The Annals of Statistics
2015, Vol. 43, No. 4, 1568–1595
DOI: 10.1214/15-AOS1316

ROBUSTNESS TO OUTLIERS IN LOCATION–SCALE PARAMETER MODEL USING LOG-REGULARLY VARYING DISTRIBUTIONS

BY ALAIN DESGAGNÉ

Université du Québec à Montréal

Estimating the location and scale parameters is common in statistics, using, for instance, the well-known sample mean and standard deviation. However, inference can be contaminated by the presence of outliers if modeling is done with light-tailed distributions such as the normal distribution. In this paper, we study robustness to outliers in location–scale parameter models using both the Bayesian and frequentist approaches. We find sufficient conditions (e.g., on tail behavior of the model) to obtain whole robustness to outliers, in the sense that the impact of the outliers gradually decreases to nothing as the conflict grows infinitely. To this end, we introduce the family of log-Pareto-tailed symmetric distributions that belongs to the larger family of log-regularly varying distributions.

1. Introduction. In Bayesian analysis, outlying observations and prior misspecification may contaminate the posterior inference. For instance, a group of observations may suggest a quite different posterior inference than that proposed by the prior and the rest of data. Using light-tailed distributions such as the normal can lead to an undesirable compromise where the posterior distribution concentrates on an area that is not supported by any sources of information. The conflict is usually resolved automatically by modeling with heavy-tailed distributions, in favor of the sources of information with the lightest tails. O'Hagan and Pericchi [16] refer to this situation as the *theory of conflict resolution in Bayesian statistics*, in their extensive review of the literature on that topic.

Conflict resolution in Bayesian analysis was first described by De Finetti [7]. The theory has mostly been developed for location parameter inference; see, for instance, Dawid [6]; O'Hagan [13–15]; Angers [5]; Desgagné and Angers [10]; Kumar and Magnus [12]; Andrade and Omey [4]; Andrade, Dorea and Guevara Otiniano [1].

The theory on pure scale parameter inference was first analyzed by Andrade and O'Hagan [2], who considered partial robustness using regularly varying distributions (see also Andrade and Omey [4] and Andrade, Dorea and Guevara Otiniano [1], who generalize their work of partial robustness), and then by Desgagné [8], who considered whole robustness using log-exponentially varying distributions.

Received August 2014; revised January 2015.
MSC2010 subject classifications. Primary 62F35; secondary 62F15.
Key words and phrases. Built-in robustness, outliers, theory of conflict resolution, Bayesian inference, log-regularly varying distributions, log-Pareto-tailed symmetric distributions.

Note that partial robustness exists if the conflicting values have a significant but limited influence on the posterior distribution, as the conflict grows infinitely. In contrast, whole robustness is achieved if the influence of the conflicting values on the posterior distribution gradually decreases to nothing. To illustrate this, consider the estimation of a location parameter for a Laplace model (with a prior of 1). Hence, the posterior mode (or the maximum likelihood estimator) is the sample median. If, for instance, the sample is $(10, 20, 30, 40, 50, x, x, x, x)$, and we let $x \to \infty$, then a wholly robust estimator of the location would be around 30 (the center of the nonoutlying observations), while the partially robust sample median estimates the location by 50, that is, the maximum of the nonoutliers.

This paper goes a step beyond the literature in that it considers robustness for both location and scale parameters in the same model. The only other paper that considers Bayesian robustness in a location–scale model is Andrade and O'Hagan [3]. The essential difference is that partial robustness to a single outlier is achieved in their paper, while whole robustness to multiple outliers for both location and scale estimation is obtained in this paper.

Another distinctive aspect of this paper is the possibility of using the results of robustness in both frequentist and Bayesian approaches. Although the model allows us to add prior information on the location and scale through a very general joint prior density $\pi(\mu, \sigma)$ [essentially, we only require that $\sigma\pi(\mu, \sigma)$ is bounded], it is also possible to choose a noninformative prior such that $\pi(\mu, \sigma) \propto 1/\sigma$. The location and scale parameters can therefore be estimated in a robust way using either the Bayesian approach or a frequentist method like maximum likelihood estimation.

This paper is organized as follows. In Section 2, we introduce the class of log-regularly varying functions because tail behavior plays a crucial role in the search of robustness. Essentially, this class includes functions with a right tail that exhibits a logarithmic decay, which can be considered a super heavy tail. As a result, we also define the family of log-regularly varying distributions.

The model with its assumptions is described in Section 3.1, and the resolution of conflicts is addressed through the main results of this paper in Section 3.2. Two simple conditions of robustness are given. Modeling with a log-regularly varying distribution is the first. In the second condition, the number of nonoutlying observations must be larger than the maximum between the number of small and large outliers. Results of robustness are asymptotic, where the outlying observations tend to $-\infty$ or $+\infty$. Note that the asymptotic nature is about the outliers and not the sample size, as is usually understood. Whole robustness is expressed through different types of convergence of quantities, based on the complete sample, to quantities based only on the nonoutlying observations, resulting in a complete rejection of outliers. We obtain the uniform convergence of the posterior densities, the convergence in L_1, the convergence in distribution and the uniform convergence of the likelihoods.

In Section 4, we introduce the family of log-Pareto-tailed symmetric distributions that belongs to the larger family of log-regularly varying distributions. It consists essentially of a symmetric density, such as the standard normal, with extremities replaced by log-Pareto tails, that is, with logarithmic decay. In the presence of outlying observations, the log-Pareto tails ensure robust inference. Otherwise, the estimation is practically unaffected by the tails and is determined mostly by the chosen symmetric density.

In Section 5, we show that even if the results are asymptotic, they are still useful in practice with data. We first illustrate the threshold feature in Section 5.1. When an observation moves away from the nonconflicting values, its influence on the inference first increases gradually up to a certain threshold. The conflict then begins, and the model resolves it by progressively reducing the influence of the moving observation (now an outlier) to nothing. This built-in feature is attractive in practice in that conflict is managed in a sensitive and automatic way. In Section 5.2, concurrent estimators are compared under different scenarios through simulations of observations to find how they perform in the presence—or absence—of outlying observations. Nonrobust, partially and wholly robust modeling is considered. We conclude in Section 6, and some proofs are given in Section 7.

2. Log-regularly varying functions. As mentioned in the Introduction, tail behavior is crucial for robust modeling. Hence, we introduce the class of log-regularly varying functions, as defined in Desgagné [8], following the idea of regularly varying functions developed by Karamata [11]. For each function in Section 2, say g, we assume that $g(z)$ is continuous and strictly positive for z larger than or equal to a certain constant.

DEFINITION 1 (Log-regularly varying function). We say that a measurable function g is *log-regularly varying* at ∞ with index $\rho \in \mathbb{R}$, written $g \in L_\rho(\infty)$, if

$$\forall \varepsilon > 0, \forall \tau \geq 1, \text{ there exists a constant } A(\varepsilon, \tau) > 0 \text{ such that}$$
$$z \geq A(\varepsilon, \tau) \text{ and } 1/\tau \leq \nu \leq \tau \Rightarrow |\nu^\rho g(z^\nu)/g(z) - 1| < \varepsilon.$$

If $\rho = 0$, g is said to be *log-slowly varying* at ∞.

In other words, $g \in L_\rho(\infty)$ if $g(z^\nu)/g(z)$ converges to $\nu^{-\rho}$ uniformly in any set $\nu \in [1/\tau, \tau]$ (for any $\tau \geq 1$) as $z \to \infty$. The pointwise convergence for any $\nu > 0$ follows.

Note that if we define the function $h(z) = g(e^z)$, or equivalently $g(z) = h(\log z)$, we have $g \in L_\rho(\infty)$ if and only if h is regularly varying at ∞ with index $-\rho$, because $\lim_{z\to\infty} h(\nu z)/h(z) = \nu^{-\rho}$. Therefore, we can obtain different results directly from the theory of regularly varying functions. For instance, the functions $\log(\log z)$ and 1 are both log-slowly varying at ∞ since $\log z$ and 1 are slowly varying.

PROPOSITION 1 (Equivalence). *For any $\rho \in \mathbb{R}$, we have $g \in L_\rho(\infty)$ if and only if there exists a constant $A > 1$ and a function $s \in L_0(\infty)$ such that for $z \geq A$, g can be written as*

$$g(z) = (\log z)^{-\rho} s(z).$$

PROOF. It is well known that if a function h is regularly varying at ∞ with index $-\rho$, it can be represented as $h(z) = z^{-\rho} l(z)$, where l is some slowly varying function. It is equivalent to say that $g \in L_\rho(\infty)$, where

$$g(z) = h(\log z) = (\log z)^{-\rho} l(\log z) = (\log z)^{-\rho} s(z),$$

with $s(z) = l(\log z) \in L_0(\infty)$. □

The next proposition establishes the asymptotic dominance of a logarithmic function over a log-slowly varying function.

PROPOSITION 2 (Dominance). *If $s \in L_0(\infty)$ and $g \in L_\rho(\infty)$, then for all $\delta > 0$, there exists a constant $A(\delta) > 1$ such that $z \geq A(\delta) \Rightarrow$*

$$(\log z)^{-\delta} < s(z) < (\log z)^{\delta} \quad \text{and} \quad (\log z)^{-\rho-\delta} < g(z) < (\log z)^{-\rho+\delta}.$$

PROOF. It is well known that if l is slowly varying, then for every $\delta > 0$, we have $z^{-\delta} l(z) \to 0$ and $z^{\delta} l(z) \to \infty$ as $z \to \infty$. It follows that $z^{-\delta} < l(z) < z^{\delta}$ for z sufficiently large. If we replace z by $\log z$ and we set $s(z) = l(\log z)$, then $s \in L_0(\infty)$, and we obtain that $(\log z)^{-\delta} s(z) \to 0$ and $(\log z)^{\delta} s(z) \to \infty$ as $\log z \to \infty$ (or equivalently $z \to \infty$) and $(\log z)^{-\delta} < s(z) < (\log z)^{\delta}$ for z sufficiently large. Since we can write $g(z) = (\log z)^{-\rho} s(z)$, the second part of the proposition follows directly. □

The index ρ can be interpreted as a measure of the tail's thickness or as a tail index, which is useful for the ordering of different tails. The function with the smallest tail index ρ has the heaviest tail. More formally, we can verify that if $g_1 \in L_{\rho_1}(\infty)$ and $g_2 \in L_{\rho_2}(\infty)$, then $\rho_1 > \rho_2 \Rightarrow g_1(z)/g_2(z) \to 0$ as $z \to \infty$. The tail index ρ is also useful to determine if $(1/z)g(z)$ is integrable, where $g(z) \in L_\rho(\infty)$, as described in the next proposition.

PROPOSITION 3 (Integrability). *If $g(z) \in L_\rho(\infty)$, then there exists a constant $A > 0$ such that $(1/z)g(z)$ is integrable on $z \geq A$, if and only if*:

(i) $\rho > 1$,

(ii) $\rho = 1$, *with the log-slowly varying part of $g(z)$ having a sufficiently fast decay* [*e.g., faster than* $(\log(\log z))^{-\beta}$, *with* $\beta > 1$].

PROOF. If we define h such that $g(z) = h(\log z)$, and we choose A sufficiently large, then h is regularly varying at ∞ with index $-\rho$, and we have

$$\int_A^\infty (1/z)g(z)\,dz = \int_A^\infty (1/z)h(\log z)\,dz = \int_{\log A}^\infty h(u)\,du = \int_{\log A}^\infty u^{-\rho}l(u)\,du,$$

where l is slowly varying. For any $\delta > 0$, if A is sufficiently large, we have $u^{-\delta} < l(u) < u^{\delta}$. Therefore, the integral exists if $\rho > 1$ and does not if $\rho < 1$.

If $\rho = 1$, we see that the decay of l determines the existence of the integral. If, for instance, $l(u) < (\log u)^{-\beta}$ or $s(z) = l(\log u) < (\log(\log u))^{-\beta}$, with $\beta > 1$ and $s \in L_0(\infty)$, then the integral exists. Instead, if $l(u) > (\log u)^{-\beta}$ or $s(z) = l(\log u) > (\log(\log u))^{-\beta}$, with $\beta < 1$ and $s \in L_0(\infty)$, then the integral does not exist. □

In particular, if f is a continuous symmetric probability density function defined on $\mathbb{R}$ such that $g(z) = zf(z) \in L_\rho(\infty)$, we know from Proposition 3 that a tail index $\rho > 1$ is sufficient to guarantee that f is proper and that $\rho \geq 1$ is a necessary condition. This leads us to the next definition.

DEFINITION 2 (Log-regularly varying distribution). A random variable Z and its distribution are said to be log-regularly varying with index $\rho \geq 1$ if their symmetric density f is such that $zf(z) \in L_\rho(\infty)$.

Using Propositions 1 and 2, this means that for all $\delta > 0$ and $|z|$ larger than a certain constant, the symmetric (with respect to 0) density f of a log-regularly varying distribution with index ρ can be written as $f(z) = (1/|z|)(\log|z|)^{-\rho}s(|z|)$, where $s \in L_0(\infty)$ can be bounded by $(\log|z|)^{-\delta}$ and $(\log|z|)^{\delta}$. Such a density with logarithmic decaying tails can be referred to as a super heavy-tailed distribution.

In the next proposition, we see the asymptotic impact of a location–scale transformation on a log-regularly varying function g and the density f of a log-regularly varying distribution. Mostly, it is another way to express tail thickness.

PROPOSITION 4 (Location–scale transformation). *If* $g(z) = zf(z) \in L_\rho(\infty)$, *then we have, as* $z \to \infty$,

$$g\big((z-\mu)/\sigma\big)/g(z) \to 1 \quad \textit{and} \quad (1/\sigma)f\big((z-\mu)/\sigma\big)/f(z) \to 1,$$

uniformly on $(\mu, \sigma) \in [-\lambda, \lambda] \times [1/\tau, \tau]$, *for any* $\lambda \geq 0$ *and* $\tau \geq 1$.

PROOF. Using $g \in L_\rho(\infty)$ with Proposition 1, there exists a function $s \in L_0(\infty)$ such that $g(z) = (\log z)^{-\rho}s(z)$, if z is large enough. Therefore, for any chosen $\lambda \geq 0$ and $\tau \geq 1$, if z is sufficiently large, we have

$$\frac{g((z-\mu)/\sigma)}{g(z)} = \left(\frac{\log((z-\mu)/\sigma)}{\log z}\right)^{-\rho} \frac{s((z-\mu)/\sigma)}{s(z)}.$$

It is purely algebraic to show that the term $(\log((z-\mu)/\sigma))/(\log z)$ converges to 1 uniformly on any set $(\mu,\sigma)\in[-\lambda,\lambda]\times[1/\tau,\tau]$ as $z\to\infty$.

Finally, we want to show that $s((z-\mu)/\sigma)/s(z)$ converges to 1 uniformly on any set $(\mu,\sigma)\in[-\lambda,\lambda]\times[1/\tau,\tau]$ as $z\to\infty$, or equivalently that $s(y)/s(z)$ converges to 1 uniformly on $y\in[(z-\lambda)/\tau,(z+\lambda)\tau]$. We observe that for any chosen $\lambda\geq 0$ and $\tau\geq 1$, if z is sufficiently large, we have

$$z^{1/2}\leq(z-\lambda)/\tau\leq(z+\lambda)\tau\leq z^2.$$

Therefore, it suffices to show that $s(y)/s(z)$ converges to 1 uniformly on $y\in[z^{1/2},z^2]$, or equivalently, that $s(z^\nu)/s(z)$ converges to 1 uniformly on any set $\nu\in[1/2,2]$, which is the case since $s\in L_0(\infty)$. The second part of the proposition follows directly. □

3. Resolution of conflicts in a location–scale parameter model.

3.1. *Model.*

(i) Let $X_1,\dots,X_n$ be n random variables conditionally independent given μ and σ with their conditional densities given by

$$X_i\mid\mu,\sigma\overset{\mathcal{D}}{\sim}(1/\sigma)f\big((x_i-\mu)/\sigma\big);$$

(ii) the joint prior density of μ and σ is given by $\mu,\sigma\overset{\mathcal{D}}{\sim}\pi(\mu,\sigma)$, where $n\geq 2, x_1,\dots,x_n,\mu\in\mathbb{R},\sigma>0$.

We assume that the prior $\pi(\mu,\sigma)$ is nonnegative on $\mathbb{R}$, and the only other required assumption is that $\sigma\pi(\mu,\sigma)$ is bounded. Note that in particular, if we have no prior information or if we use the model in a frequentist approach, then we set $\pi(\mu,\sigma)\propto 1/\sigma$, an improper joint prior density which can be considered as noninformative.

We assume that f is a proper density that is continuous and strictly positive on $\mathbb{R}$. In addition, we assume it is symmetric with respect to the origin. We also assume that both tails of $|z|f(z)$ are monotonic, which means that the tails of $f(z)$ are also monotonic. Note that monotonicity of the tails of $f(z)$ and $|z|f(z)$ means that there exists a constant $M\geq 0$ such that

$$(1)\quad |y|\geq|z|\geq M \qquad \text{implies that } f(y)\leq f(z) \quad\text{and}\quad |y|f(y)\leq|z|f(z).$$

It follows that $f(z)$ and $|z|f(z)$ are bounded on the real line, with a limit of 0 in their tails as $|z|\to\infty$. Hence, considering also the prior, we can define the constant B as follows:

$$(2)\qquad B=\max\Big\{\sup_{z\in\mathbb{R}}f(z),\sup_{z\in\mathbb{R}}|z|f(z),\sup_{\mu\in\mathbb{R},\sigma>0}\sigma\pi(\mu,\sigma)\Big\}.$$

These conditions are referred to below as the conditions of regularity on f. The density f can possess other parameters than location and scale, such as a shape parameter, but they are assumed to be known.

We study robustness of the estimation of μ and σ in the presence of outliers. The nature of the results is asymptotic, in the sense that some x_i are going to $-\infty$ or $+\infty$. We want to find sufficient conditions to obtain whole robustness, that is, a complete rejection of the outliers.

Among the n observations, denoted by $\mathbf{x_n} = (x_1, \ldots, x_n)$, we assume that $k \geq 2$ of them, denoted by the vector $\mathbf{x_k}$, form a group of nonoutlying observations. We assume that l of them are considered as left outliers (smaller than the nonoutliers) and r of them are considered as right outliers (larger than the nonoutliers), with $k + l + r = n$.

For $i = 1, \ldots, n$, we define three binary functions k_i, l_i and r_i as follows. If x_i is a nonoutlying observation, we set $k_i = 1$; if it is a left outlier, we set $l_i = 1$; and if it is a right outlier, we set $r_i = 1$. These functions are set to 0 otherwise. We have $k_i + l_i + r_i = 1$ for $i = 1, \ldots, n$, with $\sum_{i=1}^n k_i = k$, $\sum_{i=1}^n l_i = l$ and $\sum_{i=1}^n r_i = r$.

We assume that each outlier is going to $-\infty$ or $+\infty$ at its own specific rate, to the extent that the ratio of two outliers is bounded. We can write

$$x_i = a_i + b_i \omega,$$

for $i = 1, \ldots, n$, where a_i and b_i are some constants such that $a_i \in \mathbb{R}$ and:

(i) $b_i = 0$ if $k_i = 1$;
(ii) $b_i < 0$ if $l_i = 1$;
(iii) $b_i > 0$ if $r_i = 1$;

and we let $\omega \to \infty$. Note that if multiple outliers share the same b_i, they move as a block at the same rate.

Let the joint posterior density of μ and σ be denoted by $\pi(\mu, \sigma \mid \mathbf{x_n})$ and the marginal density of $X_1, \ldots, X_n$ be denoted by $m(\mathbf{x_n})$, with

$$\pi(\mu, \sigma \mid \mathbf{x_n}) = \big[m(\mathbf{x_n})\big]^{-1} \pi(\mu, \sigma) \prod_{i=1}^{n} (1/\sigma) f\big((x_i - \mu)/\sigma\big).$$

Let the joint posterior density of μ and σ considering only the nonoutlying observations $\mathbf{x_k}$ be denoted by $\pi(\mu, \sigma \mid \mathbf{x_k})$ and its corresponding marginal density be denoted by $m(\mathbf{x_k})$, with

$$\pi(\mu, \sigma \mid \mathbf{x_k}) = \big[m(\mathbf{x_k})\big]^{-1} \pi(\mu, \sigma) \prod_{i=1}^{n} \big[(1/\sigma) f\big((x_i - \mu)/\sigma\big)\big]^{k_i}.$$

The likelihood functions can be found by setting $\pi(\mu, \sigma) \propto 1/\sigma$ and letting

$$\mathcal{L}(\mu, \sigma \mid \mathbf{x_n}) \propto \sigma \pi(\mu, \sigma \mid \mathbf{x_n}) \quad \text{and} \quad \mathcal{L}(\mu, \sigma \mid \mathbf{x_k}) \propto \sigma \pi(\mu, \sigma \mid \mathbf{x_k}).$$

PROPOSITION 5. *Considering the Bayesian context given in Section 3.1, the joint posterior densities* $\pi(\mu, \sigma \mid \mathbf{x_k})$ *and* $\pi(\mu, \sigma \mid \mathbf{x_n})$ *are proper.*

The proof of Proposition 5 is given in Section 7.

3.2. *Resolution of conflicts.* The results of robustness are now given.

THEOREM 1. *Consider the model and context described in Section 3.1, and assume that the conditions of regularity on* f *are satisfied. If we have*:

(i) $zf(z) \in L_\rho(\infty)$ [$zf(z)$ *is log-regularly varying at* ∞ *with index* $\rho \geq 1$],
(ii) $k > \max(l, r)$,

then we obtain the following results:

(a)

$$\lim_{\omega \to \infty} \frac{m(\mathbf{x_n})}{\prod_{i=1}^n [f(x_i)]^{l_i + r_i}} = m(\mathbf{x_k}).$$

(b)

$$\lim_{\omega \to \infty} \pi(\mu, \sigma \mid \mathbf{x_n}) = \pi(\mu, \sigma \mid \mathbf{x_k}),$$

uniformly on $(\mu, \sigma) \in [-\lambda, \lambda] \times [1/\tau, \tau]$, *for any* $\lambda \geq 0$ *and* $\tau \geq 1$.

(c)

$$\lim_{\omega \to \infty} \int_0^\infty \int_{-\infty}^\infty \left| \pi(\mu, \sigma \mid \mathbf{x_n}) - \pi(\mu, \sigma \mid \mathbf{x_k}) \right| d\mu \, d\sigma = 0.$$

(d) *As* $\omega \to \infty$,

$$\mu, \sigma \mid \mathbf{x_n} \xrightarrow{\mathcal{D}} \mu, \sigma \mid \mathbf{x_k},$$

and in particular

$$\mu \mid \mathbf{x_n} \xrightarrow{\mathcal{D}} \mu \mid \mathbf{x_k} \quad \textit{and} \quad \sigma \mid \mathbf{x_n} \xrightarrow{\mathcal{D}} \sigma \mid \mathbf{x_k}.$$

(e)

$$\lim_{\omega \to \infty} \mathcal{L}(\mu, \sigma \mid \mathbf{x_n}) = \mathcal{L}(\mu, \sigma \mid \mathbf{x_k}),$$

uniformly on $(\mu, \sigma) \in [-\lambda, \lambda] \times [1/\tau, \tau]$, *for any* $\lambda \geq 0$ *and* $\tau \geq 1$.

Proof of result (a) is substantial and therefore is given in Section 7. This is, however, the crucial part in the proof of Theorem 1.

PROOF OF RESULT (b). Consider (μ, σ) such that $\pi(\mu, \sigma) > 0$ [the proof for the case (μ, σ) such that $\pi(\mu, \sigma) = 0$ is trivial]. We have, as $\omega \to \infty$,

$$\begin{aligned}\frac{\pi(\mu, \sigma \mid \mathbf{x_n})}{\pi(\mu, \sigma \mid \mathbf{x_k})} &= \frac{m(\mathbf{x_k})}{m(\mathbf{x_n})} \frac{\pi(\mu, \sigma) \prod_{i=1}^{n} (1/\sigma) f((x_i - \mu)/\sigma)}{\pi(\mu, \sigma) \prod_{i=1}^{n} [(1/\sigma) f((x_i - \mu)/\sigma)]^{k_i}} \\ &= \frac{m(\mathbf{x_k})}{m(\mathbf{x_n})} \prod_{i=1}^{n} \big[(1/\sigma) f\big((x_i - \mu)/\sigma\big)\big]^{l_i + r_i} \\ &= \frac{m(\mathbf{x_k}) \prod_{i=1}^{n} [f(x_i)]^{l_i + r_i}}{m(\mathbf{x_n})} \cdot \prod_{i=1}^{n} \left[\frac{(1/\sigma) f((x_i - \mu)/\sigma)}{f(x_i)}\right]^{l_i + r_i} \to 1.\end{aligned}$$

The first part of the last term does not depend on μ and σ and converges to 1 as $\omega \to \infty$, using result (a). The second part of the last term also converges to 1 uniformly in any set $(\mu, \sigma) \in [-\lambda, \lambda] \times [1/\tau, \tau]$ using Proposition 4. Furthermore, since f and $\sigma\pi(\mu, \sigma)$ are bounded, $\pi(\mu, \sigma \mid \mathbf{x_k})$ is also bounded on any set $(\mu, \sigma) \in [-\lambda, \lambda] \times [1/\tau, \tau]$. Then we have

$$\begin{aligned}&\big|\pi(\mu, \sigma \mid \mathbf{x_n}) - \pi(\mu, \sigma \mid \mathbf{x_k})\big| \\ &\quad = \pi(\mu, \sigma \mid \mathbf{x_k}) \left|\frac{\pi(\mu, \sigma \mid \mathbf{x_n})}{\pi(\mu, \sigma \mid \mathbf{x_k})} - 1\right| \to 0 \qquad \text{as } \omega \to \infty.\end{aligned}$$

□

PROOFS OF RESULTS (c) AND (d). We can use Scheffé's theorem [17] directly to prove results (c) and (d). Using Proposition 5, we know that $\pi(\mu, \sigma \mid \mathbf{x_k})$ and $\pi(\mu, \sigma \mid \mathbf{x_n})$ are proper. Using result (b), we have that $\pi(\mu, \sigma \mid \mathbf{x_n}) \to \pi(\mu, \sigma \mid \mathbf{x_k})$ pointwise as $\omega \to \infty$ for any $\mu \in \mathbb{R}$ and $\sigma > 0$, as a result of the uniform convergence. The conditions of Scheffé's theorem are then satisfied, and we obtain the convergence in L_1 given in result (c) as well as the following result:

$$\lim_{\omega \to \infty} \int_E \pi(\mu, \sigma \mid \mathbf{x_n})\, d\mu\, d\sigma = \int_E \pi(\mu, \sigma \mid \mathbf{x_k})\, d\mu\, d\sigma,$$

uniformly for all rectangles E in $\mathbb{R} \times \mathbb{R}^+$. □

PROOF OF RESULT (e). It suffices to write the likelihood functions as $\mathcal{L}(\mu, \sigma \mid \mathbf{x_n}) \propto \sigma\pi(\mu, \sigma \mid \mathbf{x_n})$ and $\mathcal{L}(\mu, \sigma \mid \mathbf{x_k}) \propto \sigma\pi(\mu, \sigma \mid \mathbf{x_k})$ with $\pi(\mu, \sigma) \propto 1/\sigma$, and result (e) follows directly from result (b). □

An attractive feature of Theorem 1 is the simplicity of its only two sufficient conditions. Condition (i) says that modeling must be done using density f of a log-regularly varying distribution with index $\rho \geq 1$; see Definition 2. Note that it involves only the tails of the function $|z| f(z)$. Essentially, the decay of the tails must be logarithmic. For that purpose, in the next section we introduce the family of log-Pareto-tailed symmetric distributions that belong to the family of log-regularly varying distributions.

Condition (ii) requires that $k > l$ and $k > r$. For instance, a group of $k = 6$ nonoutlying observations is sufficient to ensure the rejection of $l = 5$ outliers at left and $r = 5$ at right. The nonoutlying group must be the most important, which is rather intuitive. The most demanding case occurs when all outliers are on the same side (e.g., $l = 0$). Condition (ii) can then be written as $k > n/2$, which means that the nonoutliers must represent more than half of the sample. A few numerical simulations tend to confirm our expectation that a larger difference between k and $\max(l, r)$ results in a faster rejection of the outliers.

The asymptotic behavior of the marginal $m(\mathbf{x_n})$ is given in result (a). This fundamental result is probably of more theoretical than practical interest because it leads to results (b) to (e). The asymptotic behavior of the posterior density is given in result (b). The posterior considering the entire sample converges to the posterior considering only the k nonoutlying observations, uniformly in any set $(\mu, \sigma) \in [-\lambda, \lambda] \times [1/\tau, \tau]$. The outliers are then completely rejected as they are going to plus or minus infinity. We also obtain the pointwise convergence.

In result (c), we obtain the convergence in L_1 of the posterior densities considering the entire sample to the posterior considering only the nonoutlying observations. In result (d), we obtain the convergence in distribution, that is $\Pr(\mu, \sigma \in E \mid \mathbf{x_n})$ converges to $\Pr(\mu, \sigma \in E \mid \mathbf{x_k})$ as $\omega \to \infty$, uniformly for all rectangles E in $\mathbb{R} \times \mathbb{R}^+$. Because the convergence is uniform, this is actually a stronger result than the convergence in distribution, which requires only pointwise convergence. We also obtain the convergence in distribution of the posterior marginal distributions. Therefore, any estimation of μ and σ based on posterior quantiles or Bayesian credible intervals is robust to outliers.

In result (e), the likelihood considering the entire sample converges to the likelihood considering only the nonoutlying observations, uniformly in any set $(\mu, \sigma) \in [-\lambda, \lambda] \times [1/\tau, \tau]$. It follows that the maximum of $\mathcal{L}(\mu, \sigma \mid \mathbf{x_n})$ converges to the maximum of $\mathcal{L}(\mu, \sigma \mid \mathbf{x_k})$, and therefore the maximum likelihood estimates also converge, as $\omega \to \infty$.

4. The family of log-Pareto-tailed symmetric distributions. As stated in Theorem 1, modeling with a log-regularly varying distribution is one of the conditions of robustness. However, such a distribution is super heavy-tailed, and the usual densities defined on $\mathbb{R}$ are light or heavy-tailed. Therefore, we introduce in this section the family of log-Pareto-tailed symmetric distributions that belongs to the larger family of log-regularly varying distributions. Given that the conditions of robustness involve only the tails of density $f(z)$, the proposed solution consists in altering a symmetric density, such as the usual normal, uniform or Student's t distributions, by replacing its extremities with log-Pareto tails, that is, a function proportional to $|z|^{-1}(\log|z|)^{-\beta}$, with $\beta > 1$. This idea comes from the generalized exponential power (GEP) distribution, a family introduced by Angers [5] and revisited in more detail by Desgagné and Angers [9]. The GEP density is essentially a uniform density in the center with a large spectrum of tail behavior,

classified in types I to V, from light to super heavy-tailed. In particular, the GEP of type V is a log-regularly varying distribution because its density has log-Pareto tails. We propose here to generalize the GEP distribution of type V to the family of log-Pareto-tailed symmetric distributions by using any symmetric densities in the center instead of limiting the choice to the uniform density.

DEFINITION 3. A random variable Z has a log-Pareto-tailed symmetric distribution if its density is given by

$$f(z \mid \boldsymbol{\phi}, \alpha, \beta) = K_{(\boldsymbol{\phi},\alpha,\beta)}\left(g(z \mid \boldsymbol{\phi})\mathbb{1}_{[-\alpha,\alpha]}(z) + g(\alpha \mid \boldsymbol{\phi})\frac{\alpha}{|z|}\left(\frac{\log\alpha}{\log|z|}\right)^{\beta}\mathbb{1}_{(\alpha,\infty)}(|z|)\right),$$

where $z \in \mathbb{R}$, $\alpha > 1$, $\beta > 1$, $\mathbb{1}_A(\cdot)$ is an indicator function, and $g(\cdot \mid \boldsymbol{\phi})$ is any density that is symmetric with respect to the origin, continuous and strictly positive on $[-\alpha, \alpha]$, with its vector of parameters given by $\boldsymbol{\phi} \in \boldsymbol{\Phi}$. The normalizing constant is given by

$$K_{(\boldsymbol{\phi},\alpha,\beta)} = \frac{(\beta - 1)}{(2G(\alpha \mid \boldsymbol{\phi}) - 1)(\beta - 1) + 2g(\alpha \mid \boldsymbol{\phi})\alpha\log\alpha},$$

where $G(\alpha \mid \boldsymbol{\phi}) = \int_{-\infty}^{\alpha} g(u \mid \boldsymbol{\phi})\,du$.

In particular, if $g(z \mid \boldsymbol{\phi})$ is a normal density, we say that the random variable Z has a log-Pareto-tailed normal distribution. If $g(z \mid \boldsymbol{\phi})$ is a Student's t density, we say that Z has a log-Pareto-tailed Student's t distribution, and so on. The core of the density $f(z \mid \boldsymbol{\phi}, \alpha, \beta)$ is located between $-\alpha$ and α, and the tails are positioned in the area $|z| > \alpha$. Tail thickness is controlled with the parameter β. This density satisfies the condition of robustness required in Theorem 1, since for $|z| > \alpha$, we have

$$|z| f(z \mid \boldsymbol{\phi}, \alpha, \beta) \propto (\log|z|)^{-\beta} \in L_{\beta}(\infty).$$

All conditions of regularity assumed in Section 3.1 are satisfied as well. The density $f(z \mid \boldsymbol{\phi}, \alpha, \beta)$ is continuous and strictly positive on $\mathbb{R}$, proper (see Proposition 3) and symmetric with respect to the origin. Furthermore, both tails of $|z| f(z \mid \boldsymbol{\phi}, \alpha, \beta)$ are monotonic.

In practice, choosing parameters α and β directly is not necessarily an intuitive task. It could be easier to choose other indirect but related quantities. Here is an interesting strategy in five steps: a practitioner first chooses his favorite symmetric density $g(z \mid \boldsymbol{\phi})$ and its vector of parameters $\boldsymbol{\phi}$ (other than the location and scale parameters μ and σ, which will be added later), such as the $N(0, 1)$. The second step consists in setting the normalizing constant $K_{(\boldsymbol{\phi},\alpha,\beta)}$ to 1. The desirable consequence is that the core (between $-\alpha$ and α) of the density $f(z \mid \boldsymbol{\phi}, \alpha, \beta)$ becomes

exactly the density $g(z \mid \boldsymbol{\phi})$, the familiar density of the user. The third step consists in choosing the mass of the core, which is defined as

$$q = \Pr(-\alpha \le Z \le \alpha \mid \boldsymbol{\phi}, \alpha, \beta).$$

For instance, we could choose $q = 0.95$, which leaves 2.5% of the mass in each tail. Then, the density $f(z \mid \boldsymbol{\phi}, \alpha, \beta)$ would be exactly the $N(0, 1)$ density for 95% of its mass located in the center. The following steps are done automatically. Given that $K_{(\boldsymbol{\phi},\alpha,\beta)}$ has been set to 1, it follows that $q = 2G(\alpha \mid \boldsymbol{\phi}) - 1$. However, to ensure that $\alpha > 1$ as required, we must choose $q > 2G(1 \mid \boldsymbol{\phi}) - 1$. If the last equality is rearranged, it leads us to the fourth step, which consists in calculating α as follows:

$$\alpha = G^{-1}\left(\frac{1+q}{2} \,\middle|\, \boldsymbol{\phi}\right).$$

For example, a $N(0, 1)$ with $q = 0.95$ generates a value of $\alpha = 1.96$. Finally, we calculate β in the fifth step as follows:

$$\beta = 1 + \frac{2g(\alpha \mid \boldsymbol{\phi})\alpha \log \alpha}{1 - q}.$$

Note that this equation is consistent with a normalizing constant of 1, and it satisfies $\beta > 1$ since $\alpha > 1$. Our example gives a value of $\beta = 4.08$.

We compare in Figure 1 the standard normal density (dashed line) to a log-Pareto-tailed standard normal density (solid line), with $q = 0.95$, $K_{(\boldsymbol{\phi},\alpha,\beta)} = 1$, $\alpha = 1.96$ and $\beta = 4.08$. Both densities are identical between $-\alpha$ and α, but differ in the tails.

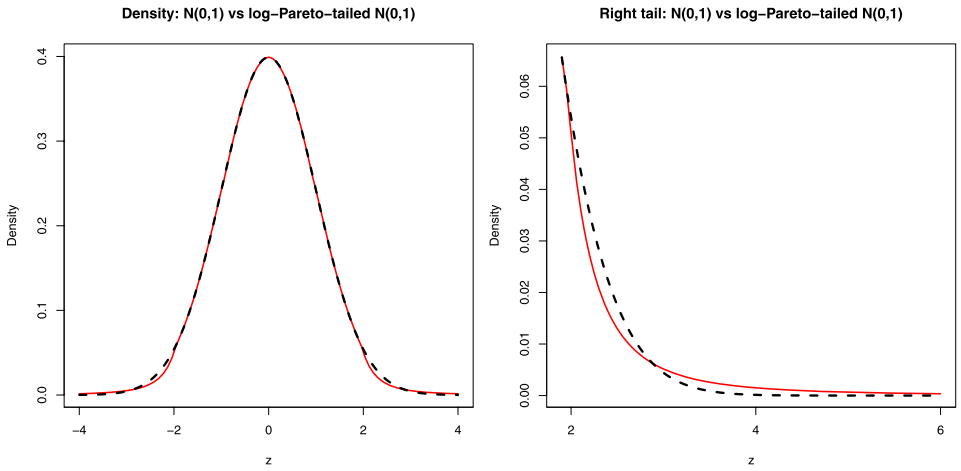

FIG. 1. *A comparison between the standard normal* (*dashed line*) *and log-Pareto-tailed standard normal* (*solid line*) *densities*.

Simulation of observations from a log-Pareto-tailed symmetric distribution is easy using the inverse transformation method. It is described in detail in Section 3.4 of Desgagné and Angers [9] for the log-Pareto-tailed uniform distribution (labeled GEP density of type V in their paper). It is straightforward to generalize it to other symmetric densities $g(\cdot \mid \boldsymbol{\phi})$.

Of course, we can add location and scale parameters, denoted, respectively, by $\mu \in \mathbb{R}$ and $\sigma > 0$, to the density $f(z \mid \boldsymbol{\phi}, \alpha, \beta)$. We obtain

$$(1/\sigma) f\big((z-\mu)/\sigma \mid \boldsymbol{\phi}, \alpha, \beta\big) = \begin{cases} K_{(\boldsymbol{\phi},\alpha,\beta)}(1/\sigma) g\big((z-\mu)/\sigma \mid \boldsymbol{\phi}\big), & \text{if } \mu - \alpha\sigma \le z \le \mu + \alpha\sigma, \\ K_{(\boldsymbol{\phi},\alpha,\beta)} g(\alpha \mid \boldsymbol{\phi}) \dfrac{\alpha}{|z-\mu|} \left(\dfrac{\log \alpha}{\log(|z-\mu|/\sigma)} \right)^{\beta}, & \\ & \text{if } |z-\mu| \ge \alpha\sigma. \end{cases}$$

Note that when this density is used in the context of robustness described in Section 3.2, the parameters $\boldsymbol{\phi}$, α and β are assumed to be known. The inference is done on the location and scale parameters only.

5. Example. In this section, the asymptotic results of robustness found in Theorem 1 are confronted with data. Without loss of generality, we choose the improper and noninformative joint prior density $\pi(\mu, \sigma) \propto 1/\sigma$. Hence, both the Bayesian and frequentist approaches can be used.

We first illustrate in Section 5.1 the behavior of different estimators of the location and scale parameters when one observation moves from 0 to 100, given that the rest of data lie between -10 and 10. For the estimator based on robust modeling provided by Theorem 1, we observe an interesting feature that we call the threshold. The influence of the moving observation on the inference increases until a certain threshold. Then the nature of this observation gradually changes to become more and more outlying, as its influence decreases and eventually completely disappears. In Section 5.2, the performances of concurrent estimators are compared for different scenarios. We consider simulation of observations from the normal as well as from contaminated normal distributions, to see how the estimators perform in the presence—or absence—of outliers. The mean square error is calculated as the measure of performance.

5.1. *Illustration of the threshold.* We consider a sample of size $n = 22$ given by $\mathbf{x_n} = (\mathbf{x_k}, \omega)$, where the $k = 21$ nonoutlying observations are represented by $\mathbf{x_k} = (-10, -9, \ldots, -1, 0, 1, \ldots, 9, 10)$. We study the impact of moving the observation ω from 0 to 100 on the location–scale parameter inference based on the maximum likelihood estimator (MLE) calculated for three different densities f, in accordance with the model described in Section 3.1. Note that results using the

Bayesian marginal posterior median are very similar. Naturally, the standard normal density has been chosen as the nonrobust model. The corresponding MLE are then the usual sample mean and (biased) sample standard deviation.

The log-Pareto-tailed standard normal density, as illustrated in Figure 1, is also studied. We have chosen $q = 0.95$, $\alpha = 1.96$ and $\beta = 4.08$, as discussed in Section 4. This modeling leads to complete rejection of the outlier, as described by Theorem 1. We also examined other values of q (the values of α and β are calculated automatically using the proposed algorithm in Section 4). If we choose a larger value of q, then the density is closer to the $N(0, 1)$, and the same goes for the inference in the absence of outliers. However, the threshold of robustness increases. The choice of 0.95 appeared to be well balanced for good inference with and without outliers.

The third density f considered is the Student's t, a common choice for robust modeling. This density satisfies the conditions of robustness given in Andrade and O'Hagan [3] (which lead to partial robustness concerning the scale parameter), but not the conditions of whole robustness given in Theorem 1. The degrees of freedom has been set to 10, again to search for balance between good inference with and without outliers. An implicit scale parameter of 0.964 (other than σ) has been added to match its interquartile range to that of the two densities considered above.

Robustness for the three models is illustrated in Figure 2. On the x-axis, the observation ω moves from 0 to 100. The estimators $\hat{\mu}$ (left graph) and $\hat{\sigma}$ (right graph) lie on the y-axis.

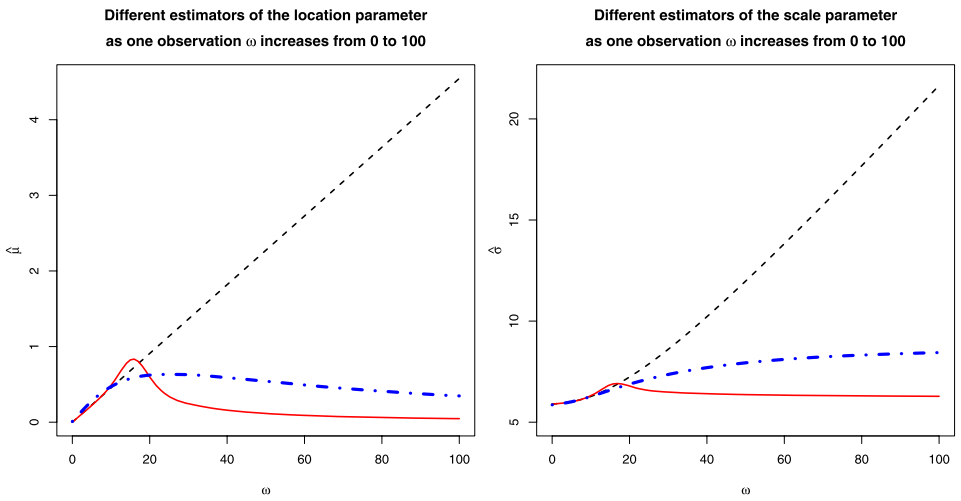

FIG. 2. *Estimation of the location* (*left graph*) *and scale* (*right graph*) *parameters for the normal model* (*dashed lines*), *the log-Pareto-tailed normal model* (*solid lines*) *and the Student's t model* (*dotted-dashed lines*), *using the MLE.*

The influence of the outlier on a nonrobust inference is clearly visible in the normal model (dashed lines) by the estimators growing indefinitely as the outlier increases. For $\omega = 100$, we find $\hat{\mu} = 4.55$ and $\hat{\sigma} = 21.65$. Using the normal quantile of 1.96, this model thus suggests that 95% of the observations should be between -37.9 and 47.0, which is barely supported by data located between -10 and 10, and not at all by the outlier $\omega = 100$.

Whole robustness is illustrated by the log-Pareto-tailed normal model (solid lines). We can see in Figure 2 that ω reaches its maximum influence around 16, where $\hat{\mu}$ and $\hat{\sigma}$ are approximately equal to 0.8 and 7. The influence of ω then begins to decrease after this threshold as $\hat{\mu}$ and $\hat{\sigma}$ eventually converge to their corresponding MLE considering only the nonoutlying observations $\mathbf{x_k}$, given by $\hat{\mu} = 0$ and $\hat{\sigma} = 6.06$. For $\omega = 100$, we find $\hat{\mu} = 0.05$ and $\hat{\sigma} = 6.28$. Using the normal quantile of 1.96 (remember that this model is a standard normal density except for the 2.5% log-Pareto tails), this model thus suggests that 95% of the observations should be between -12.3 and 12.4, which is wholly supported by data, if $\omega = 100$ is considered as an outlier generated from the log-Pareto tails.

Finally, partial robustness is illustrated by the Student's t model (dotted-dashed lines). For $\omega = 100$, we find $\hat{\mu} = 0.35$ and $\hat{\sigma} = 8.44$. Using the appropriate quantile of 2.147, this model thus suggests that 95% of the observations should be between -17.8 and 18.5, which is partially supported by data located between -10 and 10. Note that as ω continues to grow beyond 100, our calculations show that $\hat{\mu}$ decreases toward 0, and $\hat{\sigma}$ continues to grow toward an upper limit of 8.71. This indicates that location estimation using the Student's t is wholly robust. However, scale parameter estimation is only partially robust, in the sense that the inference is contaminated by the outlier, but only to a certain extent.

5.2. *Performance and simulations*. We present here a brief study of the performance of the three models described above (the robust log-Pareto-tailed normal, the partially robust Student's t and the popular but nonrobust normal distributions) under three scenarios of simulations. For each scenario and model, a sample of size $n = 30$ is simulated 25,000 times, and the location and scale parameters are estimated each time using the MLE. Note that again, results using the Bayesian marginal posterior median are very similar. The performance is then measured by the mean square error (MSE). For each scenario, the true values are $\mu = 0$ and $\sigma = 1$. The MSE for the estimation of μ and σ are given in Tables 1 and 2, respectively.

In the first scenario, the samples are simulated from a $N(0, 1)$. We see that in the absence of outliers, the three models obtain the same excellent performance both for the estimation of the location (MSE = 0.03) and the scale (MSE = 0.02). This is rather predictable, because the three densities are very similar, if not identical, except for the tails. The impact of the tails on the estimation is felt mainly in the presence of outliers.

TABLE 1
Mean square error for MLE of μ under different scenarios ($n = 30$)

	Scenario		
Model	**100% $N(0, 1)$**	**10% $N(0, 6)$**	**5% $N(8, 1)$**
Log-Pareto-tailed normal	0.03	0.05	0.07
Student's t	0.03	0.06	0.09
Normal	0.03	0.15	0.29

In the second scenario, we consider a mixture of normal distributions, where an observation has a 90% probability of being generated from a $N(0, 1)$ and 10% from a $N(0, 6)$. A mixture of normal distributions is also studied in the third scenario, where on average 95% of the observations are generated from a $N(0, 1)$ and the remaining 5% from a $N(8, 1)$.

As for the estimation of μ, we can see in Table 1 that both log-Pareto-tailed normal and Student's t models give very similar MSE for the two contaminated scenarios (0.05 to 0.09), slightly larger than those of the 100% $N(0, 1)$ scenario without outliers. However, the normal model is clearly affected by the outliers as its MSE increases to 0.15 and 0.29, respectively, for the second and third scenarios.

The picture for the estimation of σ is a bit different, as can be seen in Table 2. For both scenarios, the performance of the three models can be markedly discriminated in accordance with known theory. The MSE are around 0.10 for the robust log-Pareto normal model, around 0.30 for the partially robust Student t model and above 1 for the nonrobust normal model.

6. Conclusion. Complete rejection of outliers has been investigated in a location–scale parameter model. The analysis has been done primarily in a Bayesian context, but it has been extended to the frequentist approach with maximum likelihood estimators. Essentially, asymptotic robustness is guaranteed if modeling is done using a log-regularly varying distribution (with logarithmic tail decay) and if $k > \max(l, r)$, that is, if the number of nonoutliers is larger than both

TABLE 2
Mean square error for MLE of σ under different scenarios ($n = 30$)

	Scenario		
Model	**100% $N(0, 1)$**	**10% $N(0, 6)$**	**5% $N(8, 1)$**
Log-Pareto-tailed normal	0.02	0.11	0.09
Student's t	0.02	0.32	0.30
Normal	0.02	1.46	1.14

the number of outliers at $-\infty$ and at $+\infty$. The first condition is easy to verify because it involves only the tails of a density through a limit; there are no integrals, derivatives or distribution functions involved. The second condition is quite reasonable and intuitive.

We obtain the uniform convergence of the posterior density given the complete sample to the density considering only the nonoutlying observations. We also obtain the convergence in L_1, the convergence in distribution, as well as the uniform convergence of the likelihoods. Therefore any estimation of the location and scale parameters based on posterior quantiles or the maximum likelihood estimates is robust to outliers.

Even if the results are asymptotic, they are still useful in practice with data, as illustrated by the threshold feature in Section 5.1. When one observation moves away from the rest of data, its influence on the inference begins to increase gradually, because it brings additional information that helps us discriminate among the possible values of the parameter. However, there comes a point where this moving observation conflicts with the rest of data. When this threshold is reached, the model automatically resolves the conflict by progressively reducing the influence of the outlying observation. As the conflict grows infinitely, the impact of the outlier completely disappears. This built-in feature is attractive in practice in that conflict is managed in a sensitive and automatic way.

Estimating the location and scale parameters is common in statistics, using, for instance, the well-known sample mean and standard deviation. Results found in this paper can be readily used in practice to address this problem in a robust way, whether one prefers the Bayesian approach or maximum likelihood estimation. We consider a realistic sample of any size with multiple possible outliers in any direction. The assumption of a symmetric density f with the same tail behavior seems reasonable for most of the applications. Because we do not know beforehand which observations are going to be outlying, it is generally desirable to give each density and each tail the same weight, and to let the largest group dominate in case of conflict. The choice of the appropriate density is addressed in a practical way by introducing the family of log-Pareto-tailed symmetric distributions. Furthermore, the model allows us to add prior information on the location and scale through a very general joint prior density, which includes the possibility to choose a noninformative prior.

This paper can be generalized in different ways. For instance, we can consider asymmetric densities f with different tail behavior. The family of log-regularly varying distributions could be widened to consider, for instance, distributions with a right tail proportional to $(1/z)\exp(-\delta(\log z)^\gamma)$, with $0 < \gamma < 1$ and $\delta > 0$, which is an exponential transformation of the function $\exp(-\delta z^\gamma)$. Robustness to misspecification of the prior can also be investigated.

7. Proofs. The proof of Proposition 5 is given in Section 7.1, and the proof of result (a) of Theorem 1 is given in Section 7.2.

7.1. *Proof of Proposition 5.* To prove that $\pi(\mu, \sigma \mid \mathbf{x_n})$ is proper [the proof for $\pi(\mu, \sigma \mid \mathbf{x_k})$ is omitted because it is similar], it suffices to show that the marginal $m(\mathbf{x_n})$ is finite. Without loss of generality, we assume for convenience that $x_1 < x_2 < \cdots < x_n$. We also define the constant $\delta > 0$ as half the minimum distance between two observations, that is,

$$\delta = \min_{i \in \{1,\ldots,n-1\}} \{(x_{i+1} - x_i)/2\}.$$

We first consider $\mu \in \mathbb{R}$ and $\delta/M \le \sigma < \infty$, where M is the constant of monotonicity given in equation (1). Then we have

$$\begin{aligned}
&\int_{\delta/M}^{\infty} \int_{-\infty}^{\infty} \pi(\mu, \sigma) \prod_{i=1}^{n} (1/\sigma) f\big((x_i - \mu)/\sigma\big)\, d\mu\, d\sigma \\
&\quad \overset{a}{\le} B^n \int_{\delta/M}^{\infty} (1/\sigma)^n \int_{-\infty}^{\infty} (1/\sigma) f\big((x_1 - \mu)/\sigma\big)\, d\mu\, d\sigma \\
&\quad \overset{b}{=} B^n \int_{\delta/M}^{\infty} (1/\sigma)^n\, d\sigma \int_{-\infty}^{\infty} f\big(\mu'\big)\, d\mu' \\
&\quad \overset{c}{=} B^n (M/\delta)^{n-1}/(n-1) < \infty.
\end{aligned}$$

In step a, we bound $\sigma\pi(\mu, \sigma)$ and $n-1$ densities f by B, where B is given in (2). In step b, we use the change of variable $\mu' = (x_1 - \mu)/\sigma$. In step c, we use $n \ge 2$ as assumed in the Bayesian context given in Section 3.1.

We now consider $(x_{j-1} + x_j)/2 \le \mu \le (x_j + x_{j+1})/2$, for $j = 1, \ldots, n$ and $0 < \sigma \le \delta/M$. If we define $x_0 := -\infty$ and $x_{n+1} := \infty$, the union of these n mutually disjoint intervals constitutes the real line, that is, $-\infty < \mu < \infty$. Then we have

$$\begin{aligned}
&\pi(\mu, \sigma) \prod_{i=1}^{n} (1/\sigma) f\big((x_i - \mu)/\sigma\big) \\
&\quad \overset{a}{\le} (1/\sigma) B \prod_{i=1}^{n} (1/\sigma) f\big((x_i - \mu)/\sigma\big) \\
&\quad = (1/\sigma) B f\big((x_j - \mu)/\sigma\big) \times (1/\sigma) \prod_{i=1\ (i \ne j)}^{n} (1/\sigma) f\big((x_i - \mu)/\sigma\big) \\
&\quad \overset{b}{\le} (1/\sigma) B f\big((x_j - \mu)/\sigma\big) \times (1/\sigma) \big[(1/\sigma) f(\delta/\sigma)\big]^{n-1} \\
&\quad \overset{c}{\le} B (B/\delta)^{n-2} (1/\sigma) f\big((x_j - \mu)/\sigma\big) \times (1/\sigma)^2 f(\delta/\sigma) \\
&\quad \propto (1/\sigma) f\big((x_j - \mu)/\sigma\big) \times \big(\delta/\sigma^2\big) f(\delta/\sigma).
\end{aligned}$$

In step a, we bound $\sigma\pi(\mu, \sigma)$ by B. In step b, we use $f((x_i - \mu)/\sigma) \le f(\delta/\sigma)$ by the monotonicity of the tails of $f(z)$ since $|x_i - \mu|/\sigma \ge \delta/\sigma \ge \delta(M/\delta) = M$,

because if $i \neq j$, we have

$$|x_i - \mu| \geq \min\{(x_j - x_{j-1})/2, (x_{j+1} - x_j)/2\} \geq \delta.$$

In step c, we bound $(1/\sigma) f(\delta/\sigma)$ by B/δ for $n-2$ terms. Finally, we have

$$\int_0^{\delta/M} (\delta/\sigma^2) f(\delta/\sigma) \int_{(x_{j-1}+x_j)/2}^{(x_j+x_{j+1})/2} (1/\sigma) f\big((x_j - \mu)/\sigma\big)\, d\mu\, d\sigma$$
$$\leq \int_0^{\infty} f(\sigma')\, d\sigma' \int_{-\infty}^{\infty} f(\mu')\, d\mu' = 1/2 < \infty,$$

where we use the changes of variable $\sigma' = \delta/\sigma$ and $\mu' = (x_j - \mu)/\sigma$.

7.2. *Proof of result* (a) *of Theorem* 1. Consider the model described in Section 3.1, and assume that the conditions of regularity on f are satisfied. We also assume that $zf(z) \in L_\rho(\infty)$ and $k > \max(l, r)$, as given in Theorem 1. Two lemmas are first given, and the proof of result (a) follows.

LEMMA 1. $\forall \lambda \geq 0$, $\forall \tau \geq 1$, *there exists a constant* $D(\lambda, \tau) \geq 1$ *such that* $z \in \mathbb{R}$ *and* $(\mu, \sigma) \in [-\lambda, \lambda] \times [1/\tau, \tau] \Rightarrow$

$$1/D(\lambda, \tau) \leq (1/\sigma) f\big((z - \mu)/\sigma\big)/f(z) \leq D(\lambda, \tau).$$

PROOF. Proposition 4 states that $(1/\sigma) f((z-\mu)/\sigma)/f(z)$ converges to 1 uniformly in any set $(\mu, \sigma) \in E_{\lambda,\tau}$ as $z \to \infty$, where $E_{\lambda,\tau} = [-\lambda, \lambda] \times [1/\tau, \tau]$. Hence, $\forall \lambda \geq 0$ and $\forall \tau \geq 1$, the ratio $(1/\sigma) f((z - \mu)/\sigma)/f(z)$ can be bounded, say by $1/1.01$ and 1.01, if $|z|$ is larger than a certain constant, say $A(\lambda, \tau)$, using the symmetry of f. Therefore, we choose $D(\lambda, \tau) \geq 1.01$.

If $-A(\lambda, \tau) \leq z \leq A(\lambda, \tau)$, we observe that $|z - \mu|/\sigma$ is also bounded on $(\mu, \sigma) \in E_{\lambda,\tau}$. Therefore, since f is continuous and strictly positive on $\mathbb{R}$, it follows that $\forall \lambda \geq 0$ and $\forall \tau \geq 1$, we can find a constant $D(\lambda, \tau) \geq 1.01$ as large as we want such that the ratio $(1/\sigma) f((z - \mu)/\sigma)/f(z)$ is bounded below by $1/D(\lambda, \tau)$ and above by $D(\lambda, \tau)$, for any $(\mu, \sigma) \in E_{\lambda,\tau}$. □

LEMMA 2. *There exists a constant* $C > 0$ *such that*

$$|z| \geq 2M \quad \Longrightarrow \quad \sup_{\mu \in \mathbb{R}} \frac{f(\mu) f(z - \mu)}{f(z)} \leq C,$$

where M *is given in equation* (1).

PROOF. Let the constant $C = 2D(0, 2)B$, where B is given in equation (2), and $D(0, 2)$ comes from Lemma 1. Consider $|z| \geq 2M$.

First, consider $0 \leq |\mu| \leq |z|/2$. We have

$$\frac{f(\mu) f(z - \mu)}{f(z)} \overset{a}{\leq} \frac{f(\mu) f(z/2)}{f(z)} \overset{b}{\leq} 2D(0, 2) f(\mu) \overset{c}{\leq} 2D(0, 2)B = C.$$

In step a, we use $f(z-\mu) \le f(z/2)$ by the monotonicity of the tails of f since $|z-\mu| \ge |z|/2 \ge (2M)/2 = M$. In step b, we use $(1/2)f(z/2)/f(z) \le D(0,2)$ using Lemma 1. In step c, we bound f by B.

Second, consider $|z|/2 \le |\mu| < \infty$. We have

$$\frac{f(\mu)f(z-\mu)}{f(z)} \le \frac{f(z/2)f(z-\mu)}{f(z)} \le 2D(0,2)f(z-\mu) \le 2D(0,2)B = C,$$

using $f(\mu) \le f(z/2)$ in the first inequality by the monotonicity of the tails of f since $|\mu| \ge |z|/2 \ge (2M)/2 = M$ and the same arguments as above for the other inequalities. $\square$

We first observe that

$$\begin{aligned}
&\frac{m(\mathbf{x_n})}{m(\mathbf{x_k})\prod_{i=1}^n [f(x_i)]^{l_i+r_i}} \\
&\quad = \frac{m(\mathbf{x_n})}{m(\mathbf{x_k})\prod_{i=1}^n [f(x_i)]^{l_i+r_i}} \int_{-\infty}^{\infty}\int_0^{\infty} \pi(\mu,\sigma \mid \mathbf{x_n})\,d\sigma\,d\mu \\
&\quad = \int_{-\infty}^{\infty}\int_0^{\infty} \frac{\pi(\mu,\sigma)\prod_{i=1}^n [(1/\sigma)f((x_i-\mu)/\sigma)]^{k_i+l_i+r_i}}{m(\mathbf{x_k})\prod_{i=1}^n [f(x_i)]^{l_i+r_i}}\,d\sigma\,d\mu \\
&\quad = \int_{-\infty}^{\infty}\int_0^{\infty} \pi(\mu,\sigma \mid \mathbf{x_k}) \prod_{i=1}^n \left[\frac{(1/\sigma)f((x_i-\mu)/\sigma)}{f(x_i)}\right]^{l_i+r_i} d\sigma\,d\mu.
\end{aligned}$$

Therefore, we show that the last integral converges to 1 as $\omega \to \infty$ to prove result (a). If we use Lebesgue's dominated convergence theorem to pass the limit $\omega \to \infty$ inside the integral, we have

$$\begin{aligned}
&\lim_{\omega\to\infty} \int_{-\infty}^{\infty}\int_0^{\infty} \pi(\mu,\sigma \mid \mathbf{x_k}) \prod_{i=1}^n \left[\frac{(1/\sigma)f((x_i-\mu)/\sigma)}{f(x_i)}\right]^{l_i+r_i} d\sigma\,d\mu \\
&\quad = \int_{-\infty}^{\infty}\int_0^{\infty} \lim_{\omega\to\infty} \pi(\mu,\sigma \mid \mathbf{x_k}) \prod_{i=1}^n \left[\frac{(1/\sigma)f((x_i-\mu)/\sigma)}{f(x_i)}\right]^{l_i+r_i} d\sigma\,d\mu \\
&\quad = \int_{-\infty}^{\infty}\int_0^{\infty} \pi(\mu,\sigma \mid \mathbf{x_k})\,d\sigma\,d\mu = 1,
\end{aligned}$$

using Proposition 4 in the second equality and Proposition 5 in the last one. Note that pointwise convergence is sufficient, for any value of $\mu \in \mathbb{R}$ and $\sigma > 0$, once the limit is passed inside the integral.

However, in order to use Lebesgue's dominated convergence theorem, we need to show that $\pi(\mu,\sigma \mid \mathbf{x_k})\prod_{i=1}^n [(1/\sigma)f((x_i-\mu)/\sigma)/f(x_i)]^{l_i+r_i}$ is bounded, for any value of $\omega \ge x_0$, by an integrable function of μ and σ that does not depend on ω. The constant x_0 can be chosen as large as we want, and some minimum values for x_0 will be given throughout the proof.

To achieve this, we divide the domain of integration into four quadrants delineated by the axes $\mu = 0$ and $\sigma = 1$. Note that the proofs are only given for the two quadrants in the region of $\mu \geq 0$ because the proofs for $\mu < 0$ are similar.

We choose the constant x_0 larger than a certain threshold such that the ranking of the set $\{|x_i| : l_i + r_i = 1\}$ remain unchanged for all $\omega \geq x_0$. Given that each observation x_i can be written as $x_i = a_i + b_i\omega$, with $b_i = 0$ if $k_i = 1$, $b_i < 0$ if $l_i = 1$ and $b_i > 0$ if $r_i = 1$, the ranking is therefore primarily determined by the values of $|b_i|$. Then, without loss of generality, we assume for convenience that

$$\min_{i\,:\,l_i+r_i=1}\{|b_i|\} = 1 \quad \text{and} \quad \omega = \min_{i\,:\,l_i+r_i=1}\{|x_i|\}.$$

If $l_i + r_i = 1$, we can use Lemma 1, with $x_i = a_i + b_i\omega = b_i(\omega + a_i/b_i)$ and $|b_i| \geq 1$, to establish that the ratio $f(x_i)/f(\omega)$ is bounded, precisely by

$$1/D\big(|a_i/b_i|, |b_i|\big) \leq |b_i| f(x_i)/f(\omega) \leq D\big(|a_i/b_i|, |b_i|\big).$$

QUADRANT 1. Consider $0 \leq \mu < \infty$ and $1 \leq \sigma < \infty$. We have

$$\begin{aligned}
\pi(\mu, \sigma \mid \mathbf{x_k}) \prod_{i=1}^{n} \left[\frac{(1/\sigma) f((x_i - \mu)/\sigma)}{f(x_i)}\right]^{l_i+r_i}
&\propto \frac{\pi(\mu, \sigma)}{\sigma^n} \prod_{i=1}^{n} \frac{f((x_i - \mu)/\sigma)}{[f(x_i)]^{l_i+r_i}} \\
&\overset{a}{\leq} \frac{B}{\sigma^{n+1}} \prod_{i=1}^{n} \frac{D(|a_i|, 1) f((b_i\omega - \mu)/\sigma)}{[f(x_i)]^{l_i+r_i}} \\
&\overset{b}{\leq} \frac{1}{[f(\omega)]^{l+r}} \frac{B}{\sigma^{n+1}} \prod_{i=1}^{n} D\big(|a_i|, 1\big) f\big((b_i\omega - \mu)/\sigma\big)\big[|b_i| D\big(|a_i/b_i|, |b_i|\big)\big]^{l_i+r_i} \\
&\propto \frac{1}{[f(\omega)]^{l+r}} \frac{1}{\sigma^{n+1}} \prod_{i=1}^{n} f\big((b_i\omega - \mu)/\sigma\big) \\
&\overset{c}{=} \frac{1}{[f(\omega)]^{l+r}} \frac{[f(\mu/\sigma)]^k}{\sigma^{n+1}} \prod_{i=1}^{n} \big[f\big((b_i\omega - \mu)/\sigma\big)\big]^{l_i+r_i} \\
&= \frac{(1/\sigma) f(\mu/\sigma)}{\sigma^{k-1/2}} \left[\frac{\omega/\sigma}{\omega f(\omega)}\right]^{l+r} \frac{[f(\mu/\sigma)]^{k-1}}{\sigma^{1/2}} \prod_{i=1}^{n} \big[f\big((b_i\omega - \mu)/\sigma\big)\big]^{l_i+r_i}.
\end{aligned}$$

In step a, we use $x_i = a_i + b_i\omega$ and

$$f\big((x_i - \mu)/\sigma\big) = f\big((b_i\omega - \mu)/\sigma + a_i/\sigma\big) \leq D\big(|a_i|, 1\big) f\big((b_i\omega - \mu)/\sigma\big)$$

using Lemma 1 since $|a_i/\sigma| \leq |a_i|$. We also bound $\sigma\pi(\mu, \sigma)$ by B. In step b, we use $1/f(x_i) \leq |b_i| D(|a_i/b_i|, |b_i|)/f(\omega)$. In step c, we set $b_i = 0$ if $k_i = 1$ and we use $f(-\mu/\sigma) = f(\mu/\sigma)$ by symmetry of f.

It suffices to show that

$$\left[\frac{\omega/\sigma}{\omega f(\omega)}\right]^{l+r}\frac{[f(\mu/\sigma)]^{k-1}}{\sigma^{1/2}}\prod_{i=1}^{n}\left[f\big((b_i\omega-\mu)/\sigma\big)\right]^{l_i+r_i}<\infty,$$

since $(1/\sigma)^{k-1/2}(1/\sigma)f(\mu/\sigma)$ is an integrable function on Quadrant 1,

$$\int_1^\infty(1/\sigma)^{k-1/2}\int_0^\infty(1/\sigma)f(\mu/\sigma)\,d\mu\,d\sigma\le\int_1^\infty(1/\sigma)^{k-1/2}\,d\sigma=\frac{1}{k-3/2}\le 2,$$

since $k\ge 2$. To achieve this, we split the region of σ into three parts between $1<\omega^{1/2}<\omega/(2M)<\infty$, where M is defined in equation (1). Note that since $\omega\ge x_0$, this is well defined if $x_0>\max(1,(2M)^2)$.

Consider $0\le\mu<\infty$ and $\omega/(2M)\le\sigma<\infty$. Then we have

$$\begin{aligned}
&\left[\frac{\omega/\sigma}{\omega f(\omega)}\right]^{l+r}\frac{[f(\mu/\sigma)]^{k-1}}{\sigma^{1/2}}\prod_{i=1}^{n}\left[f\big((b_i\omega-\mu)/\sigma\big)\right]^{l_i+r_i}\\
&\quad\overset{a}{\le}B^{n-1}\left[\frac{\omega/\sigma}{\omega f(\omega)}\right]^{l+r}\frac{1}{\sigma^{1/2}}\overset{b}{\le}B^{n-1}(2M)^{l+r+1/2}\frac{(1/\omega)^{1/2}}{[\omega f(\omega)]^{l+r}}\\
&\quad\overset{c}{\le}B^{n-1}(2M)^{l+r+1/2}\frac{(1/\omega)^{1/2}}{(\log\omega)^{-(\rho+1)(l+r)}}\\
&\quad\overset{d}{\le}B^{n-1}(2M)^{l+r+1/2}\big[2(\rho+1)(l+r)/e\big]^{(\rho+1)(l+r)}<\infty.
\end{aligned}$$

In step a, we use $f(\cdot)\le B$. In step b, we use $\omega/\sigma\le 2M$ and $(1/\sigma)\le(2M)/\omega$. In step c, we use $\omega f(\omega)>(\log\omega)^{-\rho-1}$ if $\omega\ge x_0\ge A(1)$, where $A(1)$ comes from Proposition 2. In step d, it is purely algebraic to show that the maximum of $(\log\omega)^\beta/\omega^{1/2}$ is $(2\beta/e)^\beta$ for $\omega>1$ and $\beta>0$, where $\beta=(\rho+1)(l+r)$ in our equation.

Now consider the two other parts combined (we will split them in the next step), that is, $0\le\mu<\infty$ and $1\le\sigma\le\omega/(2M)$. We have

$$\begin{aligned}
&\left[\frac{\omega/\sigma}{\omega f(\omega)}\right]^{l+r}\frac{[f(\mu/\sigma)]^{k-1}}{\sigma^{1/2}}\prod_{i=1}^{n}\left[f\big((b_i\omega-\mu)/\sigma\big)\right]^{l_i+r_i}\\
&\quad\overset{a}{\le}\left[\frac{\omega/\sigma}{\omega f(\omega)}\right]^{l+r}\frac{[f(\mu/\sigma)]^{k-1}}{\sigma^{1/2}}\prod_{i=1}^{n}\left[f(b_i\omega/\sigma)\right]^{l_i}\left[f\big((b_i\omega-\mu)/\sigma\big)\right]^{r_i}\\
&\quad=\frac{[f(\mu/\sigma)]^{k-r-1}}{\sigma^{1/2}}\prod_{i=1}^{n}\left[\frac{(\omega/\sigma)f(b_i\omega/\sigma)}{\omega f(\omega)}\right]^{l_i+r_i}\left[\frac{f(b_i\omega/\sigma-\mu/\sigma)f(\mu/\sigma)}{f(b_i\omega/\sigma)}\right]^{r_i}\\
&\quad\overset{b}{\le}B^{k-r-1}C^r\frac{1}{\sigma^{1/2}}\prod_{i=1}^{n}\left[\frac{(\omega/\sigma)f(b_i\omega/\sigma)}{\omega f(\omega)}\right]^{l_i+r_i}\\
&\quad\overset{c}{\le}B^{k-r-1}C^r\frac{1}{\sigma^{1/2}}\left[\frac{(\omega/\sigma)f(\omega/\sigma)}{\omega f(\omega)}\right]^{l+r}.
\end{aligned}$$

In step a, we use $f((b_i\omega-\mu)/\sigma)\le f(b_i\omega/\sigma)$ if $l_i=1$ (which means $b_i<0$) by the monotonicity of the tails of f since $|b_i\omega-\mu|/\sigma\ge|b_i|\omega/\sigma\ge|b_i|(2M)\ge 2M\ge M$. In step b, we use $f(\mu/\sigma)\le B$ and we use Lemma 2 since $|b_i|\omega/\sigma\ge|b_i|(2M)\ge 2M$. In step c, we use $f(b_i\omega/\sigma)\le f(\omega/\sigma)$ by the monotonicity of the tails of f since $|b_i|\omega/\sigma\ge\omega/\sigma\ge 2M\ge M$.

Consider $0\le\mu<\infty$ and $\omega^{1/2}\le\sigma\le\omega/(2M)$. We have

$$\begin{aligned}\frac{1}{\sigma^{1/2}}\left[\frac{(\omega/\sigma)f(\omega/\sigma)}{\omega f(\omega)}\right]^{l+r} &\overset{a}{\le} B^{l+r}\frac{(1/\omega)^{1/4}}{[\omega f(\omega)]^{l+r}}\\ &\overset{b}{\le} B^{l+r}\frac{(1/\omega)^{1/4}}{(\log\omega)^{-(\rho+1)(l+r)}}\\ &\overset{c}{\le} B^{l+r}\big[4(\rho+1)(l+r)/e\big]^{(\rho+1)(l+r)}<\infty.\end{aligned}$$

In step a, we use $(\omega/\sigma)f(\omega/\sigma)\le B$ and $(1/\sigma)^{1/2}\le(1/\omega)^{1/4}$. In step b, we use $\omega f(\omega)>(\log\omega)^{-\rho-1}$ if $\omega\ge x_0\ge A(1)$, where $A(1)$ comes from Proposition 2. In step c, it is purely algebraic to show that the maximum of $(\log\omega)^{\beta}/\omega^{1/4}$ is $(4\beta/e)^{\beta}$ for $\omega>1$ and $\beta>0$, where $\beta=(\rho+1)(l+r)$ in our equation.

Finally consider $0\le\mu<\infty$ and $1\le\sigma\le\omega^{1/2}$. Then we have

$$\frac{1}{\sigma^{1/2}}\left[\frac{(\omega/\sigma)f(\omega/\sigma)}{\omega f(\omega)}\right]^{l+r}\overset{a}{\le}\left[\frac{\omega^{1/2}f(\omega^{1/2})}{\omega f(\omega)}\right]^{l+r}\overset{b}{\le}2^{(\rho+1)(l+r)}<\infty.$$

In step a, we use $1/\sigma\le 1$, and we use $(\omega/\sigma)f(\omega/\sigma)\le\omega^{1/2}f(\omega^{1/2})$ by the monotonicity of the tails of $|z|f(z)$ since $\omega/\sigma\ge\omega^{1/2}\ge x_0^{1/2}\ge M$ if $x_0\ge M^2$. In step b, we use $\omega^{1/2}f(\omega^{1/2})/(\omega f(\omega))\le 2(1/2)^{-\rho}=2^{\rho+1}$ if $\omega\ge x_0\ge A(1,2)$, where $A(1,2)$ comes from the definition of a log-regularly varying function.

QUADRANT 2. Consider $-\infty<\mu\le 0$ and $1\le\sigma<\infty$. The proof for Quadrant 2 is similar to that of Quadrant 1.

QUADRANT 3. Consider $-\infty<\mu\le 0$ and $0<\sigma\le 1$. The proof for Quadrant 3 is similar to that of Quadrant 4, given below. The condition $k>r$ is therefore replaced by $k>l$. Note that $k>\max(l,r)$ is assumed in Theorem 1.

QUADRANT 4. Consider $0\le\mu<\infty$ and $0<\sigma\le 1$. We need to show, actually, that

$$\begin{aligned}&\lim_{\omega\to\infty}\int_0^{\infty}\int_0^{1}\pi(\mu,\sigma\mid\mathbf{x_k})\prod_{i=1}^{n}\left[\frac{(1/\sigma)f((x_i-\mu)/\sigma)}{f(x_i)}\right]^{l_i+r_i}d\sigma\,d\mu\\ &\quad=\int_0^{\infty}\int_0^{1}\pi(\mu,\sigma\mid\mathbf{x_k})\,d\sigma\,d\mu.\end{aligned}$$

For Quadrant 1, we show this result when we integrate σ between 1 and ∞. We bound the integrand of the left term, for any value of $\omega \geq x_0$, by an integrable function of μ and σ that does not depend on ω, in order to use Lebesgue's dominated convergence theorem to pass the limit $\omega \to \infty$ inside the integral. For Quadrant 4, we proceed slightly differently. We begin by breaking down the left term into two parts as follows:

$$\begin{aligned}
&\lim_{\omega\to\infty}\int_0^\infty\int_0^1 \pi(\mu,\sigma \mid \mathbf{x_k})\prod_{i=1}^n\left[\frac{(1/\sigma)f((x_i-\mu)/\sigma)}{f(x_i)}\right]^{l_i+r_i} d\sigma\, d\mu \\
&\quad=\lim_{\omega\to\infty}\int_0^\infty\int_0^1 \pi(\mu,\sigma \mid \mathbf{x_k})\prod_{i=1}^n\left[\frac{(1/\sigma)f((x_i-\mu)/\sigma)}{f(x_i)}\right]^{l_i+r_i} \\
&\qquad\qquad\times \mathbb{1}_{[0,\omega/2]}(\mu)\, d\sigma\, d\mu \\
&\qquad+\lim_{\omega\to\infty}\int_{\omega/2}^\infty\int_0^1 \pi(\mu,\sigma \mid \mathbf{x_k})\prod_{i=1}^n\left[\frac{(1/\sigma)f((x_i-\mu)/\sigma)}{f(x_i)}\right]^{l_i+r_i} d\sigma\, d\mu,
\end{aligned}$$

where the indicator function $\mathbb{1}_A(\mu)$ is equal to 1 if $\mu \in A$, and equal to 0 otherwise. We then show that the first part is equal to the integral $\int_0^\infty \int_0^1 \pi(\mu,\sigma \mid \mathbf{x_k})\, d\sigma\, d\mu$, and the second part is equal to 0.

For the first equality, we again use Lebesgue's dominated convergence theorem to pass the limit $\omega \to \infty$ inside the integral. We have

$$\begin{aligned}
&\lim_{\omega\to\infty}\int_0^\infty\int_0^1 \pi(\mu,\sigma \mid \mathbf{x_k})\prod_{i=1}^n\left[\frac{(1/\sigma)f((x_i-\mu)/\sigma)}{f(x_i)}\right]^{l_i+r_i} \mathbb{1}_{[0,\omega/2]}(\mu)\, d\sigma\, d\mu \\
&\quad=\int_0^\infty\int_0^1 \pi(\mu,\sigma \mid \mathbf{x_k})\lim_{\omega\to\infty}\prod_{i=1}^n\left[\frac{(1/\sigma)f((x_i-\mu)/\sigma)}{f(x_i)}\right]^{l_i+r_i} \\
&\qquad\qquad\times \mathbb{1}_{[0,\omega/2]}(\mu)\, d\sigma\, d\mu \\
&\quad=\int_0^\infty\int_0^1 \pi(\mu,\sigma \mid \mathbf{x_k})\times 1\times \mathbb{1}_{[0,\infty)}(\mu)\, d\sigma\, d\mu \\
&\quad=\int_0^\infty\int_0^1 \pi(\mu,\sigma \mid \mathbf{x_k})\, d\sigma\, d\mu,
\end{aligned}$$

using Proposition 4 in the second equality. Note that pointwise convergence is sufficient, for any value of $\mu \in \mathbb{R}$ and $\sigma > 0$, once the limit is passed inside the integral. However, in order to use Lebesgue's dominated convergence theorem, we need to show that for any value of $\omega \geq x_0$, the integrand is bounded by an integrable function of μ and σ that does not depend on ω.

Consider $0 \le \mu \le \omega/2$ (the integrand is equal to 0 if $\omega/2 < \mu < \infty$) and $0 < \sigma \le 1$. We have

$$\begin{aligned}
&\pi(\mu,\sigma \mid \mathbf{x_k}) \prod_{i=1}^{n} \left[\frac{(1/\sigma) f((x_i-\mu)/\sigma)}{f(x_i)}\right]^{l_i+r_i} \mathbb{1}_{[0,\omega/2]}(\mu) \\
&\quad \overset{a}{\le} \pi(\mu,\sigma \mid \mathbf{x_k}) \prod_{i=1}^{n} \left[2D(0,2)\frac{(1/\sigma) f((x_i-\mu)/\sigma)}{f(x_i/2)}\right]^{l_i+r_i} \\
&\quad \propto \pi(\mu,\sigma \mid \mathbf{x_k}) \prod_{i=1}^{n} \left[\frac{(1/\sigma) f((x_i-\mu)/\sigma)}{f(x_i/2)}\right]^{l_i+r_i} \\
&\quad \overset{b}{\le} \pi(\mu,\sigma \mid \mathbf{x_k}) \prod_{i=1}^{n} \left[\frac{f(x_i-\mu)}{f(x_i/2)}\right]^{l_i+r_i} \\
&\quad \overset{c}{\le} \pi(\mu,\sigma \mid \mathbf{x_k}),
\end{aligned}$$

and $\pi(\mu,\sigma \mid \mathbf{x_k})$ is an integrable function. In step a, we use $\mathbb{1}_{[0,\omega/2]}(\mu) = 1$ and $(1/2) f(x_i/2)/f(x_i) \le D(0,2)$ using Lemma 1. In step b, we use $(|x_i - \mu|/\sigma) f((x_i-\mu)/\sigma) \le |x_i - \mu| f(x_i - \mu)$ by the monotonicity of the tails of $|z| f(z)$, and in step c we use $f(x_i - \mu) \le f(x_i/2)$ by the monotonicity of the tails of $f(z)$ since $|x_i - \mu|/\sigma \ge |x_i - \mu| \ge |x_i|/2 \ge \omega/2 \ge x_0/2 \ge M$, if we choose $x_0 \ge 2M$. Note that the condition $\mu \le \omega/2 (\le x_i/2)$ is used only to justify $|x_i - \mu| \ge |x_i|/2$ when $r_i = 1$.

Now we show the second equality, that is,

$$\lim_{\omega\to\infty} \int_{\omega/2}^{\infty} \int_{0}^{1} \pi(\mu,\sigma \mid \mathbf{x_k}) \prod_{i=1}^{n} \left[\frac{(1/\sigma) f((x_i-\mu)/\sigma)}{f(x_i)}\right]^{l_i+r_i} d\sigma\, d\mu = 0.$$

We first bound above the integrand, and then we show that the integral of the upper bound converges to 0 as $\omega \to \infty$.

Consider $\omega/2 \le \mu < \infty$ and $0 < \sigma \le 1$. We have

$$\begin{aligned}
&\pi(\mu,\sigma \mid \mathbf{x_k}) \prod_{i=1}^{n} \left[\frac{(1/\sigma) f((x_i-\mu)/\sigma)}{f(x_i)}\right]^{l_i+r_i} \\
&\quad \overset{a}{\le} \left[2D(0,2)\right]^{l} \pi(\mu,\sigma \mid \mathbf{x_k}) \prod_{i=1}^{n} \left[\frac{|b_i| D(|a_i/b_i|,|b_i|)(1/\sigma) f((x_i-\mu)/\sigma)}{f(\omega)}\right]^{r_i} \\
&\quad \propto \pi(\mu,\sigma) \prod_{i=1}^{n} \left[(1/\sigma) f\big((a_i-\mu)/\sigma\big)\right]^{k_i} \left[\frac{(1/\sigma) f((x_i-\mu)/\sigma)}{f(\omega)}\right]^{r_i} \\
&\quad \overset{b}{\le} (1/\sigma) B\left[4D(0,4)(1/\sigma) f(\omega/\sigma)\right]^{k} \prod_{i=1}^{n} \left[\frac{(1/\sigma) f((x_i-\mu)/\sigma)}{f(\omega)}\right]^{r_i}
\end{aligned}$$

$$\propto (1/\sigma)\big[(1/\sigma)f(\omega/\sigma)\big]^k \prod_{i=1}^{n}\Big[\frac{(1/\sigma)f((x_i-\mu)/\sigma)}{f(\omega)}\Big]^{r_i}$$

$$\overset{c}{\leq} (1/\sigma)\big[(1/\sigma)f(\omega/\sigma)\big]^{k-r} \prod_{i=1}^{n}\big[(1/\sigma)f\big((x_i-\mu)/\sigma\big)\big]^{r_i}$$

$$\overset{d}{=} (1/\sigma)\big[(1/\sigma)f(\omega/\sigma)\big]^{k-r} \prod_{i=1}^{r}(1/\sigma)f\big((x_i-\mu)/\sigma\big).$$

In step a, we use $(1/\sigma)f((x_i-\mu)/\sigma)/f(x_i) \leq 2D(0,2)$ if $l_i = 1$, using the same arguments given above for the case $0 \leq \mu \leq \omega/2$. We also use $1/f(x_i) \leq |b_i|D(|a_i/b_i|, |b_i|)/f(\omega)$ if $r_i = 1$. In step b, we bound $\sigma\pi(\mu,\sigma)$ by B. We also use

$$f\big((a_i-\mu)/\sigma\big) \leq f\big((1/4)\omega/\sigma\big) \leq 4D(0,4)f(\omega/\sigma)$$

if $k_i = 1$ using the monotonicity of the tails of $f(z)$ in the first inequality since, if we define $a_{(k)} = \max_{i:k_i=1}\{|a_i|\}$ with $\omega \geq x_0 \geq 4a_{(k)}$, we have $|a_i-\mu|/\sigma = (\mu-a_i)/\sigma \geq (\omega/2 - a_{(k)})/\sigma \geq (\omega/2-\omega/4)/\sigma = (1/4)\omega/\sigma \geq \omega/4 \geq x_0/4 \geq M$ if we choose $x_0 \geq 4M$. We use Lemma 1 in the second inequality. In step c, we use $(\omega/\sigma)f(\omega/\sigma) \leq \omega f(\omega)$, using the monotonicity of the tails of $|z|f(z)$ since $\omega/\sigma \geq \omega \geq x_0 \geq M$ if we choose $x_0 \geq M$. In step d, we assume for convenience and without loss of generality that the right outliers are denoted by $x_1 < x_2 < \cdots < x_r$.

We now split the real line (which includes the region $\omega/2 \leq \mu < \infty$) into r mutually disjoint intervals given by $(x_{j-1}+x_j)/2 \leq \mu \leq (x_j+x_{j+1})/2$, for $j = 1,\ldots,r$, where we define $x_0 := -\infty$ and $x_{r+1} := \infty$. We also define the constant $\delta > 0$ as

$$\delta = \min_{i\in\{1,\ldots,r-1\}}\big\{(x_{i+1}-x_i)/2\big\}.$$

Consider $(x_{j-1}+x_j)/2 \leq \mu \leq (x_j+x_{j+1})/2$, for $j = 1,\ldots,r$ and $0 < \sigma \leq 1$. Then we have

$$(1/\sigma)\big[(1/\sigma)f(\omega/\sigma)\big]^{k-r}\prod_{i=1}^{r}(1/\sigma)f\big((x_i-\mu)/\sigma\big)$$

$$\overset{a}{\leq} (B/\delta)^{r-1}(1/\sigma)\big[(1/\sigma)f(\omega/\sigma)\big]^{k-r}(1/\sigma)f\big((x_j-\mu)/\sigma\big)$$

$$\overset{b}{\leq} (B/\delta)^{r-1}B^{k-r-1}\omega^{-(k-r)}\big(\omega/\sigma^2\big)f(\omega/\sigma)\times(1/\sigma)f\big((x_j-\mu)/\sigma\big).$$

In step a, we use, for $i \neq j$, $(1/\sigma)f((x_i-\mu)/\sigma) \leq B/|x_i-\mu| \leq B/\delta$, where we bound $|z|f(z)$ by B, and we use $|x_i-\mu| \geq \delta$ because if $i \neq j$, we have

$$|x_i-\mu| \geq \min\big\{(x_j-x_{j-1})/2, (x_{j+1}-x_j)/2\big\} \geq \delta.$$

In step b, we use $(\omega/\sigma) f(\omega/\sigma) \le B$ for $k-r-1$ terms. Finally, we have

$$\begin{aligned}
&\omega^{-(k-r)} \int_0^1 (\omega/\sigma^2) f(\omega/\sigma) \int_{(x_{j-1}+x_j)/2}^{(x_j+x_{j+1})/2} (1/\sigma) f\big((x_j-\mu)/\sigma\big)\, d\mu\, d\sigma \\
&\quad \le \omega^{-(k-r)} \int_0^\infty (\omega/\sigma^2) f(\omega/\sigma) \int_{-\infty}^{\infty} (1/\sigma) f\big((x_j-\mu)/\sigma\big)\, d\mu\, d\sigma \\
&\quad \overset{a}{=} \omega^{-(k-r)} \int_0^\infty f(\sigma')\, d\sigma' \int_{-\infty}^{\infty} f(\mu')\, d\mu' \le \omega^{-(k-r)} \overset{b}{\to} 0 \qquad \text{as } \omega \to \infty.
\end{aligned}$$

In step a, we use the changes of variable $\sigma' = \omega/\sigma$ and $\mu' = (x_j - \mu)/\sigma$. In step b, we use the condition $k > r$.

Acknowledgments. We thank the Associate Editor and the referees for very helpful comments.

REFERENCES

[1] ANDRADE, J. A. A., DOREA, C. C. Y. and GUEVARA OTINIANO, C. E. (2013). On the robustness of Bayesian modelling of location and scale structures using heavy-tailed distributions. *Comm. Statist. Theory Methods* **42** 1502–1514. MR3211162

[2] ANDRADE, J. A. A. and O'HAGAN, A. (2006). Bayesian robustness modeling using regularly varying distributions. *Bayesian Anal.* **1** 169–188. MR2227369

[3] ANDRADE, J. A. A. and O'HAGAN, A. (2011). Bayesian robustness modelling of location and scale parameters. *Scand. J. Stat.* **38** 691–711. MR2859745

[4] ANDRADE, J. A. A. and OMEY, E. (2013). Modelling conflicting information using subexponential distributions and related classes. *Ann. Inst. Statist. Math.* **65** 491–511. MR3067435

[5] ANGERS, J.-F. (2000). P-credence and outliers. *Metron* **58** 81–108 (2001). MR1835014

[6] DAWID, A. P. (1973). Posterior expectations for large observations. *Biometrika* **60** 664–667. MR0336889

[7] DE FINETTI, B. (1961). The Bayesian approach to the rejection of outliers. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I* 199–210. Univ. California Press, Berkeley, CA. MR0133935

[8] DESGAGNÉ, A. (2013). Full robustness in Bayesian modelling of a scale parameter. *Bayesian Anal.* **8** 187–219. MR3036259

[9] DESGAGNÉ, A. and ANGERS, J.-F. (2005). Importance sampling with the generalized exponential power density. *Stat. Comput.* **15** 189–196. MR2147552

[10] DESGAGNÉ, A. and ANGERS, J.-F. (2007). Conflicting information and location parameter inference. *Metron* **65** 67–97.

[11] KARAMATA, J. (1930). Sur un mode de croissance régulière des fonctions. *Mathematica* (*Cluj*) **4** 38–53.

[12] KUMAR, K. and MAGNUS, J. R. (2013). A characterization of Bayesian robustness for a normal location parameter. *Sankhya B* **75** 216–237. MR3133925

[13] O'HAGAN, A. (1979). On outlier rejection phenomena in Bayes inference. *J. Roy. Statist. Soc. Ser. B* **41** 358–367. MR0557598

[14] O'HAGAN, A. (1988). Modelling with heavy tails. In *Bayesian Statistics* **3** (*Valencia*, 1987) (G. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.) 345–359. Oxford Univ. Press, New York. MR1008055

[15] O'HAGAN, A. (1990). Outliers and credence for location parameter inference. *J. Amer. Statist. Assoc.* **85** 172–176. MR1137363

[16] O'HAGAN, A. and PERICCHI, L. (2012). Bayesian heavy-tailed models and conflict resolution: A review. *Braz. J. Probab. Stat.* **26** 372–401. MR2949085

[17] SCHEFFÉ, H. (1947). A useful convergence theorem for probability distributions. *Ann. Math. Stat.* **18** 434–438. MR0021585

DÉPARTEMENT DE MATHÉMATIQUES
UNIVERSITÉ DU QUÉBEC À MONTRÉAL
C.P. 8888, SUCCURSALE CENTRE-VILLE
MONTRÉAL, QUÉBEC
CANADA H3C 3P8
E-MAIL: desgagne.alain@uqam.ca