Statistica Neerlandica (2003) Vol. 57, nr. 1, pp. 136–157

# Density estimation in the uniform deconvolution model

P. Groeneboom[1]

*Department of Mathematics, Vrije Universiteit, De Boelelaan 1081a, 1081 HV Amsterdam, Department of Mathematics, Delft University of Technology, Mekelweg 4, 2628 CD Delft*

G. Jongbloed[2]

*Department of Mathematics, Vrije Universiteit*

We consider the problem of estimating a probability density function based on data that are corrupted by noise from a uniform distribution. The (nonparametric) maximum likelihood estimator for the corresponding distribution function is well defined. For the density function this is not the case. We study two nonparametric estimators for this density. The first is a type of kernel density estimate based on the empirical distribution function of the observable data. The second is a kernel density estimate based on the MLE of the distribution function of the unobservable (uncorrupted) data.

*Key Words and Phrases:* maximum likelihood, kernel estimator, asymptotic distribution.

## 1 Introduction

Missing and incomplete data problems have been an important field of research in mathematical statistics during the past decades. In particular, a lot of theory has been developed for the analysis of right-censored data and martingale methods have been very successful here. Other forms of censoring have also been studied, like interval censored data. In the latter situation the data are "very incomplete" in the sense that one can never observe the variable of interest directly, but one only has information about a region or interval to which the variable of interest belongs.

In contrast with the situation for right-censored data, martingales have not been very useful for developing theory for interval censored data and one is forced to develop theory from scratch. For example, the familiar $\sqrt{n}$-convergence, still valid in the situation of right-censored data, generally does not hold any more and pointwise asymptotic normality of the maximum likelihood estimator of the unknown distribution function will generally not hold either.

[1]p.groeneboom@its.tudelft.nl
[2]geurt@cs.vu.nl

These two facts: rates of convergence slower than $\sqrt{n}$ and non-normal limit distributions for maximum likelihood estimators are characteristic for many incomplete data problems. These problems belong to the category of inverse statistical problems (see, e.g., GROENEBOOM, 1996), and are subject to "heavy loss of information". A convenient way of describing the situation is by introducing a *hidden space*, containing the information one would like to observe, an *observation space*, containing the observations one *can* actually observe, and a mapping from observation space to hidden space. We use this set-up in the sequel, in the development of asymptotic distribution theory.

We focus on one particular problem of this type: deconvolution, which poses problems of the same nature as interval censored data. In fact, in Section 3 we give an example of a situation where the two models give the same type of maximum likelihood estimator, with the same convergence rate and the same (non-normal) limit distribution.

In the deconvolution model one also never can observe the variable of interest directly, but one only has access to the sum of the variable of interest and some "noise" added to it. We specialize further to "uniform noise". In fact, the present investigation has been inspired by work on "deblurring of pictures" by ROY CHOUDHURY (1998), where methods for recovering pictures, blurred by Poisson noise are studied. Further work on the latter problem is reported in O'SULLIVAN and ROY CHOUDHURY (2001).

Another motivation for the present study has been to provide an alternative for "using EM with early stopping", which at present enjoys much popularity in certain circles. The trouble with the latter procedure is that it is very hard to determine what one actually gets if one does EM with early stopping. One gets results that will be dependent on the distribution, used as a starting point for the algorithm, and this prevents the development of distribution theory.

We show that if one wants to estimate a density in deconvolution problems, one can compute the MLE of the *distribution function* (without any early stopping!) and use a convolution of a kernel with this MLE to estimate the density. In ROY CHOUDHURY (1998) the latter procedure is discussed as a method for estimating the deblurred picture, where the picture plays the role of a density. Section 5 is based on the last part of a special topics course, given by the first author in the spring of 1998 at the University of Washington, Seattle, USA. Parts of this course are also included as an appendix in ROY CHOUDHURY (1998).

The type of deconvolution we study has also been called "boxcar deconvolution", and recent work on this model (not taking the maximum likelihood approach) can be found in JOHNSTONE and RAIMONDO (2002) and HALL *et al.* (2001).

## 2 Uniform deconvolution

Consider independent and identically distributed random variables $Z_1$, $Z_2$,... with density

$$g(z) = g_F(z) = \int_{\mathbb{R}} k(z - x)dF(x)$$

where $k$ is a *known* probability density on $\mathbb{R}$ and $F$ is an arbitrary distribution function. Estimating $F$ based on a sample from $g_F$ is a *statistical inverse problem* in the sense that the sampling distribution is the image of the distribution of interest under a known transformation $K$. Estimating $F$ can in that sense be interpreted as first to estimate $g$ and then to apply ''some inverse'' of $K$ to obtain an estimate for $F$. This inverse problem, however, can also be viewed as an *incomplete data problem*. The complete observations consist of independent pairs $(X_i, Y_i)$ where $X_i \sim F$ and $Y_i \sim k$ are independent. If the complete data were observable, the distribution function $F$ could be adequately estimated by the empirical distribution function of the $X_i$'s (or better if more information on $F$ is available). However, part of the data is missing. We do not observe the pairs $(X_i, Y_i)$ but only observe the random variables $Z_i = X_i + Y_i$. The question then is: how can we estimate $F$ and quantities related to $F$ based on the sample of $Z$'s?

In this paper we will not consider the deconvolution problem in great generality. We focus on one specific convolution kernel: the uniform density on [0,1). In that case, we have

$$g(z) = g_F(z) = \int k(z - x)dF(x) = \int_{(z-1,\, z]} \mathrm{d}F(x) = F(z) - F(z - 1), \quad z \in \mathbb{R}. \tag{1}$$

Note that, formally, the distribution function $F$ can be recovered from the density $g$ by

$$F(z) = \sum_{i=0}^{\infty} g(z - i). \tag{2}$$

Moreover, we assume $F_0(0) = 0$ throughout.

In the deblurring application considered in ROY CHOUDHURY (1998), it is necessary to estimate the probability density $f_0$ of $X$ based on data from the convolution of $f_0$ with a uniform density on $(0, b)$ for general (known) $b > 0$. Although restricting the convolution kernel to the standard uniform density is quite a loss with regard to the general deconvolution problem where $k$ is a general density, the results of this paper are also valid for the nonstandard uniform deconvolution, i.e., where $b > 0$ arbitrary (but known). The procedure then is first to transform the data by division through $b$ and use the techniques for the standard uniform deconvolution model of this paper to obtain an estimate of the rescaled density $f_0$. This estimate can then again be rescaled to get a density estimate for $f_0$. We will not restate our results for these more general $b$ since they are obvious.

Of course, in order for (2) to be valid, $g$ should belong to the image of the set of distribution functions under the convolution operator. This range is a strict subset of the class of all probability densities. In fact, this is one of the characteristic features of inverse problems. The nonparametric model on the hidden space (here the product measures that can be constructed from an arbitrary distribution function and the uniform density) transforms under the mapping from the hidden to the observation space to a specific model on the observation space. This model is smaller than the usual nonparametric model that is commonly considered in the observation space. The example below, which will be used in the sequel to illustrate the various estimators, stresses the fact that applying (2) to densities outside the range of the convolution operator, gives functions $F$ that are not distribution functions.

EXAMPLE 1. Consider the exponential model for $F$. This means that we assume $F$ to belong to the class

$$\mathcal{F} = \{F_\theta : \theta > 0\} \text{ with } F_\theta(x) = 1 - e^{-\theta x}$$

for $x > 0$. Then the sampling densities belong to the parametric model

$$\mathcal{G} = \{g_\theta : \theta > 0\} \text{ with } g_\theta(z) = \begin{cases} 1 - e^{\theta z} & \text{for } z \in (0,1) \\ (e^\theta - 1)e^{-\theta z} & \text{for } z \geqslant 1. \end{cases} \tag{3}$$

This is not one of the ''usual parametric families'' on $(0, \infty)$. It is important to note that any family $\mathcal{F}$ of distribution functions on $\mathbb{R}$ results in a family of densities on $\mathbb{R}$ using (1).

Now suppose we take the class of exponential densities as model in the observation space, so $g_\theta(z) = \theta e^{-\theta z}$ in (3). Then we get as associated model in the hidden space

$$F_\theta(x) = \theta e^{-\theta x}\left(1 + e^\theta + e^{2\theta} + \ldots + e^{\lfloor x \rfloor \theta}\right).$$

It is clear that $F_\theta(0) = \theta$ (which is not necessarily in (0,1)) and that $F_\theta$ is initially *decreasing*. This shows that the class of exponential densities is outside the range of the set of distribution functions under the uniform convolution mapping.

In Section 3 we discuss the nonparametric maximum likelihood estimator $\hat{F}$ of the distribution function $F$ and address computational issues associated with that estimator. Section 4 is devoted to a kernel estimator which is based on (2). The sampling density is estimated using a kernel estimator and this estimate is plugged into (2). This estimator has the disadvantage that it is not a probability density function due to the fact that the usual kernel density estimates are not contained in the class of densities of the form (1). However, its asymptotic behavior can be derived using the classical central limit theorem for triangular arrays. In Section 5 we introduce a natural kernel density estimator that is based on $\hat{F}$ and study its asymptotic behavior.

## 3 Nonparametric maximum likelihood estimator $\hat{F}$

In this section we consider the nonparametric maximum likelihood approach to estimating the distribution function $F$ in the uniform deconvolution model. Given data $z_1 < z_2 < \cdots < z_n$, the loglikelihood function on the class of distribution functions is given by

$$l(F) = \frac{1}{n}\sum_{i=1}^{n} \log g_F(z_i) = \frac{1}{n}\sum_{i=1}^{n} \log(F(z_i) - F(z_i - 1)).$$

This estimator was studied in GROENEBOOM and WELLNER (1992). Note that $l(F) \leqslant 0$ for all $F$, so that the loglikelihood function is bounded above. Note also that $l$ depends on $F$ only via its values at the observed points $z_i$ and the points $z_i - 1$. This means that when maximizing $l$, attention may be restricted to discrete distribution functions having their mass concentrated at the points $z_i$ and $z_i - 1$. However, any maximizer of $l$ will assign mass zero to these latter points. The loglikelihood is increased by shifting such masses to the first $z_j$ bigger than $z_i - 1$. Finally, note that $l$ can be viewed as a strictly concave function on the set of all these discrete distribution functions. All this gives that we can uniquely define the MLE $\hat{F}$ as the discrete distribution function having its mass concentrated on the observed $z_i$'s that maximizes $l$.

Another well-known feature in this setting is that if the support of the distribution with distribution function $F$ is contained in (0,1), the MLE can be computed using the fact that the loglikelihood has a structure equivalent to the loglikelihood for current status data. If the support of $F$ is bigger, this correspondence does not exist anymore and the estimator is to be computed using iterative optimization techniques. Figure 1 shows the maximum likelihood estimator of the mixing distribution $F$ and the corresponding density $g$ based on a sample of size $n = 500$. The true $F$ (mean one exponential) and $g$ are also included in the picture.

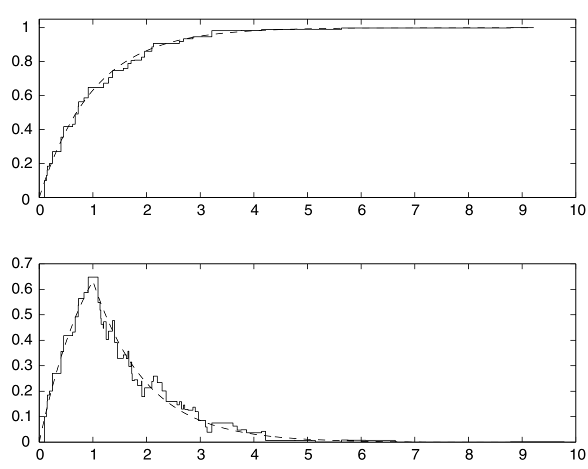


Fig. 1. Top: ML estimate of the distribution function $F$ based on a sample of size $n = 500$ with the true (exponential) distribution. Bottom: the corresponding estimate for the sampling density $g$ with the true density.

To obtain Figure 1, we used a Newton algorithm to compute $\hat{F}$, where the basic quadratic maximizations are performed using the *support reduction algorithm*. This algorithm is introduced in GROENEBOOM, JONGBLOED and WELLNER (2002), where it is also used to compute the maximum likelihood estimator of the distribution function in the Gaussian deconvolution model. It is an algorithm that belongs to the *vertex direction* family of algorithms and is closely related to the algorithm proposed in SIMAR (1976).

## 4 Inverse kernel density estimation

There are various methods that come to mind if one wants to estimate $f$ based on a sample from $g$. For example, the convolution structure of (1) suggests the use of the Fourier transform to get an estimate of $f$. However, this approach is not attractive since the characteristic function of the uniform distribution has zeroes.

In this section we study an estimator that is suggested by (2): an inverse density estimator. Indeed, assuming that $F$ has a density, we get that

$$f(z) = \sum_{i=0}^{\infty} g'(z-i),$$

so that a usual density estimate of $g$ can be used to estimate $f$. In this section we study an inverse kernel density estimator for $f$. This estimator is based on a conventional kernel estimate of $g$,

$$g_{n,h}(z) = \int_{\mathbb{R}} K_h(z-x)\mathrm{d}\mathbb{G}_n(x) = \frac{1}{h}\int_{\mathbb{R}} K\Big(\frac{z-x}{h}\Big)\mathrm{d}\mathbb{G}_n(x).$$

Here $h > 0$ is the bandwidth of the density estimate and $K$ is the kernel. Based on this estimate, we define

$$\begin{aligned}\tilde{f}_{n,h}(z) = \sum_{i=0}^{\infty} g'_{n,h}(z-i) &= \frac{1}{h^2}\sum_{i=0}^{\infty}\int_{\mathbb{R}} K'\left(\frac{z-x-i}{h}\right)\mathrm{d}\mathbb{G}_n(x)\\ &= \sum_{i=0}^{\infty}\int_{\mathbb{R}} K'_h(z-x-i)\mathrm{d}\mathbb{G}_n(x). \qquad (4)\end{aligned}$$

For this estimator we have the following theorem.

THEOREM 1. *Let $K$ be a compactly supported twice continuously differentiable kernel and let the distribution function $F_0$ be continuously differentiable at the point $t \in (0, M)$. Define the estimator $\tilde{f}_{n,h}(\mathrm{t})$ of the derivative $f_0$ of $F_0$ at $t$ by* (4) *for $t \in (0, M)$. Then:*

(i) *As $n \to \infty$, $h \downarrow 0$, and $nh \to \infty$,*

$$\{nh^3\}^{1/2}\left\{\tilde{f}_{n,h}(t) - \sum_{i=0}^{\infty}\int K'_h(t-x-i)dG_0(x)\right\} \xrightarrow{\mathcal{D}} N(0,\sigma^2), \qquad (5)$$

*where*

$$\sigma^2 = F_0(t)\int K'(u)^2 du. \tag{6}$$

*(ii) Suppose that $f_0$ is twice differentiable at t. Then, for $h \downarrow 0$,*

$$h^{-2}\left(\sum_{i=0}^{\infty}\int K_h'(t-z-i)g_0(z)dz - f_0(t)\right) \to \frac{1}{2}f_0''(t)\int u^2K(u)du =: \beta. \tag{7}$$

*(iii) If $h_n \sim c{\cdot}n^{-1/7}$, for some $c > 0$, as $n \to \infty$,*

$$n^{2/7}\{\tilde f_{n,h}(t) - f_0(t)\} \xrightarrow{\mathcal{D}} N(\beta c^2, c^{-3}\sigma^2), \quad n \to \infty, \tag{8}$$

*where $\sigma^2$ is as in (6), and $\beta$ as in (7).*

PROOF. For (*i*) write

$$\{nh^3\}^{1/2}\left\{\tilde f_{n,h}(t) - \sum_{i=0}^{\infty}\int K_h'(t-z-i)dG_0(z)\right\} = \sum_{k=1}^{n}\xi_{nk}$$

with

$$\xi_{nk} = (h^3/n)^{1/2}\sum_{i=0}^{\infty}\left(K_h'(t-Z_k-i) - \int K_h'(t-z-i)dG_0(z)\right).$$

Note that the infinite sum is in fact a finite sum due to the fact that $F_0(0) = 0$. By the central limit theorem for triangular arrays (theorem 7.2 in BILLINGSLEY, 1968), we have that

$$s_n^{-1}\sum_{k=1}^{n}\xi_{nk} \xrightarrow{\mathcal{D}} N(0,1),$$

where

$$\begin{aligned} s_n^2 &= \sum_{k=1}^{n}\mathrm{var}\,\xi_{nk} = nE\xi_{n1}^2 = h^3\mathrm{var}\left(\sum_{i=0}^{\infty}K_h'(t-Z_1-i)\right) \\ &= h^3\int_{z=0}^{\infty}\left(\sum_{i=0}^{\infty}K_h'(t-z-i)\right)^2 g_0(z)dz - h^3\left(\int_{z=0}^{\infty}\sum_{i=0}^{\infty}K_h'(t-z-i)g_0(z)dz\right)^2 \\ &= I_{1,n} - I_{2,n}^2. \end{aligned}$$

Because $K$ has compact support, for $h$ sufficiently small, the first term can be written as

$$\begin{aligned} I_{1,n} &= h^3\sum_{i=0}^{\infty}\int_{z=0}^{\infty}K_h'(t-z-i)^2 g_0(z)dz = \sum_{i=0}^{\infty}\int K'(u)^2 g_0(t-i-hu)du \\ &= \int K'(u)^2F_0(t-hu)du \to F_0(t)\int K'(u)^2du, \text{ for } h\downarrow 0, \end{aligned}$$

by dominated convergence. For the second term we have that

$$I_{2,n} = h^{3/2} \sum_{i=0}^{\infty} \int_{z=0}^{\infty} K_h'(t-z-i)g_0(z)dz$$
$$= h^{3/2} \sum_{i=0}^{\infty} \int h^{-2} K'\left(\frac{t-z-i}{h}\right) g_0(z)dz = \sqrt{h} \int K'(u) f_0(t-hu)du,$$

which is of smaller order than the first term. Hence, $s_n^2 \to F_0(t) \int K'(u)^2 du$, yielding (*i*). Finally, note that

$$P(|\xi_{nk}| > \epsilon s_n) \leq \frac{\mathrm{var}(\xi_{nk})}{\epsilon^2 s_n^2} = \frac{1}{\epsilon^2 n} \text{ and } \xi_{nk}^2 \leq \frac{c}{nh}$$

for a $c > 0$. Hence,

$$\frac{n}{s_n^2} E\xi_{nk}^2 1_{[|\xi_{nk}|>\epsilon s_n]} \leqslant \frac{n}{s_n^2} \cdot \frac{c}{nh} \cdot \frac{1}{\epsilon^2 n} = \frac{c}{\epsilon^2 s_n^2} \cdot \frac{1}{nh} \to 0$$

as $nh \to \infty$.

For (*ii*), note that

$$\sum_{i=0}^{\infty} \int K_h'(t-z-i) g_0(z)dz = h^{-2} \sum_{i=0}^{\infty} \int K'\left(\frac{t-z-i}{h}\right) g_0(z)dz$$
$$= h^{-1} \sum_{i=0}^{\infty} \int K'(u) g_0(t-i-hu)du = h^{-1} \int K'(u) F_0(t-hu)du$$
$$= h^{-1} \int K'(u)(F_0(t) - huf_0(t) + \frac{1}{2}h^2u^2 f_0'(t) - \frac{1}{6}h^3u^3 f_0''(t-h\zeta u)du$$
$$\to 0 + f_0(t) + 0 + \frac{1}{2}h^2 f_0''(t) \int u^2 K(u)du.$$

Here we use that $K$ is a smooth probability density with zero mean and partial integration.

Now, for (*iii*) note that

$$n^{2/7}(\tilde{f}_{n,h}(t) - f_0(t)) = c^{-3/2}(nh^3)^{1/2}\left(\tilde{f}_{n,h}(t) - \sum_{i=0}^{\infty} \int K_h'(t-z-i)g_0(z)dz\right)$$
$$+ n^{2/7}\left(\sum_{i=0}^{\infty} \int K_h'(t-z-i)g_0(z)dz - f_0(t)\right)$$
$$\xrightarrow{\mathcal{D}} N(\beta c^2, c^{-3}\sigma^2),$$

as $n \to \infty$ by (*i*) and (*ii*). □

COROLLARY 1. *To estimate f*(*t*) *with kernel estimator* (4)*, the asymptotic MSE optimal bandwidth is given by*

$$h = h_n = c_{opt} \cdot n^{-1/7} \text{ with } c_{opt} = \left(\frac{3\sigma^2}{4\beta^2}\right)^{1/7},$$

*with* $\sigma^2$ *and* $\beta$ *as in* (6) *and* (7).

*For the biweight kernel* $K(u) = \frac{15}{16}(1-u^2)^2 1_{[-1,1]}(u)$ *this gives*

$$\sigma^2 = \frac{15}{7}F(t_0) \text{ and } \beta = \frac{1}{14}f''(t_0) \Rightarrow c_{opt} = \left(\frac{2835F(t_0)}{64f''(t_0)^2}\right)^{1/7} \approx 1.72\left(\frac{F(t_0)}{f''(t_0)^2}\right)^{1/7}.$$

PROOF. For $c_{opt}$ we minimize the sum of the squared bias and the variance given in theorem 4 (*iii*) as a function of $c$. □

Figure 2 shows $\tilde{f}_{n,h}$ based on the same data set as Figure 1. The choice of $K$ is the biweight kernel and bandwidth $h = 0.88$ (asymptotically MSE optimal for $t_0 = 1$). Note that the optimal bandwidth in our simulation example (exponential $F_0$ and $n = 500$) depends on $t_0$ via $t_0 \mapsto 0.71(e^{2t_0} - e^{t_0})^{1/7}$. This is an increasing function, explaining to some extent that the density estimate in Figure 2 seems oversmoothed near zero and undersmoothed in the right tail. Of course, in practical situations one could first estimate $F_0$ and $f_0''$ using a pilot bandwidth and then compute a density estimate with varying bandwidth based on this pilot estimate. Also, one could apply boundary kernels to get rid of the obvious boundary problems of the estimates at zero. Boundary kernels in the spirit of Wand and Jones (1995) have a discontinuity at zero. In the current context this would be a problem since we need the density estimate to be differentiable. Although the estimate of Figure 2 could be improved near the boundary, we will not pursue this here.

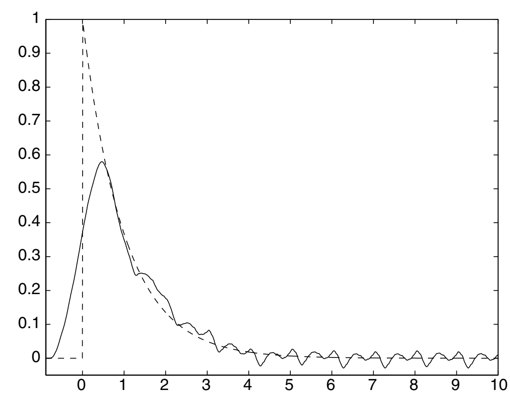


Fig. 2. The inverse kernel estimate $\tilde{f}_{n,h}$ based on a sample of size 500, using the biweight kernel bandwidth $h = 0.88$.

## 5 Density estimation based on $\hat{F}$

We want to study the asymptotic behavior of the estimator

$$\hat{f}_{n,h}(t) = \int K_h(t-y)d\hat{F}_n(y)$$

of the derivative $f_0(t)$ at $t$ of the distribution function $F_0$ of the distribution we want to estimate, where $\hat{F}_n$ is the MLE of this distribution function $F_0$ as defined in Section 3. Figure 3 shows this estimate based on the same data as Figures 1 and 2. The estimate $\hat{f}_{n,h}$ has the same undesirable behavior near zero as $\tilde{f}_{n,h}$. In this case it is straightforward to implement boundary kernels in the spirit of Section 2.11 in WAND and JONES (1995). The resulting estimate (which coincides with $\hat{f}_{n,h}$ on $[h, \infty)$) is shown in Figure 4.

We impose the following condition on $F_0$:

**(F)** The distribution corresponding to the distribution function $F_0$ has support $[0, M]$, and has a continuous derivative $f_0$ which satisfies

$$\inf_{u \in (0,M)} f_0(u) > 0.$$

Note that condition **(F)** implies that for any $\delta \in (0, M \wedge 1)$:

$$\inf_{x \in [\delta, M-\delta]} \{F_0(x) - F_0(x-1)\} \wedge \{F_0(x+1) - F_0(x)\} > 0. \tag{9}$$

For the kernel function we assume

**(K)** $K_h(u) = h^{-1}K(u/h)$, where $K$ is a fixed non-negative twice differentiable kernel with support $[-1, 1]$, and bounded second derivative, that integrates to 1 on $[-1, 1]$ and is symmetric around zero.

For our pictures we used the biweight kernel, as defined in Corollary 1. This kernel satisfies **(K)**. We show below that under these conditions the following result holds:

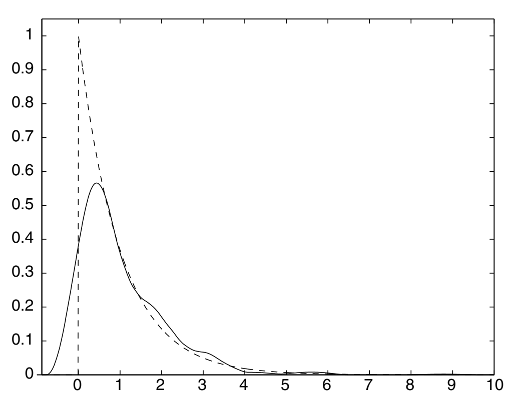


Fig. 3. The density estimate $\hat{f}_{n,h}$ based on a sample of size 500, using the biweight kernel and $h = 0.88$.

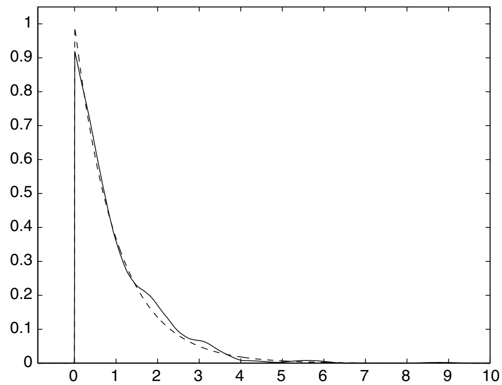


Fig. 4. The smoothed MLE density estimate based on the sample of size 500 using boundary kernels. Again, $h = 0.88$ and the kernel used is the biweight.

THEOREM 2. *Let $K$ satisfy condition* (**K**) *and let the distribution function $F_0$ satisfy condition* (**F**). *Let $t$ be a fixed point in the open interval* $(0, M)$, *such that $t \not\equiv 0 \pmod 1$ and $M - t \not\equiv 0 \pmod 1$, and let the estimator $\hat{f}_{n,h}(t)$ of the derivative $f_0$ of $F_0$ at $t$ be defined by*

$$\hat{f}_{n,h}(t) = \int K_h(t-x)\,d\hat{F}_n(x),$$

*for $t \in (0, M)$. Moreover, let $m$ be the largest integer $< M + 1$. Then:*

*(i) As $n \to \infty$, $h \downarrow 0$, and $nh^3 \to \infty$,*

$$\{nh^3\}^{1/2}\left\{\hat{f}_{n,h}(t) - \int K_h(t-x)\,dF_0(x)\right\} \xrightarrow{\mathcal{D}} N(0, \sigma^2), \tag{10}$$

*where*

$$\sigma^2 = \lim_{h\downarrow 0} h^3 \int \theta^2_{h,t,F_0}\,dG_0 > 0, \tag{11}$$

*for a function $\theta_{h,t,F_0}$, defined by*

$$\theta_{h,t,F_0}(x+k) = \begin{cases} \sum_{i=0}^{m}\{1 - F_0(x+i)\}K_h'(t-(x+i)), & \text{if } x \in [0,1], \quad k = 0, \\ -\sum_{i=0}^{k-1} K_h'(t-(x+i)) + \theta_{h,t,F_0}(x), & \text{if } x \in [0,1], \quad k = 1, \ldots, m. \end{cases} \tag{12}$$

*(ii) Suppose that $f_0$ is twice differentiable at $t$. Then, for $h \downarrow 0$,*

$$h^{-2}\left(\int K_h(t-x)\,dF_0(x) - f_0(t)\right) \to \frac{1}{2}f_0''(t)\int u^2 K(u)\,du =: \beta. \tag{13}$$

*(iii) If $h_n \sim c{\cdot}n^{-1/7}$, for some $c > 0$, as $n \to \infty$,*

$$n^{2/7}\left\{\hat{f}_{n,h}(t)-f_0(t)\right\}\xrightarrow{\mathcal{D}}N(\beta c^2,c^{-3}\sigma^2),\quad n\to\infty, \tag{14}$$

*where* $\sigma^2$ *is as in* (11), *and* $\beta$ *as in* (13).

REMARK. The statements of Theorem 2 resemble those of Theorem 1. The asymptotic bias is exactly the same for both estimators.

The definition (12) of $\theta_{h,t,F_0}$ in part (*i*) is based on a definition given in Example 11.2.3e on p. 230 of VAN DE GEER (2000). If $M \leqslant 1$, it can be seen that this is given by

$$\theta_{h,t,F_0}(x)=(1-F_0(x))K_h'(t-x)-F_0(x-1)K_h'(t-x+1)$$

if $(t-h,\ t+h)\subset(0,\ M)$. This means that $\sigma^2$ reduces to:

$$\sigma^2=F_0(t)\{1-F_0(t)\}\int K'(u)^2du. \tag{15}$$

In this case the asymptotic variances for the two estimators depend on the kernel $K$ in the same way, but the dependence on $F_0$ shows that $\hat{f}_{n,h}$ is more efficient than $\tilde{f}_{n,h}$:

$$F_0(t)\{1-F_0(t)\}<F_0(t),\quad t\in(0,M),$$

under condition (**F**) at the beginning of this section.

Apart from the pointwise asymptotic properties of $\hat{f}_{n,h}$, which we think are overall better than those of $\tilde{f}_{n,h}$, another reason to prefer $\hat{f}_{n,h}$ is that $\hat{f}_{n,h}$ is a genuine probability density function. In contrast to $\tilde{f}_{n,h}$, it only takes nonnegative values.

The assumption that $t\not\equiv 0 \pmod 1$ and $M-t\not\equiv 0 \pmod 1$ is somewhat unpleasant and probably not needed, but we saw no way to avoid this in the present proof.

The proof of Theorem 2 is much more involved than that of Theorem 1 and will be given below in a sequence of steps. The "hidden probability space" contains random variables $(X,Y)$, where $X$ has distribution function $F_0$, $Y$ has a Uniform(0,1)-distribution, and $X$ and $Y$ are independent and the mapping, relating the random variables in the hidden space to the random variables in the observation space, is:

$$T(x,y)=x+y,\quad x\in[0,M],\ y\in[0,1].$$

We will need some facts about score operators for this particular model, given in, e.g., GROENEBOOM and WELLNER (1992), Chapter 1. Let $L_2(F_0)$ denote the space of square integrable functions w.r.t. the measure $dF_0$, that is: the space of Borel measurable functions $h:\mathbb{R}\to\mathbb{R}$ such that

$$\int h^2dF_0<\infty.$$

Moreover, the space $L_2^0(F_0)$ is the subset of functions $h\in L_2(F_0)$, satisfying

$$\int hdF_0=0.$$

The score operator $L_{F_0}$, mapping scores $a\in L_2^0(F_0)$ to scores $\bar{a}$ in the observation space, only involving a Hellinger differentiable path $(F_\tau)_{\tau\in[0,\delta)}$, tending to $F_0$, is given by the conditional expectation

$$[L_{F_0}(a)](z) = E_{F_0}\left\{a(X)\middle| T(X,Y) = z\right\}.$$

Its adjoint on the space of scores contained in $L_2^0(G_0)$, where $G_0$ is defined by the convolution density $g_{F_0}$, is given by the conditional expectation

$$[L^*(\bar{a})](x) = E_{F_0}\left\{\bar{a}(Z)\middle| X = x\right\} = \int_x^{x+1} \bar{a}(z)dz.$$

Note that the adjoint operator $L^*$ does not involve the distribution function $F_0$, in contrast with the score operator $L_{F_0}$ itself (of which it is the adjoint).

The guiding principle in finding the asymptotic distribution of $\hat{f}_{n,h}(t)$ is to try to find a solution $\theta_{h,t,F}$ to the equation

$$L^*\theta_{h,t,F} = K_h(t - \cdot) - \int K_h(t-x)dF(x),$$

where $\theta_{h,t,F}$ is a score in $L_2^0(G_F)$, belonging to the (closure of the) range of $L_F$, and $G_F$ denotes the distribution with density

$$g_F(z) = F(z) - F(z-1).$$

The next step is then to prove that

$$\int K_h(t-x)d(\hat{F}_n - F_0)(x) \sim \int \theta_{h,t,F_0}(z)d(\mathbb{G}_n - G_0)(z), \tag{16}$$

in the sense that the right-hand side of (16) represents the left-hand side up to terms that are of lower order, as $n \to \infty$, and $h = h_n \downarrow 0$ (at a certain rate that will be specified below). By the central limit theorem, the right-hand side of (16), divided by

$$\sigma_{h,t} = \left\{\int \theta_{h,t,F_0}(z)^2 dG_0(z)\right\}^{1/2}$$

is asymptotically standard normal (again under some conditions on the rate at which $h$ tends to zero as a function of $n$).

Note that, since the convolution kernel is the Uniform(0,1) density, we have:

$$[L_F(a)](z) = \int_{z-1}^{z} a(x)dF(x)/\{F(z) - F(z-1)\}.$$

Defining

$$\theta(z) = \int_{z-1}^{z} a(x)dF(x)/\{F(z) - F(z-1)\}, \quad z \in \mathbb{R}, \tag{17}$$

where the ratio is defined to be zero if the denominator is zero, we have to solve the equation

$$\int_x^{x+1} \theta(z)dz = K_h(t-x) - \int K_h(t-y)dF(y),$$

for $x$ in the interior of the support of $F$, where $\theta$ will in fact depend on $t$, $h$ and $F$. Differentiation w.r.t. $x$ yields the equation

$$\theta(x+1) - \theta(x) = -K_h'(t-x), \tag{18}$$

if $\theta$ is continuous at the points $x$ and $x + 1$.

Let the function $\phi$ be defined by

$$\phi(x) = \begin{cases} \int_0^x a(u)dF(u), & \text{if } x \geq 0, \\ 0, & \text{otherwise.} \end{cases} \tag{19}$$

Then, by (18), $\phi$ should satisfy

$$\frac{\phi(x+1) - \phi(x)}{F(x+1) - F(x)} - \frac{\phi(x) - \phi(x-1)}{F(x) - F(x-1)} = -K_h'(t-x), \tag{20}$$

if $x$ and $x + 1$ are points of continuity of $\theta$. Conversely, if $\phi$ satisfies (20), then $\theta$ satisfies (18).

The following lemma is one of the stepping stones in achieving our goal (compare with Lemma 3.3 in Groeneboom, 1996).

Lemma 1. *Let the kernel $K$ satisfy condition* **(K)** *above. Furthermore, let $t \in (0,M)$ and $h \in (0,1/2)$ satisfy $(t - h,t + h) \subset (0,M)$, and let $K_h'(t - x) = \frac{\mathrm{d}}{\mathrm{du}}\mathrm{K_h}(u)\Big|_{u=t-x}$, where $K_h(u) = h^{-1}K(u/h)$. Then we have:*

(i) *Suppose that the distribution function $F$ on $[0,\infty)$ satisfies $F(M + 1) = 1$ and (9) for a $\delta \in (0, \frac{1}{2}(M \wedge 1))$ (not necessarily for all $\delta \in (0,M\wedge1)$). Then the equation*

$$\begin{aligned}&\{F(x+1) - F(x-1)\}\psi(x) - \{F(x+2) - F(x+1)\}\psi(x+1) \\ &- \{F(x-1) - F(x-2)\}\psi(x-1) \\ &= K_h'(t-x), \quad x \in \mathbb{R},\end{aligned} \tag{21}$$

*under the side condition $\psi(x) = 0$, $x \notin [0,M + 1]$, has a bounded solution $\psi_{h,t,F}$.*

(ii) *Let $F$ be as in condition (i), and let the distribution function $F_0$ satisfy condition* **(F)** *above. Moreover, let $\psi_{t,h,F}$ be defined as in (i), and let $\theta_{h,t,F}$ be defined by*

$$\theta_{h,t,F}(x) = \{F(x+1) - F(x)\}\psi(x) - \{F(x-1) - F(x-2)\}\psi(x-1), \; x \in \mathbb{R}. \tag{22}$$

*Then*

$$\int \theta_{h,t,F}(x) g_F(x) dx = 0, \tag{23}$$

*and*

$$\int K_h(t-y)d(F-F_0)(y) = -\int \theta_{h,t,F}(z)dG_0(z).$$

REMARK. The function $\phi_{h,t,F}$, defined by

$$\phi_{h,t,F}(x) = \{F(x) - F(x-1)\}\{F(x+1) - F(x)\}\psi_{h,t,F}(x), \quad x \in \mathbb{R}, \tag{24}$$

satisfies (20), if the denominators in (20) are strictly positive. The introduction of the function $\psi_{h,t,F}(x)$ is needed to cover the situation that the denominators are zero. If condition **(F)** is satisfied (as is true for the distribution function $F_0$), the introduction of the function $\psi_{h,t,F_0}$ is not necessary in the definition of $\theta_{h,t,F_0}$. We will however use the function $\psi_{h,t,F}$ in the case that $F$ equals the MLE $\hat{F}_n$. In particular we will need the fact that $\phi_{h,t,\hat{F}}(x)$ contains the factors $\hat{F}_n(x) - \hat{F}_n(x-1)$ and $\hat{F}_n(x+1) - \hat{F}_n(x)$, as is clear from its representation (24) in terms of $\psi_{h,t,\hat{F}}(x)$.

PROOF OF LEMMA 1. Fix $x \in [0,1)$ and let $m$ be the largest integer such that $x + m \le M + 1$ and $F(x + m - 1) < 1$. We define, for $j = 0, \dots, m$,

$$\alpha(x+j) = F(x+j) - F(x+j-1). \tag{25}$$

Moreover, we define for $j = 0,\dots,m$,

$$\beta(x+j) = F(x+j+1) - F(x+j-1), \tag{26}$$

and

$$\mu(x+j) = K_h'(t-x-j). \tag{27}$$

Note that $\beta(x+j) > 0$, for all $j$, $0 \le j \le m$, again by the positivity condition on the differences $F(x) - F(x-1)$ on $[\delta, M-\delta]$.

We now get the matrix equation

$$A(x)\underline{\psi(x)} = \underline{\mu(x)},$$

where $A(x)$ is the tridiagonal matrix

$$A(x) = \begin{pmatrix} \beta(x) & -\alpha(x+2) & 0 & 0 & \cdots & 0 & 0 \\ -\alpha(x) & \beta(x+1) & -\alpha(x+3) & 0 & \dots & 0 & 0 \\ 0 & -\alpha(x+1) & \beta(x+2) & -\alpha(x+4) & \dots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & -\alpha(x+m-1) & \beta(x+m) \end{pmatrix},$$

$$\underline{\psi(x)} = \begin{pmatrix} \psi(x) \\ \vdots \\ \psi(x+m) \end{pmatrix} \text{ and } \underline{\mu(x)} = \begin{pmatrix} \mu(x) \\ \vdots \\ \mu(x+m) \end{pmatrix}.$$

As is well-known, this system of equations can be solved in the following way. Let $u_j$, $j = 0, \dots, m$, be recursively defined in the following way:

$$b_0 = \beta(x), \quad u_0 = \mu(x)/b_0, \tag{28}$$

and

$$b_j = \beta(x+j) - \alpha(x+j-1)\alpha(x+j+1)/b_{j-1},$$
$$u_j = \left\{\mu(x+j) - \alpha(x+j-1)u_{j-1}\right\}/b_j, \tag{29}$$

for $j = 1, \ldots, m$. Then the solution of the system can be found by the following back-substitution:

$$\psi(x+m) = u_m, \tag{30}$$

$$\psi(x+j) = \psi(x+j+1) - \frac{\alpha(x+j+2)\psi(x+j+1)}{b_j}, \quad j = m-1, \ldots, 0. \tag{31}$$

We have, if $m > 1$,

$$b_1 = F(x+2) - F(x) - \{F(x+2) - F(x+1)\}\{F(x) - F(x-1)\}/\{F(x+1) - F(x-1)\} > F(x+2) - F(x) - \{F(x+2) - F(x+1)\} = F(x+1) - F(x) > 0$$

since $0 \leqslant \{F(x) - F(x-1)\}/\{F(x+1) - F(x-1)\} < 1$ and $F(x+2) - F(x) > F(x+1) - F(x)$. Continuing in this way, we find inductively, that

$$b_j > 0, \quad 0 \leqslant j < m.$$

Moreover, by the positivity condition on the differences $F(x) - F(x-1)$ on $[\delta, M+1-\delta]$, the coefficients $b_j$ are bounded away from zero, for any choice of $x \in [0,1)$ and $0 \leq j < m$. The solution $\psi$ is therefore bounded. The function $\phi$, defined by

$$\phi(x) = \{F(x) - F(x-1)\}\{F(x+1) - F(x)\}\psi(x)$$

is therefore also bounded.

We now have, by (22),

$$\theta_{h,t,F}(x+1) - \theta_{h,t,F}(x) = -K_h'(t-x), \quad x \in [0, M+1].$$

Furthermore, denoting $\psi_{h,t,F}$ by $\psi$ for convenience of notation, we get

$$\begin{aligned}
&\int \theta_{h,t,F}(z)g_F(z)dz \\
&= \int \psi(z)\{F(z+1) - F(z)\}\{F(z) - F(z-1)\}dz \\
&\quad - \int \psi(z-1)\{F(z-1) - F(z-2)\}\{F(z) - F(z-1)\}dz \\
&= \int \psi(z)\{F(z+1) - F(z)\}\{F(z) - F(z-1)\}dz \\
&\quad - \int \psi(z)\{F(z) - F(z-1)\}\{F(z+1) - F(z)\}dz \\
&= 0.
\end{aligned}$$

Hence

$$\begin{aligned}\int \theta_{h,t,F}(z)dG_0(z) &= \int \theta_{h,t,F}(z)g_0(z)dz = \int \theta_{h,t,F}(z)\{g_0(z) - g_F(z)\}dz \\ &= \int \theta_{h,t,F}(z)\{F_0(z) - F_0(z-1) - (F(z) - F(z-1))\}dz \\ &= -\int \big\{\theta_{h,t,F}(z+1) - \theta_{h,t,F}(z)\big\}\{F_0(z) - F(z)\}dz \\ &= \int K_h'(t-z)\{F_0(z) - F(z)\}dz = \int K_h(t-z)d(F_0 - F)(z).\end{aligned}$$

□

Since $F_0$ satisfies condition **(F)**, and since (24), with $F = F_0$, implies that $\phi_{h,t,F_0}$ is bounded and zero outside $(0,M)$, we have that $\phi_{h,t,F_0}$ is absolutely continuous w.r.t. $F_0$ on the interval $(0,M)$. So we can define $a_{h,t,F_0}$ as the Radon–Nikodym derivative of the function $\phi_{h,t,F_0}$ w.r.t. $F_0$ on $(0,M)$. The condition on the density $f_0$ also ensures that this Radon–Nikodym derivative is almost surely bounded on $(0,M)$ and hence $a_{h,t,F_0} \in L_2(F_0)$, where we use (23) of part (*ii*) of Lemma 1.

Note that it was shown in the proof of Lemma 5 that, under the conditions of this lemma, we have the following relation

$$\int \theta_{h,t,\hat{F}_n} dG_0 = \int \theta_{h,t,\hat{F}_n}(z)\big\{F_0(z) - F_0(z-1) - \{\hat{F}_n(z) - \hat{F}_n(z-1)\}\big\}dz. \quad (32)$$

In the proof of Theorem 2 we also need the following lemma.

LEMMA 2. *Let condition* (**F**) *be satisfied. Then we have for each deterministic sequence of points* $(t_n)$*, converging to a fixed point* $t \in (0,M)$*, and for each* $c > 0$*:*

$$\sup_{u\in[-c,c]} \big|\hat{F}_n(t_n + n^{-1/3}u) - F_0(t_n)\big| = \mathcal{O}_p(n^{-1/3}). \quad (33)$$

The proof of Lemma 2 proceeds along similar lines as the proof of Lemma 4.6 on p. 155 of GROENEBOOM (1996). It is omitted here for reasons of space. It shows that the MLE has so-called "cube root $n$" convergence, a property it shares with the MLE in the case of interval censoring.

PROOF OF THEOREM 2. By the consistency of $\hat{F}_n$, we may assume that condition (9) is satisfied for a $\delta \in (0, \frac{1}{2}(M \wedge 1))$, and $n$ sufficiently large, and therefore that the equation (21) has a solution $\psi_{h,t,\hat{F}_n}$ for $\hat{F}_n = F$. Hence we can define $\phi_{h,t,\hat{F}_n}$ by (24).

Let $\tau_0 = 0$, $\tau_{m+1} = M + 1$ and $\tau_1 < \cdots < \tau_m$ be the points of jump of $\hat{F}_n$, and let $J_i$ denote the interval $[\tau_i, \tau_{i+1})$, for $0 \leqslant i < m$. The interval $J_m$ will denote the closed interval $[\tau_m, \tau_{m+1}]$. Let the function $\xi_{h,t,F}$ be defined by

$$\xi_{h,t,F}(x) = \begin{cases} \frac{\phi_{h,t,F}(x)}{F(x)\{1-F(x)\}}, & \text{if } 0 < F(x) < 1, \\ 0, & \text{otherwise.}\end{cases}$$

Definition (24) implies that the function $\xi_{h,t,F}$ is bounded if $F$ satisfies the conditions of Lemma 1.

We define a piecewise constant version of $\xi_{h,t,F_0}$, constant on the intervals $J_i$, by

$$\bar{\xi}_{h,t,\hat{F}_n}(x) = \begin{cases} \xi_{h,t,F_0}(\tau_i), & \text{if } x \in J_i, \text{ and } F_0(x) > \hat{F}_n(\tau_i) \text{ for all } x \in J_i, \\ \xi_{h,t,F_0}(\tau_{i+1}-), & \text{if } x \in J_i, \text{ and } F_0(x) < \hat{F}_n(\tau_i) \text{ for all } x \in J_i, \\ \xi_{h,t,F_0}(u), & \text{if } x \in J_i, \text{ and } F_0(u) = \hat{F}_n(u) \text{ for a point } u \in J_i, \end{cases}$$

for $x \in J_i$. We note that a construction of this type was introduced under III on p. 213 of GESKUS and GROENEBOOM (1997). We here use the version in Lemma 11.10 of VAN DE GEER (2000), where instead of a piecewise constant version of $\xi_{h,t,\hat{F}_n}$ a piecewise constant version of $\xi_{h,t,F_0}$ is used.

We next define the piecewise constant function $\bar{\phi}_{h,t,\hat{F}_n}$ by

$$\bar{\phi}_{h,t,\hat{F}_n} = \hat{F}_n(x)\{1 - \hat{F}_n(x)\}\bar{\xi}_{h,t,\hat{F}_n}(x), \quad x \in \mathbb{R},$$

and the piecewise constant function $\bar{\theta}_{h,t,\hat{F}_n}$ by

$$\bar{\theta}_{h,t,\hat{F}_n}(x) = \begin{cases} \frac{\bar{\phi}_{h,t,\hat{F}_n}(x) - \bar{\phi}_{h,t,\hat{F}_n}(x-1)}{\hat{F}_n(x) - \hat{F}_n(x-1)}, & \text{if } \hat{F}_n(x) > \hat{F}_n(x-1), \\ 0, & \text{otherwise.} \end{cases}$$

Note that Lemma 2 and the assumption $t \not\equiv 0 \pmod 1$ and $t \not\equiv M \pmod 1$, we may assume that $\hat{F}_n(x) - \hat{F}_n(x-1)$ stays away from zero for all large $n$, if $\bar{\phi}_{h,t,\hat{F}_n}(x) \neq 0$ and/or $\bar{\phi}_{h,t,\hat{F}_n}(x-1) \neq 0$, implying that the function $\bar{\theta}_{h,t,\hat{F}_n}$ is bounded.

We have:

$$\begin{aligned} &\int \left\{\bar{\theta}_{h,t,\hat{F}_n} - \theta_{h,t,\hat{F}_n}\right\} dG_0 \\ &= \int \left\{\bar{\theta}_{h,t,\hat{F}_n}(z) - \theta_{h,t,\hat{F}_n}(z)\right\}\left\{F_0(z) - F_0(z-1) - \{\hat{F}_n(z) - \hat{F}_n(z-1)\}\right\} dz. \end{aligned} \tag{34}$$

This follows in the same way as (32).

We now show that this implies:

$$\int \theta_{h,t,\hat{F}_n} dG_0 = \int \bar{\theta}_{h,t,\hat{F}_n} dG_0 + o_p\left(n^{-2/3}h^{-2}\right). \tag{35}$$

Lemma 11.10 in VAN DE GEER (2000) implies that there exists a constant $c > 0$ such that

$$\left|\bar{\xi}_{h,t,\hat{F}_n}(x) - \xi_{h,t,F_0}(x)\right| \leqslant ch^{-3}\left|\hat{F}_n(x) - F_0(x)\right|, \text{ for all } x \in \mathbb{R},$$

where the factor $h^{-3}$ arises from the second derivative of $K_h$, which appears in $d\xi_{h,t,F_0}/dF_0$, using the notation in (11.61) of Lemma 11.10 in VAN DE GEER (2000). The function $\theta_{h,t,F_0}$ is defined in terms of the function $\phi_{h,t,F_0}$, with support consisting of intervals $[t - h + k, t + h + k]$ which, by the assumption $t \not\equiv 0 \pmod 1$ and $M - t \not\equiv 0 \pmod 1$, are strictly contained in the open interval $(0, M)$, if $h$ is sufficiently small.

Lemma 2 then implies that the closest point of jump of $\hat{F}_n$ to the left or right of these intervals (finite in number) have a distance $O_p(n^{-1/3})$ to the endpoints of these intervals. This means, by the condition $nh^3 \to \infty$ that

$$\left|\bar{\xi}_{h,t,\hat{F}_n}(x) - \xi_{h,t,F_0}(x)\right| \leqslant ch^{-3}\left|\hat{F}_n(x) - F_0(x)\right|,$$

for $x$ in a finite number of intervals of order $h$, shrinking to points $t + k$, for integers $k$ such that $t = k \in (0,M)$, and that

$$\bar{\xi}_{h,t,\hat{F}_n}(x) - \xi_{h,t,F_0}(x) = 0$$

outside these intervals. But this, in turn, implies by (34), the Cauchy–Schwarz inequality and the fact that we may assume that the differences $\hat{F}_n(x) - \hat{F}_n(x \pm 1)$ and $F_0(x) - F_0(x \pm 1)$ stay bounded away from zero on these intervals:

$$\int \left\{\bar{\theta}_{h,t,\hat{F}_n}(x) - \theta_{h,t,F_0}(x)\right\} dG_0(x) = O_p\left(h^{-3} d_{h,n}(\hat{F}_n, F_0)\right) = O_p\left(h^{-2}n^{-2/3}\right), \tag{36}$$

where $d_{h,n}(\hat{F}_n, F_0)^2$ is defined as

$$d_{h,n}(\hat{F}_n, F_0)^2 = \sum_{j \in \mathbb{Z}} \int_{(t-2h+j)\vee 0}^{(t+2h+j)\wedge M} \left\{\hat{F}_n(x) - F_0(x)\right\}^2 dx = O_p\left(hn^{-2/3}\right).$$

Note that the infinite sum is in fact a sum with a (uniformly) bounded number of non-zero terms. The property

$$\sum_{j \in \mathbb{Z}} \int_{(t-2h+j)\vee 0}^{(t+2h+j)\wedge M} \left\{\hat{F}_n(x) - F_0(x)\right\}^2 dx = O_p\left(hn^{-2/3}\right)$$

follows from the fact that, by the assumption $t \not\equiv 0 (\text{mod } 1)$ and $M - t \not\equiv 0 (\text{mod } 1)$, the intervals $[(t - 2h + j) \vee 0, (t + 2h + j) \wedge M]$ are strictly contained in $(0, M)$ and the fact that the processes

$$x \mapsto \hat{F}_n(x) - F_0(x), \in [(t - 2h + j) \vee 0, (t + 2h + j) \wedge M],$$

have a completely similar probabilistic behavior, together with the property

$$\|\hat{F}_n - F_0\|_2^2 = \mathcal{O}_p\left(n^{-2/3}\right).$$

On the other hand, by the structure of the solutions $\psi_{h,t,\hat{F}_n}$ and $\psi_{h,t,F_0}$, there also exists a constant $c_2 > 0$ such that

$$\left|\psi_{h,t,\hat{F}_n}(x) - \psi_{h,t,F_0}(x)\right| \leqslant c_2 h^{-2} \sum_{k \in \mathbb{Z}} \left|\hat{F}_n(x+k) - F_0(x+k)\right|,$$

for $x$ belonging to an interval $(t + k - h, t + k + h) \subset (0,M + 1)$ and that

$$\psi_{h,t,\hat{F}_n}(x) - \psi_{h,t,F_0}(x) = 0$$

outside these intervals. Note that for $x \in (t + k - h, t + k + h) \subset (M,M + 1)$, we have: $\psi_{h,t,F_0}(x) = 0$ and

$$\left|\frac{\phi_{h,t,\hat{F}_n}(x)}{\hat{F}_n(x)-\hat{F}_n(x-1)}\right| = \left\{\hat{F}_n(x+1)-\hat{F}_n(x)\right\}\left|\psi_{h,t,\hat{F}_n}(x)\right|$$

$$\left\{1-\hat{F}_n(x)\right\}\left|\psi_{h,t,\hat{F}_n}(x)\right| = \left\{F_0(x)-\hat{F}_n(x)\right\}\left|\psi_{h,t,\hat{F}_n}(x)\right| \leqslant ch^{-2}\left|F_0(x)-\hat{F}_n(x)\right|,$$

for a constant $c > 0$, and that we get in a similar way:

$$\left|\frac{\phi_{h,t,\hat{F}_n}(x)}{\hat{F}_n(x+1)-\hat{F}_n(x)}\right| \leqslant ch^{-2}\left|F_0(x)-\hat{F}_n(x)\right|,$$

for a constant $c > 0$, if $x \in (M, M+1)$.

Hence, again by (34) and the Cauchy–Schwarz inequality

$$\int\left\{\theta_{h,t,\hat{F}_n}(x)-\theta_{h,t,F_0}(x)\right\}dG_0(x) = O_p\left(h^{-2}\|\hat{F}_n-F_0\|_2^2\right) = O_p\left(h^{-2}n^{-2/3}\right). \quad (37)$$

Relation (35) now follows from (36) and (37).

We then apply, using the score equations for $\hat{F}_n$, implying $\int\bar{\theta}_{h,t,\hat{F}_n}d\mathbb{G}_n = 0$,

$$\int\bar{\theta}_{h,t,\hat{F}_n}dG_0 = \int\bar{\theta}_{h,t,\hat{F}_n}d(G_0-\mathbb{G}_n),$$

and

$$\int\left\{\bar{\theta}_{h,t,\hat{F}_n}-\theta_{h,t,F_0}\right\}d(G_0-\mathbb{G}_n) = \mathcal{O}_p\left(h^{-2}n^{-2/3}\right),$$

where the latter result again follows from an $L_2$-entropy argument. This shows, since $nh^3 \to \infty$,

$$\int K_h(t-x)d\left(\hat{F}_n-F_0\right)(x) = \int\theta_{h,t,F_0}d(\mathbb{G}_n-G_0)+o_p\left(n^{-1/2}h^{-3/2}\right),$$

and yields part (*i*) of Theorem 2.

Part (*ii*) of Theorem 2 now follows from

$$\begin{aligned}\int K_h(t-x)dF_0(x)-f_0(t) &= \int K(u)\{f_0(t+hu)-f_0(t)\}du\\ &\sim \frac{1}{2}h^2f_0''(t)\int u^2K(u)du \text{ as } h\downarrow 0,\end{aligned}$$

where we use that the kernel $K$ integrates to 1 and is symmetric around zero. Combining (*i*) and (*ii*) using that $h_n \sim c{\cdot}n^{-1/7}$, we get (*iii*). This finishes the proof of Theorem 2. □

## 6 Concluding remarks

In the preceding sections we discussed two types of nonparametric estimators for the density $f$ of the distribution function $F$ in a situation where we only have indirect information about $F$ via a sample of random variables, generated by the density

$$g_F(z) = \int_{\mathbb{R}} k(z - x)dF(x),$$

where $k$ is a uniform density. This estimation problem has been called "boxcar deconvolution".

The first estimator we discussed was the rather straightforward inverse density estimator, suggested by (2). The second estimator of $f$ was based on the maximum likelihood estimator (MLE) $\hat{F}_n$ of $F$, and had the representation

$$\hat{f}_n(x) = \int K_h(x - y)d\hat{F}_n(y), \tag{38}$$

where $K_h$ is a kernel smoother with bandwidth $h$.

It was shown that (under some conditions) both estimators of the density $f$ converge pointwise at rate $n^{-2/7}$ in the interior of the support, and also that they converge (after standardization) to a normal limit distribution. Moreover, it was shown that the bias functions have a similar behavior, but that the asymptotic variances are different. It is indicated that the estimator based on the MLE has the smaller asymptotic variance, although this is generally not so easy to verify because of the implicit nature of the asymptotic variance of the estimator, based on the MLE. The estimator (38), based on the MLE $\hat{F}_n$ of $F$, has the advantage of being a genuine density estimator, whereas the inverse density estimator can in principle also take negative values, a property it shares with estimators, based on Fourier inversion methods.

The proof of the asymptotic normality of estimator (38) is based on so-called smooth functional theory and shows that this theory can also be applied locally, on shrinking neighborhoods of a fixed point.

It may be of interest to note that the MLE $\hat{F}_n$ itself has convergence rate $n^{-1/3}$ and that $n^{1/3}\{\hat{F}_n(t) - F(t)\}$ has a non-normal limit distribution for points $t$ in the interior of the support of $f$. So $\hat{F}_n$ has a limit behavior that falls under the heading "non-standard asymptotics", whereas the convolution with the kernel $K_h$ in (38) produces an estimator which exhibits "standard asymptotic behavior", in the sense of convergence to a normal distribution.

## References

BILLINGSLEY, P. (1968), Convergence of probability measures, Wiley, New York.

GESKUS, R. B. and P. GROENEBOOM (1997), Asymptotically optimal estimation of smooth functionals for interval censoring, part 2, *Statistica Neerlandica* **51,** 201–219.

GROENEBOOM, P. (1996), Lectures on inverse problems, in: P. BERNARD (ed.), *Lectures on probability theory, Ecole d'Eté de Probabilités de Saint-Flour XXIV-1994*, Springer Verlag, Berlin.

GROENEBOOM, P., G. JONGBLOED and J.A. WELLNER (2002), The support reduction algorithm for computing nonparametric function estimates, Technical Report. http://ssor.twi.tudelft.nl/~pietg

GROENEBOOM, P. and J.A. WELLNER (1992), *Information bounds and nonparametric maximum likelihood estimation*, Birkhäuser, Basel. Springer Verlag. New York.

HALL, P., F. RUYMGAART, O. VAN GAANS and A. VAN ROOY (2001), Inverting noisy integral equations using wavelet expansions: a class of irregular convolutions, in: State of the art in Probability and Statistics: Festschrift for Willem R. van Zwet, Vol. 30 of *Lecture notes-Monograph series*, Institute of Mathematical Statistics, pp. 251–280.

JOHNSTONE, I.M. and M. RAIMONDO (2002), Periodic boxcar deconvolution and diophantine approximation, Technical Report. http://www-stat.stanford.edu/~imj/Reports/index.html

O'SULLIVAN, F. and K. ROY CHOUDHURY (2001), An analysis of the role of positivity and mixture model constraints in Poisson deconvolution problems, *Journal of Computational and Graphical Statistics* **10,** 673–696.

ROY CHOUDHURY, K. (1998), Additive mixture models for multichannel image data, Ph.D. Dissertation, University of Washington, Seattle.

SIMAR, L. (1976), Maximum likelihood estimation of a compound Poisson process, *Annals of Statistics* **4,** 1200–1209.

VAN DE GEER, S. (1996), Rates of convergence for the maximum likelihood estimator in mixture models, *Journal of Nonparametric Statistics* **6,** 293–310.

VAN DE GEER, S. (2000), *Applications of empirical process theory*, Cambridge University Press. Cambridge, U.K., 293–310.

WAND, M.P. and M.C. JONES (1995), *Kernel smoothing*, Chapman & Hall, London.

Received: June 2002. Revised: August 2002.