Full robustness to outliers in a Bayesian location-scale model

Alain Desgagné *

Abstract

The use of heavy-tailed distributions is a valuable tool in developing robust Bayesian procedures, limiting the influence of outliers on posterior inference. In this paper, the behavior of the posterior density for a location-scale model is investigated when the sample contains outliers. L-exponentially varying functions are introduced in order to characterize the tails of the densities. Simple conditions on the tails of the likelihood, using L-exponentially varying functions, are established to determine the proportion of observations that can be rejected as outliers. It is shown that the posterior distribution converges in law to the posterior that would be obtained from the reduced sample, excluding the outliers, as they tend to plus or minus infinity, at any given rate.

*Université du Québec à Montréal; Département de mathématiques, C.P. 8888, Succ. Centre-Ville; Montréal, Québec, Canada, H3C 3P8. E-mail address: desgagne.alain@uqam.ca

Key words: Bayesian inference, Conflicting information, Outlier, Heavy-tailed modeling, Location-Scale parameter, L-exponentially, L-regularly and L-slowly varying functions.

1 Introduction

The use of heavy-tailed distributions is a valuable tool in developing robust Bayesian procedures, limiting the influence of outliers on posterior inference. Outlier rejection in Bayesian analysis was first described by De Finetti (1961), where the simplest case with a single observation having mean θ was considered. Theorical results were given by Dawid (1973) and Hill (1974). O'Hagan (1979) considered outlier rejection in a sample and O'Hagan (1988) considered more general Bayesian modeling based on Student-t distributions. Outliers rejection based on the notion of credence was introduced by O'Hagan (1990). This paper was generalied by Angers (2000) and Desgagné and Angers (2007). Other authors approached outliers rejection, see for instance Meinhold and Singpurwalla (1989), Angers and Berger (1991), Carlin and Polson (1991), Angers (1992), Fan and Berger (1992), Geweke (1994) and Angers (1996).

In this paper, we consider the location-scale model on the real line, and we investigate the conditions to obtain robustness against outliers. In Section 2, we define new classes of functions resulting from generalization and transformation of the known regularly varying functions. The result is the class of L-exponentially varying functions. This family will be used to describe some conditions of robustness in the next sections. In Section 3, the mathematical context is given. We describe the location-scale model, in the context of robustness. In Section 4, we give the main theorem, with the conditions and results of robustness. We show that the posterior distribution converges in law to the posterior that would be obtained from the reduced sample, excluding the outliers, as they tend to $-\infty$ or ∞, at any given rate. The proofs are given in Section 5. In Section 6, a new density, named log-GEP, is proposed for a robust modeling.

2 L-exponentially varying functions

One main condition for robustness is to use sufficiently heavy-tailed densities for the conflicting information. To achieve it, we introduce a class of functions called L-exponentially varying functions defined. Any function is assumed to be continuous.

We say that $f(z)$ and $g(z)$ are *asymptotically equivalent* at ∞, written $f(z) \sim g(z)$ as $z \to \infty$, if

$$f(z)/g(z) \to 1 \text{ as } z \to \infty.$$

We begin with definitions of classes of functions defined on $\mathbb{R}$. We say that a function f is *slowly varying* at $\infty[-\infty]$, written $f \in E_{0,0,0}(\infty)[E_{0,0,0}(-\infty)]$, if for $\nu > 0$,

$$f(\nu x) \sim f(x) \text{ as } x \to \infty[-\infty].$$

We say that a function f is *regularly varying* at $\infty[-\infty]$ with index $\rho \in \mathbb{R}$, written $f \in E_{0,0,-\rho}(\infty)[E_{0,0,-\rho}(-\infty)]$, if for $\nu > 0$,

$$f(\nu x)/f(x) \to \nu^{\rho} \text{ as } x \to \infty[-\infty],$$

or equivalently, if for $x > 0[x < 0]$, f can be written as

$$f(x) = |x|^\rho S(x), \text{ with } S \in E_{0,0,0}(\infty)[E_{0,0,0}(-\infty)].$$

We now introduce a new class of function which generalizes the regularly varying functions as follows.

Definition 1. *We say that a function f is exponentially varying at $\infty[-\infty]$ with index (γ, δ, α), written $f \in E_{\gamma,\delta,\alpha}(\infty)[E_{\gamma,\delta,\alpha}(-\infty)]$, if for $x > 0[x < 0]$, f can be written as*

$$f(x) = \mathrm{e}^{-\delta|x|^\gamma} |x|^{-\alpha} S(x),$$

with $S \in E_{0,0,0}(\infty)[E_{0,0,0}(-\infty)]$, $\gamma \geq 0, \delta \geq 0, \alpha \in \mathbb{R}$.

By convention, we define $\gamma = 0$ iif $\delta = 0$. This class of functions includes the slowly varying functions when $\gamma = 0, \delta = 0, \alpha = 0$ and the regularly varying functions with index ρ when $\gamma = 0, \delta = 0, \alpha = -\rho$. It is easy to see that if $f_1(x) \sim f_2(x)$ as $x \to \infty[-\infty]$ and $f_1 \in E_{\gamma,\delta,\alpha}(\infty[-\infty])$, then $f_2 \in E_{\gamma,\delta,\alpha}(\infty[-\infty])$. Note that $f(x) \in E_{\gamma,\delta,\alpha}(-\infty) \Leftrightarrow f(-x) \in E_{\gamma,\delta,\alpha}(\infty)$.

Definition 2. *We say that a function f is L-slowly varying at ∞, written $f \in L_{0,0,0}(\infty)$, if for $\nu > 0$,*

$$f(z^\nu) \sim f(z) \text{ as } z \to \infty.$$

For example, $\log(\log z) \in L_{0,0,0}(\infty)$.

Definition 3. *We say that a function f is L-regularly varying at ∞ with index $\rho \in \mathbb{R}$, written $f \in L_{0,0,-\rho}(\infty)$, if for $\nu > 0$,*

$$f(z^\nu)/f(z) \to \nu^\rho \text{ as } z \to \infty,$$

or equivalently, if for $z > 1$, f can be written as

$$f(z) = (\log z)^{\rho} S(z), \text{ with } S \in L_{0,0,0}(\infty).$$

Definition 4. *We say that a function f is L-exponentially varying at ∞ with index (γ, δ, α), written $f \in L_{\gamma,\delta,\alpha}(\infty)$ if for $z > 1$, f can be written as*

$$f(z) = \mathrm{e}^{-\delta(\log z)^{\gamma}} (\log z)^{-\alpha} S(z),$$

with $S \in L_{0,0,0}(\infty)$, $\gamma \geq 0, \delta \geq 0, \alpha \in \mathbb{R}$.

By convention, we define $\gamma = 0$ iif $\delta = 0$. This class of functions includes the L-slowly varying functions when $\gamma = 0, \delta = 0, \alpha = 0$ and the L-regularly varying functions with index ρ when $\gamma = 0, \delta = 0, \alpha = -\rho$. It is easy to see that if $f(z) \sim g(z)$ as $z \to \infty$ and $f \in L_{\gamma,\delta,\alpha}(\infty)$, then $g \in L_{\gamma,\delta,\alpha}(\infty)$.

Since the vector (γ, δ, α) characterizes a L-exponentially varying function, it can be useful to define the following notation.

i) If $\gamma_1 = \gamma_2$, $\delta_1 = \delta_2$, $\alpha_1 = \alpha_2$, we write

$$(\gamma_1, \delta_1, \alpha_1) = (\gamma_2, \delta_2, \alpha_2),$$

ii) if $\gamma_1 > \gamma_2$ or $\gamma_1 = \gamma_2$, $\delta_1 > \delta_2$ or $\gamma_1 = \gamma_2$, $\delta_1 = \delta_2$, $\alpha_1 > \alpha_2$, we write

$$(\gamma_1, \delta_1, \alpha_1) > (\gamma_2, \delta_2, \alpha_2),$$

iii) if $\gamma_1 < \gamma_2$ or $\gamma_1 = \gamma_2$, $\delta_1 < \delta_2$ or $\gamma_1 = \gamma_2$, $\delta_1 = \delta_2$, $\alpha_1 < \alpha_2$, we write

$$(\gamma_1, \delta_1, \alpha_1) < (\gamma_2, \delta_2, \alpha_2).$$

The following proposition concerns the integrability of functions of this class.

Proposition 1. *For a function f such that $f \in L_{\gamma,\delta,\alpha}(\infty)$, $(1/z)f(z)$ is integrable on $z > 1$, if*

i) $(\gamma, \delta, \alpha) > (0, 0, 1)$,

ii) $(\gamma, \delta, \alpha) = (0, 0, 1)$, *with a L-slowly varying function $S(z)$ having a decay sufficiently "fast" (e.g. $S(z) = (\log(\log z))^{-\beta}$, $\beta > 1$).*

For two functions f_1 and f_2 L-exponentially varying at ∞, it is simple to order the asymptotic behavior of their tails as follows.

Proposition 2. *If $f_i \in L_{\gamma_i,\delta_i,\alpha_i}(\infty)$, $i = 1, 2$, and*

i) *if $(\gamma_1, \delta_1, \alpha_1) > (\gamma_2, \delta_2, \alpha_2)$, then*

$$f_1(z)/f_2(z) \to 0 \text{ as } z \to \infty,$$

ii) *if $(\gamma_1, \delta_1, \alpha_1) < (\gamma_2, \delta_2, \alpha_2)$, then*

$$f_1(z)/f_2(z) \to \infty \text{ as } z \to \infty,$$

iii) *if $(\gamma_1, \delta_1, \alpha_1) = (\gamma_2, \delta_2, \alpha_2)$, then the ratio of the L-slowly varying functions determines the tails dominance.*

The following proposition says that the asymptotic scale invariance is a sufficient condition for the asymptotic location invariance.

Proposition 3.

$$f(\sigma z) \sim f(z), \sigma > 0 \Rightarrow f(z + \mu) \sim f(z), \mu \in \mathbb{R}, \text{ as } z \to \infty.$$

In the next proposition, we give sufficient conditions to ensure that a scale change has no impact on the tail of a L-exponentially varying function.

Proposition 4. *If $f \in L_{\gamma,\delta,\alpha}(\infty)$ and $\gamma < 1$, then*

$$f(\sigma z) \sim f(z) \text{ as } z \to \infty, \sigma > 0,$$

that is f is a slowly varying function.

A corollary of Propositions 3 and 4 is given by the following proposition.

Proposition 5. *If $f \in L_{\gamma,\delta,\alpha}(\infty)$ and $\gamma < 1$, then*

$$f(\sigma z + \mu) \sim f(z) \text{ as } z \to \infty, \mu \in \mathbb{R}, \sigma > 0.$$

Proof.

$$\frac{f(\sigma z + \mu)}{f(z)} = \frac{f(\sigma z + \mu)}{f(\sigma z)} \frac{f(\sigma z)}{f(z)} \to 1, \text{ as } z \to \infty,$$

using Proposition 3 since $\sigma z \to \infty$ and using Proposition 4. □

A corollary of Proposition 5 is given by the following proposition.

Proposition 6. *If $zf(z) \in L_{\gamma,\delta,\alpha}(\infty)$ and if $\gamma < 1$, then*

$$(1/\sigma)f((z-\mu)/\sigma) \sim f(z) \text{ as } z \to \infty, \mu \in \mathbb{R}, \sigma > 0.$$

Proof. If we let $g(z) = |z|\, f(z)$, then using Proposition 5, we have

$$\frac{(1/\sigma)f((z-\mu)/\sigma)}{f(z)} = \frac{g((z-\mu)/\sigma)}{g(z)} \left| \frac{z}{z-\mu} \right| \to 1 \text{ as } z \to \infty.$$

□

Proposition 7. *If f is a proper density defined on $\mathbb{R}$, symmetric with respect to the origin, such that the right tail of $xf(x)$ is non-increasing and $xf(x) \in L_{\gamma,\delta,\alpha}(\infty)$ with $\gamma < 1$, then*

i)

$$\int_{-\infty}^{\infty} f(x-\mu)f(\mu)\,d\mu \sim f(x) \text{ as } x \to \pm\infty.$$

ii)

$$2\int_{0}^{\infty} (1/\sigma)f(x/\sigma)f(\sigma)\,d\sigma \sim f(x) \text{ as } x \to \pm\infty.$$

Proof. Firstly, we note that a non-increasing right tail of $xf(x)$ means also a non-increasing right tail of $f(x)$. It suffices to see that for x larger than a certain positive constant,

$$\frac{d}{dx}xf(x) \leq 0 \Rightarrow f'(x) \leq -f(x)/x \leq 0.$$

From Proposition 6 of article 1, we know that $f(x)$ is also an exponentially varying function with $\gamma < 1$. Then Proposition 7 i) of the article 1 is directly used to prove point i) of this proposition.

For point ii), if we define $h(\sigma) = 2\sigma f(\sigma)$, we can verify, using the symmetry of f, that

$$\int_{0}^{\infty} (1/\sigma)h(\sigma)\,d\sigma = 2\int_{0}^{\infty} (1/\sigma)\sigma f(\sigma)\,d\sigma = 2\int_{0}^{\infty} f(\sigma)\,d\sigma = 1.$$

Then Proposition 7 ii) of the article 1 can be used, and we have

$$\begin{aligned} h(x)^{-1} &\int_{0}^{\infty} (1/\sigma)h(x/\sigma)h(\sigma)\,d\sigma \\ &= (2xf(x))^{-1}\int_{0}^{\infty} (1/\sigma)2(x/\sigma)f(x/\sigma)2\sigma f(\sigma)\,d\sigma \\ &= f(x)^{-1}\cdot 2\int_{0}^{\infty} (1/\sigma)f(x/\sigma)f(\sigma)\,d\sigma \\ &\to 1 \text{ as } x \to \infty. \end{aligned}$$

By symmetry of f, the result is also true for $x \to -\infty$. □

3 Context

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

$$X_i \mid \mu, \sigma \sim (1/\sigma) f_i((x_i - \mu)/\sigma),$$

ii) the joint prior density of μ and σ is assumed improper as follows $\mu, \sigma \sim 1/\sigma$,

where $x_1, \ldots, x_n, \mu \in \mathbb{R}$, $\sigma > 0$, and $f_i(x)$ are continuous, positive everywhere, symmetric ($f_i(-x) = f_i(x)$) and proper densities. The tails of $|x| f_i(x)$ are assumed to be non-increasing, which necessarily means that the tails of $f_i(x)$ are also non-increasing. It follows that the functions $|x| f_i(x)$ are symmetric, bounded above for all $x \in \mathbb{R}$ and they have a limit of 0 in their tails as $x \to \pm\infty$. The same is necessarily true for the density $f_i(x)$. Any other parameters are also assumed to be known and are implicitly included in the densities.

We study robustness of the inference on μ and σ in presence of extreme observations x_i. The nature of the results is asymptotic, in the sense that we consider cases where some x_i are going to $\pm\infty$. Among the n observations, denoted by $\underset{\sim}{x}_n = (x_1, \ldots, x_n)$, we assume that k of them, denoted by the vector $\underset{\sim}{x}_k$, form a group of non-outlier or fixed values. We assume that l of them, denoted by the vector $\underset{\sim}{x}_l$, are considered as left outliers (smaller than the fixed values) and r of them, denoted by the vector $\underset{\sim}{y}_r$, are considered as right outliers (larger than the fixed values), where $0 \leq k, l, r \leq n$ and $k + l + r = n$.

We define three binary functions l_i, k_i and r_i as follows. If x_i is a left outlier, we set $l_i = 1$, if it is a fixed value, we set $k_i = 1$ and if it is a right outlier, we set $r_i = 1$. The functions are 0 otherwise. We have $l_i + k_i + r_i = 1$ for any x_i. Note that $\sum_{i=1}^n l_i = l$, $\sum_{i=1}^n k_i = k$ and $\sum_{i=1}^n r_i = r$.

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

$$\pi(\mu, \sigma \mid \underset{\sim}{x}_n) = m(\underset{\sim}{x}_n)^{-1}(1/\sigma) \prod_{i=1}^n (1/\sigma) f_i((x_i - \mu)/\sigma),$$

$$m(\underset{\sim}{x}_n) = \int_{-\infty}^{\infty} \int_0^{\infty} (1/\sigma) \prod_{i=1}^n (1/\sigma) f_i((x_i - \mu)/\sigma)\, d\sigma\, d\mu.$$

We also define

$$\pi(\mu, \sigma \mid \underset{\sim}{x}_k) = m(\underset{\sim}{x}_k)^{-1}(1/\sigma) \prod_{i=1}^n [(1/\sigma) f_i((x_i - \mu)/\sigma)]^{k_i},$$

$$m(\underset{\sim}{x}_k) = \int_{-\infty}^{\infty} \int_0^{\infty} (1/\sigma) \prod_{i=1}^n [(1/\sigma) f_i((x_i - \mu)/\sigma)]^{k_i}\, d\sigma\, d\mu.$$

Similarly, we can define

$$\pi(\mu, \sigma \mid \underset{\sim}{x}_l) \propto (1/\sigma) \prod_{i=1}^n [(1/\sigma) f_i((x_i - \mu)/\sigma)]^{l_i},$$

$$\pi(\mu, \sigma \mid \underset{\sim}{x}_r) \propto (1/\sigma) \prod_{i=1}^n [(1/\sigma) f_i((x_i - \mu)/\sigma)]^{r_i},$$

$$\pi(\mu, \sigma \mid \underset{\sim}{x}_l, \underset{\sim}{x}_r) \propto (1/\sigma) \prod_{i=1}^n [(1/\sigma) f_i((x_i - \mu)/\sigma)]^{l_i + r_i}.$$

We can interpret $\pi(\mu \mid \underset{\sim}{x}_k)$ as a posterior density considering only the fixed observations $\underset{\sim}{x}_k$ and $m(\underset{\sim}{x}_k)$ as the corresponding marginal density of

$\underset{\sim}{X}_k$, with $1/\sigma$ as the prior joint density of μ and σ. Similar interpretation for $\pi(\mu \mid \underset{\sim}{x}_l)$, $\pi(\mu \mid \underset{\sim}{x}_r)$ and $\pi(\mu, \sigma \mid \underset{\sim}{x}_l, \underset{\sim}{x}_r)$ can be done. It can be seen that

$$\sigma\pi(\mu \mid \underset{\sim}{x}_n) \propto \sigma\pi(\mu \mid \underset{\sim}{x}_k) \times \sigma\pi(\mu \mid \underset{\sim}{x}_l) \times \sigma\pi(\mu \mid \underset{\sim}{x}_r)$$

or

$$\sigma\pi(\mu \mid \underset{\sim}{x}_n) \propto \sigma\pi(\mu \mid \underset{\sim}{x}_k) \times \sigma\pi(\mu \mid \underset{\sim}{x}_l, \underset{\sim}{x}_r).$$

4 Resolution of conflicts

Using the Bayesian context described in Section 3, the main theorem of this paper is now presented. We denote

$$\omega = \min(-\underset{\sim}{x}_l, \underset{\sim}{x}_r).$$

If we let $\omega \to \infty$, it means that each component of the vector is going to ∞ at any given rate.

Theorem 1. *If the following conditions are satisfied:*

i) $xf_i(x) \in L_{\gamma,\delta,\alpha}(\infty)$ *with* $\gamma < 1$, $i = 1\ldots, n$,

ii) $k - (l + r) \geq 2$,

then we have the following results:

a) $m(\underset{\sim}{x}_n) \sim m(\underset{\sim}{x}_k) \prod_{i=1}^{n} [f_i(x_i)]^{l_i + r_i}$ *as* $\omega \to \infty$,

b) $\pi(\mu, \sigma \mid \underset{\sim}{x}_n) \to \pi(\mu, \sigma \mid \underset{\sim}{x}_k)$, $\mu \in \mathbb{R}, \sigma > 0$, *as* $\omega \to \infty$,

c) $\sigma\pi(\mu, \sigma \mid \underset{\sim}{x}_n) \to 0$, *as* $|\mu| \to \infty$ *and/or* $|\log \sigma| \to \infty$, $\omega \to \infty$,

d) $\mu, \sigma \mid \underset{\sim}{x}_n \xrightarrow{\mathcal{L}} \mu, \sigma \mid \underset{\sim}{x}_k$ *as* $\omega \to \infty$.

Result a) gives the asymptotic behavior of the marginal, as $\omega \to \infty$. Result b) gives the asymptotic behavior of the posterior at any fixed value of μ and σ. We see that the conflicting information is completely rejected and the posterior considering the entire information behaves as the the posterior considering only the non-conflicting values. Result c) says that $\sigma\pi(\mu, \sigma \mid \underset{\sim}{x}_n)$ converges to 0, if ω are going to ∞ and if at least one of the following occurs: $\mu \to \infty$, $\mu \to -\infty$, $\sigma \to \infty$, $\sigma \to 0$, independently and at any given rate. It means that $\sigma\pi(\mu, \sigma \mid \underset{\sim}{x}_n)$ (and the posterior as well) converges to 0 for any area around the conflicting values ($\underset{\sim}{x}_l$, $\underset{\sim}{x}_r$), and that an eventual mode at these values will also converge to 0. Note that $\pi(\mu, \sigma \mid \underset{\sim}{x}_n)/\pi(\mu, \sigma \mid \underset{\sim}{x}_k)$, as $|\mu| \to \infty$ and/or $|\log \sigma| \to \infty$, $\omega \to \infty$, has a form of 0/0 and its limit can be anywhere between 0 and ∞, depending on the relation between μ, σ and ω. In result d), the convergence in law is understood as $\Pr[\mu \leq d_1, \sigma \leq d_2 \mid \underset{\sim}{x}_n] \to \Pr[\mu \leq d_1, \sigma \leq d_2 \mid \underset{\sim}{x}_k]$, for any $d_1 \in \mathbb{R}, d_2 > 0$, as $\omega \to \infty$.

5 Proof of Theorem 1

5.1 Proof of result a) of Theorem 1

From condition i), it follows that Propositions 5, 6 and 7 can be applied on the right tail of f_i. Note that by symmetry, we have $f_i(-x) = f_i(x)$ and all properties valid for the right tail of $f_i(x)$ are also valid for the right tail of $f_i(-x)$.

More explicitly, Propositions 5 and 6 say that $\forall \epsilon > 0, \forall \lambda > 0, \forall \tau > 1$, there exist $A_1(\epsilon, \lambda, \tau) > 0$ such that $x > A_1(\epsilon, \lambda, \tau), -\lambda \leq \mu \leq \lambda, 1/\tau \leq \sigma \leq \tau \Rightarrow$

$$1 - \epsilon \leq (\sigma x + \mu) f(\sigma x + \mu)/(x f(x)) \leq 1 + \epsilon$$

and

$$1 - \epsilon \leq (1/\sigma) f((x - \mu)/\sigma)/f(x) \leq 1 + \epsilon.$$

Since the densities f_i are continuous and proper, it is easy to verify that $\pi(\mu, \sigma \mid \underset{\sim}{x}_k)$ and $\pi(\mu, \sigma \mid \underset{\sim}{x}_n)$ are also proper densities. The following lemma follows directly.

Lemma 1. $\forall \epsilon > 0$, *there exists a constant* $A_5(\epsilon) > 1$ *such that* $\lambda, \tau \geq A_5(\epsilon) \Rightarrow$

i) $\int_{-\lambda}^{\lambda} \int_{1/\tau}^{\tau} \pi(\mu, \sigma \mid \underset{\sim}{x}_k)\, d\sigma\, d\mu \geq 1 - \epsilon$,

ii) $\int \int_{[\mathbb{R} \times \mathbb{R}^+] \setminus [(-\lambda, \lambda) \times (1/\tau, \tau)]} \pi(\mu, \sigma \mid \underset{\sim}{x}_k)\, d\sigma\, d\mu < \epsilon$.

Now we define two functions as follows.

$$H(\mu, \sigma, \underset{\sim}{x}_n) = \pi(\mu, \sigma \mid \underset{\sim}{x}_k) \prod_{i=1}^{n} \left[\frac{(1/\sigma) f_i((x_i - \mu)/\sigma)}{f_i(x_i)} \right]^{l_i + r_i} \tag{1}$$

and

$$M(\underset{\sim}{x}_n) = \int_{-\infty}^{\infty} \int_{0}^{\infty} H(\mu, \sigma, \underset{\sim}{x}_n)\, d\sigma\, d\mu. \tag{2}$$

It is easy to verify that

$$m(\underset{\sim}{x}_n) \pi(\mu, \sigma \mid \underset{\sim}{x}_n) = m(\underset{\sim}{x}_k) \pi(\mu, \sigma \mid \underset{\sim}{x}_k) \prod_{i=1}^{n} \left[(1/\sigma) f_i((x_i - \mu)/\sigma)\right]^{l_i + r_i}.$$

By dividing each side of this equation by $\prod_{i=1}^{n} f_i(x_i)^{l_i + r_i}$ and using equation (1), we obtain

$$H(\mu, \sigma, \underset{\sim}{x}_n) = \frac{m(\underset{\sim}{x}_n) \pi(\mu, \sigma \mid \underset{\sim}{x}_n)}{m(\underset{\sim}{x}_k) \prod_{i=1}^{n} f_i(x_i)^{l_i + r_i}}. \tag{3}$$

Using equations (2) and (3) with $\int_{-\infty}^{\infty}\int_0^{\infty}\pi(\mu,\sigma\mid \underset{\sim}{x}_n)\,d\sigma\,d\mu=1$, we obtain

$$M(\underset{\sim}{x}_n)=\frac{m(\underset{\sim}{x}_n)}{m(\underset{\sim}{x}_k)\prod_{i=1}^{n}f_i(x_i)^{l_i+r_i}}. \tag{4}$$

Dividing (3) by (4), we see that

$$\pi(\mu,\sigma\mid \underset{\sim}{x}_n)=H(\mu,\sigma,\underset{\sim}{x}_n)/M(\underset{\sim}{x}_n). \tag{5}$$

Finally, from equations (1) and (5), it follows that

$$\frac{\pi(\mu,\sigma\mid \underset{\sim}{x}_n)}{\pi(\mu,\sigma\mid \underset{\sim}{x}_k)}=(1/M(\underset{\sim}{x}_n))\prod_{i=1}^{n}\left[\frac{(1/\sigma)f_i((x_i-\mu)/\sigma)}{f_i(x_i)}\right]^{l_i+r_i}. \tag{6}$$

Equation (6) will be useful for the proof of result b) of Theorem 1. From Equation (4), we see that Result a) is equivalent to $M(\underset{\sim}{x}_n)\to 1$ as $\omega\to\infty$. And using Equation (2), Result a) can be written as follows. $\forall\epsilon>0$, there exists $A_0(\epsilon)$ such that

$$\omega>A_0(\epsilon)\Rightarrow 1-\epsilon\le\int_{-\infty}^{\infty}\int_0^{\infty}H(\mu,\sigma,\underset{\sim}{x}_n)\,d\sigma\,d\mu\le 1+\epsilon.$$

Now choose any $0<\epsilon<1$. Note that if the result is true for $0<\epsilon<1$, it is necessarily true for any larger $\epsilon\ge 1$. Then define

$$\begin{aligned}\epsilon_0&=\min\left[1-(1-\epsilon/3)^{1/n},(1+\epsilon/3)^{1/n}-1\right]\\ \lambda_1&=\max\left[A_5(\epsilon_0)\right],\\ \tau_1&=\max\left[A_5(\epsilon_0)\right],\\ A_0(\epsilon)&=\max\left[A_1(\epsilon_0,\lambda_1,\tau_1)\right].\end{aligned}$$

Note that $0<\epsilon_0<1$. The integral of result a) is divided into nine parts.

Firstly consider the integral on $\mu,\sigma\in[-\lambda_1,\lambda_1]\times[1/\tau_1,\tau_1]$.

$$\int_{-\lambda_1}^{\lambda_1}\int_{1/\tau_1}^{\tau_1}H(\mu,\sigma,\underset{\sim}{x}_n)\,d\sigma\,d\mu\ge\int_{-\lambda_1}^{\lambda_1}\int_{1/\tau_1}^{\tau_1}\pi(\mu,\sigma\mid \underset{\sim}{x}_k)(1-\epsilon_0)^{l+r}\,d\sigma\,d\mu$$

$$\geq (1-\epsilon_0)^{l+r+1} \geq (1-\epsilon_0)^n \geq 1-\epsilon/3.$$

Note that the location-scale invariance of f_i (Proposition 6) is used in the first inequality since $A_0(\epsilon) \geq A_1(\epsilon_0, \lambda_1, \tau_1)$ and Lemma 1 is used in the second inequality since $\lambda_1, \tau_1 \geq A_5(\epsilon_0)$. In a similar way, it can be shown that

$$\int_{-\lambda_1}^{\lambda_1} \int_{1/\tau_1}^{\tau_1} H(\mu, \sigma, \underset{\sim}{x}_n)\, d\sigma\, d\mu \leq \ldots \leq (1+\epsilon_0)^{l+r+1} \leq (1+\epsilon_0)^n \leq 1+\epsilon/3.$$

The proof is yet to be completed by showing that

$$\int\int_{\mathbb{R}\times\mathbb{R}^+ \setminus(-\lambda_1,\lambda_1)\times(1/\tau_1,\tau_1)} H(\mu, \sigma, \underset{\sim}{x}_n)\, d\sigma\, d\mu \leq \ldots \leq \epsilon/3.$$

5.2 Proof of result b) of Theorem 1

From equation (6), it is clear that

$$\frac{\pi(\mu, \sigma \mid \underset{\sim}{x}_n)}{\pi(\mu, \sigma \mid \underset{\sim}{x}_k)} \to 1 \text{ as } \omega \to \infty,$$

for any fixed $\mu \in \mathbb{R}$ and $\sigma > 0$, using result a) of Theorem 1 and the location-scale invariance (Proposition 6).

5.3 Proofs of results c) and d) of Theorem 1

The proof for result c) is yet to be completed. For the proof of result d), we need first to write Result b) explicitly as follows. $\forall\epsilon > 0, \forall\lambda > 0, \forall\tau > 1$, there exists a constant $A_6(\epsilon, \lambda, \tau) > 0$ such that $-\lambda \leq \mu \leq \lambda, 1/\tau \leq \sigma \leq \tau$ and $\omega > A_6(\epsilon, \lambda, \tau) \Rightarrow$

$$1-\epsilon < \pi(\mu, \sigma \mid \underset{\sim}{x}_n)/\pi(\mu, \sigma \mid \underset{\sim}{x}_k) < 1+\epsilon.$$

We want to show result d), which can be written as follows. $\forall \epsilon > 0, \forall d_1 \in \mathbb{R}, \forall d_2 > 0$ there exists a constant $A_0(\epsilon, d_1, d_2) > 0$ such that $\omega > A_0(\epsilon, d_1, d_2) \Rightarrow$

$$\left| \int_{d_1}^{\infty} \int_{d_2}^{\infty} \pi(\mu, \sigma \mid \underset{\sim}{x}_n) \, d\sigma \, d\mu - \int_{d_1}^{\infty} \int_{d_2}^{\infty} \pi(\mu, \sigma \mid \underset{\sim}{x}_k) \, d\sigma \, d\mu \right| < \epsilon.$$

Choose any $\epsilon > 0$, any $d_1 \in \mathbb{R}$ and $d_2 > 0$, let $\lambda = \max(|d_1|, A_5(\epsilon/3))$ and $\tau = \max(1/d_2, d_2, A_5(\epsilon/3))$, where the constant A_5 comes from Lemma 1, and let $A_0(\epsilon, d_1, d_2) = A_6(\epsilon/3, \lambda, \tau)$. Notice that $\lambda \geq |d_1| \Leftrightarrow -\lambda \leq d_1 \leq \lambda$ and $\tau \geq \max(1/d_2, d_2) \Leftrightarrow 1/\tau \leq d_2 \leq \tau$. Let

$$Z = [(d_1, \infty) \times (d_2, \infty)] \setminus [(d_1, \lambda) \times (d_2, \tau)] \, .$$

Notice that

$$[(d_1, \lambda) \times (d_2, \tau)] \in [(-\lambda, \lambda) \times (1/\tau, \tau)]$$

and

$$Z \in \left[\mathbb{R} \times \mathbb{R}^+\right] \setminus [(-\lambda, \lambda) \times (1/\tau, \tau)] \, .$$

Consider $\omega > A_0(\epsilon, d_1, d_2)$.

Firstly, using Lemma 1, we have

$$\int \int_Z \pi(\mu, \sigma \mid \underset{\sim}{x}_k) \, d\sigma \, d\mu \leq \int \int_{\left[\mathbb{R} \times \mathbb{R}^+\right] \setminus [(-\lambda, \lambda) \times (1/\tau, \tau)]} \pi(\mu, \sigma \mid \underset{\sim}{x}_k) \, d\sigma \, d\mu \leq \epsilon/3,$$

since $\lambda, \tau \geq A_5(\epsilon/3)$.

Secondly,

$$\begin{aligned} \int \int_Z \pi(\mu, \sigma \mid \underset{\sim}{x}_n) \, d\sigma \, d\mu &\leq \int \int_{\left[\mathbb{R} \times \mathbb{R}^+\right] \setminus [(-\lambda, \lambda) \times (1/\tau, \tau)]} \pi(\mu, \sigma \mid \underset{\sim}{x}_n) \, d\sigma \, d\mu \\ &= 1 - \int_{-\lambda}^{\lambda} \int_{1/\tau}^{\tau} \pi(\mu, \sigma \mid \underset{\sim}{x}_n) \, d\sigma \, d\mu \end{aligned}$$

$$\leq 1-(1-\epsilon/3)\int_{-\lambda}^{\lambda}\int_{1/\tau}^{\tau}\pi(\mu,\sigma\mid \underset{\sim}{x}_k)\,d\sigma\,d\mu$$
$$\leq 1-(1-\epsilon/3)^2 \leq 1-(1-\epsilon/3)=\epsilon/3.$$

In the equality, we use the fact that $\pi(\mu,\sigma \mid \underset{\sim}{x}_n)$ is a proper density. In the second inequality, we used result b) since $\omega > A_6(\epsilon/3,\lambda,\tau)$. In the third inequality, Lemma 1 is used since $\lambda,\tau \geq A_5(\epsilon/3)$.

Thirdly,

$$\left|\int_{d_1}^{\lambda}\int_{d_2}^{\tau}\pi(\mu,\sigma\mid \underset{\sim}{x}_n)\,d\sigma\,d\mu-\int_{d_1}^{\lambda}\int_{d_2}^{\tau}\pi(\mu,\sigma\mid \underset{\sim}{x}_k)\,d\sigma\,d\mu\right|$$
$$\leq \int_{d_1}^{\lambda}\int_{d_2}^{\tau}\left|\pi(\mu,\sigma\mid \underset{\sim}{x}_n)-\pi(\mu,\sigma\mid \underset{\sim}{x}_k)\right|\,d\sigma\,d\mu$$
$$=\int_{d_1}^{\lambda}\int_{d_2}^{\tau}\pi(\mu,\sigma\mid \underset{\sim}{x}_k)\left|\pi(\mu,\sigma\mid \underset{\sim}{x}_n)/\pi(\mu,\sigma\mid \underset{\sim}{x}_k)-1\right|\,d\sigma\,d\mu$$
$$\leq \epsilon/3\int_{d_1}^{\lambda}\int_{d_2}^{\tau}\pi(\mu,\sigma\mid \underset{\sim}{x}_k)\,d\sigma\,d\mu \leq \epsilon/3.$$

Result b) is used in the second inequality since $\omega > A_6(\epsilon/3,\lambda,\tau)$.

Combining the three last inequalities, it follows that

$$\left|\int_{d_1}^{\infty}\int_{d_2}^{\infty}\pi(\mu,\sigma\mid \underset{\sim}{x}_n)\,d\sigma\,d\mu-\int_{d_1}^{\infty}\int_{d_2}^{\infty}\pi(\mu,\sigma\mid \underset{\sim}{x}_k)\,d\sigma\,d\mu\right|$$
$$\leq \left|\int_{d_1}^{\lambda}\int_{d_2}^{\tau}\pi(\mu,\sigma\mid \underset{\sim}{x}_n)\,d\sigma\,d\mu-\int_{d_1}^{\lambda}\int_{d_2}^{\tau}\pi(\mu,\sigma\mid \underset{\sim}{x}_k)\,d\sigma\,d\mu\right|$$
$$+\int\int_{Z}\pi(\mu,\sigma\mid \underset{\sim}{x}_n)\,d\sigma\,d\mu+\int\int_{Z}\pi(\mu,\sigma\mid \underset{\sim}{x}_k)\,d\sigma\,d\mu$$
$$\leq \epsilon/3+\epsilon/3+\epsilon/3=\epsilon.$$

6 Family of D-GEP and log-GEP densities

The Generalized Exponential Power (GEP) density has been introduced by Desgagné and Angers (2007). It is a symmetric density around the origin, defined on the real line and made of a constant part in the center. Its interest lies in its large spectrum of tail behavior. In this section, we propose two families of densities, the D-GEP and the log-GEP densities. The D-GEP density (D stands for double), defined on the real line, is built with the right tail of the GEP density translated to the origin and doubled. The log-GEP, defined on the positive real line, is simply given by a logarithmic/exponential transformation of the D-GEP density. The D-GEP and log-GEP densities are respectively exponentially (at $-\infty$ and ∞) and L-exponentially (at 0 and ∞) varying functions. These densities are then useful for robust modeling.

Definition 5. *A random variable X has a D-GEP distribution, written $X \sim$ D-GEP$(\gamma, \delta, \alpha, \beta, \theta, \tau)$, if its density is given by*

$$f(x) = \frac{K(\gamma, \delta, \alpha, \beta, \theta)}{2\tau} \exp[-\delta(|x/\tau| + \theta)^{\gamma}](|x/\tau| + \theta)^{-\alpha}(\log(|x/\tau| + \theta))^{-\beta},$$

where $x \in \mathbb{R}$, $\gamma \geq 0, \delta \geq 0, \alpha \in \mathbb{R}, \beta \in \mathbb{R}, \theta \geq 0$. The scale parameter is $\tau > 0$. A part of the normalizing constant is given by

$$1/K(\gamma, \delta, \alpha, \beta, \theta) = \int_{\theta}^{\infty} \exp[-\delta z^{\gamma}] z^{-\alpha} (\log z)^{-\beta} dz. \tag{7}$$

By convention, we set $\gamma = 0$ if and only if $\delta = 0$. In order for f to be strictly positive, continuous and proper, these additional constraints must be satisfied: ***i)*** *$\theta > 1$ if $\beta \neq 0$,* ***ii)*** *$\theta > 0$ if $\beta = 0, \alpha \neq 0$,* ***iii)*** *$\alpha > 1$ or $\alpha = 1, \beta > 1$ if $\gamma = \delta = 0$.*

The D-GEP density is in general unimodal, except maybe if $\alpha < 0$ and/or $\beta < 0$. In this case, it suffices to choose θ large enough to guarantee unimodality. More precisely, θ must be chosen such that $\gamma\delta\theta^{\gamma} + \alpha + \beta/\log\theta \geq 0$. Furthermore, it can be verified that $f(x) \in E_{\gamma,\delta,\alpha}(\pm\infty)$. It suffices to choose $\gamma < 1$ to satisfy the condition of robustness relative to the tails.

Definition 6. *A random variable Y has a log-GEP distribution, written $Y \sim$ log-GEP$(\gamma, \delta, \alpha, \beta, \theta, \tau)$, if its density is given by*

$$g(y) = K(\gamma, \delta, \alpha, \beta, \theta)(0.5/\tau)(1/y)\exp[-\delta(|\log y|/\tau + \theta)^{\gamma}] \times (|\log y|/\tau + \theta)^{-\alpha}(\log(|\log y|/\tau + \theta))^{-\beta},$$

where $y > 0$, $\gamma \geq 0, \delta \geq 0, \alpha \in \mathbb{R}, \beta \in \mathbb{R}, \theta \geq 0, \tau > 0$. The constant $K(\gamma, \delta, \alpha, \beta, \theta)$ is given by equation (7). By convention, we set $\gamma = 0$ if and only if $\delta = 0$. In order for g to be strictly positive, continuous and proper, these additional constraints must be satisfied: ***i)*** *$\theta > 1$ if $\beta \neq 0$,* ***ii)*** *$\theta > 0$ if $\beta = 0, \alpha \neq 0$,* ***iii)*** *$\alpha > 1$ or $\alpha = 1, \beta > 1$ if $\gamma = \delta = 0$.*

The median of $g(y)$ is 1. If a scale parameter σ is added to the density, the median of $(1/\sigma)g(y/\sigma)$ is σ. We can see there is a symmetry with respect to the median, in the sense that $(y/\sigma)g(y/\sigma) = (\sigma/y)g(\sigma/y)$ or $yg(y) = (1/y)g(1/y)$ if $\sigma = 1$. The left tail of the density $g(y)$, as $y \to 0$, can be anything from 0 to ∞. More precisely, it is i) a positive constant if $(\gamma, \delta, \alpha, \beta) = (1, \tau, 0, 0)$, ii) 0 if $(\gamma, \delta, \alpha, \beta) > (1, \tau, 0, 0)$ and iii) ∞ if $(\gamma, \delta, \alpha, \beta) < (1, \tau, 0, 0)$.

7 Conclusion

In this paper, the behavior of the posterior density of a location-scale model has been investigated when the sample contains outliers. The families of L-exponentially varying functions have been introduced. Simple conditions on the tails of the likelihood, using L-exponentially varying functions, are established to determine the proportion of observations that can be rejected as outliers. We have shown that the posterior distribution converges in law to the posterior that would be obtained from the reduced sample, excluding the outliers.

8 References

ANGERS, J.-F. (1992) Use of Student-t prior for the estimation of normal means: A computational approach, *Bayesian Statistic IV*, eds. Bernardo, J.M., Berger, J.O., David, A.P., and Smith, A.F.M., New York: Oxford University Press, pp. 567-575.

ANGERS, J.-F. (1996) Protection against outliers using a symmetric stable law prior, *IMS Lecture Notes - Monograph Series*, Vol. **29**, pp. 273-283.

ANGERS, J.-F. (2000) P-Credence and outliers, *Metron*, **58**, 81-108.

ANGERS, J.-F. and BERGER, J.O. (1991) Robust hierarchical Bayes estimation of exchangeable means, *The Canadian Journal of Statistics*, **19**, 39-56.

CARLIN, B. and POLSON, N. (1991) Inference for nonconjugate Bayesian models using the Gibbs sampler, *The Canadian Journal of Statistics*, **19**, 399-405.

DAWID, A.P. (1973) Posterior expectations for large observations, *Biometrika*, **60**, 664-667.

DE FINETTI, B. (1961) The Bayesian approach to the rejection of outliers, *Proceedings of the Fourth Berkeley Symposium on Probability and Statistics*, **(Vol. 1)**, Berkeley: University of California Press, pp. 199-210.

DESGAGNÉ, A. and ANGERS, J.-F. (2003) *Computational aspect of the generalized exponential power density*, Technical report CRM-2918, Université de Montréal (http://www.crm.umontreal.ca/pub/Rapports/2900-2999/2918.pdf).

DESGAGNÉ, A. et ANGERS, J.-F. (2005) Importance Sampling with the Generalized Exponential Power Density, *Statistics and Computing*, **15**, 189-195.

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

FAN, T.H. and BERGER, J.O. (1992) Behaviour of the posterior distribution and inferences for a normal means with t prior distributions, *Statistics & Decisions*, **10**, 99-120.

GEWEKE, J. (1994) Priors for macroeconomic time series and their applications, *Econometric Theory*, **10**, 609-632.

HILL, B.M. (1974) On coherence, inadmissibility and inference about many parameters in the theory of least squares, *Studies, Bayesian Econometrics and Statistics*, eds. S. E. Fienberg and A. Zellner, Amsterdam: North-

Holland, pp. 555-584.

MEINHOLD, R. and SINGPURWALLA, N. (1989) Robustification of Kalman filter models, *Journal of the American Statistical Association*, **84**, 479-486.

O'HAGAN, A. (1979) On outlier rejection phenomena in Bayes inference, *Journal of the Royal Statistical Society*, Ser. B, **41**, 358-367.

O'HAGAN, A. (1988) Modelling with heavy tails, *Bayesian Statistic III*, eds. Bernardo, J.M., DeGroot M.H., Lindley, D.V., and Smith, A.F.M., Oxford: Clarendon Press, pp. 345-359.

O'HAGAN, A. (1990) Outliers and credence for location parameter inference, *Journal of the American Statistical Association*, **85**, 172-176.