

Estimating the upper support point in deconvolution

LUCIE AARTS

PIET GROENEBOOM

GEURT JONGBLOED

Stanislas College Westplantsoen Delft University of Technology

Vrije Universiteit

Vrije Universiteit

Abstract

We consider estimation of the upper boundary point $F^{-1}(1)$ of a distribution function F with finite upper boundary or “frontier” in deconvolution problems, primarily focusing on deconvolution models where the noise density is decreasing on the positive halfline. Our estimates are based on the (nonparametric) maximum likelihood estimator \hat{F}_n (MLE) of F . We show that $\hat{F}_n^{-1}(1)$ is asymptotically never too small. If the convolution kernel has bounded support the estimator $\hat{F}_n^{-1}(1)$ can generally be expected to be consistent. In this case we establish a relation between the extreme value index of F and the rate of convergence of $\hat{F}_n^{-1}(1)$ to the upper support point for the “boxcar” deconvolution model. If the convolution density has unbounded support, $\hat{F}_n^{-1}(1)$ can be expected to overestimate the upper support point. We define consistent estimators $\hat{F}_n^{-1}(1 - \beta_n)$, for appropriately chosen vanishing sequences (β_n) and study these in a particular case.

Keywords: (de)convolution, extreme value index, kernel estimation, non parametric maximum likelihood estimators, upper support point

Running heading: support estimation in deconvolution models

1 Introduction

Let X_1, X_2, \dots be a sequence of independent identically distributed random variables, generated by a distribution function F with finite upper support point

$$\vartheta_0 = F^{-1}(1) = \inf\{x \in \mathbb{R} : F(x) = 1\}.$$

We consider the problem of estimating ϑ_0 based on a sample Z_1, Z_2, \dots where $Z_i = X_i + Y_i$ and Y_1, Y_2, \dots are independent identically distributed random variables with known density function k , independent of the X_i 's. Stated otherwise, we estimate ϑ_0 , based on a sample, generated by the density

$$g(z) = \int k(z - x)dF(x). \quad (1)$$

The problem has received a lot of attention the past few years. As stated in Hall & Simar (2002) a range of problems in economics and statistics involve calculations of the boundary or frontier of a distribution. Meister (2006) suggests a procedure for estimating the support by constructing a sequence of estimators of moments of the convolution distribution. Moments of the unknown distribution are estimated by solving sequences of convolution equations. Since the following type of property is used:

$$(EX^j)^{1/j} \uparrow R, j \rightarrow \infty, \text{ if } X \text{ has support } [0, R], \quad (2)$$

and since we also need

$$E(X + Y)^j = \sum_{l=0}^j \binom{j}{l} EX^l EY^{j-l}, \quad (3)$$

(the latter relations represent the convolution equations for the moments EX^j which have to be solved, after estimating the moments $E(X + Y)^j$), one also has to decide how many moments $E(X + Y)^j$ have to be estimated and, correspondingly, how many equations in l variables (3) (with moments replaced by estimates of moments) have to be solved. The whole procedure seems rather computer intensive.

Goldenshluger & Tsybakov (2004) study estimators for the case where the convolution density is Gaussian and also for the case where the convolution density has bounded support from a minimax point of view. We will briefly come back to the relation of our results with their work in Section 4.

Hall & Simar (2002) study estimators for the situation that the convolution density has a variance, tending to zero with the sample size, which is a situation rather different from the situation we consider, where the convolution density is kept fixed.

Delaigle & Gijbels (2006) study estimators based on kernel density estimation. In their sim-

ulation study they consider Laplace and Gaussian error. We apply their methodology to the exponential deconvolution problem and compare their results with ours in Section 6.

We motivate the choice of considering the nonparametric maximum likelihood estimator (MLE) when estimating the upper support point in deconvolution models in Section 2, and also discuss algorithmic aspects of the MLE in general deconvolution problems in this section. Since the MLE of a distribution function is a natural estimator in a whole class of deconvolution models, the MLE is also a natural candidate to use as plug-in estimator for the upper support point of a distribution.

In Section 3 we consider general locally consistent estimators \hat{F}_n of F . We prove that the associated plug-in estimator

$$\hat{\vartheta}_n = \hat{F}_n^{-1}(1) = \inf\{x \in \mathbb{R} : \hat{F}_n(x) = 1\}$$

of ϑ_0 is asymptotically never too small. This is proved without putting any additional conditions on the distribution of X or Y and may perhaps come a bit as a surprise in view of the fact that $\mathbb{F}_n^{-1}(1)$ is always a (finite sample) negatively biased estimate of the support point θ_0 in non-contaminated estimation, where we can consider the empirical distribution function \mathbb{F}_n as the nonparametric MLE. Furthermore we derive consistency and rate results for the estimator $\hat{F}_n^{-1}(1 - \beta_n)$ of ϑ_0 for appropriately chosen vanishing sequences $\{\beta_n\}$.

In Sections 4 and 5 we specialize our results to two prototypes of decreasing (i.e., non-increasing) kernels. As a prototype of a kernel with compact support we study the uniform kernel, and as a prototype of a kernel with unbounded support we study the exponential kernel. The convolution model with the uniform kernel is in fact the periodic boxcar deconvolution model described in Johnstone & Raimondo (2004). For this model we establish a relation between the rate of convergence of our estimator and the extreme value index of F . For the exponential kernel we show in Section 5, not surprisingly, that $\hat{F}_n^{-1}(1)$ does not estimate ϑ_0 consistently. However, we introduce a consistent modification of this estimator. The approach is likelihood-based and can be adopted in more general deconvolution models.

As mentioned above, the relation with Delaigle & Gijbels (2006) is discussed in Section 6.

2 Deconvolution model and algorithmic aspects of the MLE

Many nonparametric methods have been studied to estimate the distribution function F based on a sample from density g in (1). If F is assumed to have a density function f , kernel methods are often used to estimate f . See e.g. Carroll & Hall (1988) and Stefanski & Carroll (1990). Without putting more restrictions on the density f , straightforward maximum likelihood estimation of f as a proper density is not feasible. Penalizing or smoothing the likelihood can yield consistent estimators for f . See e.g. Eggermont & LaRiccia (1997).

However, maximizing the log likelihood function directly in terms of the distribution function F , where the log likelihood function is given by

$$F \mapsto \sum_{i=1}^n \log g(z_i) = \sum_{i=1}^n \log \int k(z_i - x) dF(x)$$

does lead to a well defined estimator under fairly general conditions. The log likelihood function attains its maximum over all distribution functions at a discrete distribution function having at most n points of jump, see e.g. Lindsay (1995) Section 1.5.

Groeneboom & Wellner (1992) show that if k is decreasing on $[0, \infty)$, the nonparametric maximum likelihood estimator (MLE) \hat{F}_n has all its mass concentrated on the set of observed Z_i values. Although it is only rarely possible to construct the MLE of F explicitly without using an iterative algorithm, several algorithms are available for computing the MLE. For example, the EM algorithm and variations of it, the iterative convex minorant algorithm (Groeneboom & Wellner, 1992, and Jongbloed, 1998) and various types of vertex direction methods (Böhning, 1995).

So computing the MLE is not really a problem. In the case of the uniform and exponential deconvolution problem there even exists a one step convex minorant algorithm for computing the MLE. These problems will serve as examples in subsequent sections. Figure 1 shows the MLE based on a sample of size 100 generated from the convolution of the uniform distribution on $(0, 1)$ and the density $k(x) = \sqrt{\frac{2}{\pi}} e^{-\frac{1}{2}x^2}$ with $x \geq 0$. The observed data are tick marked on the x -axis. To obtain Figure 1, we used a particular vertex direction method, the support reduction algorithm described in Groeneboom *et al.* (2003). Clearly, the MLE is a step function.

HERE FIGURE 1

It is often felt that the discreteness of the estimator of the distribution function, provided by the MLE is unpleasant, which has motivated people to do EM with early stopping, etc. In this connection it might be useful to note that the MLE can in fact be used for density estimation, without using any early stopping of EM or a similar device. The point is that, if one wants to assume certain smoothness properties of the underlying distribution, one can smooth the MLE to achieve faster rates of convergence or to construct estimates of the density. To estimate the density f , one simply can use a kernel density estimate of the form

$$\hat{f}_{n,h}(x) = \int K_h(x-y) d\hat{F}_n(y), \quad K_h(u) = h^{-1}K(u/h),$$

where \hat{F}_n is the MLE and K is one of the usual kernels for density estimation, e.g., the triweight kernel

$$K(u) = \frac{35}{32} (1-u^2)^3.$$

or the triweight kernel, transformed to a boundary kernel. This method was used in Groeneboom & Jongbloed (2003), where it was shown that, under certain conditions on F ,

$$n^{2/7} \left\{ \hat{f}_{n,h}(x) - f(x) \right\} \xrightarrow{\mathcal{D}} N(\mu_F, \sigma_F^2),$$

where the constants μ_F and σ_F depend on F and a bandwidth of order $cn^{-1/7}$ is chosen, see Theorem 5.1 of Groeneboom & Jongbloed (2003). This is the typical rate of convergence one can expect in this situation for estimates of the density (note that the deconvolution step causes a slower rate of convergence than we would get in direct density estimation), and indeed also other types of estimators, for example based on Fourier inversion, can be found having the same rate of convergence.

However, from a methodological point of view, smoothing the distribution did not seem to us the most natural way to go, if one wants to estimate the upper support point of a distribution. This has motivated us to study estimators, directly based on the MLE, without performing any

additional smoothing.

3 Consistency

In this section we derive three results for general F and general noise density k . We consider the behavior of $\hat{F}_n^{-1}(1)$ and $\hat{F}_n^{-1}(1 - \beta_n)$. In the latter case, we have a consistency and rate result under certain conditions on the vanishing sequence β_n . The following theorem shows that, if \hat{F}_n is a locally consistent estimator of F in a neighborhood of the upper support point, the associated plug-in estimator $\hat{F}_n^{-1}(1)$ of ϑ_0 is asymptotically never too small.

Theorem 1 *Let \hat{F}_n be a sequence of distribution functions and F a distribution function with $F^{-1}(1) = \vartheta_0 < \infty$. Moreover, suppose that for some $\varepsilon_0 > 0$ the following holds:*

$$|\hat{F}_n(x) - F(x)| \rightarrow 0 \quad \text{in probability} \quad \text{for all } x \in [\vartheta_0 - \varepsilon_0, \vartheta_0].$$

Let $\hat{\vartheta}_n = \hat{F}_n^{-1}(1)$. Then for all $\varepsilon \in (0, \varepsilon_0)$, $\lim_{n \rightarrow \infty} P\{\hat{\vartheta}_n \leq \vartheta_0 - \varepsilon\} = 0$.

PROOF: Fix $\varepsilon \in (0, \varepsilon_0)$. Then

$$\begin{aligned} P\{\hat{\vartheta}_n \leq \vartheta_0 - \varepsilon\} &= P\{\hat{F}_n(\vartheta_0 - \varepsilon) = 1\} \\ &= P\{\hat{F}_n(\vartheta_0 - \varepsilon) - F(\vartheta_0 - \varepsilon) = 1 - F(\vartheta_0 - \varepsilon)\} \\ &\leq P\{|\hat{F}_n(\vartheta_0 - \varepsilon) - F(\vartheta_0 - \varepsilon)| \geq 1 - F(\vartheta_0 - \varepsilon)\} \rightarrow 0 \end{aligned}$$

The last statement is true since $F(\vartheta_0 - \varepsilon) < 1$ and $\hat{F}_n(\vartheta_0 - \varepsilon) \rightarrow F(\vartheta_0 - \varepsilon)$ in probability. \square

Remark 1 Maximum likelihood estimators and isotonic inverse estimators (see Van Es *et al.*, 1998) are estimators that satisfy the condition given in Theorem 1.

We now show that under the additional assumption that \hat{F}_n is α_n -consistent at each $x > \vartheta_0$, $\hat{F}_n^{-1}(1 - \beta_n)$ converges in probability to θ_0 whenever $\{\beta_n\}$ tends to zero more slowly than $\{\alpha_n\}$.

Theorem 2 Suppose $\hat{F}_n(x) \rightarrow F(x)$ in probability $\forall x \in [\vartheta_0 - \varepsilon_0, \vartheta_0]$ and $\alpha_n^{-1}(\hat{F}_n(x) - 1)$ is tight for every $x > \vartheta_0$ with $0 \leq \alpha_n \rightarrow 0$. Then

$$\hat{F}_n^{-1}(1 - \beta_n) \rightarrow \vartheta_0 \quad \text{in probability}$$

whenever $0 \leq \beta_n \rightarrow 0$ and $\frac{\beta_n}{\alpha_n} \rightarrow \infty$.

PROOF: We first show that for every $\epsilon > 0$

$$P(\hat{F}_n^{-1}(1 - \beta_n) > \vartheta_0 + \varepsilon) \rightarrow 0$$

as $n \rightarrow \infty$. Choose $\varepsilon > 0$ and $\delta > 0$ and choose $K > 0$ such that $P(\alpha_n^{-1}|\hat{F}_n(\vartheta_0 + \varepsilon) - 1| > K) < \delta$ for all n . Then for n sufficiently large such that $\frac{\beta_n}{\alpha_n} > K$ it follows that

$$\begin{aligned} P(\hat{F}_n^{-1}(1 - \beta_n) > \vartheta_0 + \varepsilon) &= P(\hat{F}_n(\vartheta_0 + \varepsilon) < 1 - \beta_n) = P(1 - \hat{F}_n(\vartheta_0 + \varepsilon) > \beta_n) \\ &= P\left(\alpha_n^{-1}(1 - \hat{F}_n(\vartheta_0 + \varepsilon)) > \frac{\beta_n}{\alpha_n}\right) \\ &\leq P\left(\alpha_n^{-1}(1 - \hat{F}_n(\vartheta_0 + \varepsilon)) > K\right) < \delta. \end{aligned}$$

Now note that

$$\begin{aligned} P\left\{\hat{F}_n^{-1}(1 - \beta_n) < \vartheta_0 - \varepsilon\right\} &\leq P\left\{\hat{F}_n(\vartheta_0 - \varepsilon) \geq 1 - \beta_n\right\} = \\ &P\left\{\hat{F}_n(\vartheta_0 - \varepsilon) - F(\vartheta_0 - \varepsilon) \geq 1 - F(\vartheta_0 - \varepsilon) - \beta_n\right\} \rightarrow 0 \end{aligned}$$

since $F(\vartheta_0 - \varepsilon) < 1$, $\beta_n \rightarrow 0$ and $\hat{F}_n(\vartheta_0 - \varepsilon) \rightarrow F(\vartheta_0 - \varepsilon)$ in probability. \square

Strengthening the pointwise rate assumption for \hat{F}_n to a local uniform rate assumption in a left neighborhood, we can derive the following rate result for $\hat{F}_n^{-1}(1 - \beta_n)$.

Theorem 3 Suppose

$$\sup_{x \in [\vartheta_0 - \varepsilon, \vartheta_0]} |\hat{F}_n(x) - F(x)| = O_p(\alpha_n)$$

with $0 \leq \alpha_n \rightarrow 0$. Furthermore assume there exists $\eta > 0$ and $c > 0$ such that

$$1 - F(x) \sim c(\vartheta_0 - x)^\eta \quad \text{for } x \uparrow \vartheta_0.$$

Then for all β_n and γ_n such that $\gamma_n \rightarrow 0$, $\frac{\beta_n}{\alpha_n} \rightarrow \infty$ and $\gamma_n \geq \beta_n^{\frac{1}{\eta}}$ it holds that

$$\hat{F}_n^{-1}(1 - \beta_n) \leq \vartheta_0$$

with probability tending to one and

$$\gamma_n^{-1} \left(\hat{F}_n^{-1}(1 - \beta_n) - \vartheta_0 \right) = O_p(1).$$

PROOF: First we show that

$$P \left(\hat{F}_n^{-1}(1 - \beta_n) > \vartheta_0 \right) \rightarrow 0$$

as $n \rightarrow \infty$. This holds since

$$\begin{aligned} P \left(\hat{F}_n^{-1}(1 - \beta_n) > \vartheta_0 \right) &= P \left(1 - \hat{F}_n(\vartheta_0) > \beta_n \right) \\ &\leq P \left(\alpha_n^{-1} |\hat{F}_n(\vartheta_0) - F(\vartheta_0)| > \frac{\beta_n}{\alpha_n} \right) \rightarrow 0 \end{aligned}$$

as $n \rightarrow \infty$.

On the other hand we have for $K > \left(\frac{2}{c}\right)^{\frac{1}{\eta}}$

$$\begin{aligned} &P \left(\hat{F}_n^{-1}(1 - \beta_n) < \vartheta_0 - K\gamma_n \right) \leq P \left(\hat{F}_n(\vartheta_0 - K\gamma_n) \geq 1 - \beta_n \right) \\ &= P \left(\hat{F}_n(\vartheta_0 - K\gamma_n) - F(\vartheta_0 - K\gamma_n) \geq 1 - F(\vartheta_0 - K\gamma_n) - \beta_n \right) \\ &\leq P \left(\sup_{x \in [\vartheta_0 - \varepsilon, \vartheta_0]} \alpha_n^{-1} |\hat{F}_n(x) - F(x)| \geq \alpha_n^{-1} (1 - F(\vartheta_0 - K\gamma_n) - \beta_n) \right) \\ &\leq P \left(\sup_{x \in [\vartheta_0 - \varepsilon, \vartheta_0]} \alpha_n^{-1} |\hat{F}_n(x) - F(x)| \geq \left\{ \frac{\beta_n}{\alpha_n} \left[\frac{c}{2} K^\eta - 1 \right] \right\} \right) \rightarrow 0. \end{aligned}$$

□

The theorem above implies that for e.g. $\eta = 1$, the case considered in Delaigle & Gijbels (2006), any rate slower than α_n can be achieved.

4 Noise density with compact support and extreme value index

In this section we consider a particular estimator, the MLE \hat{F}_n , in case that the noise density has bounded support. Without loss of generality (because of rescaling), we restrict ourselves to the situation that this support is equal to $[0, 1]$. Then a natural estimator is $Z_{(n)} - 1$, where $Z_{(n)}$ is the largest order statistic of the observations Z_i . We first derive a general result that $\hat{F}_n^{-1}(1) \geq Z_{(n)} - 1 - O_p(n^{-1} \log n)$. Then we show for the uniform deconvolution context, where $k(y) = 1_{[0,1]}(y)$, that also $\hat{F}_n^{-1}(1) \leq Z_{(n)} - 1 + O_p(n^{-1} \log n)$. Using this asymptotic equivalence of $Z_{(n)} - 1$ and $\hat{F}_n^{-1}(1)$, we show that the asymptotic behavior of $\hat{F}_n^{-1}(1)$ and its rate of convergence in the uniform deconvolution model depend on F via its extreme value index.

Lemma 1 *For the deconvolution problem where the noise has compact support $[0, 1]$ and general (nondegenerate) distribution functions F we have:*

$$\hat{F}_n^{-1}(1) \geq Z_{(n)} - 1 - O_p\left(\frac{\log n}{n}\right).$$

where the O_p term is nonnegative.

PROOF: Note that with probability tending to one (for $n \rightarrow \infty$),

$$\max \{Z_j : Z_j < Z_{(n)} - 1\} = Z_{(n^*)}$$

exists. We show that then $\hat{F}_n^{-1}(1) > Z_{(n^*)}$. Indeed, suppose that $\hat{F}_n^{-1}(1) \leq Z_{(n^*)}$. Then $F(Z_{(n^*)}) = 1$ and

$$\hat{g}_n(Z_{(n)}) = \int_{Z_{(n)}-1}^{Z_{(n)}} k(Z_{(n)} - x) d\hat{F}_n(x) = 0,$$

since k has support $[0, 1]$. This, however, implies that the log likelihood is minus infinity, a contradiction. We conclude by using Lemma 4

$$0 \leq Z_{(n)} - 1 - Z_{(n^*)} \leq Z_{(n^*+1)} - Z_{(n^*)} \leq \max(\text{spacing near } \vartheta_0) = O_p\left(\frac{\log n}{n}\right).$$

□

If k is the uniform $(0, 1)$ density we can write

$$g(z) = \int_{z-1}^z 1dF(x) = F(z) - F(z-1),$$

and we get the following result.

Lemma 2 *For uniform $(0, 1)$ noise distribution and general distribution functions F ,*

$$\hat{F}_n^{-1}(1) \leq Z_{(n)} - 1 + O_p\left(\frac{\log n}{n}\right),$$

where the O_p term is nonnegative.

PROOF: Define

$$Z_{(i^*)} = \min\{Z_i : Z_i \geq Z_{(n)} - 1\}.$$

We first show that $\hat{F}_n(Z_{(i^*)}) = 1$. Suppose that $\hat{F}_n(Z_{(i^*)}) < 1$ and define

$$\bar{F}_n(x) = \begin{cases} \hat{F}_n(x) & \text{if } x < Z_{(i^*)} \\ 1 & \text{if } x \geq Z_{(i^*)}. \end{cases}$$

Now, defining $\bar{g}(z) = \int k(z-x) d\bar{F}_n(x)$, we have

$$\bar{g}(z) - \hat{g}_n(z) = \bar{F}(z) - \hat{F}_n(z) - \bar{F}(z-1) + \hat{F}_n(z-1)$$

and thus

$$\bar{g}(z) - \hat{g}_n(z) = \begin{cases} 0 & \text{if } z < Z_{(i^*)} \\ 1 - \hat{F}_n(z) & \text{if } Z_{(i^*)} \leq z \leq \vartheta_0 + 1 \end{cases}$$

This implies that the likelihood function at \bar{F} is strictly larger than the likelihood function at \hat{F}_n , contradicting that \hat{F}_n is the MLE. Hence, $\hat{F}_n(Z_{(i^*)}) = 1$ and by Lemma 4

$$\hat{F}_n^{-1}(1) \leq Z_{(i^*)} \leq Z_{(n)} - 1 + \max(\text{spacing near } \vartheta_0) = Z_{(n)} - 1 + O_p\left(\frac{\log n}{n}\right).$$

□

In the following lemma we assume that $1 - F(x) \sim c(\vartheta_0 - x)^{\frac{-1}{\gamma}}$ with $\gamma < 0$ and $c > 0$. This can be rewritten as

$$\frac{1 - F(\vartheta_0 - xh)}{1 - F(\vartheta_0 - h)} = \frac{c(xh)^{\frac{-1}{\gamma}}}{c(h)^{\frac{-1}{\gamma}}} = x^{\frac{-1}{\gamma}}.$$

By e.g. Gnedenko (1943), we see that γ is the extreme value index of F . For convenience we replace $\frac{-1}{\gamma}$ by η in the following lemma. If F is the uniform $(0, 1)$ distribution, we see that $\eta = 1$.

Theorem 4 *Suppose that for some $\eta > 0$*

$$1 - F(x) \sim c(\vartheta_0 - x)^\eta$$

when $x \uparrow \vartheta_0$. I.e.

$$\frac{1 - F(x)}{c(\vartheta_0 - x)^\eta} \rightarrow 1. \tag{4}$$

Then, for the uniform noise distribution

$$\left(\frac{cn}{\eta + 1}\right)^{\frac{1}{\eta+1}} \left(\vartheta_0 - \hat{F}_n^{-1}(1)\right) \rightarrow W_\eta \tag{5}$$

in distribution where $W_\eta \sim \text{Weibull}(\eta + 1)$.

PROOF: By Lemma 1 and Lemma 2 we have that $Z_{(n)} - 1 - \hat{F}_n^{-1}(1) = O_p\left(\frac{\log n}{n}\right)$. Furthermore,

$$\frac{1}{\eta+1} = \frac{\eta+1-\eta}{\eta+1} = 1 - \frac{\eta}{\eta+1}, \text{ so } n^{\frac{1}{\eta+1}} \frac{\log n}{n} = n^{-\frac{\eta}{\eta+1}} \log n \rightarrow 0$$

for $n \rightarrow \infty$. By Slutsky it now suffices to prove (5) with $\hat{F}_n^{-1}(1)$ replaced by $Z_{(n)} - 1$. Let $w > 0$ and write $w_n = w \left(\frac{\eta+1}{cn}\right)^{\frac{1}{\eta+1}}$

Then

$$\begin{aligned} P \left\{ \vartheta_0 - Z_{(n)} + 1 \leq w \left(\frac{\eta+1}{cn}\right)^{\frac{1}{\eta+1}} \right\} &= P\{\vartheta_0 - Z_{(n)} + 1 \leq w_n\} = \\ P\{Z_{(n)} \geq \vartheta_0 + 1 - w_n\} &= 1 - P\{Z_{(n)} < \vartheta_0 + 1 - w_n\} = \\ 1 - (P\{Z_i < \vartheta_0 + 1 - w_n\})^n &= 1 - (1 - P\{Z_i \geq \vartheta_0 + 1 - w_n\})^n. \end{aligned}$$

If $1 + \vartheta_0 - w_n \geq 1$ (which holds for n sufficiently large), then

$$\begin{aligned} P\{Z_i \geq \vartheta_0 + 1 - w_n\} &= \int_{\vartheta_0+1-w_n}^{\vartheta_0+1} g(z) dz \\ &= \int_{\vartheta_0+1-w_n}^{\vartheta_0+1} (1 - F(z-1)) dz = \int_{\vartheta_0-w_n}^{\vartheta_0} (1 - F(z)) dz. \end{aligned}$$

By (4), we see that for each $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$(1 - \varepsilon)c(\vartheta_0 - x)^\eta \leq 1 - F(x) \leq (1 + \varepsilon)c(\vartheta_0 - x)^\eta \quad \forall x \in [\vartheta_0 - \delta, \vartheta_0].$$

We therefore have that

$$\begin{aligned} \int_{\vartheta_0-w_n}^{\vartheta_0} (1 - F(z)) dz &\sim \frac{-c}{\eta+1} (\vartheta_0 - x)^{\eta+1} \Big|_{\vartheta_0-w_n}^{\vartheta_0} \\ &= \frac{c}{\eta+1} w_n^{\eta+1} = \frac{c}{\eta+1} \cdot \frac{\eta+1}{cn} w^{\eta+1} = \frac{w^{\eta+1}}{n}. \end{aligned}$$

Then, for $n \rightarrow \infty$

$$P \left\{ \left(\frac{cn}{\eta+1}\right)^{\frac{1}{\eta+1}} (\vartheta_0 - Z_{(n)} + 1) \leq w \right\} \sim 1 - \left(1 - \frac{w^{\eta+1}}{n}\right)^n \rightarrow 1 - e^{-w^{\eta+1}}.$$

□

Remark 2 Goldenshluger & Tsybakov (2004) show that the simple estimator $Z_{(n)} - r$ attains an optimal minimax convergence rate $n^{1/(\alpha+\beta+2)}$ if the support of k is $[-r, r]$, for some $r > 0$, and if

$$f(x) \geq c_1(\theta_0 - x)^\alpha, k(y) \geq c_2(r - y)^\beta, \theta_0 - h < x < \theta_0, r - h < y < r,$$

for some $\alpha, \beta > 0$ and $h > 0$, see their Theorem 3 on page 44. It is clear that in our particular case $\hat{F}_n^{-1}(1)$ will also attain the required rate, by the asymptotic equivalence of $\hat{F}_n^{-1}(1)$ and $Z_{(n)} - 1$. In fact, we could also have deduced Theorem 4 from their Theorem 3, using this asymptotic equivalence. An intriguing question is whether this asymptotic equivalence holds more generally.

5 Noise density with unbounded support

In this section we take a closer look at situations where the noise density has unbounded support. One particular example of this is the exponential deconvolution problem, i.e. $k(y) = e^{-y}1_{[0,\infty)}(y)$.

By studying this specific problem, we show how other situations where the noise density has infinite support could be handled. A consistent estimator of F in the exponential deconvolution problem is the MLE \hat{F}_n described in Jongbloed (1998). This estimator can be explicitly constructed.

We show that the MLE $\hat{F}_n^{-1}(1)$ is not a consistent estimator of ϑ_0 . Then we consider the general ‘high quantile’ estimator $\hat{F}_n^{-1}(1 - \beta_n)$ as introduced in Section 3.

To prove inconsistency of $\hat{F}_n^{-1}(1)$ as an estimator of the upper support point in case of the exponential deconvolution model, we exploit the relation between the exponential deconvolution problem and the uniform random fraction problem by considering $W_i = e^{-Z_i}$. The largest Z values in the exponential deconvolution problem then correspond to the smallest W values in the uniform random fraction problem.

Then we apply known asymptotic results for the Grenander (maximum likelihood) estimator of a decreasing density, to obtain the inconsistency of $\hat{F}_n^{-1}(1)$ as an estimator of ϑ_0 . The Grenander estimator $\hat{g}_{n,W}(w)$ of $g_W(w)$, is the left derivative of the concave majorant $\hat{G}_{n,W}$ of the empirical

distribution function of the W_i 's evaluated at w . For the Grenander estimator see Grenander (1956) and Groeneboom & Lopuhaä (1993).

First we rewrite the problem and derive several results we use to prove inconsistency of $\hat{F}_n^{-1}(1)$.

Lemma 3 Define $W = e^{-Z}$, $V = e^{-X}$ and $U = e^{-Y}$. Then

$$Z = X + Y \Rightarrow W = VU,$$

where U has the uniform $(0, 1)$ density. Furthermore

1. $g_W(e^{-z})e^{-z} = g_Z(z)$, $z \in (0, \infty)$.
2. $g_W(w)$ is constant for $w < e^{-\vartheta_0}$.
3. $1 - F(x) = G_W(e^{-x}) - e^{-x}g_W(e^{-x})$.
4. $\vartheta_0 = F^{-1}(1) = -\log(\sup\{w \geq 0 : g_W(w) = g_W(0+)\})$.
5. $\hat{\vartheta}_n = \hat{F}_n^{-1}(1) = -\log(\sup\{w \geq 0 : \hat{g}_{n,W}(w) = \hat{g}_{n,W}(0+)\})$, where $\hat{g}_{n,W}$ is the Grenander estimator.

PROOF: By definition we have that $P(W \leq w) = 1 - G_Z(-\log(w))$, where $w \in (0, 1)$. Taking the derivative of both sides and replacing w by e^{-z} , we obtain 1.

By rewriting $g_Z(z)$ and using that $-\log w > \vartheta_0$ implies $w < e^{-\vartheta_0}$, we obtain that $g_W(w) = \int_0^{\vartheta_0} e^u dF(u)$ for $w < e^{-\vartheta_0}$ and thus 2.

Now 3. follows by Fubini, see e.g. Jongbloed (1998). We derive that

$$1 - F(x) = 1 - G_Z(x) - g_Z(x) = G_W(e^{-x}) - e^{-x}g_W(e^{-x}).$$

Furthermore we define

$$\tilde{F}(v) = \frac{\int_0^v \frac{1}{y} dF_V(y)}{\int_0^1 \frac{1}{y} dF_V(y)}, \text{ so that } \frac{dF_V}{d\tilde{F}}(v) = cv,$$

where $c = \int_0^1 \frac{1}{y} dF_V(y) < \infty$ since $F_V(e^{-\vartheta_0}) = 0$. Using integration by parts we get

$$P(W \leq w) = c \int_0^w (1 - \tilde{F}(v)) dv, \text{ so that } g_W(w) = c (1 - \tilde{F}(w)). \quad (6)$$

Furthermore $F(x) = 1 - F_V((e^{-x})^-)$. By (6) we see that $g_W(w)$ is decreasing as function of w . Note that if $\vartheta_0 \in [0, \infty)$ all bounded non-increasing densities can occur. Furthermore the support of \tilde{F} is the same as the support of F_V . By using (6) we can rewrite ϑ_0 as follows and obtain 4.:

$$\begin{aligned} \vartheta_0 &= F^{-1}(1) = \inf\{x \geq 0 : F(x) = 1\} = -\log(\sup\{v \geq 0 : F_V(v) = 0\}) \\ &= -\log(\sup\{v \geq 0 : 1 - F_V(v) = 1\}) = -\log(\sup\{w \geq 0 : g_W(w) = g_W(0+)\}). \end{aligned}$$

For 5. we use that there is a one-to-one relation between g_W and F . Given a realization z_1, z_2, \dots the log likelihood as a function of g (or, equivalently of F) is given by

$$l(F) = \frac{1}{n} \sum_{i=1}^n \log(g_W(w_i)) - \bar{z}_n.$$

The log likelihood $l(F)$ is maximized if we take the Grenander estimator $\hat{g}_{n,W}$ of $g_W(w)$. For a proof of this fact see Grenander (1956) or Groeneboom & Lopuhaä (1993). \square

We now prove that $\hat{F}_n^{-1}(1)$ is an inconsistent estimator of ϑ_0 .

Theorem 5 *The estimator $\hat{F}_n^{-1}(1)$ is inconsistent for θ_0 .*

PROOF: First we notice that $ng(0)W_{(1:n)}$ has asymptotically the standard exponential distribution.

Then

$$\begin{aligned} P(\hat{g}_{n,W}(0+) \geq 2g(0)) &\geq P\left(\frac{1}{nW_{(1:n)}} \geq 2g(0)\right) \\ &= P\left(nW_{(1:n)}g(0) \leq \frac{1}{2}\right) \rightarrow 1 - e^{-\frac{1}{2}}. \end{aligned}$$

Consistency of $\hat{g}_{n,W}(w)$ for $w > 0$ (see Woodroffe & Sun, 1993) gives

$$P\left(\hat{g}_{n,W}\left(e^{-2\vartheta_0}\right) \leq \frac{3}{2}g(0)\right) \rightarrow 1.$$

Furthermore, using that $P(A \cap B) = P(A) + P(B) - P(A \cup B) \geq P(A) + P(B) - 1$, we obtain

$$\begin{aligned} P\left(\hat{F}_n^{-1}(1) \geq 2\vartheta_0\right) &= P\left(\sup\{w : \hat{g}_{n,W}(w) = \hat{g}_{n,W}(0+)\} \leq e^{-2\vartheta_0}\right) \\ &\geq P\left(\hat{g}_{n,W}(0+) \geq 2g_W(0) \wedge \hat{g}_{n,W}\left(e^{-2\vartheta_0}\right) \leq \frac{3}{2}g_W(0)\right) \\ &\geq P(\hat{g}_{n,W}(0+) \geq 2g_W(0)) + P\left(\hat{g}_{n,W}\left(e^{-2\vartheta_0}\right) \leq \frac{3}{2}g_W(0)\right) - 1 \\ &\geq 1 - e^{-\frac{1}{2}} + o(1) \text{ as } n \rightarrow \infty. \quad \square \end{aligned}$$

Inspired by work on the Grenander estimator, we now prove consistency of a likelihood based estimator for ϑ_0 : $\hat{\vartheta}_n = \hat{F}_n^{-1}(1 - \beta_n)$, with $\beta_n \downarrow 0$. Consistency follows from the general result of Theorem 2.

Theorem 6 For $0 < \delta < \frac{1}{2}$

$$\hat{F}_n^{-1}\left(1 - n^{-\delta}\right) \rightarrow \vartheta_0$$

in probability.

PROOF: By Theorem 2 it suffices to show that

$$n^{\frac{1}{2}}\left(1 - \hat{F}_n(x)\right) = O_p(1)$$

for every $x > \vartheta_0$. From Marshall's lemma (Robertson *et al.*, 1988, page 329), and the Glivenko Cantelli theorem, we obtain for the concave majorant $\hat{G}_{n,W}$ of the empirical distribution function of the sample W_1, W_2, \dots, W_n that

$$\sup_{w \geq 0} |\hat{G}_{n,W}(w) - G_W(w)| = O_p\left(n^{-\frac{1}{2}}\right). \quad (7)$$

By Lemma 3 it follows that for $x > \vartheta_0$

$$\begin{aligned}
n^{\frac{1}{2}} \left(1 - \hat{F}_n(x)\right) &= n^{\frac{1}{2}} \left(\hat{G}_{n,W}(e^{-x}) - e^{-x} \hat{g}_{n,W}(e^{-x})\right) \\
&= n^{\frac{1}{2}} \left(\hat{G}_{n,W}(e^{-x}) - G_W(e^{-x})\right) + n^{\frac{1}{2}} \left(G_W(e^{-x}) - e^{-x} \hat{g}_{n,W}(e^{-x})\right) \\
&= n^{\frac{1}{2}} \left(\hat{G}_{n,W}(e^{-x}) - G_W(e^{-x})\right) + n^{\frac{1}{2}} \left(e^{-x} g_W(e^{-x}) - e^{-x} \hat{g}_{n,W}(e^{-x})\right) \\
&= n^{\frac{1}{2}} \left(\hat{G}_{n,W}(e^{-x}) - G_W(e^{-x})\right) + n^{\frac{1}{2}} e^{-x} \left(g_W(e^{-x}) - \hat{g}_{n,W}(e^{-x})\right).
\end{aligned}$$

Here we use that $g_W(e^{-x}) = g_W(0)$ for every $x > \vartheta_0$. The first term is $O_p(1)$ by (7). The second term is $O_p(1)$ by Lemma 5. \square

To conclude this section, we state the following result about the rate of convergence.

Corollary 1 *Assume $1 - F(x) \sim c(\vartheta_0 - x)^\eta$ for $x \uparrow \vartheta_0$ with $\eta = 1$. Then for any sequence β_n such that $n^{1/3}\beta_n/\log n \rightarrow \infty$ and $\beta_n = \gamma_n$,*

$$\gamma_n^{-1} \left(\hat{F}_n^{-1}(1 - \beta_n) - \vartheta_0\right) = O_p(1).$$

Note that one could take, for any $\epsilon > 0$, $\beta_n = \gamma_n = n^{-\frac{1}{3}}(\log n)^{1+\epsilon}$.

PROOF: It suffices to show that Theorem 3 holds for $\alpha_n = n^{-\frac{1}{3}} \log n$. First we notice that

$$\sup_{x \in [\vartheta_0 - \varepsilon, \vartheta_0]} |\hat{F}_n(x) - F(x)| = \sup_{x \in [e^{-\vartheta_0}, e^{-\vartheta_0 + \delta}]} |\hat{g}_{n,W}(x) - g_W(x)|.$$

Then by defining

$$U_n = n^{\frac{1}{3}}(\log n)^{-1} \sup_{x \in [e^{-\vartheta_0}, e^{-\vartheta_0 + \delta}]} |\hat{g}_{n,W}(x) - g_W(x)|,$$

we know by Kulikov & Lopuhaä (2005) that $U_n = O_p(1)$. \square

6 Other methods and simulations

In this section we compare our ‘high quantile of the MLE’-method described in Section 5: $\hat{\vartheta}_n = \hat{F}_n^{-1}(1 - \beta_n)$, $\beta_n \downarrow 0$ to the kernel estimation method of the upper support point described in Delaigle & Gijbels (2006). There only the case of symmetric noise density k is considered in the simulation study. We apply the kernel estimation method to the exponential deconvolution problem. Hereby we use a simulation-based approach to find an approximately MSE-optimal band-width. The kernel estimation method cannot be used in the uniform deconvolution problem.

We first give a short description of the kernel estimation method proposed by Delaigle & Gijbels (2006). Assume that the density of X , f , is smooth except for a discontinuity at the boundary points of the support and construct a smooth kernel estimate of f . First define $\phi_U(t) = E(e^{itU}) = \int e^{itx} f_U(x) dx$ as the characteristic function of a random variable U having density f_U . In Stefanski & Carroll (1990) the following estimator $\hat{f}_X(x; h_n)$ is derived for the density f of X :

$$\hat{f}_X(x; h_n) = \frac{1}{nh_n} \sum_{j=1}^n K_Y \left(\frac{x - Z_j}{h_n}; h_n \right) \quad (8)$$

with

$$K_Y(x; h_n) = \frac{1}{2\pi} \int e^{-itx} \frac{\phi_K(t)}{\phi_Y \left(\frac{t}{h_n} \right)} dt. \quad (9)$$

Here K is a kernel function, supported on \mathbb{R} and real-valued, and h_n is the bandwidth, depending on n , such that $h_n \rightarrow 0$ as $n \rightarrow \infty$. For simplicity we denote h_n by h .

Remark 3 Note that in Delaigle & Gijbels (2006) Z is the noise variable instead of Y in this paper. Also note that a condition that is needed for (9) to be well defined, is that $\phi_Y(t) \neq 0$ for all t , which does not hold in the uniform deconvolution problem.

The method of Delaigle & Gijbels (2006) relies on derivative estimation. In the case of a single boundary point ϑ_0 , we define the estimator of ϑ_0 by

$$\hat{\tau}_n = \operatorname{argmax}_x |\hat{J}(x)|, \quad (10)$$

where the diagnostic function $\hat{J}(x)$ is given by

$$\hat{J}(x) = \frac{1}{nh} \sum_{i=1}^n K'_Y \left(\frac{x - Z_i}{h} \right) = h \frac{d}{dx} \hat{f}_X(x; h). \quad (11)$$

Here the derivative of $K_Y(\cdot; h)$ is denoted by K'_Y . In (11) we first compute $K'_Y(u)$ and then substitute $\frac{x-Z_i}{h}$ for u .

Since the density of X , f , is assumed to be discontinuous in the upper support point and the kernel estimate of f is a smooth function, it is to be expected that \hat{f}_X will have a large (negative) derivative at the upper support point. This is the motivation behind estimator (10). Figure 5 shows a typical picture of \hat{J} .

We now derive an explicit expression of K'_Y in case of the exponential deconvolution problem. We choose K to be the standard normal density, i.e.

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}.$$

It is well known (see e.g. Ross, 1993), that $\phi_K(t) = e^{-\frac{1}{2}t^2}$ and $\phi_Y(t) = (1 - it)^{-1}$. Substituting these expressions in (9) we obtain

$$K_Y(x, h) = \frac{1}{\sqrt{2\pi}} \left(e^{-\frac{1}{2}x^2} - \frac{1}{h} x e^{-\frac{1}{2}x^2} \right).$$

By differentiation it follows that

$$K'_Y(x, h) = \frac{1}{\sqrt{2\pi}} \left\{ -x e^{-\frac{1}{2}x^2} - \frac{1}{h} e^{-\frac{1}{2}x^2} + \frac{1}{h} x^2 e^{-\frac{1}{2}x^2} \right\}. \quad (12)$$

Now h must be chosen appropriately. In Delaigle & Gijbels (2006) asymptotically optimal bandwidths are studied. In their simulation study Y is symmetrically distributed (Laplace and Gaussian error).

We therefore compare the two estimators in a simulation study to be described below. For the

simulation, we use the following underlying distribution function F :

$$F(x) = \frac{27}{26} - \frac{(4-x)^3}{26} \text{ for } x \in [1, 3]. \quad (13)$$

This distribution function corresponds to density 5 proposed in Delaigle & Gijbels (2006) and has a jump continuity at ϑ_0 as required for their method to work.

We first experimentally approximate the optimal choices for the parameters needed to compute the estimators by simulation. More specifically, for a fixed sample size n , we repeat the following procedure 500 times. First, generate a sample of size n from the convolution of the standard exponential density and F given in (13). Then, for a whole range of values of p (for our method) and h (for the kernel based method), compute the associated estimates of $\vartheta_0 = 3$. Then compute the squared error of all these estimates. Having 500 squared errors for each value of p and h , compute the mean over these 500 values to obtain an estimate of the mean squared errors (MSE) of the estimators, depending on p and h respectively. Finally, define as ‘simulation based optimal parameters’ those p and h that minimize the estimated MSE.

Having followed this procedure for $n = 1000$, we obtained as approximate MSE optimal parameters $p = 0.965$ and $h = 0.90$. Using these parameters, we computed both estimators for 500 newly generated data sets (for one generated data set, this is illustrated in Figure 4 and 5). Figure 2 shows boxplots of the estimators thus obtained. From this picture it is clear that our estimator outbeats the kernel-based estimator in terms of the MSE criterion. It is the bias of the kernel-based estimator that dominates the MSE, whereas the variance of our estimator dominates the MSE. For $n = 2000$ we followed the same procedure, resulting in optimal parameters $p = 0.975$ and $h = 0.90$. Figure 3 shows a similar picture as Figure 2.

HERE FIGURE 2, 3, 4 AND 5

7 Discussion

The maximum likelihood estimator of a distribution function is a natural estimator in a whole class of deconvolution models. As such, it provides a natural candidate to use as plug-in estimator for

the upper support point of a distribution in such models.

In this paper, we study this plug-in estimator and variants thereof, proving general results of consistency. The two specific problems considered in more detail are prototypes of deconvolution problems where the kernel has bounded or unbounded support. For these problems, where the MLE can be explicitly computed, we derive more precise asymptotic results. A simulation study shows that the method can compete with other methods used in practice.

Acknowledgement

The first author's work is supported by the "Leraar in Onderzoek"-project of NWO ("High-School-Teacher in Research"-project of The Netherlands Organization for Scientific Research). We also thank two anonymous referees for suggestions that improved the presentation of this paper.

References

- Böhning, D. (1995). A review of reliable maximum likelihood algorithms for semiparametric mixture models. *J. Statist. Plann. Inference*, **47**, 5-28.
- Carroll, R.J. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.*, **83**, 1184-1186.
- Delaigle, A. and Gijbels, I. (2006). Data-driven boundary estimation in deconvolution problems. *Comput. Statist. and Data Anal.*, **50**, 1965-1994.
- Eggermont P.P.B. and LaRiccia, V.N. (1997). Nonlinearly smoothed EM estimation with automated smoothing parameter selection for nonparametric deconvolution problems. *J. Amer. Statist. Assoc.*, **92**, 1451-1458.
- Es, A. van, Jongbloed, G. and Zuijlen, M.C.A. van (1998). Isotonic inverse estimators for nonparametric deconvolution. *Ann. Statist.*, **26:6**, 2395-2406.
- Gnedenko, B.V. (1943). Sur la distribution limite du terme maximum d' une série aléatoire. *Ann.*

Math., **44**, 423-453.

Goldenshluger, A. & Tsybakov, A. (2004). Estimating the endpoint of a distribution in the presence of additive observation errors. *Statist. Probab. Lett.*, **68**, 39-49.

Grenander, U. (1956). On the theory of mortality measurement. Part II, *Skandinavisk Aktuarietidskrift* **39**, 125-153.

Groeneboom, P. and Jongbloed, G. (2003). Density estimation in the uniform deconvolution model, *Statist. Neerlandica*, **57**, 136-157.

Groeneboom, P., Jongbloed, G. and Wellner, J.A. (2003). The support reduction algorithm for computing nonparametric function estimates in mixture models. Submitted.

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

Groeneboom, P. and Wellner, J.A. (1992). *Information bounds and nonparametric maximum likelihood estimation*, Birkäuser, Basel.

Hall, P. & Simar, L. (2002). Estimating a changepoint, boundary, or frontier in the presence of observation error. *J. Amer. Statist. Assoc.*, **97**, 523-534.

Johnstone, I.M. and Raimondo M. (2004). Periodic boxcar deconvolution and Diophantine approximation. *Ann. of Statist.*, **32**, 1781-1804.

Jongbloed, G. (1998). Exponential deconvolution: two asymptotically equivalent estimators. *Statist. Neerlandica*, **52**, 6-17.

Kulikov, V.N. and Lopuhaä H.P. (2005). Asymptotic normality of the L_k - error of the Grenander estimator. *Ann. Statist.*, **33**, 2228-2255.

Lindsay, B.G. (1995). Mixture models: theory, geometry and applications. *NSF-CBMS Regional Conference Series in Probability and Statistics*, Vol **5**.

Meister, A. (2006). Support estimation via moment estimation in presence of noise. *Statistics*, **40**, 259-275.

Robertson, T., Wright, F.T. and Dykstra, R.L. (1988). *Order restricted statistical inference*, Wiley, New York.

Ross, S.M. (1993). *Introduction to probability models, Fifth Edition*, Academic Press, Inc., San Diego.

Stefanski, L. and Carroll, R.J. (1990). Deconvoluting kernel density estimators. *Statistics*, **2**, 169-184.

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

Corresponding author:

G. Jongbloed, Department of Mathematics, Vrije Universiteit Amsterdam, De Boelelaan 1081a, 1081 HV Amsterdam, The Netherlands.

E-mail: G.Jongbloed@few.vu.nl

Appendix

Lemma 4 *Let Z_1, Z_2, \dots be independent identically distributed with distribution function G . Suppose that $g = G'$ exists on an interval $I_\delta = (z_0 - \delta, z_0 + \delta)$ and that there is a $c > 0$ such that $\frac{1}{c} < g(z) < c$ for all $z \in I_\delta$. Then*

$$\max\{Z_{(i)} - Z_{(i-1)} : Z_{(i)}, Z_{(i-1)} \in I_\delta\} = O_p\left(\frac{\log n}{n}\right).$$

PROOF: Let E_1, E_2, \dots be independent identically distributed random variables with the standard exponential density and U_1, U_2, \dots i.i.d. uniformly distributed random variables on $(0, 1)$. Note that

$$\left(\frac{E_j}{\sum_{i=1}^{n+1} E_i}\right)_{j=1}^n \stackrel{\mathcal{D}}{=} (U_{(1:n)}, U_{(2:n)} - U_{(1:n)}, \dots, U_{(n:n)} - U_{((n-1):n)}),$$

where $U_{(i:n)}$ is the i -th order statistic in the sample U_1, \dots, U_n . Defining $M_n = \max_{1 \leq i \leq n} E_i$, we

get that

$$P(M_n - \log n \leq k) = P(E_i \leq \log n + k)^n \rightarrow e^{-e^{-k}}$$

implying $M_n = O_p(\log n)$ and hence

$$\max_{1 \leq i \leq n} (U_{(i)} - U_{(i-1)}) = O_p(n^{-1} \log n).$$

Using that

$$(Z_{(i:n)})_{i=1}^n \stackrel{\mathcal{D}}{=} (G^{-1}(U_{(i:n)}))_{i=1}^n,$$

we obtain that

$$\begin{aligned} & \max\{Z_{(i:n)} - Z_{((i-1):n)} : Z_{(i:n)}, Z_{((i-1):n)} \in I_\delta\} \stackrel{\mathcal{D}}{=} \\ & \max\{G^{-1}(U_{(i:n)}) - G^{-1}(U_{((i-1):n)}) : G^{-1}(U_{(i:n)}), G^{-1}(U_{((i-1):n)}) \in I_\delta\} \leq \\ & c \max\{U_{(i:n)} - U_{((i-1):n)} : 1 \leq i \leq n\} = O_p(n^{-1} \log n). \quad \square \end{aligned}$$

Lemma 5 *Let W_1, W_2, \dots be independent identically distributed variables with decreasing density g_W and $g_W(w) = g_W(0)$ for every $w \in (0, \delta)$. Let $\hat{g}_{n,W}$ be the Grenander estimator based on W_1, \dots, W_n . Then for every $w \in (0, \delta)$:*

$$\sqrt{n}(\hat{g}_{n,W}(w) - g_W(w)) = O_p(1).$$

PROOF: Fix $w \in (0, \delta)$. Suppose that for a certain M ,

$$\hat{g}_{n,W}(w) > g_W(w) + \frac{M}{\sqrt{n}} = g_W(0) + \frac{M}{\sqrt{n}}.$$

This implies that

$$\hat{G}_{n,W}(w) = \int_0^w \hat{g}_{n,W}(u) du \geq \int_0^w \hat{g}_{n,W}(w) du > g_W(0)w + \frac{M}{\sqrt{n}}w = G_W(w) + \frac{M}{\sqrt{n}}w.$$

Hence,

$$P\left(\hat{g}_{n,W}(w) > g_W(w) + \frac{M}{\sqrt{n}}\right) \leq P\left(\hat{G}_{n,W}(w) - G_W(w) > \frac{Mw}{\sqrt{n}}\right)$$

which, by (7), can be made arbitrarily small uniformly in n , by choosing M sufficiently large. On the other hand, suppose for some $w \in (0, \delta)$ that

$$\hat{g}_{n,W}(w) < g_W(w) - \frac{M}{\sqrt{n}} = g_W(0) - \frac{M}{\sqrt{n}}.$$

Then

$$\begin{aligned} \hat{G}_{n,W}(\delta) - \hat{G}_{n,W}(w) &\leq \hat{g}_{n,W}(w)(\delta - w) < g_W(0)(\delta - w) - (\delta - w)\frac{M}{\sqrt{n}} \\ &= G_W(\delta) - G_W(w) - (\delta - w)\frac{M}{\sqrt{n}}. \end{aligned}$$

Now either

$$\hat{G}_{n,W}(\delta) - G_W(\delta) < -\frac{1}{2}(\delta - w)\frac{M}{\sqrt{n}} \text{ or } G_W(w) - \hat{G}_{n,W}(w) < -\frac{1}{2}(\delta - w)\frac{M}{\sqrt{n}}.$$

Hence, again using (7),

$$\begin{aligned} P\left(\hat{g}_{n,W}(w) < g_W(w) - \frac{M}{\sqrt{n}}\right) &\leq P\left(\hat{G}_{n,W}(\delta) - G_W(\delta) < -\frac{1}{2}(\delta - w)\frac{M}{\sqrt{n}}\right) + \\ &P\left(G_W(w) - \hat{G}_{n,W}(w) < -\frac{1}{2}(\delta - w)\frac{M}{\sqrt{n}}\right) \end{aligned}$$

which can be made arbitrarily small uniformly in n , by choosing M sufficiently large. □

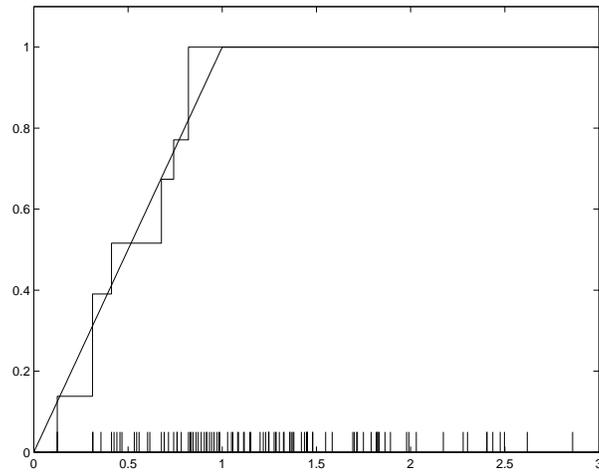


Figure 1: The true standard uniform F and the MLE \hat{F}_n of F based on a sample of size 100.

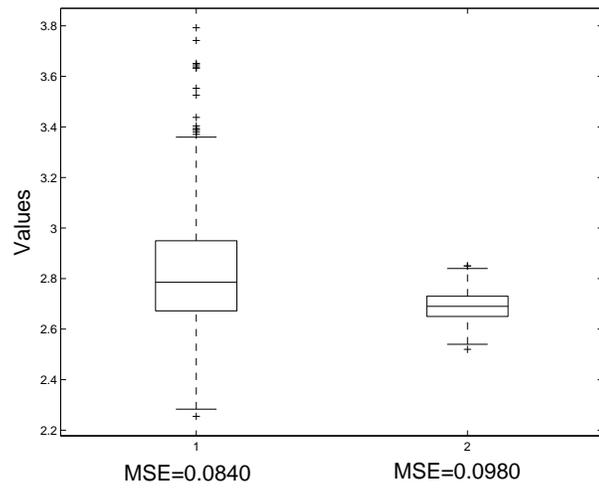


Figure 2: The upper support point is $\vartheta_0 = 3$. The samples have size 1000, $p_n = 0.965$ and $h_n = 0.90$. The left boxplot corresponds to the high quantile estimator proposed here and the right boxplot to the kernel based method of Delaigle & Gijbels (2006)

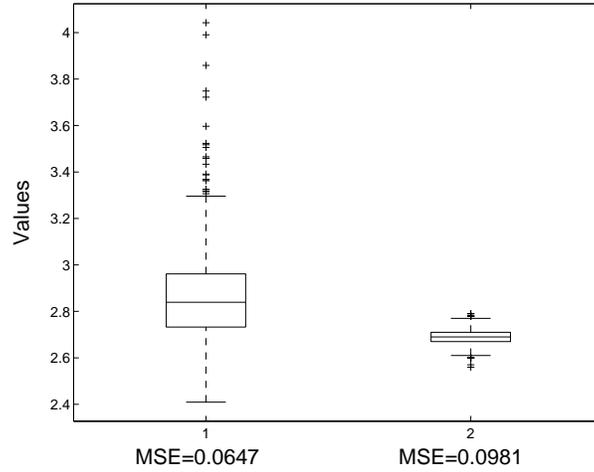


Figure 3: The upper support point is $\vartheta_0 = 3$. The samples have size 2000, $p_n = 0.975$ and $h_n = 0.90$. The left boxplot corresponds to the high quantile estimator proposed here and the right boxplot to the kernel based method of Delaigle & Gijbels (2006)

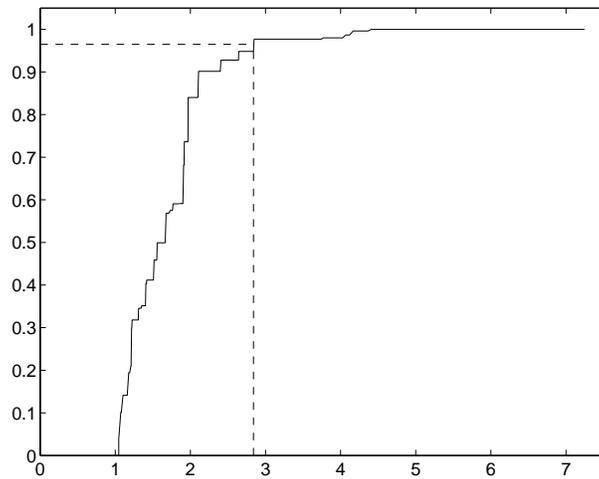


Figure 4: The MLE \hat{F}_n of F based on a sample of size 1000. We detect $\hat{\vartheta}_n$ for a given p_n . Here $p_n = 0.965$.

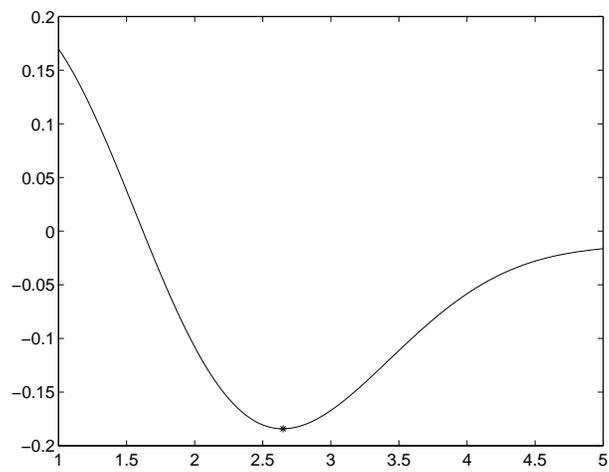


Figure 5: The diagnostic function $\hat{J}(x)$ for $h_n = 0.90$ based on the same sample used for Figure 3.