}_n(x_0) - r_0(x_0)) \\ n^{1/5}d_2(r_0)(\hat{r}'_n(x_0) - r'_0(x_0)) \end{pmatrix} \rightarrow_d \begin{pmatrix} H''(0) \\ H^{(3)}(0) \end{pmatrix}$$

where  $(H''(0), H^{(3)}(0))$  are the second and third derivatives at 0 of the envelope  $H$  of  $Y$  as described in Theorem 6.1, and

$$d_1(r_0) = \left( \frac{24}{\sigma^4 r_0''(x_0)} \right)^{1/5}, \quad d_2(r_0) = \left( \frac{24^3}{\sigma^2 r_0''(x_0)^3} \right)^{1/5}. \quad (6.7)$$

**Proof of Theorem 6.2.** We begin with the least squares estimator. First some notation. Define the local  $Y_n$ -process by

$$\tilde{Y}_n^{loc}(t) \equiv n^{4/5} \int_{x_0}^{x_0+n^{-1/5}t} \left\{ \mathbb{F}_n(v) - \mathbb{F}_n(x_0) - \int_{x_0}^v (f_0(x_0) + (u-x_0)f'_0(x_0)) du \right\} dv \quad (6.8)$$

and the local  $H_n$ -process by

(6.9)

$$\tilde{H}_n^{loc}(t) \equiv n^{4/5} \int_{x_0}^{x_0+n^{-1/5}t} \int_{x_0}^v \left\{ \tilde{f}_n(u) - f_0(x_0) - (u-x_0)f'_0(x_0) \right\} du dv + \tilde{A}_n t + \tilde{B}_n$$

where

$$\tilde{A}_n = n^{3/5} \left\{ \tilde{F}_n(x_0) - \mathbb{F}_n(x_0) \right\} \quad \text{and} \quad \tilde{B}_n = n^{4/5} \left\{ \tilde{H}_n(x_0) - Y_n(x_0) \right\}.$$

Noting that

$$\tilde{A}_n = n^{3/5} \left\{ \tilde{F}_n(x_0) - \tilde{F}_n(x_n^-) - (\mathbb{F}_n(x_0) - \mathbb{F}_n(x_n^-)) \right\},$$

where

$$x_n^- \equiv \max\{t \leq x_0 : \tilde{H}_n(t) = Y_n(t) \quad \text{and} \quad \tilde{H}'_n(t) = Y_n(t)\},$$

it follows by Lemmas 4.2 and 4.4 that  $\{\tilde{A}_n\}$  is tight. Indeed,

$$\begin{aligned} |\tilde{A}_n| &= n^{3/5} \left| \int_{x_n^-}^{x_0} \tilde{f}_n(u) - f_0(x_0) - (u-x_0)f'_0(x_0) du \right. \\ &\quad \left. - \int_{x_n^-}^{x_0} f_0(u) - f_0(x_0) - (u-x_0)f'_0(x_0) du - \int_{x_n^-}^{x_0} d(\mathbb{F}_n - F_0)(u) \right| \\ &\leq n^{3/5} (x_0 - x_n^-) \sup_{u \in [x_n^-, x_0]} |\tilde{f}_n(u) - f_0(x_0) - (u-x_0)f'_0(x_0)| \\ &\quad + n^{3/5} f''(x_0) (1+o(1)) (x_0 - x_n^-)^3 + n^{3/5} \left| \int_{x_n^-}^{x_0} d(\mathbb{F}_n - F_0)(u) \right| \end{aligned}$$

which is  $\mathcal{O}_P(1)$  by the lemmas mentioned. For  $\tilde{B}_n$  a similar calculation works.

Now we can write

$$\begin{aligned} \tilde{H}_n^{loc}(t) - \tilde{Y}_n^{loc}(t) &= n^{4/5} \int_{x_0}^{x_0+n^{-1/5}t} \left\{ \tilde{F}_n(u) - \tilde{F}_n(x_0) - (\mathbb{F}_n(u) - \mathbb{F}_n(x_0)) \right\} du + \tilde{A}_n t + \tilde{B}_n \\ &= n^{4/5} \int_{x_0}^{x_0+n^{-1/5}t} \left\{ \tilde{F}_n(u) - \mathbb{F}_n(u) \right\} du + \tilde{B}_n \\ &= n^{4/5} \left\{ \tilde{H}_n(x_0 + n^{-1/5}t) - Y_n(x_0 + n^{-1/5}t) \right\} \geq 0 \end{aligned}$$

with equality if  $x_0 + n^{-1/5}t \in \mathcal{T}_n$ .

Using the identity

$$\begin{aligned} F_0(v) - F_0(x_0) &= \int_{x_0}^v f_0(u) du = \int_{x_0}^v \left\{ f_0(x_0) + f'_0(x_0)(u-x_0) + \frac{1}{2} f''_0(u^*)(u-x_0)^2 \right\} du \\ &= \int_{x_0}^v \left\{ f_0(x_0) + f'_0(x_0)(u-x_0) \right\} du + \frac{1}{2} (f''_0(x_0) + o(1)) \int_{x_0}^v (u-x_0)^2 du \end{aligned}$$

as  $v \rightarrow x_0$ , and letting  $\mathbb{U}_n = \sqrt{n}(\mathbb{G}_n - I)$  denote the empirical process of i.i.d. uniform(0, 1) random variables with empirical distribution function  $\mathbb{G}_n$  (as in SHORACK AND WELLNER (1986)), we can rewrite  $\tilde{Y}_n^{loc}$  as

$$\begin{aligned} \tilde{Y}_n^{loc}(t) &= n^{4/5} \int_{x_0}^{x_0+n^{-1/5}t} \{\mathbb{F}_n(v) - \mathbb{F}_n(x_0) - (F_0(v) - F_0(x_0))\} dv \\ &\quad + n^{4/5} \int_{x_0}^{x_0+n^{-1/5}t} \frac{1}{6} f_0''(x_0)(v - x_0)^3 dv + o(1) \\ &=_{d} n^{3/10} \int_{x_0}^{x_0+n^{-1/5}t} \{\mathbb{U}_n(F_0(v)) - \mathbb{U}_n(F_0(x_0))\} dv + \frac{1}{24} f_0''(x_0)t^4 + o(1) \\ &\Rightarrow \sqrt{f_0(x_0)} \int_0^t W(s) ds + \frac{1}{24} f_0''(x_0)t^4 \end{aligned}$$

uniformly in  $|t| \leq c$ ; see Theorem 3.1.1, page 93, SHORACK AND WELLNER (1986), together with the representation of a Brownian bridge process  $\mathbb{U}$  in terms of Brownian motion  $B$  as  $\mathbb{U}(t) = B(t) - tB(1)$ . Alternatively, this follows easily from Theorem 2.11.22 or 2.11.23, VAN DER VAART AND WELLNER (1996), pages 220-221.

Now we will line up the argument to match with Theorem 6.1. For any  $k_1, k_2 > 0$ , we see that

$$\tilde{H}_n^l(t) - \tilde{Y}_n^l(t) := k_1 \tilde{H}_n^{loc}(k_2 t) - k_1 \tilde{Y}_n^{loc}(k_2 t) \geq 0 \quad (6.10)$$

with equality if and only if  $x_0 + k_2 n^{-1/5}t \in \mathcal{T}_n$ . Using the scaling property of Brownian motion, saying that  $\alpha^{-1/2}W(\alpha t)$  is Brownian motion for all  $\alpha > 0$  if  $W$  is, we see that choosing

$$k_1 = 24^{-3/5} f_0(x_0)^{-4/5} f_0''(x_0)^{3/5} \text{ and } k_2 = 24^{2/5} f_0(x_0)^{1/5} f_0''(x_0)^{-2/5} \quad (6.11)$$

yields that  $\tilde{Y}_n^l \Rightarrow Y$  as defined in Theorem 6.1. Also note, using  $c_1$  and  $c_2$  as defined in (6.6) that

$$(\tilde{H}_n^l)''(0) = k_1 k_2^2 (\tilde{H}_n^{loc})''(0) = n^{2/5} c_1(f_0)(\tilde{f}_n(x_0) - f_0(x_0))$$

and

$$(\tilde{H}_n^l)'''(0) = k_1 k_2^3 (\tilde{H}_n^{loc})'''(0) = n^{1/5} c_2(f_0)(\tilde{f}'_n(x_0) - f'_0(x_0)).$$

We take  $\tilde{f}'_n$  to be the right derivative below, but this is not essential. Hence, what remains to be shown is that along with the process  $\tilde{Y}_n^l$ , the ‘‘invelopes’’  $\tilde{H}_n^l$  converge in such a way that the second and third derivative of this envelope at zero converge in distribution to the corresponding quantities of  $H$  in Theorem 6.1.

Define, for  $c > 0$ , the space  $E[-c, c]$  of vector-valued functions as follows:

$$E[-c, c] = (C[-c, c])^4 \times (D[-c, c])^2$$

and endow  $E[-c, c]$  by the product topology induced by the uniform topology on the spaces  $C[-c, c]$  and the Skorohod topology on  $D[-c, c]$ . The space  $E[-c, c]$  supports vector-valued stochastic process

$$\{Z_n\} \equiv \{(\tilde{H}_n^l, (\tilde{H}_n^l)', (\tilde{H}_n^l)'', \tilde{Y}_n^l, (\tilde{H}_n^l)''', (\tilde{Y}_n^l)')\}$$

Note that the subset of  $D[-c, c]$  consisting of increasing functions, absolutely bounded by  $M < \infty$  is compact in the Skorohod topology. Hence, Lemma 4.4 together with the monotonicity of  $(\tilde{H}_n^l)'''$ , gives that the sequence  $(\tilde{H}_n^l)'''$  is tight in  $D[-c, c]$  endowed with the Skorohod topology. Moreover, since the set of continuous functions, with its values as well as its derivative absolutely bounded by  $M$ , is compact in  $C[-c, c]$  with the uniform topology, the sequences  $(\tilde{H}_n^l)''$ ,  $(\tilde{H}_n^l)'$  and  $\tilde{H}_n^l$  are also tight in  $C[-c, c]$ . This follows from Lemma 4.4. Since  $Y_n$  and  $Y_n'$  both converge weakly, they are also tight in  $C[-c, c]$  and  $D[-c, c]$  with their topologies respectively. This means that for each  $\epsilon > 0$  we can construct a compact product set in  $E[-c, c]$  such that the vector  $Z_n$  will be contained in that set with probability at least  $1 - \epsilon$  for all  $n$ . This means that the sequence  $Z_n$  is tight in  $E[-c, c]$ .

Fix an arbitrary subsequence  $Z_{n'}$ . Then we can construct a subsequence  $\{Z_{n''}\}$  such that  $\{Z_{n''}\}$  converges weakly to some  $Z_0$  in  $E[-c, c]$ , for each  $c > 0$ . By the continuous mapping theorem, it follows that the limit  $Z_0 = (H_0, H_0', H_0'', Y_0, H_0''', Y_0')$  satisfies both

$$\inf_{t \in [-c, c]} (H_0(t) - Y_0(t)) \geq 0, \text{ for each } c > 0 \quad (6.12)$$

and

$$\int_{[-c, c]} \{H_0(t) - Y_0(t)\} dH_0^{(3)}(t) = 0 \quad (6.13)$$

almost surely. The inequality (6.12) can e.g. be seen by using convergence of expectations of the nonpositive continuous function  $\phi : E[-c, c] \rightarrow \mathbb{R}$  defined by

$$\phi(z_1, z_2, \dots, z_6) = \inf_t (z_1(t) - z_4(t)) \wedge 0$$

using that  $\phi(Z_n) \equiv 0$  a.s. This gives  $\phi(Z_0) = 0$  a.s., and hence (6.12). Note also that  $H_0''$  is convex and decreasing. The equality (6.13) follows from considering the function

$$\phi(z_1, z_2, \dots, z_6) = \int_{-c}^c (z_1(t) - z_4(t)) dz_5(t)$$

which is continuous on the subset of  $E[-c, c]$  consisting of functions with  $z_5$  increasing.

Now, since  $Z_0$  satisfies (6.12) for all  $c > 0$ , and  $Y_0 = Y$  as defined in Theorem 6.1, we see that condition (6.4) of Theorem 6.1 is satisfied by the first and fourth component of  $Z_0$ . Moreover, also condition (6.5) of Theorem 6.1 is satisfied by  $Z_0$ .

Hence it follows that the limit  $Z_0$  is in fact equal to  $Z = (H, H', H'', Y, H''', Y')$  involving the unique function  $H$  described in Theorem 6.1. Since the limit is the same for any such subsequence, it follows that the full sequence  $\{Z_n\}$  converges weakly and has the same limit, namely  $Z$ . In particular  $Z_n(0) \rightarrow_d Z(0)$ , and this yields the least squares part of Theorem 6.2.

Now consider the MLE. Define the local  $H_n$ -process as

$$\hat{H}_n^{loc}(t) \equiv n^{4/5} f_0(x_0) \int_{x_0}^{x_0 + n^{-1/5}t} \int_{x_0}^v \left\{ \frac{\hat{f}_n(u) - f_0(x_0) - (u - x_0)f_0'(x_0)}{\hat{f}_n(u)} \right\} du dv + \hat{A}_n t + \hat{B}_n$$

where

$$\hat{A}_n = -n^{3/5} f_0(x_0) \left\{ \hat{H}'_n(x_0) - x_0 \right\} \text{ and } \hat{B}_n = -n^{4/5} f(x_0) \left\{ \hat{H}_n(x_0) - \frac{1}{2} x_0^2 \right\}.$$

Tightness of these variables can be shown similarly to that of  $\tilde{A}_n$ . Define the local  $Y_n$ -process as

$$\begin{aligned} \hat{Y}_n^{loc}(t) &\equiv n^{4/5} f_0(x_0) \int_{x_0}^{x_0+n^{-1/5}t} \int_{x_0}^v \left\{ \frac{f_0(u) - f_0(x_0) - (u-x_0)f'_0(x_0)}{\hat{f}_n(u)} \right\} du dv \\ &\quad + n^{4/5} f_0(x_0) \int_{x_0}^{x_0+n^{-1/5}t} \int_{x_0}^v \frac{1}{\hat{f}_n(u)} d(\mathbb{F}_n - F_0)(u) dv. \end{aligned}$$

Then we have that

$$\begin{aligned} \hat{H}_n^{loc}(t) - \hat{Y}_n^{loc}(t) &= n^{4/5} f_0(x_0) \int_{x_0}^{x_0+n^{-1/5}t} \int_{x_0}^v \left\{ \frac{\hat{f}_n(u) - f_0(u)}{\hat{f}_n(u)} \right\} du dv \\ &\quad - n^{4/5} f_0(x_0) \int_{x_0}^{x_0+n^{-1/5}t} \int_{x_0}^v \frac{1}{\hat{f}_n(u)} d(\mathbb{F}_n - F_0)(u) dv + \hat{A}_n t + \hat{B}_n \\ &= n^{4/5} f_0(x_0) \left( \frac{1}{2} n^{-2/5} t^2 - \int_{x_0}^{x_0+n^{-1/5}t} \int_{x_0}^v \frac{1}{\hat{f}_n(u)} dF_0(u) dv \right) \\ &\quad - n^{4/5} f_0(x_0) \int_{x_0}^{x_0+n^{-1/5}t} \int_{x_0}^v \frac{1}{\hat{f}_n(u)} d(\mathbb{F}_n - F_0)(u) dv + \hat{A}_n t + \hat{B}_n \\ &= n^{4/5} f_0(x_0) \left( \frac{1}{2} n^{-2/5} t^2 - \int_{x_0}^{x_0+n^{-1/5}t} \int_{x_0}^v \frac{1}{\hat{f}_n(u)} d\mathbb{F}_n(u) dv \right) + \hat{A}_n t + \hat{B}_n \\ &= n^{4/5} f_0(x_0) \left( \frac{1}{2} n^{-2/5} t^2 - \hat{H}_n(x_0 + n^{-1/5}t) + \hat{H}_n(x_0) + n^{-1/5}t \hat{H}'_n(x_0) \right) + \hat{A}_n t + \hat{B}_n \\ &= n^{4/5} f_0(x_0) \left( \frac{1}{2} n^{-2/5} t^2 - \hat{H}_n(x_0 + n^{-1/5}t) + \frac{1}{2} x_0^2 + n^{-1/5}t x_0 \right) \\ &= n^{4/5} f_0(x_0) \left( \frac{1}{2} (x_0 + n^{-1/5}t)^2 - \hat{H}_n(x_0 + n^{-1/5}t) \right) \geq 0 \end{aligned}$$

with equality if  $x_0 + n^{-1/5}t \in \mathcal{T}_n$ .

Now rescale the processes  $\hat{Y}_n^{loc}$  and  $\hat{H}_n^{loc}$  as in (6.10), with  $k_1$  and  $k_2$  as defined in (6.11) and note that  $\tilde{Y}_n^l - \hat{Y}_n^l \rightarrow 0$  in probability uniformly on compacta by consistency Theorem 3.2. Also note that by the same theorem

$$|(\hat{H}_n^l)''(0) - n^{2/5} c_1(f_0)(\hat{f}_n(x_0) - f_0(x_0))| \rightarrow 0$$

and

$$|(\hat{H}_n^l)'''(0) - n^{1/5} c_2(f_0)(\hat{f}'_n(x_0) - f'_0(x_0))| \rightarrow 0$$

in probability. Applying the same arguments as in case of the least squares estimator, we obtain our result.  $\square$

**Proof of Theorem 6.3.** First some notation. Denote by  $\hat{r}_n : [0, 1] \rightarrow \mathcal{R}$  the piecewise linear function through the points  $(x_{n,i}, \hat{r}_{n,i})$  such that  $\hat{r}_n$  is linear with minimal absolute slope for  $x \in [0, x_{n,1}] \cup [x_{n,n}, 1]$ . Then define

$$\mathbb{S}_n(t) = \frac{1}{n} \sum_{i=1}^n Y_{n,i} 1_{[x_{n,i} \leq t]}, \quad \mathbb{R}_n(t) = \frac{1}{n} \sum_{i=1}^n \hat{r}_{n,i} 1_{[x_{n,i} \leq t]} = \int_0^t \hat{r}_n(s) dF_n(s), \quad \text{and} \quad \tilde{\mathbb{R}}_n(t) = \int_0^t \hat{r}_n(s) ds.$$

Hence,

$$\mathbb{S}_n(x_{n,k}) = n^{-1} S_k = n^{-1} (Y_{n,1} + \cdots + Y_{n,k}), \quad \text{and} \quad \mathbb{R}_n(x_{n,k}) = n^{-1} \hat{R}_k = n^{-1} (\hat{r}_{n,1} + \cdots + \hat{r}_{n,k}).$$

Inspired by the notation in the density estimation context, we define the processes

$$Y_n(x) = \int_0^x \mathbb{S}_n(v) dv, \quad H_n(x) = \int_0^x \mathbb{R}_n(v) dv, \quad \tilde{H}_n(x) = \int_0^x \tilde{\mathbb{R}}_n(v) dv,$$

and their ‘local counterparts’

$$Y_n^{loc}(t) = n^{4/5} \int_{x_0}^{x_0 + n^{-1/5}t} \left\{ \mathbb{S}_n(v) - \mathbb{S}_n(x_0) - \int_{x_0}^v (r_0(x_0) + (u - x_0)r'_0(x_0)) dF_n(u) \right\} dv,$$

$$H_n^{loc}(t) = n^{4/5} \int_{x_0}^{x_0 + n^{-1/5}t} \left\{ \mathbb{R}_n(v) - \mathbb{R}_n(x_0) - \int_{x_0}^v \{r_0(x_0) + (u - x_0)r'_0(x_0)\} dF_n(u) \right\} dv + A_n t + B_n$$

and

$$\tilde{H}_n^{loc}(t) = n^{4/5} \int_{x_0}^{x_0 + n^{-1/5}t} \left\{ \tilde{\mathbb{R}}_n(v) - \tilde{\mathbb{R}}_n(x_0) - \int_{x_0}^v \{r_0(x_0) + (u - x_0)r'_0(x_0)\} du \right\} dv + A_n t + B_n.$$

Here

$$A_n = n^{3/5} \{\mathbb{R}_n(x_0) - \mathbb{S}_n(x_0)\} \quad \text{and} \quad B_n = n^{4/5} \{H_n(x_0) - Y_n(x_0)\}.$$

For  $\tilde{H}_n^{loc}$  we have

$$(\tilde{H}_n^{loc})''(t) = n^{2/5} (\hat{r}_n(x_0 + n^{-1/5}t) - r_0(x_0) - r'_0(x_0)n^{-1/5}t)$$

and

$$(\tilde{H}_n^{loc})'''(t) = n^{1/5} (\hat{r}'_n(x_0 + n^{-1/5}t) - r'_0(x_0)).$$

Noting that

$$A_n = n^{3/5} \{\mathbb{R}_n(x_0) - \mathbb{R}_n(x_n^-) - (\mathbb{S}_n(x_0) - \mathbb{S}_n(x_n^-))\},$$

where

$$x_n^- = \max\{v \leq x_0 : H_n(v) = Y_n(v) \text{ and } \mathbb{R}_n(v) = \mathbb{S}_n(v)\},$$

it follows by Lemma 8, page 757 of MAMMEN (1991) and Lemma 4.5 that  $\{A_n\}$  is tight. Indeed, writing  $R_0(t) = \int_0^t r_0(u) du$ ,

$$\begin{aligned}
|A_n| &= n^{3/5} \left| \mathbb{R}_n(x_0) - \mathbb{R}_n(x_0^-) - \int_{x_n^-}^{x_0} r_0(x_0) + (u - x_0)r_0'(x_0) du \right. \\
&\quad \left. - \int_{x_n^-}^{x_0} r_0(u) - r_0(x_0) - (u - x_0)r_0'(x_0) du - \int_{x_n^-}^{x_0} d(\mathbb{S}_n - R_0)(u) \right| \\
&\leq n^{3/5}(x_0 - x_n^-) \sup_{u \in [x_n^-, x_0]} |\hat{r}_n(u) - r_0(x_0) - (u - x_0)r_0'(x_0)| + \\
&\quad + n^{3/5}r_0''(x_0)(x_0 - x_n^-)^3 + n^{3/5} \left| \int_{x_n^-}^{x_0} d(\mathbb{S}_n - R_0)(u) \right|
\end{aligned}$$

which is  $\mathcal{O}_P(1)$  by the lemmas mentioned. For  $B_n$  a similar calculation works.

Now we can write

$$\begin{aligned}
H_n^{loc}(t) - Y_n^{loc}(t) &= n^{4/5} \int_{x_0}^{x_0 + n^{-1/5}t} \{ \mathbb{R}_n(u) - \mathbb{R}_n(x_0) - (\mathbb{S}_n(u) - \mathbb{S}_n(x_0)) \} du + A_n t + B_n \\
&= n^{4/5} \int_{x_0}^{x_0 + n^{-1/5}t} \{ \mathbb{R}_n(u) - \mathbb{S}_n(u) \} du + B_n \\
&= n^{4/5} \{ H_n(x_0 + n^{-1/5}t) - Y_n(x_0 + n^{-1/5}t) \} \geq 0
\end{aligned} \tag{6.14}$$

with equality if  $x_0 + n^{-1/5}t \in \mathcal{T}_n$ ; here  $\mathcal{T}_n$  is the collection of  $x_{n,i}$ 's where equality occurs in (2.23) of Lemma 2.6.

We will show that

$$Y_n^{loc}(t) \Rightarrow \sigma \int_0^t W(s) ds + \frac{1}{24} r_0''(x_0) t^4 \tag{6.15}$$

uniformly in  $|t| \leq c$ . To prove (6.15) we decompose  $Y_n^{loc}$  as follows:

$$\begin{aligned}
Y_n^{loc}(t) &= n^{4/5} \int_{x_0}^{x_0 + n^{-1/5}t} \left\{ \mathbb{S}_n(v) - \mathbb{S}_n(x_0) - \int_{x_0}^v (r_0(x_0) + (u - x_0)r_0'(x_0)) dF_n(u) \right\} dv \\
&= n^{4/5} \int_{x_0}^{x_0 + n^{-1/5}t} \{ \mathbb{S}_n(v) - \mathbb{S}_n(x_0) - (R_0(v) - R_0(x_0)) \} dv \\
&\quad + n^{4/5} \int_{x_0}^{x_0 + n^{-1/5}t} \left\{ R_0(v) - R_0(x_0) - \int_{x_0}^v (r_0(x_0) + (u - x_0)r_0'(x_0)) dF_n(u) \right\} dv \\
&= n^{4/5} \int_{x_0}^{x_0 + n^{-1/5}t} \left\{ n^{-1} \sum_{i=1}^n Y_{n,i} 1_{(x_0, v]}(x_{n,i}) - \int_{x_0}^v r_0(u) du \right\} dv \\
&\quad + n^{4/5} \int_{x_0}^{x_0 + n^{-1/5}t} \left\{ R_0(v) - R_0(x_0) - \int_{x_0}^v (r_0(x_0) + (u - x_0)r_0'(x_0)) dF_n(u) \right\} dv
\end{aligned}$$



$$\begin{aligned}
&= n^{4/5} \int_{x_0}^{x_0+n^{-1/5}t} \left\{ n^{-1} \sum_{i=1}^n \epsilon_{n,i} 1_{(x_0,v]}(x_{n,i}) \right\} dv \\
&\quad + n^{4/5} \int_{x_0}^{x_0+n^{-1/5}t} \left\{ n^{-1} \sum_{i=1}^n r_0(x_{n,i}) 1_{(x_0,v]}(x_{n,i}) - \int_{x_0}^v r_0(u) du \right\} dv \\
&\quad - n^{4/5} \int_{x_0}^{x_0+n^{-1/5}t} \left\{ \int_{x_0}^v (r_0(x_0) + (u-x_0)r'_0(x_0)) d(F_n(u) - u) \right\} dv \\
&\quad + n^{4/5} \int_{x_0}^{x_0+n^{-1/5}t} \left\{ R_0(v) - R_0(x_0) - \int_{x_0}^v (r_0(x_0) + (u-x_0)r'_0(x_0)) du \right\} dv \\
&= I_n(t) + II_n(t) + III_n(t),
\end{aligned}$$

where  $II_n(t)$  is given by the two middle terms. Now first note that

$$\begin{aligned}
III_n(t) &= n^{4/5} \int_{x_0}^{x_0+n^{-1/5}t} \left\{ R_0(v) - R_0(x_0) - \int_{x_0}^v (r_0(x_0) + (u-x_0)r'_0(x_0)) du \right\} dv \\
&= n^{4/5} \int_{x_0}^{x_0+n^{-1/5}t} \frac{1}{6} r_0''(x_0) (v-x_0)^3 dv + o(1) = \frac{1}{24} r_0''(x_0) t^4 + o(1)
\end{aligned}$$

uniformly in  $|t| \leq c$ . The term  $II_n(t)$  is  $o(1)$  uniformly in  $|t| \leq c$ . This is seen as follows. Define  $G_n$  by

$$G_n(x) = n^{1/5} \left( \frac{1}{n} \sum_{i=1}^n 1_{[t_n, i \leq x]} - x_0 \right) = n^{1/5} (F_n(x_0 + n^{-1/5}x) - x_0).$$

Under Assumption 6.1 it follows that  $G_n(x) \rightarrow x$  uniformly for  $|x| \leq c$ . By use of the changes of variables  $u = x_0 + n^{-1/5}u'$ ,  $v = x_0 + n^{-1/5}v'$ ,

$$\begin{aligned}
II_n(t) &= n^{4/5} \int_{x_0}^{x_0+n^{-1/5}t} \left\{ \int_{(x_0,v]} r_0(u) dF_n(u) - \int_{x_0}^v r_0(u) du \right\} dv \\
&\quad - n^{4/5} \int_{x_0}^{x_0+n^{-1/5}t} \left\{ \int_{x_0}^v (r_0(x_0) + (u-x_0)r'_0(x_0)) d(F_n(u) - u) \right\} dv \\
&= n^{4/5} \int_{x_0}^{x_0+n^{-1/5}t} \left\{ \int_{(x_0,v]} (r_0(u) - r_0(x_0) - r'_0(x_0)(u-x_0)) d(F_n(u) - u) \right\} dv \\
&= n^{3/5} \int_0^t \int_0^{v'} \left( r_0(x_0 + n^{-1/5}u') - r_0(x_0) - r'_0(x_0)n^{-1/5}u' \right) \\
&\quad \quad \quad d(F_n(x_0 + n^{-1/5}u') - (x_0 + n^{-1/5}u')) dv' \\
&= n^{2/5} \int_0^t \int_0^{v'} \left( r_0(x_0 + n^{-1/5}u') - r_0(x_0) - r'_0(x_0)n^{-1/5}u' \right) d(G_n(u') - u') dv' \\
&= \frac{1}{2} \int_0^t \int_0^{v'} r_0''(u^*) u'^2 d(G_n(u') - u') dv' \rightarrow 0 \quad \text{uniformly in } |t| \leq c, \tag{6.16}
\end{aligned}$$

and this yields the convergence in (6.16). Finally, note that

$$\begin{aligned} I_n(t) &= n^{4/5} \int_{x_0}^{x_0+n^{-1/5}t} \left\{ n^{-1} \sum_{i=1}^n \epsilon_{n,i} 1_{(x_0,v]}(x_{n,i}) \right\} dv = n^{-1/5} \sum_{i=1}^n \epsilon_{n,i} 1_{[x_0 < x_{n,i}]} \int_{x_0}^{x_0+n^{-1/5}t} 1_{[x_{n,i} \leq v]} dv \\ &= n^{-1/5} \sum_{i=1}^n \epsilon_{n,i} 1_{[x_0 < x_{n,i} \leq x_0+n^{-1/5}t]} (x_0 + n^{-1/5}t - x_{n,i}). \end{aligned}$$

Thus we have, writing  $t_{n,i} = n^{1/5}(x_{n,i} - x_0)$ ,

$$\begin{aligned} \text{Var}(I_n(t)) &= \frac{\sigma^2}{n^{2/5}} \sum_{i=1}^n 1_{[x_0 < x_{n,i} \leq x_0+n^{-1/5}t]} (x_0 + n^{-1/5}t - x_{n,i})^2 \\ &= \frac{\sigma^2}{n^{4/5}} \sum_{i=1}^n 1_{[0 < t_{n,i} \leq t]} (t - t_{n,i})^2 = \sigma^2 \int_0^t (t-x)^2 dG_n(x) \rightarrow \sigma^2 \int_0^t (t-x)^2 dx = \frac{\sigma^2}{3} t^3, \end{aligned}$$

the variance of  $\sigma \int_0^t W(s) ds$ . By similar calculations the hypotheses of Theorem 2.11.1 of VAN DER VAART AND WELLNER (1996) can easily be shown to hold, and this completes the proof of (6.15).

The next step is to show that  $\tilde{H}_n^{loc}$  and  $H_n^{loc}$  are asymptotically the same and thereby show that  $\tilde{H}_n^{loc}$  satisfies the characterizing conditions (asymptotically). Note that by the change of variables  $u = x_0 + n^{-1/5}u'$ ,  $v = x_0 + n^{-1/5}v'$ ,

$$\begin{aligned} H_n^{loc}(t) - \tilde{H}_n^{loc}(t) &= n^{4/5} \int_{x_0}^{x_0+tn^{-1/5}} \int_{(x_0,u]} (\hat{r}_n(u) - r_0(x_0) - (u-x_0)r'_0(x_0)) d(F_n(u) - u) dv \\ &= \int_0^t \int_0^{v'} n^{2/5} (\hat{r}_n(x_0 + n^{-1/5}u') - r_0(x_0) - n^{-1/5}u'r'_0(x_0)) d(G_n(u') - u') dv' \\ &= o_p(1) \quad \text{uniformly in } |t| \leq c \end{aligned}$$

since the integrand is uniformly bounded in probability by Lemma 4.5, and  $G_n(u) \rightarrow u$  uniformly in  $|u| \leq c$  by Assumption 6.1.

Now we will line up the argument to match with Theorem 6.1. For any  $k_1, k_2 > 0$ , using (6.14), we see that

$$H_n^l(t) - Y_n^l(t) := k_1 \tilde{H}_n^{loc}(k_2 t) - k_1 Y_n^{loc}(k_2 t) \geq 0 - o_p(1)$$

uniformly in  $|t| \leq c$  with equality if and only if  $x_0 + k_2 n^{-1/5}t \in \mathcal{T}_n$ . Using the scaling property of Brownian motion, saying that  $\alpha^{-1/2}W(\alpha t)$  is Brownian motion for all  $\alpha > 0$  if  $W$  is, we see that choosing

$$k_1 = 24^{-3/5} \sigma^{-8/5} r_0''(x_0)^{3/5} \quad \text{and} \quad k_2 = 24^{2/5} \sigma^{2/5} r_0''(x_0)^{-2/5}$$

yields that  $Y_n^l \Rightarrow Y$  as defined in Theorem 6.1. Also note that

$$(H_n^l)''(0) = k_1 k_2^2 (\tilde{H}_n^{loc})''(0) = n^{2/5} d_1(r_0)(\hat{r}_n(x_0) - r_0(x_0))$$

and

$$(H_n^l)'''(0) = k_1 k_2^3 (\tilde{H}_n^{loc})'''(0) = n^{1/5} d_2(r_0)(\hat{r}'_n(x_0) - r'_0(x_0)),$$

where  $d_1$  and  $d_2$  are as defined in (6.7). Hence, what remains to be shown is that along with the process  $Y_n^l$ , the “invelopes”  $H_n^l$  converge in such a way that the second and third derivative of this envelope at zero converge in distribution to the corresponding quantities of  $H$  in Theorem 6.1. Defining a vector-valued process, arguing along subsequences and using Theorem 6.1, the result follows along the same lines as the proof of Theorem 6.2.  $\square$

**ACKNOWLEDGEMENTS:** We thank two anonymous referees and the associate editor for their valuable remarks and suggestions. We owe thanks to Richard Gill for pointing out Hampel (1987) and the resulting discussions about estimation of a convex density function, Enno Mammen for discussions about estimation of convex regression functions and Yazhen Wang for sending us copies of unpublished Ph.D. dissertation work and subsequent technical reports.

## References

- Anevski, D. (1994). Estimating the derivative of a convex density. *Technical Report 1994:8*, Department of Mathematical Statistics, University of Lund.
- Barlow, R.E., Bartholomew, R. J., Bremner, J. M., and Brunk, H.D. (1972). *Statistical Inference Under Order Restrictions*. Wiley, New York.
- Brunk, H. D. (1958). On the estimation of parameters restricted by inequalities. *Ann. Math. Statist.* **29**, 437 - 454.
- Brunk, H.D. (1970). Estimation of isotonic regression. *Nonparametric Techniques in Statistical Inference* Puri, M.L. ed.
- Donoho, D.L. and Liu, R.C. (1987). Geometrizing rates of convergence, I. Technical Report 137, Department of Statistics, University of California, Berkeley.
- Grenander, U. (1956). On the theory of mortality measurement, part II. *Skandinavisk Aktuarietidskrift* **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.
- Groeneboom, P. (1988). Brownian motion with a parabolic drift and Airy functions. *Probability Theory and Related Fields* **81**, 79 - 109.
- Groeneboom, P. (1996). Inverse problems in statistics. Proceedings of the St. Flour Summer School in Probability, 1994. *Lecture Notes in Math.* **1648**, 67 - 164. Springer Verlag, Berlin.
- Groeneboom, P., Jongbloed, G., and Wellner, J. A. (2001a). A canonical process for estimation of convex functions:  
the “envelope” of integrated Brownian motion  $+t^4$ . Tentatively accepted by The Annals of Statistics.

- Groeneboom, P., Jongbloed, G., and Wellner, J. A. (2001b). Vertex direction algorithms for computing nonparametric function estimates. Manuscript in progress.
- Groeneboom, P. and Wellner, J. A. (2000). Computing Chernoff's distribution. *J. Computational and Graphical Statistics*, to appear.
- Hampel, F.R. (1987). Design, modelling and analysis of some biological datasets. In *Design, data and analysis, by some friends of Cuthbert Daniel*, C.L. Mallows, editor, 111- 115. Wiley, New York.
- Hanson, D. L. and Pledger, G. (1976). Consistency in concave regression. *Ann. Statist.* **4**, 1038 - 1050.
- Hildreth, C. (1954). Point estimators of ordinates of concave functions. *J. Amer. Statist. Assoc.* **49**, 598-619.
- Jongbloed, G. (1995). *Three Statistical Inverse Problems*. Unpublished Ph.D. dissertation, Delft University of Technology, Delft.
- Jongbloed, G. (1998). The iterative convex minorant algorithm for nonparametric estimation. *J. of Computational and Graphical Stat.* **7**, 310-321
- Jongbloed, G. (2000). Minimax lower bounds and moduli of continuity. *Stat. and Prob. Letters* **50**, 279-284
- Kim, J. and Pollard, D. (1990). Cube root asymptotics. *Ann. Statist.* **18**, 191-219.
- Lavee, D., Safrie, U.N., and Meilijson, I. (1991). For how long do trans-Saharan migrants stop over at an oasis? *Ornis Scandinavica* **22**, 33-44.
- Mammen, E. (1991). Nonparametric regression under qualitative smoothness assumptions. *Ann. Statist.* **19**, 741-759.
- Meyer, M. C. (1997). *Shape Restricted Inference with Applications to Nonparametric Regression, Smooth Nonparametric Function estimation, and Density Estimation*. Ph.D. Dissertation, Department of Statistics, University of Michigan, Ann Arbor, Michigan.
- Prakasa Rao, B. L. S. (1969). Estimation of a unimodal density. *Sankya Series A* **31**, 23 - 36.
- Robertson, T., Wright, F. T. and Dykstra, R. L. (1988). *Order Restricted Statistical Inference*. Wiley, New York.
- Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.
- Van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer Verlag, New York.
- Van Eeden, C. (1956). Maximum likelihood estimation of ordered probabilities. *Proc. Kon. Ned. Akad. Wet.* **A 59**, 444-455.
- Van Eeden, C. (1957). Maximum likelihood estimation of partially ordered probabilities, I. *Proc. Kon. Ned. Akad. Wet.* **A 60**, 128 - 136.

- Wang, Y. (1994). The limiting distribution in concave regression. *Preprint*, University of Missouri, Columbia.
- Woodroffe, M. and Sun, J. (1993). A penalized maximum likelihood estimate of  $f(0+)$  when  $f$  is non-increasing. *Statistica Sinica* **3**, 501 - 515.
- Wright, F. T. (1981). The asymptotic behavior of monotone regression estimates. *Ann. Statist.* **9**, 443 - 448.

DEPARTMENT OF MATHEMATICS  
DELFT UNIVERSITY OF TECHNOLOGY  
P.O. Box 356  
2600 AJ DELFT  
THE NETHERLANDS  
*e-mail: P.Groeneboom@twi.tudelft.nl*

DEPARTMENT OF MATHEMATICS  
VRIJE UNIVERSITEIT AMSTERDAM  
1081 HV AMSTERDAM  
THE NETHERLANDS  
*e-mail: geurt@cs.vu.nl*

UNIVERSITY OF WASHINGTON  
STATISTICS  
BOX 354322  
SEATTLE, WASHINGTON 98195-4322  
U.S.A.  
*e-mail: jaw@stat.washington.edu*