Annals of the Institute of Statistical Mathematics manuscript No.
(will be inserted by the editor)

Maximizing leave-one-out likelihood for the location parameter of unbounded densities

Krzysztof Podgórski · **Jonas Wallin**

Received: date / Revised: date

Abstract We propose simple estimation of the location parameter for a density that is unbounded at the mode. The estimator maximizes a modified likelihood in which the singular term in the full likelihood is left out, whenever the parameter value approaches a neighborhood of the singularity location. The consistency and super-efficiency of this maximum leave-one-out likelihood estimator is shown through a direct argument. The importance for estimation within parametric families is discussed and illustrated by an example involving the gamma mixture of normal distributions.

Keywords unbounded likelihood · location parameter · super-efficiency · generalized asymmetric Laplace distribution

1 Introduction

The classical problem of the location parameter estimation frequently serves as an illustration of how the asymptotic theory can be used to identify an estimator with some optimal properties. In particular, the asymptotics for the maximum likelihood estimators (MLE) has been established not only under the so-called regular conditions but also when the density has a cusp at its mode. The history here goes back to the Ph.D. Thesis of Prakasa Rao, Rao (1966), and the subsequent related paper Rao

Krzysztof Podgórski
Department of Statistics
Lund University
Box 743
220 07 Lund, Sweden
E-mail: Krzysztof.Podgorski@stat.lu.se

Jonas Wallin
Mathematical Statistics
Lund University
Box 118
221 00 Lund, Sweden
E-mail: wallin@maths.lth.se

(1968), where consistency and super-efficiency of the MLE of the location parameter have been demonstrated for a bounded density with a cusp at the mode.

Estimation of the location can be also considered for an unbounded density. This case has been first approached in Ibragimov and Khasminskii (1981a) and later summarized in the influential monograph Ibragimov and Khasminskii (1981b), where, to deal with the unboundedness of the likelihood, Bayesian estimation has been considered. There, as well as in Rao (1966), weak convergence of the log-likelihood ratio process to an appropriately defined Gaussian process has been established yielding the consistency for the MLE, whenever this is well defined, or otherwise for Bayesian-type estimators.

This work also deals with the unbounded density case but instead of resorting to the Bayesian approach we modify the likelihood approach. A modification is needed since the likelihood is unbounded at each data point and the classical MLE is not even properly defined. To remedy this issue, we propose to leave a singular term out from the full likelihood in a neighborhood of the datum location which leads to the concept of the *leave-one-out* likelihood function, for a formal definition see Section 3.2, Eq. (1). The estimator $\hat{\delta}$ is defined as the maximizer of the leave-one-out likelihood. Under rather natural conditions it is shown that $\hat{\delta}$ is consistent. Moreover, a lower bound for the rate of convergence is established showing, in particular, that the estimator is super-efficient, i.e. its rate is faster than in the classical case of $n^{-1/2}$. The proof presented is completely self-contained, direct, and uses only elementary arguments. Consequently, it is formally independent of any other asymptotic results, including these for the convergence of the likelihood ratio process. Nevertheless, the intuitive reason for the supper-efficiency is the rate of convergence of the likelihood ratio process (or its moments as exploited in this work). Namely, for the densities that are unbounded this rate is faster than under the standard regular conditions, see Lemma 5 (this faster rate is tied to the asymptotics of the density around the location parameter as presented in Lemma 4).

The idea of leaving out a trouble causing factor in the likelihood seems to be quite natural and, in fact, has been recently proposed in the problem of estimation of parameters for a finite mixtures of normal densities in Seo and Kim (2012). Despite general similarities between the approaches, neither the estimators nor the results of that work translate to the setup of this paper.

The paper is organized as follows. Section 2 motivates the problem and, in particular, points at convenience of the method when used in a general multi-parameter setup. In Section 3, we present the assumptions and the main result which is Theorem 1. In Section 4, we formulate and prove the lemmas that eventually lead to the proof of Theorem 1 presented in Section 5. Finally, in the Appendix, we present an example illustrating how a version of the EM algorithm can be applied to maximize the leave-one-out likelihood.

2 Motivation

Although in this work we concentrate on the location parameter, the applicability of the approach extends to the multiparameter context. The leave-one out likelihood

function presents only a slightly modified likelihood and thus the maximizers over other than location parameters would have the asymptotic properties dictated by the classical MLE theory given, of course, that appropriate assumptions of the likelihood are satisfied. For this reason, the proposed estimation of location in the unbounded density case is not only of a theoretical interest but also have important implications for actual estimation problems. In fact, there are natural parametric families for which estimation in the presence of unboundedness becomes an important practical issue. This study was inspired by investigation of applicability of the EM algorithm to parameter estimation for linear models involving the generalized Laplace distributions.

Recall a generalized Laplace random variable X admits the representation $X = \delta + \mu\Gamma + \sigma\sqrt{\Gamma}Z$, where Γ has Gamma distribution with the shape τ and scale one, while Z has the standard normal distribution, see Kotz et al (2001) for details. This class is made of infinitely divisible distributions, is closed under the convolutions and the corresponding Lévy motions are referred to as the Laplace motions (in mathematical finance, specially in the symmetric case, these models are naturally known as the gamma variance processes). The density of X is of the form $p(x)|x|^{\alpha}$, where $\alpha = 2\tau - 1$ and $p(x)$ being a function that is bounded and non-negative around zero.

The explicit form of the density involves one of the Bessel functions so the distribution is also refered to as the Bessel function distribution. To maximize the likelihood one has to resort to numerical methods and, for example, the EM (expectation-maximization) algorithm can be conveniently employed to evaluate the MLE of the parameters $(\delta, \mu, \sigma, \tau)$. We refer to Protassov (2004) for a presentation of such an approach applied to a subclass of the generalized hyperbolic distributions (the latter were introduced by Barndorff-Nielsen (1978) and include also the generalized Laplace distributions). Since the range of values of τ is a'priori not known, one can not exclude a possibility of an unbounded density, which occurs when $\tau < 1/2$, i.e. $-1 < \alpha < 0$. In fact, the value of τ is tied to the grid of sampling for spatial or temporal models involving the Laplace motion – the finer grid the smaller value of τ which typically leads to an unbounded density.

The EM algorithm can be adopted to the leave-one-out likelihood by not accounting in each loop for the observation that is closest to the evaluated values of the location parameter. This is actually the EM algorithm applied to a penalized log-likelihood where the penalty term is $-\log f(x_{k(\hat{\delta})})$, in which it resembles the method of Chen et al (2008). In these applications, the EM algorithm preserves the fundamental monotonicity property entertained by the original EM method of Dempster et al (1977). In the resulting approximations, the estimate of δ has the same super-efficient asymptotic behavior as demonstrated in this work, while the estimates of μ, σ and τ behave asymptotically in the same way as the MLE under the standard regularity conditions. The formal argument supporting these statements in full generality is left for another occasion. However in the Appendix we do discuss main steps in such an EM approach when applied to the maximizing for the leave-one-out likelihood for the generalized Laplace distributions.

It should be mentioned that the proposed method is useful also in the case when the densities are bounded for all values in the interior of the parameters range but may become unbounded if the parameters reach boundaries of the range. Let us mention two examples when this is of importance. Firstly, for the generalized Laplace distri-

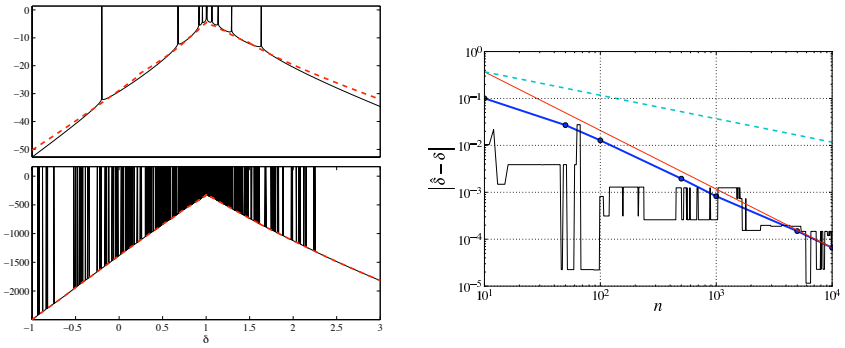

Fig. 1 *Left:* The full log-likelihood (solid line) vs. the leave-one-out log-likelihood (dashed line) used for the sample of the size $n = 10$ (*Top*) and $n = 500$ (*Bottom*). In the bottom figure the dashed line cannot be distinguished from the lower envelope of the log-likelihood. *Right:* Asymptotics and super-efficiency of the estimator: the optimal rate – straight thin line, the estimated rate from Monte Carlo simulation – thick line, a trajectory of the absolute estimation error $|\hat{\delta}_n - \delta_0|$ with increasing sample size – thin line. For comparison the rate of MLE under regular assumptions is given by the dashed line.

bution, if $\tau \in [1/2, 1)$ and $\sigma > 0$, then the generalized Laplace density is bounded. However, if the parameter value for σ reaches the boundary $\sigma = 0$, then the distribution approaches the gamma distribution with the shape $\tau \in [1/2, 1)$ which constitutes an example of unbounded density. In consequence, using the leave-one-out method allows to avoid ensuing problems. The second case relates to the fact that the generalized Laplace distributions represent a special and the only unbounded density case of the generalized hyperbolic distributions. Here again the leave-one-out method can be applied to deal with the unboundedness due to the parameters approaching the values corresponding to a generalized Laplace distribution.

For illustration of the leave-one-out likelihood and the discussed properties of the estimator, we performed a small Monte Carlo (MC) study based on samples generated from an asymmetric generalized Laplace distribution with $(\delta_0, \mu, \sigma, \tau) = (1, -0.5, 1, 0.4)$. In Figure 1 *(Left)*, the full likelihood is compared to the leave-one-out one (dashed line) in a small sample size case ($n = 10$, top) and a large sample size case ($n = 500$, bottom) cases. We can clearly observe the smoothing effect offered by the method.

The asymptotic behavior of the estimator is illustrated in Figure 1 *(Right)*, where, on the logarithmic scale, we see the optimal rate (straight thin line) and the rate for the proposed estimator obtained through MC simulations. The latter is represented here by 90% MC-sample quantiles of $|\hat{\delta}_n - \delta_0|$ computed for 1000 MC samples and for a number of sizes n (thick line). For comparison, a trajectory of $|\hat{\delta}_n - \delta_0|$ evaluated for the subsequently increased n values of a single large sample is represented by the thin line. Finally, the dashed line on the graph corresponds to the regular rate of convergences $n^{-1/2}$, from which we clearly see a super-efficient rate of the estimator.

3 The maximum leave-one-out likelihood estimator and its supereefficiency

3.1 *Assumptions*

Through the remainder of the paper, let $X_1,\dots,X_n$ be an iid sample from a distribution given by a density $f(x-\delta_0)$ that is differentiable everywhere except for δ_0. Recall that the Fisher information for a location parameter associated with a density f is defined as $\mathscr{I}_f = \mathbb{E}[((\log f)'(X))^2] = \mathbb{E}[f'^2/f^2(X)]$, where X is a random variable with the distribution defined by f. In our case the Fisher information is not finite due to the assumed unbounded behavior of f around zero so instead we use the incomplete Fisher information defined for $\varepsilon > 0$ as $\mathscr{I}_f(\varepsilon) = \mathbb{E}[f'^2/f^2(X)\big||X| > \varepsilon]$. We assume that

A1 $f(x) = p(x)|x|^\alpha,\ \alpha \in (-1,0)$, p has bounded derivative on $\mathbb{R}\setminus\{0\}$ and, for some $\varepsilon_0 > 0$, is non-zero and continuous either on $[-\varepsilon_0,0]$ or on $[0,\varepsilon_0]$.
A2 There exists $b > 0$ such that $f(x) = O(|x|^{-b-1})$ when $|x| \to \infty$.
A3 For some (and thus for all) $\varepsilon > 0$ the Fisher information $\mathscr{I}_f(\varepsilon)$ is finite.

3.2 *Maximum leave-one-out likelihood estimator*

Here we introduce the estimator and present several convenient representations of the leave-one-out likelihood ratio process.

Let us denote

$$k(\delta) = \operatorname*{argmin}_{k\in\{1,\dots,n\}} |X_k - \delta|,$$

with the convention that if there are two indices we take the one for which corresponding $X_{k(\delta)}$ is on the right hand side of δ. Define the estimator $\hat\delta = \hat\delta_n$ as the argument that maximizes

$$l(\delta) = l_n(\delta) = \frac{\prod_{i=1}^n f(X_i-\delta)}{f(X_{k(\delta)}-\delta)}. \tag{1}$$

Note here that $l(\delta)$ is a cadlag function (the left hand side continuous) and converging to zero at infinity so there is a maximizer (if there are more than one maximizer, we choose, for example, the smallest one). We also observe that $\hat u_n = \hat\delta_n - \delta_0$ is the maximizer of

$$Z(u) = Z_n(u) = \frac{l(u+\delta_0)}{l(\delta_0)} = \frac{f(X_{k(\delta_0)}-u-\delta_0)}{f(X_{k(u+\delta_0)}-\delta_0)} \prod_{i\neq k(\delta_0), i\neq k(u+\delta_0)} \frac{f(X_i-u-\delta_0)}{f(X_i-\delta_0)}.$$

By introducing the event $C_{i,\delta} = \{k(\delta)\neq i\}$ and its indicator function $I_{C_{i,\delta}}$, we obtain the following convenient representations of the above functions

$$l(\delta) = \prod_{i=1}^n f(X_i-\delta)^{I_{C_{i,\delta}}} = \sum_{k=1}^n I_{C^c_{k,\delta}} \prod_{i=1, i\neq k}^n f(X_i-\delta), \tag{2}$$

and

$$Z(u) = \prod_{i=1}^{n} f(X_i - \delta_0)^{-I_{C_{i,\delta_0}}} \prod_{i=1}^{n} f(X_i - u - \delta_0)^{I_{C_{i,u+\delta_0}}} = \\ = \Big(\sum_{l=1}^{n} I_{C^c_{l,\delta_0}} \prod_{i=1, i\neq l}^{n} 1/f(X_i - \delta_0) \Big) \cdot \Big(\sum_{k=1}^{n} I_{C^c_{k,\delta_0+u}} \prod_{j=1, j\neq k}^{n} f(X_j - u - \delta_0) \Big). \tag{3}$$

3.3 *The main result*

The purpose of this paper is to establish consistency of $\hat{\delta}_n$ which is done together with getting a super-efficient rate of convergence in the following result.

Theorem 1 *Let f satisfy the above assumptions and let $\hat{\delta}_n$ be the maximizer of l_n given by (1). Then $\hat{\delta}_n$ is a consistent estimator of δ_0 and for any $\beta < 1/(1+\alpha)$:*

$$\lim_{n\to\infty} n^{\beta}(\hat{\delta}_n - \delta_0) \stackrel{p}{=} 0. \tag{4}$$

4 Lemmas and the proof of the theorem

Additionally to the notation and assumptions of the previous section, we also use what follows. For $\lambda > 0$ and $L > 0$:

$$A_\lambda = A_{n,\lambda} = \{ \min_{\substack{i,j=1,\ldots,n \\ i\neq j}} |X_i - X_j| > \lambda \}, \tag{5}$$

$$B_L = B_{n,L} = \{ \max_{i=1,\ldots,n} |X_i - \delta_0| < L \}. \tag{6}$$

In our argument the variable L is eventually increasing without bound so whenever the symbol $O(L^\rho)$ is used for some ρ, it means that $\limsup_{L\to\infty} |O(L^\rho)|/L^\rho < \infty$. Finally, for compactness of our formulations, we define $S_r(u_0) = [u_0 - r, u_0 + r]$.

We start with a result about the rate of convergence of the minimal distance between X_i's.

Lemma 1 *Assume that a sequence of positive numbers λ_n has the following asymptotics for a certain $c > 0$:*

$$\lambda_n = O\left(n^{-1-\frac{1}{\alpha+1}-c} \right).$$

Then for $A_n = A_{n,\lambda_n}$ defined through (5) we have

$$\lim_{n\to\infty} \mathbb{P}(A_n) = 1.$$

Proof Since $\lambda_n \leq D \cdot n^{-1-\frac{1}{\alpha+1}-c}$ for some $D > 0$, it is enough to show the result for $\lambda_n = D \cdot n^{-1-\frac{1}{\alpha+1}-c}$. Define

$$C_n = \{ X_{n+1} \in \bigcup_{i=1}^{n} [X_i - \lambda_n, X_i + \lambda_n] \}.$$

We first demonstrate that for a proof it is sufficient to show that $C = \limsup_{n\to\infty} C_n$ is of probability zero, which is equivalent to saying that with probability one the number of times that an observation X_{n+1} is inside of $\cup_{i=1}^{n}[X_i - \lambda_n, X_i + \lambda_n]$ is finite.

To see this consider an outcome ω from C^c. Then there exists n_0 such that for $n > n_0$:

$$|X_{n+1}(\omega) - X_i(\omega)| > \lambda_n,\ i = 1,\ldots,n.$$

For such n_0, let

$$\varepsilon_0 = \min_{\substack{i,j=1,\ldots,n_0 \\ i\neq j}} |X_i(\omega) - X_j(\omega)|$$

while n_1 be such that for $n > n_1 > n_0$ we have $\lambda_n < \varepsilon_0$. Take $n > n_1$ and note that the minimum of $|X_i(\omega) - X_j(\omega)|$ over all pairs (i,j) such that $i,j = 1,\ldots,n, i \neq j$ is obtained as the minimum of the numbers standing on the left hand side of the following inequalities

$$\begin{aligned}
\min_{\substack{i,j=1,\ldots,n_0 \\ i\neq j}} |X_i(\omega) - X_j(\omega)| &> \lambda_n, \\
\min_{i=1,\ldots,n_0} |X_i(\omega) - X_{n_0+1}(\omega)| &> \lambda_{n_0} \geq \lambda_n, \\
\min_{i=1,\ldots,n_0+1} |X_i(\omega) - X_{n_0+2}(\omega)| &> \lambda_{n_0+1} \geq \lambda_n, \\
&\vdots \\
\min_{i=1,\ldots,n-1} |X_i(\omega) - X_n(\omega)| &> \lambda_{n-1} \geq \lambda_n.
\end{aligned}$$

Consequently the outcome ω has to be in A_n for each $n > n_1$, which proves that

$$C^c \subset \liminf_{n\to\infty} A_n.$$

Thus if A denotes the right hand side event in the above and $\mathbb{P}(C^c) = 1$, then

$$1 = \mathbb{P}(C^c) \leq \mathbb{P}(A) = \lim_{n\to\infty} \mathbb{P}\Big(\bigcap_{k\geq n} A_k\Big) \leq \liminf_{n\to\infty} \mathbb{P}(A_n) \leq \limsup_{n\to\infty} \mathbb{P}(A_n) \leq 1$$

and consequently it is indeed enough to show that $\mathbb{P}(C) = 0$.

To prove the latter, by the Borel-Canteli lemma, it is enough to show that $\mathbb{P}(C_n)$'s form a convergent series. To this end notice that by Assumption A1, the density of X_i is bounded except at δ_0. Hence there exists sufficiently small $u > 0$ and an interval neighborhood I of zero and of the diameter not exceeding u such that $f(x) = p(x)|x|^\alpha$ for $x \in I$ is larger than the value $f(y)$ for any $y \notin I$. Thus if a subset $D \subset \mathbb{R}$ has measure at most u, then

$$\begin{aligned}
\mathbb{P}(X_{n+1} \in D) &= \int_D p(x-\delta_0)|x-\delta_0|^\alpha\, dx \\
&\leq \int_I p(x)|x|^\alpha\, dx \leq \mathbb{P}(X \in [-u+\delta_0, u+\delta_0]).
\end{aligned}$$

Using this fact, the convergence of $n\lambda_n$ to zero, and independence of X_{n+1} from $\mathbf{X}_n = (X_1,\ldots,X_n)$, we obtain for sufficiently large n:

$$\mathbb{P}(C_n) = \mathbb{P}\big(X_{n+1} \in \bigcup_{i=1}^{n}[X_i-\lambda_n, X_i+\lambda_n]\big) = \mathbb{E}\big(\mathbb{P}\big(X_{n+1} \in \bigcup_{i=1}^{n}[X_i-\lambda_n, X_i+\lambda_n]\big|\mathbf{X}_n\big)\big)$$
$$\leq \mathbb{P}(X \in [-n\lambda_n+\delta_0, n\lambda_n+\delta_0]).$$

Note that there exists $K>0$ such that for sufficiently small u we have $\mathbb{P}(X\in[-u+\delta_0,u+\delta_0]) \leq Ku^{\alpha+1}$, so for sufficiently large n:

$$\mathbb{P}(C_n) \leq K(n\lambda_n)^{\alpha+1} \leq K\big(n^{-1/(\alpha+1)-c}\big)^{\alpha+1} = Kn^{-1-c(\alpha+1)}$$

and thus convergence of the series holds. □

The next lemma is a quite obvious consequence of Assumption A2.

Lemma 2 *If n/L_n^b converges to zero, then for $B_n = B_{n,L_n}$ given in (6):*

$$\lim_{n\to\infty}\mathbb{P}(B_n) = 1.$$

Proof By A2, the following inequality holds for some $K>0$ and sufficiently large L:

$$\mathbb{P}\left(|X-\delta_0|\leq L\right) \leq 1-KL^{-b}$$

and the result follows immediately from

$$\mathbb{P}(B_n) = \mathbb{P}\{\max_{i=1,\ldots,n}|X_i-\delta_0| < L_n\} \leq \left(1-KL_n^{-b}\right)^n,$$

which holds for sufficiently large n. □

In the proof of the next result we use Assumption A3, i.e. the finiteness of the partial Fisher information. Let us introduce the following function that is also used in the proof of Lemma 4:

$$v(x) = \frac{p'(x)|x|}{2p^{1/2}(x)} + \frac{\alpha}{2}\mathrm{sign}(x)p^{1/2}(x), \tag{7}$$

and note that it is bounded in neighborhood of zero. Moreover for $x\neq 0$:

$$(f^{1/2})'(x) = \frac{f'(x)}{2f^{1/2}(x)} = |x|^{\alpha/2-1}v(x). \tag{8}$$

Lemma 3 *There exists $K>0$ such that for each $x_0\in\mathbb{R}$, $c<1$ and $r\in(0,c/2)$:*

$$\int_{[-c,c]^c}\sup_{|h|<r}|f^{1/2}(x)-f^{1/2}(x-h)|\cdot f^{1/2}(x+x_0)\,dx \leq Kr^{(\alpha+1)/2}. \tag{9}$$

Proof First by the Schwartz inequality

$$\int_{[-c,c]^c} \sup_{|h|<r} |f^{1/2}(x) - f^{1/2}(x-h)| f^{1/2}(x+x_0)\, dx$$
$$\leq \left(\int_{[-c,c]^c} \sup_{|h|<r} (f^{1/2}(x) - f^{1/2}(x-h))^2\, dx \right)^{1/2}$$
$$= \frac{1}{2} \left(\int_{[-c,c]^c} \sup_{|h|<r} \left(\int_0^h \frac{f'}{f^{1/2}}(x-y)\, dy \right)^2 dx \right)^{1/2}.$$

By the Jensen inequality and then by the Fubini theorem

$$\int_{[-c,c]^c} \sup_{|h|<r} \left(\int_0^h \frac{f'}{f^{1/2}}(x-y)\, dy \right)^2 dx$$
$$\leq \int_{[-c,c]^c} \sup_{|h|<r} \left(h \int_0^h \frac{f'^2}{f}(x-y)\, dy \right) dx$$
$$= r \int_0^r 4 \int_{[-c,c]^c} |x-y|^{\alpha-2} v^2(x-y)\, dx\, dy$$
$$= r \int_0^r 4 \int_{[-c+y,y+c]^c} |s|^{\alpha-2} v^2(s)\, ds\, dy.$$

Note that for $y \in [0,r]$ we have $-c+y < -c+r < -r$ and $y+c > c > r$. Combining this with the boundedness of v in a neighborhood of zero, we obtain that for some K_0 and for each $\varepsilon > 0$:

$$\int_{[-c+y,y+c]^c} 2|s|^{\alpha-2} v^2(s)\, ds \leq \int_{[-r,r]^c} 2|s|^{\alpha-2} v^2(s)\, ds$$
$$= \int_r^\varepsilon 2|s|^{\alpha-2} v^2(s)\, ds + \int_{-\varepsilon}^{-r} 2|s|^{\alpha-2} v^2(s)\, ds + \frac{1}{2} \mathscr{I}_f(\varepsilon)$$
$$\leq K_0 \left| \int_r^\varepsilon s^{\alpha-2}\, ds \right| + \frac{1}{2} \mathscr{I}_f(\varepsilon)$$
$$\leq \frac{K_0}{2-\alpha} |r^{\alpha-1} - \varepsilon^{\alpha-1}| + \frac{1}{2} \mathscr{I}_f(\varepsilon)$$
$$\leq K^2 r^{\alpha-1},$$

where K is some positive constant independent of r and c. From these inequalities we obtain

$$\int_{[-c,c]^c} \sup_{|h|<r} |f^{1/2}(x) - f^{1/2}(x-h)| f^{1/2}(x+x_0)\, dx \leq K r^{(\alpha+1)/2},$$

which concludes the proof. □

The following result stands behind a super-efficient rate of convergence that is eventually obtained in the proof of the main theorem.

Lemma 4 *There exist $B>0$ and $K>0$ such that for each $s\in\mathbb{R}$:*

$$\mathbb{E}\left[\frac{f^{1/2}(X-s)}{f^{1/2}(X)}\right]\leq 1-K\min(|s|^{\alpha+1},B). \tag{10}$$

Proof Let us set $r(x,s)=(f^{1/2}(x+s)-f^{1/2}(x))^2$ and note

$$\begin{aligned}\mathbb{E}\left[\frac{f^{1/2}(X-s)}{f^{1/2}(X)}\right]&=\frac{1}{2}\left(\int f(x)dx+\int f(x-s)dx-\int r(x,s)dx\right)\\&=1-\frac{1}{2}\int r(x,s)\,dx.\end{aligned}$$

Note that $r(s)=\int r(x,s)\,dx$ is a continuous non-negative function taking value 2 at infinity, zero at $s=0$, which is also its unique global minimum. Consequently, it is enough to show that $r(s)$ is $O(s^{\alpha+1})$.

Consider be a one-sided neighborhood of zero, say $[0,\varepsilon_0]$, where v being negative is separated from zero by, say, $-L$, $L>0$. Then for positive s and x such that $x+s\in[0,\varepsilon_0]$ we have

$$\begin{aligned}r(x,s)&=\left(\int_0^s(f^{1/2})'(t+x)\,dt\right)^2=\left(\int_0^s(x+t)^{\alpha/2-1}v(x+t)\,dt\right)^2\\&\geq L^2\left(\int_0^s(x+t)^{\alpha/2-1}\,dt\right)^2=\frac{4L^2}{\alpha^2}s^\alpha\left(\left(\frac{x}{s}+1\right)^{\alpha/2}-\left(\frac{x}{s}\right)^{\alpha/2}\right)^2.\end{aligned}$$

Using this we get for positive $s<\varepsilon_0/2$:

$$\begin{aligned}\int r(x,s)\,dx&\geq\int_0^{\varepsilon_0/2}r(x,s)\,dx\\&\geq\frac{4L^2}{\alpha^2}s^{\alpha+1}\int_0^{\varepsilon_0/(2s)}((y+1)^{\alpha/2}-y^{\alpha/2})^2\,dy\\&\geq\frac{4L^2}{\alpha^2}\int_0^1((y+1)^{\alpha/2}-y^{\alpha/2})^2\,dy\cdot s^{\alpha+1}.\end{aligned}$$

The argument for negative s follows the same way. □

The preceding result is explicitly used only in the following lemma, which plays a central role in our proof of the main result.

Lemma 5 *There exist positive constants K_1, K_2 such that for all $n\in\mathbb{N}$, γ and λ both in $(0,1)$, if $r\in(0,\lambda/6)$ and $|u_0|>\gamma$, then*

$$\mathbb{E}\left[I_{A_\lambda\cap B_L}\sup_{u\in S_r(u_0)}Z^{1/2}(u)\right]\leq O(L^a)r^{\frac{\alpha}{2}}n^2(1-K_1\gamma^{1+\alpha}+K_2r^{\frac{1+\alpha}{2}})^{n-2}, \tag{11}$$

where $a=\max(0,(1-b)/2)$.

Proof We note that the left hand side does not depend on δ_0 so let us assume that $\delta_0 = 0$. Let us take arbitrary values λ, r, γ and u_0 that satisfy the required conditions (K_1, K_2 will come later). By (3)

$$\sup_{u\in S_r(u_0)} Z^{1/2}(u) \le \sum_{l=1}^{n} I_{C_{l,0}^c} \prod_{\substack{i=1\\ i\neq l}}^{n} f^{-1/2}(X_i) \cdot \sum_{k=1}^{n} \sup_{u\in S_r(u_0)} I_{C_{k,u}^c} \prod_{\substack{j=1\\ j\neq k}}^{n} f^{1/2}(X_j - u). \tag{12}$$

Let us note that

$$C_{k,u}^c = \left(\bigcup_{i\neq k}^{n} C_{i,u}^c\right)^c = \bigcap_{i\neq k}^{n} C_{i,u}.$$

Moreover, since in A_λ all observations are at least λ apart and in $C_{i,u}$ the value X_i is not the closest to u the distance between X_i and u must be at least $\lambda/2$ which gives

$$\{|X_i - u| \ge \lambda/2\} \supseteq A_\lambda \cap C_{i,u}.$$

For $u \in S_r(u_0)$, by the triangle inequality

$$C_{i,u_0,r} \stackrel{def}{=} \{|X_i - u_0| \ge \lambda/2 - r\} \supseteq \{|X_i - u| \ge \lambda/2\}.$$

Thus for each $k = 1,\dots,n$:

$$I_{A_\lambda} \sup_{u\in S_r(u_0)} I_{C_{k,u}^c} \prod_{\substack{i=1,\\ i\neq k}}^{n} f^{1/2}(X_i - u) \le \sup_{u\in S_r(u_0)} \prod_{\substack{i=1,\\ i\neq k}}^{n} f^{1/2}(X_i - u) I_{C_{i,u_0,r}} \tag{13}$$

and for each $l = 1,\dots,n$ we have

$$I_{A_\lambda} I_{C_{k,0}^c} \prod_{\substack{i=1,\\ i\neq k}}^{n} f^{-1/2}(X_i) \le \prod_{\substack{i=1,\\ i\neq k}}^{n} \frac{I_{|X_i|>\lambda/2}}{f^{1/2}(X_i)}. \tag{14}$$

Combining (12), (13), and (14) we obtain

$$I_{A_\lambda} \sup_{u\in S_r(u_0)} Z^{1/2}(u) \le \sum_{k,l=1}^{n} \sup_{u\in S_r(u_0)} \prod_{\substack{i=1,\\ i\neq k}}^{n} f^{1/2}(X_i - u) I_{C_{i,u_0,r}} \prod_{\substack{j=1,\\ j\neq l}}^{n} \frac{I_{|X_j|>\lambda/2}}{f^{1/2}(X_j)}.$$

For $i = 1,\dots,n$ let us define

$$\tilde{Y}_i = \frac{I_{L>|X_i|>\lambda/2}}{f^{1/2}(X_i)},$$
$$\bar{Y}_i(u) = f^{1/2}(X_i - u) I_{C_{i,u_0,r}}.$$

Then we obtain

$$I_{A_\lambda \cap B_L} \sup_{u\in S_r(u_0)} Z^{1/2}(u) \le \sum_{k,l=1}^{n} \tilde{Y}_k \sup_{u\in S_r(u_0)} \bar{Y}_l(u) \prod_{\substack{i=1,\\ i\neq k,\\ i\neq l}}^{n} \bar{Y}_i(u) \tilde{Y}_i.$$

As a result and by independence, we obtain

$$\int_{A_\lambda \cap B_L} \sup_{u\in S_r(u_0)} Z^{1/2}(u)\, d\mathbb{P} \le \sum_{k,l=1}^{n} \mathbb{E}[\tilde{Y}_k]\cdot \mathbb{E}\Big[\sup_{u\in S_r(u_0)} \bar{Y}_l(u) \prod_{\substack{i=1,\\ i\neq k,\\ i\neq l}}^{n} \bar{Y}_i(u)\tilde{Y}_i\Big]$$

$$= n^2\, \mathbb{E}[\tilde{Y}_1]\cdot \mathbb{E}\Big[\sup_{u\in S_r(u_0)} \bar{Y}_1(u) \prod_{i=3}^{n} \bar{Y}_i(u)\tilde{Y}_i\Big]$$

$$\le n^2\, \mathbb{E}[\tilde{Y}_1]\cdot \mathbb{E}\Big[\sup_{u\in S_r(u_0)} \bar{Y}_1(u)\Big]\cdot \mathbb{E}\Big[\sup_{u\in S_r(u_0)} \bar{Y}_1(u)\tilde{Y}_1\Big]^{n-2}. \quad (15)$$

In what follows, we bound each of the three expectations on the right hand side of the above inequality.

First, by Assumption A2, $\mathbb{E}[\tilde{Y}_1] \le \int_{-L}^{L} f^{1/2}(x)\, dx = O(L^a)$, where $a = \max(0,(1-b)/2)$. To deal with the second expectation, notice that by Assumption A1 on $f(x)$ there is a constant $K_0 > 0$ such that $f(x) \le K_0 \min(|x^\alpha|, 1) \le K_0(\lambda/2 - 2r)^\alpha$, since $0 < \lambda/2 - 2r < 1$. Therefore, if $|u-u_0| \le r$ and $|x-u_0| \ge \lambda/2$, then $|x-u| \ge \lambda/2 - 2r$ and thus

$$\sup_{u\in S_r(u_0)} \bar{Y}_1(u) \le K_0(\lambda/2 - 2r)^{\alpha/2} \le K_0 r^{\alpha/2},$$

where the last inequality holds since $\lambda > 6r$.

The final expectation requires a few more steps. First, using the triangle inequality yields

$$\tilde{Y}_1 \cdot \sup_{u\in S_r(u_0)} \bar{Y}_1(u) \le \tilde{Y}_1 \cdot \left(\bar{Y}_1(u_0) + \sup_{|h|<r} |\bar{Y}_1(u_0+h) - \bar{Y}_1(u_0)| \right).$$

Then from Lemma 3 there exists K_2 such that

$$\mathbb{E}\Big[\tilde{Y}_1 \cdot \sup_{|h|<r} |\bar{Y}_1(u_0+h) - \bar{Y}_1(u_0)|\Big]$$

$$\le \int_{[-\lambda/2+r,\lambda/2-r]^c} \sup_{|h|<r} |f^{1/2}(s-h) - f^{1/2}(s)| \cdot f^{1/2}(s+u_0)\, ds$$

$$\le K_2 r^{(1+\alpha)/2}$$

and from Lemma 4:

$$\mathbb{E}[\tilde{Y}_1 \cdot \bar{Y}_1(u_0)] \le 1 - K_1 \min(\gamma^{1+\alpha}, b).$$

Putting all the three bounds together in (15) completes the proof. □

Chebyshev's inequality combined with the inequality $1 + a \le e^a$ yields the following corollary to the above lemma.

Corollary 1 *There exist positive constants K_1 and K_2 such that for all $n \in \mathbb{N}$, γ and λ both in $(0,1)$, if $r \in (0, \lambda/6)$ and $|u| > \gamma$, then*

$$\mathbb{P}(I_{A_\lambda \cap B_L} \sup_{u \in S_r(u)} Z(u) \geq 1) \leq O(L^a) r^{\frac{\alpha}{2}} n^2 e^{-(n-2)\left(K_1 \gamma^{\alpha+1} - K_2 r^{(1+\alpha)/2}\right)},$$

where $a = \max(0, (1-b)/2)$.

Lemma 5 will enter the proof of the main theorem through the following result, which is a consequence of the above corollary.

Lemma 6 *Let $\hat{\delta}_L$ be the maximizer of $l(\delta)$ over $[-L+\delta_0, L+\delta_0]$. There exist positive constants K_1 and K_2 such that for all $n \in \mathbb{N}$, γ and λ both in $(0,1)$, if $r \in (0, \lambda/6)$, then*

$$\mathbb{P}(A_\lambda \cap B_L \cap \{|\hat{\delta}_L - \delta_0| > \gamma\}) \leq O(L^{a+1}) r^{\frac{\alpha}{2}-1} n^2 e^{-(n-2)\left(K_1 \gamma^{\alpha+1} - K_2 r^{(1+\alpha)/2}\right)}, \quad (16)$$

where $a = \max(0, (1-b)/2)$.

Proof From the definition of $\hat{\delta}_L$, $\hat{u}_L = \hat{\delta}_L - \delta_0$ maximizes $Z(u)$ over $[-L, L]$ and thus $Z(\hat{u}_L) \geq Z(0) = 1$. Consequently, if $|\hat{u}_L| > \gamma$, then

$$\sup_{u \in [-\gamma,\gamma]^c \cap [-L,L]} Z(u) \geq 1.$$

This leads to

$$\mathbb{P}(A_\lambda \cap B_L \cap \{|\hat{\delta}_L - \delta_0| > \gamma\}) \leq \mathbb{P}(I_{A_\lambda \cap B_L} \sup_{u \in [-\gamma,\gamma]^c \cap [-L,L]} Z(u) \geq 1).$$

Let $S_r(u_k), k = 1, \ldots, 2[L/r]+1$ be a cover of $[-\gamma,\gamma]^c \cap [-L,L]$, such that $|u_k| > \gamma$. By Corollary 1:

$$\begin{aligned}
&\mathbb{P}(A_\lambda \cap B_L \cap \{\sup_{u \in [-\gamma,\gamma]^c \cap [-L,L]} Z(u) \geq 1\}) \\
&\quad = \mathbb{P}(\bigcup_{k=1}^{2[L/r]+1} \{I_{A_\lambda \cap B_L} \sup_{u \in [-\gamma,\gamma]^c \cap S_r(u_k)} Z(u) \geq 1\}) \\
&\quad \leq \sum_{k=1}^{2[L/r]+1} \mathbb{P}(I_{A_\lambda \cap B_L} \sup_{u \in S_r(u_k)} Z(u) \geq 1) \\
&\quad \leq O(L^a) r^{\frac{\alpha}{2}-1} n^2 e^{-(n-2)\left(K_1 \gamma^{\alpha+1} - K_2 r^{(1+\alpha)/2}\right)}.
\end{aligned}$$

□

5 Proof of Theorem 1

Here we present our proof of the main theorem.

Proof Set $\beta < 1/(1+\alpha)$. Let $L_n = n^{s\frac{2}{1+b}}$, with s being a positive constant that will be set later but at the moment we require only that $L_n > n^{3/b}$. Further, let λ_n be set so that Lemma 1 is satisfied.

Because of Lemmas 1 and 2, the events A_{n,λ_n} and B_n=$B_{n,n^{3/b}}$ that are defined through (5) and (6), respectively, have probabilities converging to one. Consequently, it is sufficient to show that for each $\gamma > 0$:

$$\lim_{n\to\infty} \mathbb{P}(A_{n,\lambda_n} \cap B_n \cap \{n^{\beta}|\hat{\delta}_n - \delta_0| > \gamma\}) = 0.$$

Let $\gamma_n = \gamma n^{-\beta}$ and note that since $B_n \subseteq B_{n,L_n}$:

$$\begin{aligned}\mathbb{P}(A_{n,\lambda_n} \cap B_n \cap \{|\hat{\delta}_n - \delta_0| > \gamma_n\}) \leq \\ \leq \mathbb{P}(A_{n,\lambda_n} \cap B_{n,L_n} \cap \{\gamma_n < |\hat{\delta}_n - \delta_0| \leq L_n\}) + \mathbb{P}(B_n \cap \{|\hat{\delta}_n - \delta_0| > L_n\}). \quad (17)\end{aligned}$$

Let us consider the first term on the right hand side and take a sequence r_n so that $r_n \leq \lambda_n/6$. Then, by Lemma 6, for $a = \max(0,(1-b)/2)$:

$$\begin{aligned}\limsup_{n\to\infty} \mathbb{P}(A_{n,\lambda_n} \cap B_{n,L_n} \cap \{\gamma_n < |\hat{\delta}_n - \delta_0| \leq L_n\}) \\ \leq \limsup_{n\to\infty} O(L_n^{\beta+1}) n^2 r_n^{\frac{\alpha}{2}-1} e^{-(n-2)\left(K_1\gamma_n^{1+\alpha} - K_2 r_n^{(1+\alpha)/2}\right)}.\end{aligned}$$

By choosing r_n so that $n r_n^{(1+\alpha)/2} \leq n^{-d}$ for some $d > 0$, we have for suitably chosen $h > 0$, ε, and $K > 0$:

$$\limsup_{n\to\infty} \mathbb{P}(A_{n,\lambda_n} \cap B_{n,L_n} \cap \{\gamma_n < |\hat{\delta}_n - \delta_0| \leq L_n\}) \leq \lim_{n\to\infty} n^h e^{-n^{\varepsilon} + Kn^{-d}} = 0.$$

The second term on the right hand side of (17) also converges to zero as shown next. Since $\{|\hat{\delta}_n - \delta_0| > L_n\} \subseteq \{\sup_{|u|>L_n} Z(u) \geq 1\}$ and by a direct application of Chebyshev's inequality it is enough to show

$$\lim_{n\to\infty} \mathbb{E}\left[I_{B_n} \sup_{|u|>L_n} Z^{1/2}(u)\right] = 0. \quad (18)$$

To this end note that on B_n, $|X_i| \leq n^{3/b} + |\delta_0|$, thus for sufficiently large n, for $|u| > L_n$ and on B_n:

$$|X_i - u| \geq |u| - |X_i| \geq L_n - n^{3/b} - |\delta_0| = O(L_n).$$

From this, Assumption A2, and by the choice of L_n:

$$I_{B_n} f^{1/2}(X_i - u) \leq I_{B_n} K |X_i - u|^{-(b+1)/2} \leq O(L_n^{-(b+1)/2}) = O(n^{-s}).$$

In consequence,

$$I_{B_n} \sup_{|u|>L_n} Z^{1/2}(u) \le O^{n-1}(n^{-s}) I_{B_n} \prod_{i\neq k(\delta_0)} f^{-1/2}(X_i-\delta_0). \tag{19}$$

Using the representation (2), we have

$$\begin{aligned}\prod_{i\neq k(\delta_0)} f^{-1/2}(X_i-\delta_0) &= \sum_{k=1}^{n} I_{C^c_{k,\delta_0}} \prod_{i=1,i\neq k}^{n} f^{-1/2}(X_i-\delta_0)\\ &\le \sum_{k=1}^{n} \prod_{i=1,i\neq k}^{n} f^{-1/2}(X_i-\delta_0).\end{aligned}$$

By Assumption A2, we also have

$$\mathbb{E}(I_{|X-\delta_0|<L} f^{-1/2}(X-\delta_0)) = O(L^c),$$

where $c=(1-b)^+/2$, which along with the mutual independence of X_i's yields

$$\begin{aligned}\mathbb{E}\Big[I_{B_n} \prod_{i\neq k(\delta_0)} f^{-1/2}(X_i-\delta_0)\Big] &\le n\left(\mathbb{E}\left[I_{|X-\delta_0|\le n^{3/b}} f^{-1/2}(X-\delta_0)\right]\right)^{n-1}\\ &\le nO^{n-1}(n^{3c/b}).\end{aligned}$$

Putting this together with (19), for sufficiently large n we obtain

$$\mathbb{E}\Big[I_{B_n} \sup_{|u|>L_n} Z^{1/2}(u)\Big] \le nO^{n-1}(n^{3c/b-s}),$$

where s as of now was not set yet. Thus by taking $s>3c/b+1$ we make the right hand side converging to zero, which concludes the proof. □

6 Concluding remarks

We have demonstrated that the maximum *leave-one-out likelihood* estimator is consistent and has a superefficient rate of convergence. The rate of convergence does not differ by a power factor from $n^{-1/(1+\alpha)}$ which would be the optimal rate of convergence. In fact, the proof of the main theorem yields a bit stronger conclusion stating that the lower bound on the rate of convergence differs from the optimal rate only by a certain power-of-logarithm factor. However, the presented proof does not yield the optimal rate and an open question is if this rate is actually reached by the estimator. In Polfeldt (1970b) and Polfeldt (1970a), this rate was proven optimal for the minimal variance estimation of the location. There an estimator achieving this rate is constructed. This optimal rate is also obtained in Ibragimov and Khasminskii (1981b) for the Pitman estimators.

It is worth stressing again that the leave-one-out estimator unlike the other estimators has the advantage that it can be easily implemented through the MLE approach in a general multi-parameter setup, for example, when scale or/and shape parameters are present. Optimization algorithms such as the gradient based methods, see

Lange (1995), and/or a modified EM algorithm are well suited for maximization of the leave-one-out likelihood. In the Appendix we demonstrate how the monotonicity property can be obtained for a modification of the EM algorithm for the leave-one-out likelihood. Investigation of effectiveness of such algorithms deserves a separate study.

Appendix

Here we present a formalized approach to the maximizing the leave-one-out likelihood by means of the EM algorithm and in the presence of an other than location parameter. Although the majority of the presented argument is valid in a fairly general setup, we focus an example of a symmetric generalized Laplace distribution.

Specifically, we consider estimation of a vector of parameters $\theta_0 = (\delta_0, \sigma_0)$ of a symmetric ($\mu = 0$) generalized Laplace distribution with some known shape parameter $\tau < 0.5$. See Section 2 for the definitions and the notation. In our setup, the observed values are $Y_i = \sigma_0\sqrt{\Gamma_i}Z_i + \delta_0$, $i = 1,\dots,n$ and the complete set of variables is $\mathbf{X} = (\Gamma_1,\dots,\Gamma_n,Y_1,\dots,Y_n)$. The density $f_{\theta_0}(y)$ of Y_i's is having the form $p_{\sigma_0}(y-\delta_0)|y-\delta_0|^{2\tau-1}$ for some bounded and non-vanishing around zero function p_{σ_0} (cf. Kotz et al (2001)).

We need some additional notation and definitions. For a vector $\mathbf{y} = (y_1,\dots,y_n) \in \mathbb{R}^n$, let $(|y_{i_1}-\delta|,\dots,|y_{i_n}-\delta|)$ be the order statistics of $(|y_1-\delta|,\dots,|y_n-\delta|)$ and consider the permutation $\pi^\delta(1,\dots,n) = (i_1,\dots,i_n)$. By slightly abusing notation we also write

$$\pi^\delta(\mathbf{y}) = (y_{i_1},\dots,y_{i_n}) = (y_0^\delta,\dots,y_{n-1}^\delta).$$

Thus for a given $\delta \in \mathbb{R}$, π^δ becomes a mapping from the Euclidean space $\mathbb{R}^n$ into to its subset $\mathbf{R}^\delta = \pi^\delta(\mathbb{R}^n)$. Clearly, if $\mathbf{y} \in \mathbf{R}^\delta$, then $|y_1-\delta| \le |y_2-\delta| \le \cdots \le |y_n-\delta|$. We further extent this notation by writing $\pi^\delta(\mathbf{x})$ for $(\gamma_1,\dots,\gamma_n,\pi^\delta(\mathbf{y}))$.

For $\mathbf{Y}^\delta = \pi^\delta(\mathbf{Y})$, we consider the conditional distributions of the vector $\tilde{\mathbf{Y}}^\delta = (Y_1^\delta,\dots,Y_{n-1}^\delta)$ given $Y_0^\delta = y_0$. This distribution has the density denoted by $g_\theta(\tilde{\mathbf{y}}|y_0)$ and is defined on $\mathbf{R}_{y_0}^\delta = \{\tilde{\mathbf{y}} \in \mathbb{R} : (y_0,\tilde{\mathbf{y}}) \in \mathbf{R}^\delta\}$. Similarly, we consider the conditional density of the vector $\mathbf{X}^\delta = \pi^\delta(\mathbf{X})$ conditionally on $\mathbf{Y}^\delta = \mathbf{y} \in \mathbf{R}^\delta$, denoted by $k_\theta(\mathbf{x}|\mathbf{y})$ for $\mathbf{x} \in \mathbb{R}_+^n \times \{\mathbf{y}\}$; and, finally, the density of the distribution of $\mathbf{X}^\delta$ given $Y_0^\delta = y_0$ for $\mathbf{x} \in \mathbb{R}_+^n \times \{y_0\} \times \mathbf{R}_{y_0}^\delta$ is denoted by $h_\theta(\mathbf{x}|y_0)$.

We note the key relation that connects the leave-one-out likelihood $l(\theta) = l(\delta,\sigma)$, as defined in (1) but in the presence of the additional parameter σ, with the conditional distribution g_θ:

$$g_\theta(y_1,\dots,y_{n-1}|y_0) = \frac{(n-1)!}{F(\theta,y_0)^{n-1}} f_\theta(y_1)\cdots f_\theta(y_{n-1}) = C_{n,\theta,y_0}\cdot l(\theta), \tag{20}$$

where $F(\theta,y_0) = 1 - \int_{-|y_0-\delta|}^{|y_0-\delta|} f_\theta(s)\,ds$, $\tilde{\mathbf{Y}}^\delta = \mathbf{y} = (y_1,\dots,y_{n-1})$, and $C_{n,\theta,y_0} = (n-1)!/F(\theta,y_0)^{n-1}$. In our following discussion, we simply consider the maximization of $l(\theta)$ as it would be equivalent to the maximization of $L^{\mathbf{y}}(\theta) = g_\theta(y_1,\dots,y_{n-1}|y_0)$.

Formally, it is not entirely obvious since the constant C_{n,θ,y_0} is dependent on θ. However, we note that the y_0 is closer to δ than any other y_i's and, by our assumptions, if y_0 is in the vicinity of δ, then $F(\theta,y_0) \approx 1-2|y_0-\delta|^{2\tau} p_\sigma(y-\delta)$, which can be separated from zero. In the result, the value of the maximizer will not be significantly effected by the presence of C_{n,θ,y_0}. It is beyond the scope of this note to analytically investigate the above claim and thus from now on we discuss maximization of $L^{\mathbf{y}}(\theta) = g_\theta(y_1,\ldots,y_{n-1}|y_0)$.

The function $L^{\mathbf{y}}(\theta)$ can be viewed as the likelihood function given that $\tilde{\mathbf{Y}}^\delta = \mathbf{y}$ is observed. Formally $\mathbf{y}$ is obtained from the actual observations by excluding the observation that is closest to δ, where δ is not the true parameter of the distribution; thus the properties of the maximizer of $L^{\mathbf{y}}(\theta)$ can not be deduced directly from the properties of the MLE for such likelihood. Consequently, the main result of this paper (or rather its version in which $l(\theta)$ is replaced by $L^{\mathbf{y}}(\theta)$) is needed to justify the consistency of such estimator.

To obtain the EM algorithm for maximizing $L^{\mathbf{y}}(\theta)$ we repeat the standard steps as presented for example in Wu (1983). Namely, the following two fundamental facts hold for any fixed value of incomplete observations $\mathbf{y} \in \mathbf{R}^{\delta'}$:

$$\begin{aligned} L^{\mathbf{y}}(\theta) = &\int_{\mathbb{R}^n_+\times\{\mathbf{y}\}} \log h_\theta\left(\pi^\delta(\mathbf{x})|y_0^\delta\right) k_{\theta'}(\mathbf{x}|\mathbf{y})\,d\mathbf{x} \\ &- \int_{\mathbb{R}^n_+\times\{\mathbf{y}\}} \log k_\theta(\pi^\delta(\mathbf{x})|\pi^\delta(\mathbf{y}))\cdot k_{\theta'}(\mathbf{x}|\mathbf{y})\,d\mathbf{x} \end{aligned} \tag{21}$$

and

$$\begin{aligned} &\int_{\mathbb{R}^n_+\times\{\mathbf{y}\}} \log k_\theta(\pi^\delta(\mathbf{x})|\pi^\delta(\mathbf{y}))\cdot k_{\theta'}(\mathbf{x}|\mathbf{y})\,d\mathbf{x} \\ &\qquad\qquad \le \int_{\mathbb{R}^n_+\times\{\mathbf{y}\}} \log k_{\theta'}(\pi^{\delta'}(\mathbf{x})|\pi^{\delta'}(\mathbf{y}))\cdot k_{\theta'}(\mathbf{x}|\mathbf{y})\,d\mathbf{x}. \end{aligned}$$

The proof is standard and thus we omit it here.

These two conditions guarantee the monotonicity of $L^{\mathbf{y}}(\hat{\theta}_i)$ in i, where $\hat{\theta}_i$ are the updates of the algorithm based on the maximizing the first term of the right hand side of (21), which we denote as $Q^{\mathbf{y}}(\theta|\theta')$.

Let us now discuss how this maximization procedure avoids $\hat{\delta}_i$ being trapped at one of the $Y_1,\ldots,Y_n$ – the problem that the EM algorithm does not protect against; this is because the likelihood is infinite at δ's equal to any of the observations. The local maxima are equal to infinity due to the unboundedness of the likelihood. In our discussion we use the explicit form of a symmetric generalized Laplace density f_θ. Let $s(\gamma)$ be the density of gamma distribution with the shape parameter $\tau < 0.5$ and the scale equal to one and define

$$\begin{aligned} M(y,y';\theta|\theta') &= \frac{\displaystyle\int_0^\infty \left(\log\frac{s(\gamma)}{\sqrt{2\pi\sigma^2\gamma}} - \frac{(y-\delta)^2}{2\sigma^2\gamma}\right)\frac{s(\gamma)}{\sqrt{2\pi\sigma'^2\gamma}} e^{-\frac{(y'-\delta')^2}{2\sigma'^2\gamma}}\,d\gamma}{f_{\theta'}(y')} \\ &= P(y',\theta') - \frac{\log(2\pi\sigma^2)}{2} - \frac{(y-\delta)^2}{2\sigma^2} N(y',\theta'), \end{aligned}$$

where

$$N(y',\theta') = \frac{\int_0^\infty \frac{1}{\gamma}\frac{s(\gamma)}{\sqrt{2\pi\sigma'^2\gamma}} e^{-\frac{(y'-\delta')^2}{2\sigma'^2\gamma}}\, d\gamma}{f_{\theta'}(y')},$$

$$P(y',\theta') = \frac{\int_0^\infty \log\frac{s(\gamma)}{\sqrt{\gamma}}\cdot\frac{s(\gamma)}{\sqrt{2\pi\sigma'^2\gamma}} e^{-\frac{(y'-\delta')^2}{2\sigma'^2\gamma}}\, d\gamma}{f_{\theta'}(y')}.$$

Straight computations lead us to

$$Q^{\mathbf{y}}(\theta|\theta') = \sum_{i=0}^{n-1} M(y_i^\delta, y_i^{\delta'};\theta|\theta') - \log f_\theta(y_0^\delta) - (n-1)\log\left(F(\theta,y_0^\delta)/(n-1)!\right) \tag{22a}$$

$$= \sum_{i=0}^{n-1} P(y_i^{\delta'},\theta') - \frac{n}{2}\log(2\pi\sigma^2) - \sum_{i=0}^{n-1}\frac{(y_i^\delta-\delta)^2}{2\sigma^2} N(y_i^{\delta'},\theta') \tag{22b}$$

$$- \log f_\theta(y_0^\delta) - (n-1)\log\left(F(\theta,y_0^\delta)/(n-1)!\right). \tag{22c}$$

If we would not consider the leave one out algorithm, the maximization would be based solely on the function of δ that is listed in (22b); this is a simple quadratic function of δ and the maximum is easily found in the explicit form. However, in the unbounded density case, the algorithm would reach a value $\hat{\delta}_i = y_0$ in the ith step and in the next step the solution would favor the same $\hat{\delta}_{i+1} = y_0$. Consequently the algorithm updates would be trapped in the local maximum.

In the version of the EM algorithm that is discussed above, the term in (22c) punishes choosing the value $\hat{\delta}_{i+1}$ close to y_0 because $-\log f_\theta(y_0^\delta)$ converges to minus infinity at δ approaching y_0^δ. It effectively pushes δ_i away from any particular observation. This has a similar effect to taking out the term $M(y_0^\delta, y_0^{\delta'};\theta|\theta')$ from the right hand side of the first line of (22a) and thus explains why the algorithm relates to the leave-one-out approach.

Acknowledgment

We are endebted to anonymous referees for their comments and suggestions that enhanced and improved the original version of the paper. The research has been supported by the Swedish Research Council Grant 2008-5382.

References

Barndorff-Nielsen O (1978) Hyperbolic distributions and distributions on hyperbolae. *Scandinavian Journal of statistics* 5:151–157

Chen J, Tan X, Zhang R (2008) Inference for normal mixtures in mean and variance. *Statistica Sinica* 18(2):443–465

Dempster A, Laird N, Rubin D (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B (Methodological)* 39(1):1–38

Ibragimov IA, Khasminskii RZ (1981a) Asymptotic behavior of statistical estimates of the shift parameter for samples with unbounded density. *Journal of Mathematical Sciences* 16:1035–1041, translation from Russian, the original date of publication: 1976

Ibragimov IA, Khasminskii RZ (1981b) *Statistical estimation, Asymptotic theory*, Applications of Mathematics, vol 16. Springer, New York, translation from Russian, the original date of publication 1979

Kotz S, Kozubowski T, Podgórski K (2001) *The Laplace distribution and generalizations: a revisit with applications to communications, economics, engineering, and finance.* Progress in Mathematics Series, Birkhäuser, Boston

Lange K (1995) A gradient algorithm locally equivalent to the em algorithm. *Journal of the Royal Statistical Society Series B (Methodological)* 57(2): 425–437, URL http://www.jstor.org/stable/2345971

Polfeldt T (1970a) Minimum variance order when estimating the location of an irregularity in the density. *The Annals of Mathematical Statistics* 41:673–679

Polfeldt T (1970b) The order of the minimum variance in a non-regular case. *The Annals of Mathematical Statistics* 41:667–672

Protassov R (2004) Em-based maximum likelihood parameter estimation for multivariate generalized hyperbolic distributions with fixed λ. *Statistics and Computing* 14(1):67–77

Rao B (1966) *Asymptotic distributions in some nonregular statistical problems.* PhD thesis, Michigan State University

Rao B (1968) Estimation of the location of the cusp of a continuous density. *The Annals of Mathematical Statistics* 39:76–87

Seo B, Kim D (2012) Root selection in normal mixture models. *Computational Statistics & Data Analysis* 56(8):2454–2470

Wu C (1983) On the convergence properties of the EM algorithm. *The Annals of Statistics* 11:95–103