Statistica Sinica **21** (2011), 873-899

ON THE GRENANDER ESTIMATOR AT ZERO

Fadoua Balabdaoui[1], Hanna Jankowski[2], Marios Pavlides[3],
Arseni Seregin[4] and Jon Wellner[4]

[1] *Université Paris-Dauphine,* [2] *York University,* [3] *Frederick University Cyprus and* [4] *University of Washington*

Abstract: We establish limit theory for the Grenander estimator of a monotone density near zero. In particular we consider the situation when the true density f_0 is unbounded at zero, with different rates of growth to infinity. In the course of our study we develop new switching relations using tools from convex analysis. The theory is applied to a problem involving mixtures.

Key words and phrases: Convex analysis, inconsistency, limit distribution, maximum likelihood, mixture distributions, monotone density, nonparametric estimation, Poisson process, rate of growth, switching relations.

1. Introduction and Main Results

Let $X_1, \ldots, X_n$ be a sample from a decreasing density f_0 on $(0, \infty)$, and let $\widehat{f}_n$ denote the Grenander estimator (i.e. the maximum likelihood estimator) of f_0. Thus $\widehat{f}_n \equiv \widehat{f}_n^L$ is the *left derivative* of the least concave majorant $\widehat{F}_n$ of the empirical distribution function $\mathbb{F}_n$; see e.g., Grenander (1956a,b), Groeneboom (1985), and Devroye (1987, Chap. 8).

The Grenander estimator $\widehat{f}_n$ is a uniformly consistent estimator of f_0 on sets bounded away from 0 if f_0 is continuous:

$$\sup_{x \geq c} |\widehat{f}_n(x) - f_0(x)| \to_{a.s.} 0$$

for each $c > 0$. It is also known that $\widehat{f}_n$ is consistent with respect to the L_1 ($\|p - q\|_1 \equiv \int |p(x) - q(x)| dx$) and Hellinger ($h^2(p,q) \equiv 2^{-1} \int \left[\sqrt{p(x)} - \sqrt{q(x)}\right]^2 dx$) metrics: that is,

$$\|\widehat{f}_n - f_0\|_1 \to_{a.s.} 0 \qquad \text{and} \qquad h(\widehat{f}_n, f_0) \to_{a.s.} 0;$$

see e.g. Devroye (1987, Thm. 8.3) and van de Geer (1993).

However, it is also known that $\widehat{f}_n(0) \equiv \widehat{f}_n(0+)$ is an inconsistent estimator of $f_0(0) \equiv f_0(0+) = \lim_{x \searrow 0} f_0(x)$, even when $f_0(0) < \infty$. In fact, Woodroofe and Sun (1993) showed that

$$\widehat{f}_n(0) \to_d f_0(0) \sup_{t>0} \frac{\mathbb{N}(t)}{t} \stackrel{d}{=} f_0(0) \frac{1}{U} \tag{1.1}$$

as $n \to \infty$, where $\mathbb{N}$ is a standard Poisson process on $[0, \infty)$ and $U \sim \text{Uniform}(0, 1)$. Woodroofe and Sun (1993) introduced penalized estimators $\widetilde{f}_n$ of f_0 which yield consistency at 0: $\widetilde{f}_n(0) \to_p f_0(0)$. Kulikov and Lopuhaä (2006) study estimation of $f_0(0)$ based on the Grenander estimator $\widehat{f}_n$ evaluated at points of the form $t = cn^{-\gamma}$. Among other things, they show that $\widehat{f}_n(n^{-1/3}) \to_p f_0(0)$ if $|f_0'(0+)| > 0$.

Our view in this paper is that the inconsistency of $\widehat{f}_n(0)$ as an estimator of $f_0(0)$ exhibited in (1.1) can be regarded as a simple consequence of the fact that the class of all monotone decreasing densities on $(0, \infty)$ includes many densities f which are unbounded at 0, so that $f(0) = \infty$, and the Grenander estimator $\widehat{f}_n$ simply has difficulty deciding which is true, even when $f_0(0) < \infty$. From this perspective we seek answers to three questions under some reasonable hypotheses concerning the growth of $f_0(x)$ as $x \searrow 0$.

Q1: How fast does $\widehat{f}_n(0)$ diverge as $n \to \infty$?

Q2: Do the stochastic processes $\{b_n \widehat{f}_n(a_n t) : 0 \le t \le c\}$ converge for some sequences a_n, b_n, and $c > 0$?

Q3: What is the behavior of the relative error

$$\sup_{0 \le x \le c_n} \left| \frac{\widehat{f}_n(x)}{f_0(x)} - 1 \right|$$

for some constant c_n?

It turns out that answers to questions **Q1 - Q3** are intimately related to the limiting behavior of the minimal order statistic $X_{n:1} \equiv \min\{X_1, \ldots, X_n\}$. By Gnedenko (1943) or de Haan and Ferreira (2006, Thm. 1.1.2)), it is well-known that there exists a sequence $\{a_n\}$ such that

$$a_n^{-1} X_{n:1} \to_d Y, \tag{1.2}$$

where Y has a nondegenerate limiting distribution G if and only if

$$nF_0(a_n x) \to x^{\gamma}, \qquad x > 0, \tag{1.3}$$

for some $\gamma > 0$, and hence $a_n \to 0$. One possible choice of a_n is $a_n = F_0^{-1}(1/n)$, but any sequence $\{a_n\}$ satisfying $nF_0(a_n) \to 1$ also works. Since F_0 is concave the convergence in (1.3) is uniform on any interval $[0, K]$. Concavity of F_0 and existence of f_0 also implies convergence of the derivative:

$$na_n f_0(a_n x) \to \gamma x^{\gamma-1}. \tag{1.4}$$

By Gnedenko (1943), (1.2) is equivalent to

$$\lim_{x\to 0+} \frac{F_0(cx)}{F_0(x)} = c^\gamma, \qquad c > 0. \tag{1.5}$$

Thus (1.2), (1.3), and (1.5) are equivalent. In this case we have

$$G(x) = 1 - e^{-x^\gamma}, \quad x \geq 0. \tag{1.6}$$

Since F_0 is concave, the power $\gamma \in (0, 1]$.

As illustrations of our general result, we consider three hypotheses on f_0:

G0: the density f_0 is bounded at zero, $f_0(0) < \infty$;

G1: for some $\beta \geq 0$ and $0 < C_1 < \infty$, $(\log(1/x))^{-\beta} f_0(x) \to C_1$, as $x \searrow 0$;

G2: for some $0 \leq \alpha < 1$ and $0 < C_2 < \infty$, $x^\alpha f_0(x) \to C_2$, as $x \searrow 0$.

Note that in **G2** the value $\alpha = 1$ is not possible for a positive limit C_2, since $xf(x) \to 0$ as $x \to 0$ for any monotone density f; see e.g. Devroye (1986, Thm. 6.2). Below we assume that F_0 satisfies the condition (1.5). Our cases **G0** and **G1** correspond to $\gamma = 1$ and **G2** to $\gamma = 1 - \alpha$.

One motivation for considering monotone densities which are unbounded at zero comes from the study of mixture models. An example of this type, as discussed by Donoho and Jin (2004), is as follows. Suppose $X_1, \ldots, X_n$ are i.i.d. with distribution function F where,

$$\begin{aligned} &\text{under } H_0: \ F = \Phi, && \text{the standard normal d.f.,} \\ &\text{under } H_1: \ F = (1-\epsilon)\Phi + \epsilon\Phi(\cdot - \mu), && \epsilon \in (0,1), \quad \mu > 0. \end{aligned}$$

If we transform to $Y_i \equiv 1 - \Phi(X_i) \sim G$, then for $0 \leq y \leq 1$

$$\begin{aligned} &\text{under } H_0: G(y) = y, \qquad \text{the Uniform}(0,1) \text{ d.f.,} \\ &\text{under } H_1: G = G_{\epsilon,\mu}(y) = (1-\epsilon)y + \epsilon(1 - \Phi(\Phi^{-1}(1-y) - \mu)). \end{aligned}$$

It is easily seen that the density $g_{\epsilon,\mu}$ of $G_{\epsilon,\mu}$, given by

$$g_{\epsilon,\mu}(y) = (1-\epsilon) + \epsilon \frac{\phi(\Phi^{-1}(1-y) - \mu)}{\phi(\Phi^{-1}(1-y))},$$

is monotone decreasing on $(0,1)$ and is unbounded at zero. We show in Section 4 that $G_{\epsilon,\mu}$ satisfies our key hypothesis (1.5) with $\gamma = 1$. Moreover, we show that the whole class of models of this type with Φ replaced by the generalized Gaussian (or Subbotin) distribution, also satisfy (1.5), and hence the behavior of the Grenander estimator at zero gives information about the behavior of the contaminating component of the mixture model (in the transformed form) at zero.

Another motivation for studying these questions in the monotone density framework is to gain insights for a study of the corresponding questions in the context of nonparametric estimation of a monotone spectral density. In that setting, singularities at the origin correspond to the interesting phenomena of long-range dependence and long-memory processes; see e.g. Cox (1984), Beran (1994), Martin and Walker (1997), Gneiting (2000), and Ma (2002). Although our results here do not apply directly to the problem of nonparametric estimation of a monotone spectral density function, it seems plausible that similar results hold in that setting; note that when f is a spectral density, **G1** and **G2** correspond to long-memory processes (with the usual description being in terms of $\beta = 1-\alpha \in (0,1)$ or the Hurst coefficient $H = 1-\beta/2 = 1-(1-\alpha)/2 = (1+\alpha)/2$). See Anevski and Soulier (2011) for recent work on nonparametric estimation of a monotone spectral density.

Let $\mathbb{N}$ denote the standard Poisson process on $\mathbb{R}^+$. When (1.5), and hence (1.6) hold, it follows from Miller (1976, Thm. 2.1) together with Jacod and Shiryaev (2003, Thm. 2.15(c)(ii)), that

$$n\mathbb{F}_n(a_n t) \Rightarrow \mathbb{N}(t^\gamma) \qquad \text{in } D[0,\infty), \tag{1.7}$$

which should be compared to (1.3).

Since we are studying the estimator $\widehat{f}_n$ near zero, and because the value of $\widehat{f}_n$ at zero is defined as the right limit $\lim_{x\searrow 0} \widehat{f}_n(x) \equiv \widehat{f}_n(0)$, it is sensible to study instead the right-continuous modification of $\widehat{f}_n$, and this of course coincides with the right derivative $\widehat{f}_n^R$ of the least concave majorant $\widehat{F}_n$ of the empirical distribution function $\mathbb{F}_n$. Therefore we change notation for the rest of this paper and write $\widehat{f}_n$ for $\widehat{f}_n^R$ throughout. We write $\widehat{f}_n^L$ for the left-continuous Grenander estimator.

Theorem 1.1. *Suppose that* (1.5) *holds. Let* a_n *satisfy* $nF_0(a_n) \sim 1$*, let* $\widehat{h}_\gamma$ *denote the right derivative of the least concave majorant of* $t \mapsto \mathbb{N}(t^\gamma)$*,* $t \geq 0$*. Then*

(i) $na_n \widehat{f}_n(ta_n) \Rightarrow \widehat{h}_\gamma(t)$ *in* $D[0,\infty)$*,*

(ii) *for all* $c \geq 0$,

$$\sup_{0<x\leq ca_n}\left|\frac{\widehat{f}_n(x)}{f_0(x)} - 1\right| \to_d \sup_{0<t\leq c}\left|\frac{t^{1-\gamma}\widehat{h}_\gamma(t)}{\gamma} - 1\right|.$$

The behavior of $\widehat{f}_n$ near zero under the different hypotheses **G0**, **G1**, and **G2** now follows as corollaries to Theorem 1.1. Let $Y_\gamma \equiv \widehat{h}_\gamma(0)$. We then have

$$Y_\gamma = \sup_{t>0}(\mathbb{N}(t^\gamma)/t) = \sup_{s>0}(\mathbb{N}(s)/s^{1/\gamma}). \tag{1.8}$$

Here we note that $Y_1 =_d 1/U$, where $U \sim \text{Uniform}(0,1)$ has distribution function $H_1(x) = 1 - 1/x$ for $x \geq 1$. The distribution of Y_γ for $\gamma \in (0,1]$ is given in Proposition 1.5 below. The first part of the following corollary was established by Woodroofe and Sun (1993).

Corollary 1.2. *Suppose that* **G0** *holds. Then* $\gamma = 1$, $a_n^{-1} = nf_0(0+)$ *satisfies* $nF_0(a_n) \to 1$, *and it follows that*

(i) $\widehat{f}_n(0) \to_d f_0(0)\widehat{h}_1(0) = f_0(0)Y_1$,

(ii) *the processes* $\{t \mapsto \widehat{f}_n(tn^{-1}) : n \geq 1\}$ *satisfy*

$$\widehat{f}_n(tn^{-1}) \Rightarrow f_0(0)\widehat{h}_1(f_0(0)t) \qquad \textit{in} \;\; D[0,\infty),$$

(iii) *for* $c_n = c/n$ *with* $c > 0$,

$$\sup_{0<x\leq c_n}\left|\frac{\widehat{f}_n(x)}{f_0(x)} - 1\right| \to_d Y_1 - 1,$$

which has distribution function $H_1(x+1) = 1 - 1/(x+1)$ *for* $x \geq 0$.

Corollary 1.3. *Suppose that* **G1** *holds. Then* $F_0(x) \sim C_1 x(\log(1/x))^\beta$ *so* $\gamma = 1$, *and* $a_n^{-1} = C_1 n(\log n)^\beta$ *satisfies* $nF_0(a_n) \to 1$. *It follows that*

(i) $\widehat{f}_n(0)/(\log n)^\beta \to_d C_1 Y_1$,

(ii) *the processes* $\{t \mapsto (\log n)^{-\beta}\widehat{f}_n(t/(n(\log n)^\beta)) : n \geq 1\}$ *satisfy*

$$\frac{1}{(\log n)^\beta}\widehat{f}_n\left(\frac{t}{n(\log n)^\beta}\right) \Rightarrow C_1\widehat{h}_1(C_1 t) \qquad \text{in} \;\; D[0,\infty),$$

(iii) *for* $c_n = c/(n(\log n)^\beta)$ *with* $c > 0$,

$$\sup_{0<x\leq c_n}\left|\frac{\widehat{f}_n(x)}{f_0(x)} - 1\right| \to_d Y_1 - 1.$$

Corollary 1.4. *Suppose that* **G2** *holds and set* $\widetilde{C}_2 = (C_2/(1-\alpha))^{1/(1-\alpha)}$. *Then* $F_0(x) \sim C_2 x^{1-\alpha}/(1-\alpha)$ *so* $\gamma = 1-\alpha$, $a_n^{-1} = \widetilde{C}_2 n^{1/(1-\alpha)}$ *satisfies* $nF_0(a_n) \to 1$, *and it follows that*

(i)
$$\frac{\widehat{f}_n(0)}{n^{\alpha/(1-\alpha)}} \to_d \widetilde{C}_2 Y_{1-\alpha}, \tag{1.9}$$

(ii) *the processes* $\{t \mapsto n^{-\alpha/(1-\alpha)} \widehat{f}_n(tn^{-1/(1-\alpha)}) : \ n \geq 1\}$ *satisfy*

$$\frac{\widehat{f}_n(tn^{-1/(1-\alpha)})}{n^{\alpha/(1-\alpha)}} \Rightarrow \widetilde{C}_2 \widehat{h}_{1-\alpha}(\widetilde{C}_2 t) \qquad \text{in} \quad D[0,\infty),$$

(iii) *for* $c_n = c/n^{1/(1-\alpha)}$ *with* $c > 0$,

$$\sup_{0<x\leq c_n} \left| \frac{\widehat{f}_n(x)}{f_0(x)} - 1 \right| \to_d \sup_{0<t\leq c\widetilde{C}_2} \left| \frac{t^\alpha \widehat{h}_{1-\alpha}(t)}{1-\alpha} - 1 \right|.$$

Taking $\beta = 0$ in (i) of Corollary 1.3 yields the limit theorem (1.1) of Woodroofe and Sun (1993) as a corollary; in this case $C_1 = f_0(0)$. Similarly, taking $\alpha = 0$ in (ii) of Corollary 1.4 yields the limit theorem (1.1) of Woodroofe and Sun (1993) as a corollary; in this case $C_2 = f_0(0)$. Note that Theorem 1.1 yields further corollaries when assumptions **G1** and **G2** are modified by other slowly varying functions.

Recall the definition (1.8) of Y_γ. The following proposition gives the distribution of Y_γ for $\gamma \in (0,1]$.

Proposition 1.5. *For fixed* $0 < \gamma \leq 1$ *and* $x > 0$,

$$\Pr\left(\sup_{s>0} \left\{ \frac{\mathbb{N}(s)}{s^{1/\gamma}} \right\} \leq x \right) = \begin{cases} 1 - \frac{1}{x}, & \text{if } \gamma = 1, \ x \geq 1, \\ 1 - \sum_{k=1}^{\infty} a_k(x,\gamma), & \text{if } \gamma < 1, \ x > 0, \end{cases}$$

where the sequence $\{a_k(x,\gamma)\}_{k\geq 1}$ *is constructed recursively as follows:*

$$a_1(x,\gamma) = p\left(\left(\frac{1}{x}\right)^\gamma ; 1 \right),$$

and, for $j \geq 1$,

$$a_k(x,\gamma) = p\left(\left(\frac{k}{x}\right)^\gamma ; k \right) - \sum_{i=1}^{k-1} \left\{ a_i(x,\gamma) \cdot p\left(\left(\frac{k}{x}\right)^\gamma - \left(\frac{i}{x}\right)^\gamma ; k-i \right) \right\},$$

where $p(m;k) \equiv e^{-m} m^k / k!$.

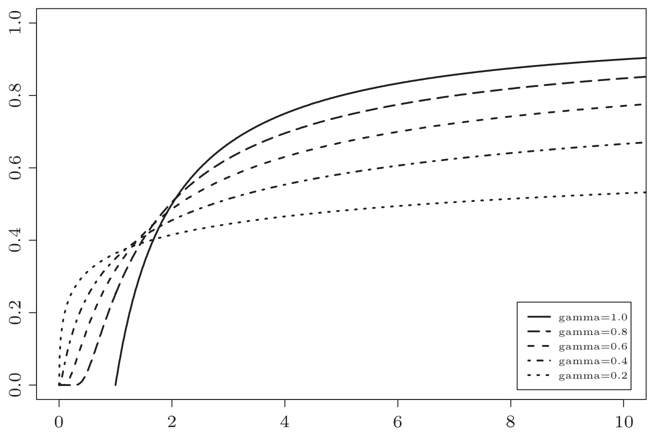

Figure 1. The distribution functions of Y_γ, $\gamma \in \{0.2, 0.4, 0.6, 0.8, 1.0\}$.

Remark 1.6. The random variables Y_γ are increasingly heavy-tailed as γ decreases; cf. Figure 1. Let $T_1, T_2, \ldots$ be the event times of the Poisson process $\mathbb{N}$, i.e., $\mathbb{N}(t) = \sum_{j=1}^{\infty} 1_{[T_j \le t]}$. Then note that

$$Y_\gamma \stackrel{d}{=} \sup_{j \ge 1} \frac{j}{T_j^{1/\gamma}} \ge \frac{1}{T_1^{1/\gamma}},$$

where $T_1 \sim$ Exponential(1). On the other hand

$$Y_\gamma = \left(\sup_{t>0} \frac{\mathbb{N}(t)^\gamma}{t} \right)^{1/\gamma} \le \left(\sup_{t>0} \frac{\mathbb{N}(t)}{t} \right)^{1/\gamma} \stackrel{d}{=} \frac{1}{U^{1/\gamma}},$$

where $U \sim$ Uniform$(0,1)$. Thus it is easily seen that $E(Y_\gamma^r) < \infty$ if and only if $r < \gamma$, and that the distribution function F_γ of Y_γ is bounded above and below by the distribution functions G_γ^L and G_γ^U of $1/T_1^{1/\gamma}$ and $1/U^{1/\gamma}$, respectively.

The proofs of the above results appear in Appendix A. They rely heavily on a set equality known as the "switching relation". We study this relation using convex analysis in Section 2. Section 3 gives some numerical results that accompany the results presented here, and Section 4 studies applications to the estimation of mixture models.

2. Switching Relations

In this section we consider several general variants of the so-called switching relation first given in Groeneboom (1985), and used repeatedly by other authors, including Kulikov and Lopuhaä (2005, 2006), and van der Vaart and Wellner (1996). Other versions of the switching relation were studied by van der Vaart and van der Laan (2006, Lemma 4.1). In particular, we provide a novel proof of the result using convex analysis. This approach also allows us to restate the relation without restricting the domain to compact intervals. Throughout this section we make use of definitions from convex analysis (cf., Rockafellar (1970); Rockafellar and Wets (1998); Boyd and Vandenberghe (2004)) that are given in Appendix B.

Suppose that Φ is a function, $\Phi : D \to \mathbb{R}$, defined on the (possibly infinite) closed interval $D \subset \mathbb{R}$. The least concave majorant $\widehat{\Phi}$ of Φ is the pointwise infimum of all closed concave functions $g : D \to \mathbb{R}$ with $g \geq \Phi$. Since $\widehat{\Phi}$ is concave, it is continuous on D^o, the interior of D. Furthermore, $\widehat{\Phi}$ has left and right derivatives on D^o, and is differentiable with the exception of at most countably many points. Let $\widehat{\phi}_L$ and $\widehat{\phi}_R$ denote the left and right derivatives, respectively, of $\widehat{\Phi}$.

If Φ is upper semicontinuous, then so is $\Phi_y(x) = \Phi(x) - yx$ for each $y \in \mathbb{R}$. If D is compact, then Φ_y attains a maximum on D, and the set of points achieving the maximum is closed. Compactness of D was assumed by van der Vaart and van der Laan (2006, see their Lemma 4.1). One of our goals here is to relax this assumption.

Assuming they are defined, we consider the argmax functions

$$\begin{aligned}\kappa_L(y) \equiv \text{argmax}^L \Phi_y &\equiv \text{argmax}_x^L\{\Phi(x) - yx\}\\ &= \inf\{x \in D : \ \Phi_y(x) = \sup_{z \in D} \Phi_y(z)\},\\ \kappa_R(y) \equiv \text{argmax}^R \Phi_y &\equiv \text{argmax}_x^R\{\Phi(x) - yx\}\\ &= \sup\{x \in D : \ \Phi_y(x) = \sup_{z \in D} \Phi_y(z)\}.\end{aligned}$$

Theorem 2.1. *Suppose that Φ is a proper upper-semicontinuous real-valued function defined on a closed subset $D \subset \mathbb{R}$. Then $\widehat{\Phi}$ is proper if and only if $\Phi \leq l$ for some linear function l on D. Furthermore, if conv(hypo(Φ)) is closed, then the functions κ_L and κ_R are well defined and for $x \in D$ and $y \in \mathbb{R}$,*

S1 $\widehat{\phi}_L(x) < y$ *if and only if* $\kappa_R(y) < x$.

S2 $\widehat{\phi}_R(x) \leq y$ *if and only if* $\kappa_L(y) \leq x$.

When Φ is the empirical distribution function $\mathbb{F}_n$ as in Section 1, then $\widehat{\Phi} = \widehat{F}_n$ is the least concave majorant of $\mathbb{F}_n$, and $\widehat{\phi}_L = \widehat{f}_n^L$ the Grenander estimator, while $\widehat{\phi}_R = \widehat{f}_n = \widehat{f}_n^R$ is the right continuous version of the estimator. In this situation the argmax functions κ_R, κ_L correspond to

$$\begin{aligned}\widehat{s}_n^R(y) &= \sup\{x \geq 0 : \mathbb{F}_n(x) - yx = \sup_{z \geq 0}(\mathbb{F}_n(z) - yz)\},\\ \widehat{s}_n^L(y) &= \inf\{x \geq 0 : \mathbb{F}_n(x) - yx = \sup_{z \geq 0}(\mathbb{F}_n(z) - yz)\}.\end{aligned}$$

The switching relation given by Groeneboom (1985) says that, with probability one,

$$\{\widehat{f}_n^L(x) \le y\} = \{\widehat{s}_n^R(y) \le x\}. \tag{2.1}$$

van der Vaart and Wellner (1996, p.296), say that (2.1) holds for every x and y; see also Kulikov and Lopuhaä (2005, p.2229), and Kulikov and Lopuhaä (2006, p.744). The advantage of (2.1) is immediate: the MLE is related to a continuous map of a process whose behavior is well-understood.

The following corollary gives the conclusion of Theorem 2.1 when Φ is the empirical distribution function $\mathbb{F}_n$.

Corollary 2.2. *Let $\widehat{F}_n$ be the least concave majorant of the empirical distribution function $\mathbb{F}_n$, and let $\widehat{f}_n^L$ and $\widehat{f}_n^R$ denote its left and right derivatives, respectively. Then*

$$\{\widehat{f}_n^L(x) < y\} = \{\widehat{s}_n^R(y) < x\}, \tag{2.2}$$

$$\{\widehat{f}_n^R(x) \le y\} = \{\widehat{s}_n^L(y) \le x\}. \tag{2.3}$$

The following example shows, however, that the set identity (2.1) can fail.

Example 2.3. Suppose that we observe $(X_1, X_2, X_3) = (1, 2, 4)$. Then the MLE is

$$\widehat{f}_n^L(x) = \begin{cases} \frac{1}{3}, & 0 < x \le 2, \\ \frac{1}{6}, & 2 < x \le 4, \\ 0, & 4 < x < \infty. \end{cases}$$

The process $\widehat{s}_n^R$ is given by

$$\widehat{s}_n^R(y) = \begin{cases} 4, & 0 < y \le \frac{1}{6}, \\ 2, & \frac{1}{6} < y \le \frac{1}{3}, \\ 0, & \frac{1}{3} < y < \infty. \end{cases}$$

Note that (2.1) fails if $x = 4$ and $0 < y < 1/6$, since in this case $\widehat{f}_n^L(x) = \widehat{f}_n^L(4) = 1/6$ and the event $\{\widehat{f}_n^L(x) \le y\}$ fails to hold, while $\widehat{s}_n^R(y) = 4$ and the event $\{\widehat{s}_n^R(y) \le x\}$ holds. However, (2.2) does hold: with $x = 4$ and $0 < y < 1/6$, both of the events $\{\widehat{f}_n^L(x) < y\}$ and $\{\widehat{s}_n^R(y) < x\}$ fail to hold. Some checking shows that (2.2) and (2.3) hold for all other values of x and y.

Our proof of Theorem 2.1 is based on a proposition that is a consequence of general facts concerning convex functions, as given in Rockafellar (1970) and

Rockafellar and Wets (1998). Let h be a closed proper convex function on $\mathbb{R}$, and let f be its conjugate, $f(y) = \sup_{x\in\mathbb{R}}\{yx - h(x)\}$. Let h'_- and h'_+ be the left and right derivatives of h, and define functions s_- and s_+ by

$$s_-(y) = \inf\{x \in \mathbb{R} : \ yx - h(x) = f(y)\}, \tag{2.4}$$

$$s_+(y) = \sup\{x \in \mathbb{R} : \ yx - h(x) = f(y)\}. \tag{2.5}$$

Proposition 2.4. *The following set identities hold:*

$$\{(x, y) : \ h'_-(x) \le y\} = \{(x, y) : \ s_+(y) \ge x\}; \tag{2.6}$$

$$\{(x, y) : \ h'_+(x) < y\} = \{(x, y) : \ s_-(y) > x\}. \tag{2.7}$$

Proof. All references are to Rockafellar (1970). By Theorem 24.3 the set $\Gamma = \{(x, y) \in \mathbb{R}^2 : \ y \in \partial h(x)\}$ is a maximal complete non-decreasing curve. By Theorem 23.5, the closed proper convex function h and its conjugate f satisfy $h(x) + f(y) \ge xy$, and equality holds if and only if $y \in \partial h(x)$, or equivalently if $x \in \partial f(y)$ where ∂h and ∂f denote the subdifferentials of h and f, respectively. Thus we have $\Gamma = \{(x, y) \in \mathbb{R}^2 : \ x \in \partial f(x)\}$ and, by the definitions of s_- and s_+, $\Gamma = \{(x, y) : \ s_-(y) \le x \le s_+(y)\}$. By Theorem 24.1, the curve Γ is defined by the left and right derivatives of h:

$$\Gamma = \{(x, y) : \ h'_-(x) \le y \le h'_+(x)\}. \tag{2.8}$$

Using the dual representation we obtain

$$\Gamma = \{(x, y) : \ f'_-(y) \le x \le f'_+(y)\}, \tag{2.9}$$

so $s_- \equiv f'_-$ and $s_+ \equiv f'_+$. Moreover, the functions h'_- and f'_- are left-continuous, the functions h'_+ and f'_+ are right continuous, and all of these functions are nondecreasing.

From (2.8) and (2.9) it follows that $\{h'_-(x) \le y\} = \{f'_+(y) \ge x\}$, which implies (2.6). Since the functions h and f are conjugate to each other, the relations between them are symmetric. Thus we have $\{f'_-(y) \le x\} = \{h'_+(x) \ge y\}$ or, equivalently, $\{f'_-(y) > x\} = \{h'_+(x) < y\}$, which implies (2.7).

Before proving Theorem 2.1 we need two lemmas.

Lemma 2.5. *Let* $S = \operatorname{argmax}_D \Phi$ *and* $\widehat{S} = \operatorname{argmax}_D \widehat{\Phi}$ *be the maximal superlevel sets of* Φ *and* $\widehat{\Phi}$*. Then the set* $\widehat{S}$ *is defined if and only if the set* S *is defined and, in this case,* $\operatorname{conv}(S) \subseteq \widehat{S}$*.*

Lemma 2.6. *If* $\operatorname{conv}(\operatorname{hypo}(\Phi))$ *is a closed convex set then* $\operatorname{conv}(S) = \widehat{S}$*.*

Proof of Lemma 2.5. Since $\mathrm{cl}(\Phi) \le \widehat{\Phi}$ the set S is defined if $\widehat{S}$ is defined. On the other hand, if S is defined then Φ is bounded from above on D. Since

$$\sup_D \Phi = \sup_D \widehat{\Phi},$$

the function $\widehat{\Phi}$ is also bounded from above on D, i.e. the set $\widehat{S}$ is defined.

By (2.10) we have $S \subseteq \widehat{S}$. Since Φ and $\widehat{\Phi}$ are upper semicontinuous the sets S and $\widehat{S}$ are closed. Since $\widehat{S}$ is convex we have $\mathrm{conv}(S) \subseteq \widehat{S}$.

Proof of Lemma 2.6. Indeed, we have $\mathrm{conv}(\mathrm{hypo}(\Phi)) \equiv \mathrm{conv}(\mathrm{cl}(\mathrm{hypo}(\Phi)))$, and $\mathrm{conv}(\mathrm{hypo}(\Phi)) \subseteq \mathrm{hypo}(\widehat{\Phi})$. Therefore $\mathrm{conv}(\mathrm{hypo}(\Phi))$ is a hypograph of some closed concave function H such that $\Phi \le H \le \widehat{\Phi}$. Thus $H = \widehat{\Phi}$. The set $\widehat{S}$ is a face of $\mathrm{hypo}(\widehat{\Phi})$ and the set $\mathrm{conv}(S)$ is a face of $\mathrm{conv}(\mathrm{hypo}(\Phi))$. The statement now follows from Rockafellar (1970, Thm. 18.3).

Proof of Theorem 2.1. To prove the first statement, start with $\widehat{\Phi}$ proper. We have

$$\mathrm{hypo}(\Phi) \subseteq \mathrm{hypo}(\mathrm{cl}(\Phi)) \equiv \mathrm{cl}(\mathrm{hypo}(\Phi)) \subseteq \mathrm{cl}(\mathrm{conv}(\mathrm{hypo}(\Phi))) \equiv \mathrm{hypo}(\widehat{\Phi}), \quad (2.10)$$

and therefore $\mathrm{hypo}(\Phi)$ is bounded by any support plane of $\mathrm{hypo}(\widehat{\Phi})$. This implies that there exists a linear function l such that $\Phi \le l$.

Now suppose that there is a linear function l such that $\Phi \le l$ on D. Then $\mathrm{cl}(\Phi) \le l$ and, from (2.10), we have $\mathrm{hypo}(\Phi) \subseteq \mathrm{hypo}(l)$, $\mathrm{conv}(\mathrm{hypo}(\Phi)) \subseteq \mathrm{hypo}(l)$, and $\mathrm{hypo}(\widehat{\Phi}) \equiv \mathrm{cl}(\mathrm{conv}(\mathrm{hypo}(\Phi))) \subseteq \mathrm{hypo}(l)$. Thus $\widehat{\Phi} < +\infty$ on D. Since $\mathrm{hypo}(\Phi) \subseteq \mathrm{hypo}(\widehat{\Phi})$ there exists a finite point in $\mathrm{hypo}(\widehat{\Phi})$.

To show that the two switching relations hold, first consider the convex function $h = -\widehat{\Phi}$. Then $\widehat{\phi}_L(x) = -h'_-(x), \widehat{\phi}_R(x) = -h'_+(x), \kappa_L(y) = s_-(-y)$, and $\kappa_R(y) = s_+(-y)$. By the properness of $\widehat{\Phi}$ proved above and Proposition 2.4, it suffices to show that

$$\mathrm{argmax}^L_x(\Phi(x) - yx) = \mathrm{argmax}^L_x(\widehat{\Phi}(x) - yx),$$
$$\mathrm{argmax}^R_x(\Phi(x) - yx) = \mathrm{argmax}^R_x(\widehat{\Phi}(x) - yx).$$

To accomplish this, it suffices, without loss of generality, to prove the equalities in the last display when $y = 0$, and this in turn follows if we relate the maximal superlevel sets of Φ and $\widehat{\Phi}$. This follows from Lemmas 2.5 and 2.6.

Remark 2.7. Note that $\mathrm{conv}(S) \ne \widehat{S}$ in general. To see this, consider the function

$$\Phi(x) = \begin{cases} 0 & x \ne 0, \\ 1 & x = 0. \end{cases}$$

We have that Φ is upper-semicontinuous, $S = \{0\}$ and $\widehat{\Phi} \equiv 1$, so $\widehat{S} = \mathbb{R}$.

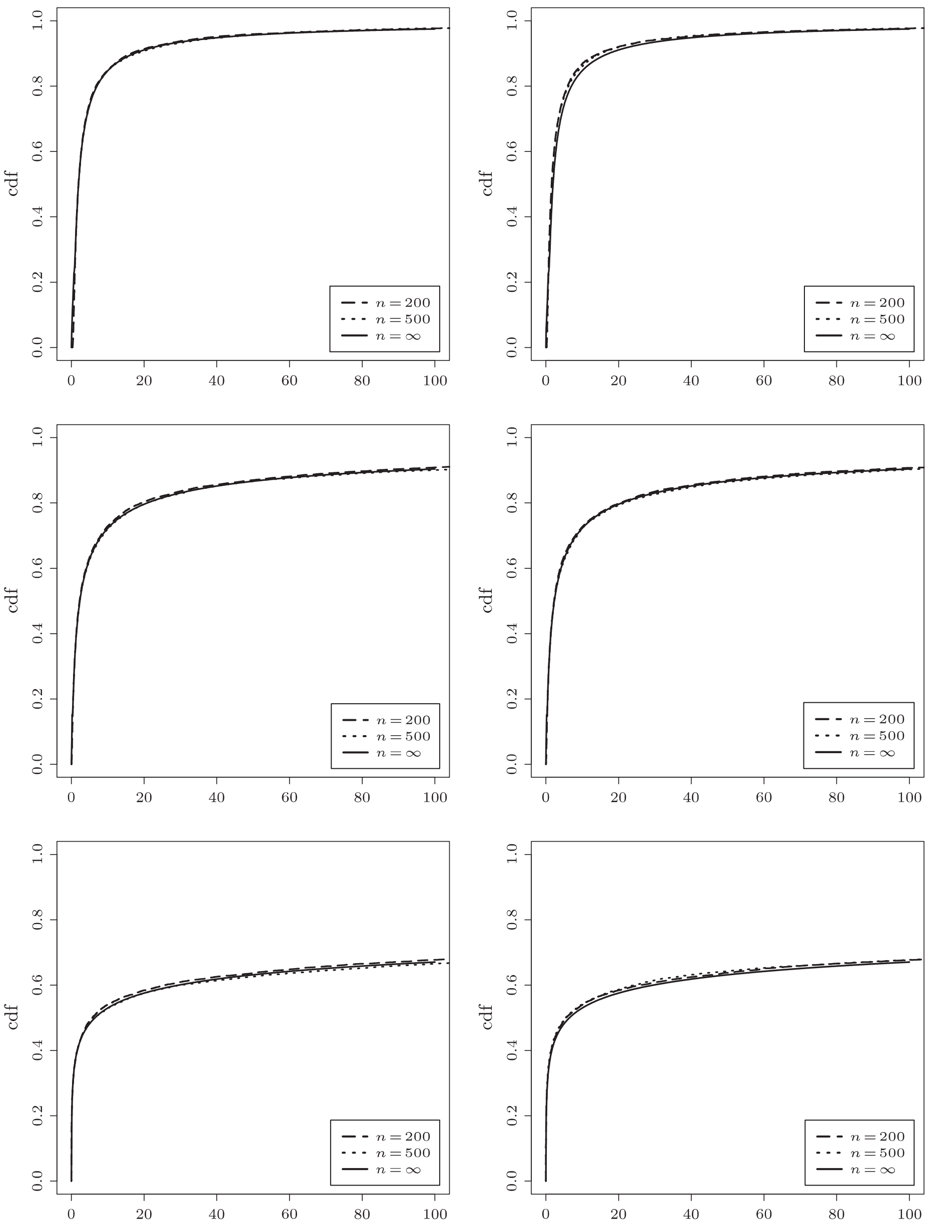

Figure 2. Empirical distributions of the re-scaled MLE at zero when sampling from the Beta distribution (left) and the Gamma distribution (right): from top to bottom we have $\alpha = 0.2, 0.5, 0.8$.

Remark 2.8. Note that if conv(hypo(Φ)) is a polyhedral set, then it is closed (see e.g., Rockafellar (1970, Corollary 19.1.2)). This is the case in our applications.

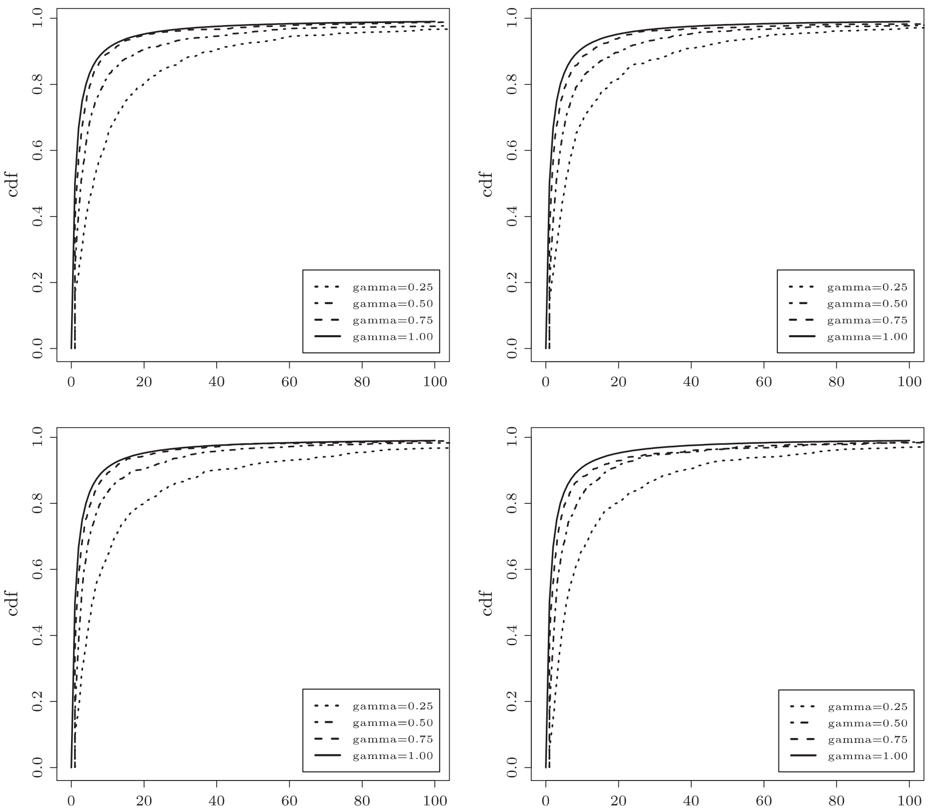

Figure 3. Empirical distributions of the supremum measure: the cutoff values shown are $c = 5$ (top left), $c = 25$ (top right), $c = 100$ (bottom left), $c = 1{,}000$ (bottom right).

3. Some Numerical Results

Figure 2 gives plots of the empirical distributions of $m = 10{,}000$ Monte Carlo samples from the distributions of $\widehat{f}_n(0)/(C_2 n^{\alpha}/(1-\alpha))^{1/(1-\alpha)})$ when $n = 200$ and $n = 500$, together with the limiting distribution function obtained in (1.9). The true density f_0 on the right side in Figure 2 is

$$f_0(x) = \int_0^{\infty} \frac{1}{y} 1_{[0,y]}(x) \frac{y^{c-1}}{\Gamma(c)} \exp(-y) dy. \tag{3.1}$$

For $c \in (0, 1)$, this family satisfies (G2) with $\alpha = 1 - c$ and $C_2 = 1/(\alpha\Gamma(1-\alpha))$. (Note that for $c = 1$, $f_0(x) \sim \log(1/x)$ as $x \searrow 0$.)

The true density f_0 on the left side in Figure 2 is

$$f_0(x) = \frac{1}{\mathrm{Beta}(1-a, 2)} x^{-a} (1-x) 1_{(0,1]}(x). \tag{3.2}$$

Table 1. Simulation of (3.4) for different values of γ and c.

	$c = 0.5$	$c = 5$	$c = 25$	$c = 100$	$c = 1{,}000$
$\gamma = 0.25$	0.361	0.171	0.140	0.092	0.060
$\gamma = 0.50$	0.422	0.249	0.190	0.162	0.148
$\gamma = 0.75$	0.489	0.387	0.349	0.358	0.367

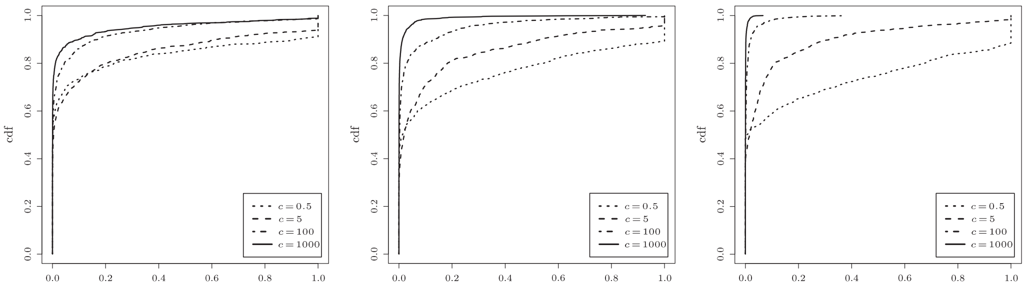

Figure 4. Empirical distributions of the location where the supremum occurs: from left to right we have $\gamma = 0.25, 0.50, 0.75$. Recall that for $\gamma = 1$, the (non-unique) location of the supremum is always zero by Corollary 1.2. The data were re-scaled to lie within the interval $[0, 1]$.

For $a \in [0, 1)$, this family satisfies (G2) with $\alpha = a$ and $C_2 = 1/\text{Beta}(1 - \alpha, 2)$.

Figure 3 shows simulations of the limiting distribution

$$\sup_{0 \le t \le c} \left| t^{1-\gamma} \frac{\widehat{h}(t)}{\gamma} - 1 \right| \tag{3.3}$$

for different values of c and γ. Recall that if $\gamma = 1$ the supremum occurs at $t = 0$ regardless of the value of c, and the limiting distribution (3.3) has cumulative distribution function $1 - 1/(x + 1)$. However, for $\gamma < 1$, the distribution of (3.3) depends both on γ and on c, although the dependence on c is not visually prominent in Figure 3. Table 1 shows estimated values of

$$P \left(\sup_{0 \le t \le c} \left| t^{1-\gamma} \frac{\widehat{h}(t)}{\gamma} - 1 \right| = 1 \right) \tag{3.4}$$

for different c and $\gamma < 1$, which clearly depends on the cutoff value c (upper bound on the standard deviation in each case is 0.016). Note that (3.3) is equal to one if the location of the supremum occurs at $t = 0$ (with probability one).

Cumulative distribution functions for the location of the supremum in (3.3) are shown in Figure 4; these depend on both γ and c.

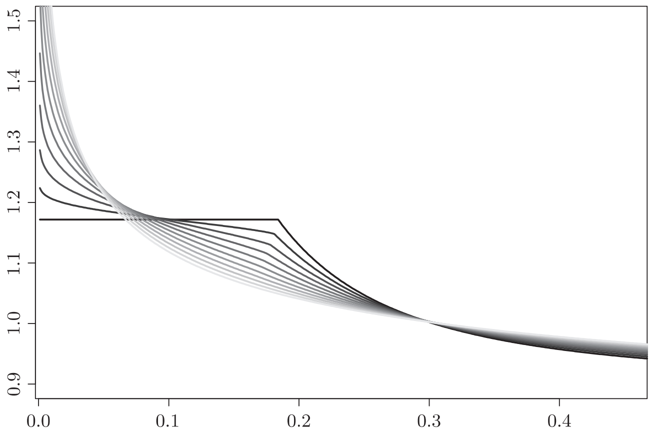

Figure 5. Generalized Gaussian (or Subbotin) mixture densities with $\epsilon = .1$, $\mu = 1$, $r \in \{1.0, 1.2, \ldots, 2.0\}$ (black to light grey, respectively) as given by (4.1).

4. Application to Mixtures

4.1. Behavior near zero

Suppose $X_1, \ldots, X_n$ are i.i.d. with distribution function F, where

$$\text{under } H_0: \ F = \Phi_r, \qquad \text{the generalized normal distribution,}$$
$$\text{under } H_1: \ F = (1-\epsilon)\Phi_r + \epsilon\Phi_r(\cdot - \mu), \quad \epsilon \in (0,1), \quad \mu > 0,$$

where $\Phi_r(x) \equiv \int_{-\infty}^{x} \phi_r(y)dy$ with $\phi_r(y) \equiv \exp(-|y|^r/r)/C_r$ for $r > 0$ gives the generalized normal (or Subbotin) distribution; here $C_r \equiv 2\Gamma(1/r)r^{(1/r)-1}$ is the normalizing constant. If we transform to $Y_i \equiv 1 - \Phi_r(X_i) \sim G$, then, for $0 \le y \le 1$,

$$\text{under } H_0: G(y) = y, \qquad \text{the Uniform}(0,1) \text{ d.f.},$$
$$\text{under } H_1: G(y) = G_{\epsilon,\mu,r}(y) = (1-\epsilon)y + \epsilon(1 - \Phi_r(\Phi_r^{-1}(1-y) - \mu)).$$

Let $g_{\epsilon,\mu,r}$ denote the density of $G_{\epsilon,\mu,r}$; thus

$$g_{\epsilon,\mu,r}(y) = 1 - \epsilon + \epsilon \exp\left\{-\frac{1}{r}\left(|\Phi_r^{-1}(1-y) - \mu|^r - |\Phi_r^{-1}(1-y)|^r\right)\right\}. \quad (4.1)$$

It is easily seen that $g_{\epsilon,\mu,r}$ is monotone decreasing on $(0,1)$ and is unbounded at zero if $r > 1$. Figure 5 shows plots of these densities for $\epsilon = .1$, $\mu = 1$, and $r \in \{1.0, 1.1, \ldots, 2.0\}$. Note that $g_{\epsilon,\mu,1}$ is bounded at 0, in fact $g_{\epsilon,\mu,1}(y) = 1 - \epsilon + \epsilon e^{\mu}$ for $0 \le y \le 2^{-1}e^{-\mu}$.

Proposition 4.1. *The distribution* $F_{\mu,r}(y) \equiv 1 - \Phi_r(\Phi_r^{-1}(1-y) - \mu)$ *is regularly varying at* 0 *with exponent* 1*. That is, for any* $c > 0$*,*

$$\lim_{y \to 0+} \frac{F_{\mu,r}(cy)}{F_{\mu,r}(y)} = c.$$

Proof. Let $\kappa_r(y) = \Phi_r^{-1}(1-y)$. Our first goal is to show that

$$\lim_{y \to 0} \frac{\kappa_r(y)}{\tilde{\kappa}_r(y)} = 1, \tag{4.2}$$

where (for y small)

$$\tilde{\kappa}_r(y) = \left(-r \log \left(C_r \; y \; \left\{ r \log \left(\frac{1}{C_r y} \right) \right\}^{(r-1)/r} \right) \right)^{1/r}.$$

To prove (4.2), it is enough to show that

$$\lim_{y \to 0} \tilde{\kappa}_r(y)^{r-1} (\kappa_r(y) - \tilde{\kappa}_r(y)) = 0. \tag{4.3}$$

This result follows from de Haan and Ferreira (2006, Thm. 1.1.2). Let $b_n = \tilde{\kappa}_r(1/n)$, $a_n = 1/b_n^{r-1}$, and choose $F = \Phi_r$ in the statement of Theorem 1.1.2. Then, if we can show that

$$n(1 - \Phi_r(a_n x + b_n)) \to \log G(x) \equiv e^{-x}, \qquad x \in \mathbb{R}, \tag{4.4}$$

it would follow from de Haan and Ferreira (2006, Thm. 1.1.2 and Sec. 1.1.2) that for all $x \in \mathbb{R}$,

$$\lim_{y \to 0} \frac{U(x/y) - b_{\lfloor 1/y \rfloor}}{a_{\lfloor 1/y \rfloor}} = G^{-1}(e^{-1/x}) = \log(1/x),$$

where $U(t) = (1/(1-\Phi_r))^{-1}(t) = \Phi_r^{-1}(1 - 1/t)$. Choosing $x = 1$ yields (4.3).

To prove (4.4), we make use of the following, a generalization of Mills' ratio to the generalized Gaussian family,

$$1 - \Phi_r(z) \sim \frac{\phi_r(z)}{z^{r-1}} \qquad \text{as} \quad z \to \infty. \tag{4.5}$$

This follows from l'Hôpital's rule:

$$\begin{aligned} \lim_{z \to \infty} \frac{\int_z^\infty \phi_r(y) dy}{z^{1-r} \phi_r(z)} &= \lim_{z \to \infty} \frac{-\phi_r(z)}{(1-r) z^{-r} \phi_r(z) + z^{1-r} \phi_r(z)(-z^{r-1})} \\ &= \lim_{z \to \infty} \frac{1}{1 - (1-r) z^{-r}} = 1. \end{aligned}$$

Now,

$$
\begin{aligned}
n(1-\Phi_r(a_nx+b_n)) &\sim n\frac{\phi_r(a_nx+b_n)}{(a_nx+b_n)^{r-1}} \\
&= \frac{n}{C_r b_n^{r-1}} \frac{\exp\left(-(b_n^r/r)\,(1+a_nx/b_n)^r\right)}{(1+a_nx/b_n)^{r-1}} \\
&\sim \frac{n}{C_r b_n^{r-1}} \exp\left(-\frac{b_n^r}{r}\left(1+\frac{rx}{b_n^r}\right)\right) \\
&= \exp\left(-\left(\frac{b_n^r}{r}+(r-1)\log b_n-\log n+\log C_r\right)\right)\exp(-x) \\
&\to \exp(-0)\cdot\exp(-x)
\end{aligned}
$$

by using the definition of b_n. We have thus shown that (4.2) holds. Then, for $y \to 0$, by (4.5) and (4.2),

$$
\begin{aligned}
F_{\mu,r}(y) &= 1-\Phi_r(\kappa_r(y)-\mu) \sim 1-\Phi_r(\tilde{\kappa}_r(y)-\mu) \\
&\sim \frac{\phi_r(\tilde{\kappa}_r(y)-\mu)}{(\tilde{\kappa}_r(y)-\mu)^{r-1}}.
\end{aligned}
$$

Plugging in the definition of ϕ_r, we find that

$$
\begin{aligned}
F_{\mu,r}(y) &\sim \frac{1/C_r}{(\tilde{\kappa}_r(y)-\mu)^{r-1}} \exp\left(-\frac{\tilde{\kappa}_r(y)^r}{r}\left|1-\frac{\mu}{\tilde{\kappa}_r(y)}\right|^r\right) \\
&= \frac{1/C_r}{(\tilde{\kappa}_r(y)-\mu)^{r-1}} \exp\left\{\left(\log(C_r y)+\log(r\log(\frac{1}{C_r y}))\right)\left|1-\frac{\mu}{\tilde{\kappa}_r(y)}\right|^r\right) \\
&= \frac{1/C_r}{(\tilde{\kappa}_r(y)-\mu)^{r-1}}\ (C_r y)^{|1-\mu/\tilde{\kappa}_r(y)|^r}\cdot\left\{r\log\frac{1}{C_r y}\right\}^{[(r-1)/r]|1-\mu/\tilde{\kappa}_r(y)|^r}.
\end{aligned}
$$

Note that $\lim_{y\to 0}\tilde{\kappa}_r(cy)/\tilde{\kappa}_r(y)=1$. Therefore,

$$
\begin{aligned}
\frac{F_{\mu,r}(cy)}{F_{\mu,r}(y)} &\sim c^{|1-\mu/\tilde{\kappa}_r(cy)|^r}\cdot(C_r y)^{|1-\mu/\tilde{\kappa}_r(cy)|^r-|1-\mu/\tilde{\kappa}_r(y)|^r}\cdot\left(\frac{\tilde{\kappa}_r(y)-\mu}{\tilde{\kappa}_r(cy)-\mu}\right)^{r-1} \\
&\quad\cdot\frac{\left\{r\log\frac{1}{C_r cy}\right\}^{[(r-1)/r]|1-\mu/\tilde{\kappa}_r(cy)|^r}}{\left\{r\log\frac{1}{C_r y}\right\}^{[(r-1)/r]|1-\mu/\tilde{\kappa}_r(y)|^r}} \\
&\to c\cdot 1\cdot 1\cdot 1 = c.
\end{aligned}
$$

Thus (1.5) holds with $\gamma = 1$.

By the theory of regular variation (see e.g., Bingham, Goldie and Teugels (1989, p. 21)), $F_{\mu,r}(y) = y\ell(y)$ where ℓ is slowly varying at 0. It then follows easily that (1.5) holds for $F_0 = G_{\epsilon,\mu,r}$ with exponent 1. Thus the theory of Section 1 applies with a_n of Theorem 1.1 taken to be $a_n = G_{\epsilon,\mu,\gamma}(1/n)$; i.e.

$$\frac{1}{n} = G_{\epsilon,\mu,r}(a_n) = (1-\epsilon)a_n + \epsilon F_{\mu,r}(a_n) \doteq \epsilon F_{\mu,r}(a_n),$$

where the last approximation is valid for $r > 1$, but not for $r = 1$. When $r = 1$, the first equality can be solved explicitly, and we find

$$a_n = \begin{cases} 1 - \Phi_r(\Phi_r^{-1}(1-(\frac{1}{n\epsilon}) + \mu), \text{when} \quad r > 1, \\ n^{-1}(1-\epsilon+\epsilon e^{\mu})^{-1}, \qquad \text{when} \quad r = 1. \end{cases} \tag{4.6}$$

We conclude that Theorem 1.1 holds for a_n as in the last display, where $\widehat{f}_n$ is the Grenander estimator of $g_{\epsilon,\mu,r}$ based on $Y_1, \ldots, Y_n$.

Another interesting mixture family is as follows: suppose that Φ_1, Φ_2 are two fixed distribution functions, then

$$\begin{aligned} &\text{under } H_0: \ F = \Phi_1, \\ &\text{under } H_1: \ F = (1-\epsilon)\Phi_1 + \epsilon\Phi_2, \quad \epsilon \in (0,1). \end{aligned}$$

Using $Y_i \equiv 1 - \Phi_1(X_i) \sim G$, then, for $0 \le y \le 1$, we find that under H_1 the distribution of the Y_i's is given by

$$\begin{aligned} G(y) &= (1-\epsilon)y + \epsilon(1 - \Phi_2(\Phi_1^{-1}(1-y))), \\ g(y) &= (1-\epsilon) + \epsilon\frac{\phi_2(\Phi_1^{-1}(1-y))}{\phi_1(\Phi_1^{-1}(1-y))}. \end{aligned}$$

For Φ_2 given in terms of Φ_1 by the (Lehmann alternative) distribution function $\Phi_2(y) = 1 - (1 - \Phi_1(y))^{\gamma}$, this becomes

$$G(y) = (1-\epsilon)y + \epsilon y^{\gamma} \quad \text{and} \quad g(y) = (1-\epsilon) + \epsilon\gamma y^{\gamma-1}.$$

When $0 < \gamma < 1$ this family fits into the framework of our condition **G2** with $\alpha = 1 - \gamma$ and $C_2 = \epsilon\gamma$.

4.2. Estimation of the contaminating density

Suppose that $G_{\epsilon,F}(y) = (1-\epsilon)y + \epsilon F(y)$ where F is a concave distribution on $[0,1]$ with monotone decreasing density f. Thus the density $g_{\epsilon,F}$ of $G_{\epsilon,F}$ is given by $g_{\epsilon,F}(y) = (1-\epsilon) + \epsilon f(y)$. Note that $g_{\epsilon,F}$ is also monotone decreasing, and $g_{\epsilon,F}(y) \ge 1 - \epsilon + \epsilon f(1) = 1 - \epsilon = g_{\epsilon,F}(1)$ if $f(1) = 0$. For $\epsilon > 0$ we can write

$$f(y) = \frac{g_{\epsilon,F}(y) - (1-\epsilon)}{\epsilon}.$$

If $Y_1, \ldots, Y_n$ are i.i.d. $g_{\epsilon,F}$, then we can estimate $g_{\epsilon,F}$ by the Grenander estimator $\widehat{g}_n$, and we can estimate ϵ by $\widehat{\epsilon}_n = 1 - \widehat{g}_n(1)$. This results in an estimator of the contaminating density f,

$$\widehat{f}_n(y) = \frac{\widehat{g}_n(y) - (1 - \widehat{\epsilon}_n)}{\widehat{\epsilon}_n} = \frac{\widehat{g}_n(y) - \widehat{g}_n(1)}{1 - \widehat{g}_n(1)},$$

which is quite similar in spirit to a setting studied by Swanepoel (1999). Here, however, we propose using the shape constraint of monotonicity, and hence the Grenander estimator, to estimate both ϵ and f. We will study this estimator elsewhere.

Appendix A: Proofs for Section 1

For the proof of Theorem 1.1 we need two lemmas. Together, they show that argmax^R and argmax^L are continuous. We assume that (1.5) holds and that $nF_0(a_n) \sim 1$. Thus both (1.3) and (1.7) also hold.

Lemma A.1. (i) *When $\gamma = 1$ and $x > 1$, $argmax_v^{L,R}\{n\mathbb{F}_n(a_n v) - xv\} = O_p(1)$.*
(ii) *When $\gamma \in (0,1)$ and $x > 0$, $argmax_v^{L,R}\{n\mathbb{F}_n(a_n v) - xv\} = O_p(1)$.*

Proof. It suffices to show that $\limsup_{n\to\infty} P(\sup_{v\geq K}\{n\mathbb{F}_n(a_n v) - xv\} \geq 0) \to 0$, as $K \to \infty$ under the conditions specified. Let $h(x) = x(\log x - 1) + 1$ and recall the inequality

$$P(\mathrm{Bin}(n,p)/(np) \geq t) \leq \exp(-nph(t))$$

for $t \geq 1$, where $\mathrm{Bin}(n,p)$ denotes a Binomial(n,p) random variable; see e.g. Shorack and Wellner (1986, p.415). It follows that

$$\begin{aligned}
&P(\sup_{v\geq K}\{n\mathbb{F}_n(a_n v) - xv\} \geq 0) \\
&\quad = P(\cup_{j=K}^{\infty}\{n\mathbb{F}_n(a_n v) - xv \geq 0 \ \text{ for some } \ v \in [j, j+1)\}) \\
&\quad \leq \sum_{j=K}^{\infty} P(n\mathbb{F}_n(a_n(j+1)) - xj \geq 0) \\
&\quad = \sum_{j=K}^{\infty} P\left(\frac{n\mathbb{F}_n(a_n(j+1))}{nF_0(a_n(j+1))} \geq \frac{xj}{nF_0(a_n(j+1))}\right) \\
&\quad \leq \sum_{j=K}^{\infty} \exp\left(-nF_0(a_n(j+1))h\left(\frac{xj}{nF_0(a_n(j+1))}\right)\right). \qquad \text{(A.1)}
\end{aligned}$$

Next, since F_0 is concave,

$$nF_0(a_n(j+1)) \leq nF_0(a_n(K+1))\frac{j+1}{K+1}$$

for $j \geq K$ and $nF_0(a_n(K+1)) \to (K+1)^\gamma$ and $n \to \infty$. Therefore, for all $j \geq K$ and sufficiently large n, we have

$$\frac{xj}{nF_0(a_n(j+1))} \geq \delta(K+1)^{1-\gamma}\frac{xj}{j+1}$$

for any fixed $\delta < 1$. We need to handle the two cases $\gamma = 1$ and $\gamma < 1$ separately. Note that if $\gamma < 1$, then the above display shows that K, n can be chosen sufficiently large so that $(xj)/nF_0(a_n(j+1))$ is uniformly large. On the other hand if $\gamma = 1$ and $x > 1$, then we can pick δ, K, n large enough so that $(xj)/nF_0(a_n(j+1))$ is strictly greater than $1+\epsilon$ for some $\epsilon > 0$, again uniformly in j.

Suppose first that $\gamma < 1$. Then for K, n large, since $h(x) \sim x \log x$ as $x \to \infty$, there exists a constant $0 < C < 1$ such that for all $j \geq K$

$$\begin{aligned} nF_0(a_n(j+1))h\left(\frac{xj}{nF_0(a_n(j+1))}\right) &\geq C(xj)\log\left(\frac{xj}{j+1}\right) \\ &\geq C_x(xj), \end{aligned}$$

for some other constant $C_x > 0$. This shows that the sum in (A.1) converges to zero as $K \to \infty$, as required.

Suppose next that $\gamma = 1$. Note that the function $h(x) > 0$ for $x > 1$. Therefore, combining our arguments above, we find that for all $j \geq K$

$$\begin{aligned} nF_0(a_n(j+1))h\left(\frac{xj}{nF_0(a_n(j+1))}\right) &\geq \delta(j+1)h\left(\frac{xj}{nF_0(a_n(j+1))}\right) \\ &\geq C_{x,\delta}(j+1), \end{aligned}$$

again for some $C_{x,\delta} > 0$. This again implies that the sum in (A.1) converges to zero as $K \to \infty$, and completes the proof.

Lemma A.2. *Suppose that* $\gamma \in (0,1]$. *Then*

$$V_x^L \equiv \operatorname*{argmax}^L_v \{\mathbb{N}(v^\gamma) - xv\} = \operatorname*{argmax}^R_v \{\mathbb{N}(v^\gamma) - xv\} \equiv V_x^R \qquad a.s..$$

Proof. Suppose that $V_x^L < V_x^R$. Then it follows that $\mathbb{N}((V_x^L)^\gamma) - xV_x^L = \mathbb{N}((V_x^R)^\gamma) - xV_x^R$ or, equivalently,

$$\mathbb{N}((V_x^R)^\gamma) - \mathbb{N}((V_x^L)^\gamma) = x\{V_x^R - V_x^L\}.$$

Now $(V_x^R)^\gamma$, $(V_x^L)^\gamma \in J(\mathbb{N}) \equiv \{t > 0 : \ \mathbb{N}(t) - \mathbb{N}(t-) \geq 1\}$, so the left side here takes values in the set $\{1, 2, \ldots\}$ while the right side takes values in $x \cdot \{r^{1/\gamma} - s^{1/\gamma} : r, s \in J(\mathbb{N}), r > s\}$. But it is well-known that all the (joint) distributions of the points in $J(\mathbb{N})$ are absolutely continuous with respect to Lebesgue measure, and hence the equality in the last display holds only for sets with probability 0.

Proof of Theorem 1.1. We first prove convergence of the one-dimensional distributions of $na_n\widehat{f}_n(a_nt)$. Fix $K > 0$, and let $x > 1_{\{\gamma=1\}}$ and $t \in (0, K]$. By the switching relation (2.3),

$$\begin{aligned} P(na_n\widehat{f}_n(a_nt) \le x) &= P(\widehat{s}_n^L(\frac{x}{na_n}) \le a_nt) \\ &= P(\mathrm{argmax}_s^L\{\mathbb{F}_n(s) - \frac{xs}{na_n}\} \le a_nt) \\ &= P(\mathrm{argmax}_v^L\{\mathbb{F}_n(va_n) - x(\frac{v}{n})\} \le t) \\ &= P(\mathrm{argmax}_v^L\{n\mathbb{F}_n(va_n) - xv\} \le t) \\ &\to P(\mathrm{argmax}_v^L\{\mathbb{N}(v^\gamma) - xv\} \le t) \\ &= P(\widehat{h}_\gamma(t) \le x), \end{aligned}$$

where the convergence follows from (1.7), and the argmax continuous mapping theorem for $D[0,\infty)$ applied to the processes $\{v \mapsto n\mathbb{F}_n(va_n) - xv : \ v \ge 0\}$; see e.g. Ferger (2004, Thm. 3 and Corollary 1). Note that Lemma A.1 yields the $O_p(1)$ hypothesis of Ferger's Corollary 1, while Lemma A.2 shows that equality holds in the limit.

Convergence of the finite-dimensional distributions of $\widehat{h}_n(t) \equiv na_n\widehat{f}_n(a_nt)$ follows in the same way by using the process convergence in (1.7) for finitely many values $(t_1, x_1), \ldots, (t_m, x_m)$, where each $t_j \in \mathbb{R}^+$ and $x_j > 1_{\{\gamma=1\}}$.

To verify tightness of $\widehat{h}_n$ in $D[0,\infty)$, we use Billingsley (1999, Thm. 16.8). Thus, it is sufficient to show that for any $K > 0$, and any $\epsilon > 0$,

$$\lim_{M\to\infty} \limsup_n P\left(\sup_{0\le t\le K} |\widehat{h}_n(t)| \ge M\right) = 0, \tag{A.2}$$

$$\lim_{\delta\to 0} \limsup_n P\left(w_{\delta,K}(\widehat{h}_n) \ge \epsilon\right) = 0. \tag{A.3}$$

Here $w_{\delta,K}(h)$ is the modulus of continuity in the Skorohod topology,

$$w_{\delta,K}(h) = \inf_{\{t_i\}_r} \max_{0<i\le r} \sup\left\{|h(t) - h(s)| : s, t \in [t_{i-1}, t_i) \cap [0, K]\right\},$$

where $\{t_i\}_r$ is a partition of $[0, K]$ such that $0 = t_0 < t_1 < \ldots < t_r = K$ and $t_i - t_{i-1} > \delta$. Suppose then that h is a piecewise constant function with discontinuities occurring at the (ordered) points $\{\tau_i\}_{i\ge 0}$. Then if $\delta \le \inf_i |\tau_i - \tau_{i-1}|$ we necessarily have that $w_{\delta,K}(h) = 0$.

First, note that since $\widehat{h}_n$ is non-increasing,

$$\|\widehat{h}_n\|_0^m \equiv \sup_{0\le t\le m} |\widehat{h}_n(t)| = \widehat{h}_n(0),$$

and hence (A.2) follows from the finite-dimensional convergence proved above.

Next, fix $\epsilon > 0$. Let $0 = \tau_{n,0} < \tau_{n,1} < \cdots < \tau_{n,K_n} < K$ denote the (ordered) jump points of $\widehat{h}_n$, and let $0 = T_{n,0} < T_{n,1} < \cdots < T_{n,J_n} < K$ denote the (again, ordered) jump points of $n\mathbb{F}_n(a_n t)$. Because $\{\tau_{n,1}, \ldots, \tau_{n,K_n}\} \subset \{T_{n,1}, \ldots, T_{n,J_n}\}$, it follows that $\inf\{\tau_{i,n} - \tau_{i-1,n}\} \ge \inf\{T_{i,n} - T_{i-1,n}\}$, and hence

$$P\left(w_{\delta,K}(\widehat{h}_n) \ge \epsilon\right) \le P\left(\inf_{i=1,\ldots,J_n}\{T_{i,n} - T_{i-1,n}\} < \delta\right).$$

Now, by (1.7) and continuity of the inverse map (see e.g., Whitt (2002, Thm. 13.6.3))

$$(T_{n,1}, \ldots, T_{n,J_n}, 0, 0, \ldots) \Rightarrow (T_1^{1/\gamma}, \ldots, T_J^{1/\gamma}, 0, 0, \ldots),$$

where $T_1, \ldots, T_J$ denote the successive arrival times on $[0, K]$ of a standard Poisson process. Thus

$$\lim_{\delta\to 0} P\left(\inf_{i=1,\ldots,J}\{T_i^{1/\gamma} - T_{i-1}^{1/\gamma}\} < \delta\right) = 0,$$

and therefore (A.3) holds. This completes the proof of (i).

To prove (ii), fix $0 < c < \infty$. Write

$$\sup_{0<x\le ca_n}\left|\frac{\widehat{f}_n(x)}{f_0(x)} - 1\right| = \sup_{0<t\le c}\left|\frac{na_n\widehat{f}_n(ta_n)}{na_n f_0(ta_n)} - 1\right|. \tag{A.4}$$

Suppose we could show that the ratio process $na_n\widehat{f}_n(a_n t)/na_n f_0(a_n t)$ converges to the process $t^{1-\gamma}\widehat{h}_\gamma(t)/\gamma$ in $D[0,\infty)$. Then the conclusion follows by noting that the functional $h \mapsto \sup_{0<t\le c}|h|$ is continuous in the Skorohod topology as long as c is not a point of discontinuity of h (Jacod and Shiryaev (2003, Prop. VI 2.4)). Since $\mathbb{N}(t^\gamma)$ is stochastically continuous (i.e. $P(\mathbb{N}(t^\gamma) - \mathbb{N}(t^\gamma -) > 0) = 0$ for each fixed $t > 0$), $t^{1-\gamma}\widehat{h}_\gamma(t)/\gamma$ is almost surely continuous at c.

It remains to prove convergence of the ratio. Fix $K > c$, and again we may assume that K is a continuity point. Consider the term in the denominator, $na_n f_0(a_n t)$: it follows from (1.4) that $g_n(t) \equiv (na_n f_0(a_n t))^{-1} \to g(t) \equiv \gamma^{-1}t^{1-\gamma}$, where g is monotone increasing and uniformly continuous on $[0, K]$. Thus $g_n \to g$ in $C[0, K]$. Since the term in the numerator satisfies $h_n(t) \equiv na_n\widehat{f}_n(a_n t) \Rightarrow \widehat{h}_\gamma(t) \equiv h(t)$ in $D[0, K]$, it follows that $g_n h_n \Rightarrow gh$ in $D[0, K]$, as required. Here, we have again used the continuity of the supremum. This completes the proof of (ii).

Lemma A.3. *Suppose that* $a_n = p(1/n)$ *for some function with* $p(0) = 0$ *satisfying* $\lim_{x\to 0+} p'(x)f_0(p(x)) = 1$. *Then* $nF_0(a_n) \to 1$.

Proof. This follows easily from l'Hôpital's rule, since

$$\lim_{n\to\infty} nF_0(a_n) = \lim_{x\to 0+} \frac{F_0(p(x))}{x} = \lim_{x\to 0+} f_0(p(x))p'(x).$$

Proof of Corollary 1.2. Under the assumption **G0** we see that $F_0(x) \sim f_0(0+)x$ as $x \to 0$, so (1.5) holds with $\gamma = 1$. The claim that $a_n = 1/(nf_0(0+))$ satisfies $nF_0(a_n) \to 1$ follows from Lemma A.3 with $p(x) = x/f_0(0+)$. For (i) note that $\widehat{h}_1(0) = \widehat{h}_1(0+) = \sup_{t>0}(\mathbb{N}(t)/t)$, and the indicated equality in distribution follows from Pyke (1959); see Proposition 1.5 and its proof. (ii) follows directly from (i) of Theorem 1.1. To prove (iii), note that from (ii) of Theorem 1.1 that it suffices to show that

$$\sup_{0<t\le c} \left|\widehat{h}_1(t) - 1\right| = \left|\widehat{h}_1(0+) - 1\right| = \widehat{h}_1(0+) - 1 = Y_1 - 1 \tag{A.5}$$

for each $c > 0$, where $\widehat{h}_1(t)$ is the right derivative of the LCM of $\mathbb{N}(t)$. The equality in (A.5) holds if $\widehat{h}_1(c) > 1$, since $\widehat{h}_1$ is decreasing by definition. By the switching relation (2.3), we have the equivalence $\{\widehat{h}_1(c) > 1\} = \{\widehat{s}^L(1) > c\}$. The equality in (A.5) thus follows if $\widehat{s}^L(1) = \infty$. That is, if $\mathbb{N}(t) - t < \sup_{y\ge 0}\{\mathbb{N}(y) - y\}$ for all finite t. Let $W = \sup_{y\ge 0}\{\mathbb{N}(y) - y\}$. Pyke (1959, pp. 570-571) showed that $P(W \le x) = 0$ for $x \ge 0$, i.e. $P(W = \infty) = 1$.

Proof of Corollary 1.3. Under the assumption **G1** we see that $F_0(x) \sim C_1 x(\log(1/x))^\beta$ as $x \to 0$, so (1.5) holds with $\gamma = 1$. The claim that $a_n = 1/(C_1 n(\log n)^\beta)$ satisfies $nF_0(a_n) \to 1$ follows from Lemma A.3 with $p(x) = x/(C_1\log(1/x))^\beta$. For (i), note that $\widehat{h}_1(0) = \widehat{h}_1(0+) = \sup_{t>0}(\mathbb{N}(t)/t)$, as in the proof of Corollary 1.2. (ii) again follows directly from (i) of Theorem 1.1, and the proof of (iii) is the same as the proof of Corollary 1.2.

Proof of Corollary 1.4. Under the assumption **G2** we see that $F_0(x) \sim C_2 x^{1-\alpha}/(1-\alpha)$ as $x \to 0$, so (1.5) holds with $\gamma = 1 - \alpha$. The claim that $a_n = \{(1-\alpha)/(nC_2)\}^{1/(1-\alpha)}$ satisfies $nF_0(a_n) \to 1$ follows from Lemma A.3 with $p(x) = ((1-\alpha)x/C_2)^{1/(1-\alpha)}$. For (i), note that

$$\widehat{h}_{1-\alpha}(0) = \widehat{h}_{1-\alpha}(0+) = \sup_{t>0}\Big(\frac{\mathbb{N}(t^{1-\alpha})}{t}\Big) = \sup_{s>0}\Big(\frac{\mathbb{N}(s)}{s^{1/(1-\alpha)}}\Big),$$

much as in the proof of Corollary 1.2. (ii) and (iii) follow directly from (i) and (ii) of Theorem 1.1.

Proof of Proposition 1.5. The part of the proposition with $\gamma = 1$ follows from Pyke (1959, pp. 570-571); this is closely related to a classical result of Daniels (1945) for the empirical distribution function, see e.g. Shorack and Wellner (1986, Thm. 9.1.2).

The proof for the case $\gamma < 1$ proceeds along the lines of Mason (1983, pp. 103-105). Fix $x > 0$ and $\gamma < 1$. We aim to establish an expression for the distribution function of $Y_\gamma \equiv \sup_{s>0}(\mathbb{N}(s)/s^{1/\gamma})$ at $x > 0$. First, observe that

$$\begin{aligned} P(Y_\gamma \le x) &= P\left(\sup_{s>0}\left\{\frac{\mathbb{N}(s)}{s^{1/\gamma}}\right\} \le x\right) \\ &= P(\mathbb{N}(t) \le U(t) \quad \text{for all } t > 0), \end{aligned} \tag{A.6}$$

where the function $U(t) = xt^{1/\gamma}$. For $j \in \mathbb{N}$, let $t_j := (j/x)^\gamma$ and note that $t_1 < t_2 < \ \ldots$ and $U(t_j) = j$.

Let $B \equiv [\mathbb{N}(t_k) \ne k\,;$ for all $k \ge 1]$ and $C \equiv [\mathbb{N}(s) > U(s)\,;$ for some $s > 0]$. Then $P(B \cap C) = 0$ as a consequence of the following argument. Suppose that there exists some $t > 0$ and $k \in \mathbb{N}$ such that $k = \mathbb{N}(t) > U(t)$ and $\mathbb{N}(t_i) \ne i$, for all $i \ge 1$. It then follows that $t_k > t$, for otherwise $k = U(t_k) \le U(t)$, as $U(\cdot)$ is increasing, which is a contradiction. Therefore, $t_k > t$ implies that $\mathbb{N}(t_k) > \mathbb{N}(t) = k$, as $\mathbb{N}(\cdot)$ is non–decreasing and $\mathbb{N}(t_k) = k$ is disallowed by hypothesis. Hence $\mathbb{N}(t_i) > i$ holds for all $i \ge k$, for otherwise there would exist some $j \ge k$ such that $\mathbb{N}(t_j) = j$, since $\mathbb{N}(\cdot)$ is a counting process. Therefore, for each $i \ge k$ we have that $\mathbb{N}(s) \ge i + 1$ for all $t_i \le s \le t_{i+1}$ and, consequently, that $\mathbb{N}(s) \ge U(s)$ for all $s \ge t_k$. This implies that $B \cap C \subseteq [\liminf_{s\to\infty}\{\mathbb{N}(s)/s^{1/\gamma}\} \ge x]$ and therefore $P(B\cap C) = 0$, since the SLLN implies that $\mathbb{N}(s)/s^{1/\gamma} \to 0$ holds almost surely, for fixed $\gamma < 1$. Thus $P(B \cap C) = 0$.

Now $P(C) = P(C \cap B^c)$. Furthermore, since U is a strictly increasing function and since $\mathbb{N}$ has jumps at the points $\{t_k\}$ with probability zero, we also find that $P(C \cap B^c) = P(B^c)$. Finally, write $B^c = \cup_{k=1}^\infty A_k$ for the disjoint sets $A_k \equiv [\mathbb{N}(t_k) = k, \mathbb{N}(t_j) \ne j$ for all $1 \le j < k]$, $k \ge 1$. Combining the arguments above,

$$P(Y_\gamma \le x) = 1 - P(C) = 1 - \sum_{k=1}^\infty P(A_k),$$

where $P(A_1) = P(\mathbb{N}(t_1) = 1) = p(t_1; 1)$ and, for $k \ge 2$, $P(A_k)$ may be written as

$$\begin{aligned} &P(\mathbb{N}(t_k) = k) - P(\{\mathbb{N}(t_k) = k\} \cap \{\mathbb{N}(t_i) \ne i,\, i < k\}^c) \\ &= P(\mathbb{N}(t_k) = k) - \sum_{j=1}^{k-1} P(\mathbb{N}(t_k) = k,\, \mathbb{N}(t_j) = j,\, \mathbb{N}(t_i) \ne i,\, i < j) \\ &= P(\mathbb{N}(t_k) = k) - \sum_{j=1}^{k-1} P(\mathbb{N}(t_k) - \mathbb{N}(t_j) = k - j) P(\mathbb{N}(t_j) = j, \mathbb{N}(t_i) \ne i,\, i < j). \end{aligned}$$

The result follows.

Appendix B: Definitions from Convex Analysis

The *epigraph* (*hypograph*) of a function f from a subset S of $\mathbb{R}^d$ to $[-\infty, +\infty]$ is the subset $\text{epi}(f)$ ($\text{hypo}(f)$) of $\mathbb{R}^{d+1}$ defined by

$$\text{epi}(f) = \{(x,t) : x \in S, t \in \mathbb{R}, t \geq f(x)\},$$
$$\text{hypo}(f) = \{(x,t) : x \in S, t \in \mathbb{R}; t \leq f(x)\}.$$

The function f is *convex* if $\text{epi}(f)$ is a convex set. The *effective domain* of a convex function f on S is

$$\text{dom}(f) = \{x \in \mathbb{R}^d : (x,t) \in \text{epi}(f) \text{ for some } t\} = \{x \in \mathbb{R}^d : f(x) < \infty\}.$$

The t-sublevel set of a convex function f is the set $C_t = \{x \in \text{dom}(f) : f(x) \leq t\}$, and the t-superlevel set of a concave function g is the set $S_t = \{x \in \text{dom}(g) : g(x) \geq t\}$. The sets C_t, S_t are convex. The *convex hull* of a set $S \subset \mathbb{R}^d$, denoted by $\text{conv}(S)$, is the intersection of all the convex sets containing S.

A convex function f is said to be *proper* if its epigraph is non-empty and contains no vertical lines, i.e., if $f(x) < +\infty$ for at least one x and $f(x) > -\infty$ for every x. Similarly, a concave function g is *proper* if the convex function $-g$ is proper. The *closure of a concave function* g, denoted by $\text{cl}(g)$, is the pointwise infimum of all affine functions $h \geq g$. If g is proper, then $\text{cl}(g)(x) = \limsup_{y \to x} g(y)$. For every proper convex function f there exists closed proper convex function $\text{cl}(f)$ such that $\text{epi}(\text{cl}(f)) \equiv \text{cl}(\text{epi}(f))$. The *conjugate function* g^* of a concave function g is defined by $g^*(y) = \inf\{\langle x, y\rangle - g(x) : x \in \mathbb{R}^d\}$, and the conjugate function f^* of a convex function f is defined by $f^*(y) = \sup\{\langle x, y\rangle - f(x) : x \in \mathbb{R}^d\}$. If g is concave, then $f = -g$ is convex and f has conjugate $f^*(y) = -g^*(-y)$.

A *complete non-decreasing curve* is a subset of $\mathbb{R}^2$ of the form

$$\Gamma = \{(x,y) : x \in \mathbb{R}, y \in \mathbb{R}, \varphi_-(x) \leq y \leq \varphi_+(x)\}$$

for some non-decreasing function φ from $\mathbb{R}$ to $[-\infty, +\infty]$ that is not everywhere infinite. Here φ_+ and φ_- denote the right and left continuous versions of φ, respectively. A vector $y \in \mathbb{R}^d$ is said to be a *subgradient* of a convex function f at a point x if $f(z) \geq f(x) + \langle y, z - z\rangle$ for all $z \in \mathbb{R}^d$. The set of all subgradients of f at x is called the *subdifferential of f at x*, and is denoted by $\partial f(x)$.

A *face* of a convex set C is a convex subset B of C such that every closed line segment in C with a relative interior point in B has both endpoints in B. If B is the set of points where a linear function h achieves its maximum over C, then B is a face of C. If the maximum is achieved on the relative interior of a line segment $L \subset C$, then h must be constant on L and $L \subset B$. A face B of this type is called an *exposed face*.

References

Anevski, D. and Soulier, P. (2011). Monotone spectral density estimation. *Ann. Statist.* **39**, to appear. arXiv:0901.3471v1.

Beran, J. (1994). *Statistics for Long-Memory Processes*, Vol. 61 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, New York.

Billingsley, P. (1999). *Convergence of Probability Measures*. 2nd edition. Wiley, New York.

Bingham, N. H., Goldie, C. M. and Teugels, J. L. (1989). *Regular Variation*. Vol. 27 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge.

Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press, Cambridge.

Cox, D. R. (1984). Long-range dependence: A review. In *Statistics: An Appraisal* (H. A. David and H. T. David, eds.). Iowa State University.

Daniels, H. E. (1945). The statistical theory of the strength of bundles of threads. I. *Proc. Roy. Soc. London. A* **183**, 405–435.

de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory*. Springer, New York.

Devroye, L. (1986). *Nonuniform Random Variate Generation*. Springer-Verlag, New York.

Devroye, L. (1987). *A Course in Density Estimation*. Vol. 14 of *Progress in Probability and Statistics*. Birkhäuser Boston Inc., Boston, MA.

Donoho, D. and Jin, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.* **32**, 962–994.

Ferger, D. (2004). A continuous mapping theorem for the argmax-functional in the non-unique case. *Statist. Neerlandica* **58**, 83–96.

Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. of Math.* **44**, 423–453.

Gneiting, T. (2000). Power-law correlations, related models for long-range dependence and their simulation. *J. Appl. Probab.* **37**, 1104–1109.

Grenander, U. (1956a). On the theory of mortality measurement. I. *Skand. Aktuarietidskr.* **39**, 70–96.

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

Groeneboom, P. (1985). Estimating a monotone density. In *Proceedings of the Berkeley conference in honor of Jerzy Neyman and Jack Kiefer, Vol. II.* Wadsworth, Belmont, CA.

Jacod, J. and Shiryaev, A. N. (2003). *Limit Theorems for Stochastic Processes*. Vol. 288 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*, 2nd ed. Springer-Verlag, Berlin.

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

Kulikov, V. N. and Lopuhaä, H. P. (2006). The behavior of the NPMLE of a decreasing density near the boundaries of the support. *Ann. Statist.* **34**, 742–768.

Ma, C. (2002). Correlation models with long-range dependence. *J. Appl. Probab.* **39**, 370–382.

Martin, R. J. and Walker, A. M. (1997). A power-law model and other models for long-range dependence. *J. Appl. Probab.* **34**, 657–670.

Mason, D. M. (1983). The asymptotic distribution of weighted empirical distribution functions. *Stochastic Process. Appl.* **15**, 99–109.

Miller, D. R. (1976). Order statistics, Poisson processes and repairable systems. *J. Appl. Probability* **13**, 519–529.

Pyke, R. (1959). The supremum and infimum of the Poisson process. *Ann. Math. Statist.* **30**, 568–576.

Rockafellar, R. T. (1970). *Convex Analysis.* Princeton University Press, Princeton, N.J.

Rockafellar, R. T. and Wets, R. J.-B. (1998). *Variational Analysis.* Vol. 317 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences].* Springer-Verlag, Berlin.

Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics.* John Wiley & Sons Inc., New York.

Swanepoel, J. W. H. (1999). The limiting behavior of a modified maximal symmetric $2s$-spacing with applications. *Ann. Statist.* **27**, 24–35.

van de Geer, S. (1993). Hellinger-consistency of certain nonparametric maximum likelihood estimators. *Ann. Statist.* **21,** 14–44.

van der Vaart, A. and van der Laan, M. J. (2006). Estimating a survival distribution with current status data and high-dimensional covariates. *Int. J. Biostat.* **2**, 42.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer-Verlag, New York.

Whitt, W. (2002). *Stochastic Process Limits.* Springer-Verlag, New York.

Woodroofe, M. and Sun, J. (1993). A penalized maximum likelihood estimate of $f(0+)$ when f is nonincreasing. *Statist. Sinica* **3**, 501–515.

Centre de Recherche en Mathématiques de la Décision, Université Paris-Dauphine, Paris, 75775, France.

E-mail: fadoua@ceremade.dauphine.fr

Department of Mathematics and Statistics, York University, Toronto, Ontario M3J 1P3, Canada.

E-mail: hkj@mathstat.yorku.ca

School of Engineering and Applied Sciences, Frederick University Cyprus, Nicosia, 1036, Cyprus.

E-mail: m.pavlides@frederick.ac.cy

Department of Statistics, University of Washington, Seattle, WA 98195, USA

E-mail: arseni@stat.washington.edu

Department of Statistics, University of Washington, Seattle, WA 98195, USA.

E-mail: jaw@stat.washington.edu

(Received March 2009; accepted November 2009)