Bayesian robustness modelling using subexponential distributions and related classes of distributions

Andrade J.A.A.
Omey E.

HUB RESEARCH PAPERS 2011/22
ECONOMICS & MANAGEMENT
NOVEMBER 2011

Bayesian robustness modelling using subexponential distributions and related classes of distributions

J.A.A. Andrade[a,1], Omey, E.[b]

[a] Federal University of Ceara, 60455-670 Fortaleza - CE, Brazil.
[b] HUB-EHSAL, Stormstraat 2, 1000 Brussels, Belgium.

1 Introduction

In general, we model a certain phenomenon of interest by translating each item of evidence that we have about it (data or prior knowledge) into a probability law. In practical situations our belief about which probability law should be used is inevitably subject to error. Firstly, modelling is only an attempt to describe the reality which is much more complex than our whole knowledge and experience can account for. This is a natural distinguishing mark between theory and practice. Not to mention that usually we do not make precise judgements about the accuracy of the model in explaining the real world, subconsciously we accept that our chosen model is *adequate* to solve a given problem. Notwithstanding the fact that statistical models are quite successful in solving problems in real world, in some practical situations those models may be disturbed by *surprising events* such as outliers, gross errors (such as copying), or even the initial assumptions, which led us to choose a particular model, may be mistaken. Those problems, of course, may potentially lead to inaccurate results.

As in Bayesian analysis there are more sources of information to model than the classical approach we may have more potential problems concerning the robustness of the posterior distribution. In addition to the outlier problem, the prior information, like the data, is also subject to modelling problems. In general, it is easier to model data than prior knowledge, since usually data are generated by a random process and most of the times we have good evidences (such as histograms) about which probability law could be chosen to represent them. On the other hand, prior information is often

[1]Corresponding author. E-mail address: ailton@ufc.br.

regarded as a drawback in Bayesian statistics due to the complexity that is found in translating experts' opinions into a probability density.

In Bayesian context the idea of surprising events is associated with the presence of conflicting information. A detailed discussion about conflicts and their potential effects on the posterior distribution can be found in O'Hagan & Forster (2004, §3.35). Broadly speaking, we say that two sources of information *conflict* if they disagree; that is, the functions (likelihood/densities) concerning the parameter of interest are concentrated far away from each other. As extensively discussed in the literature (e.g. Finetti (1961), Lindley (1968) and O'Hagan & Forster (2004)), problems of conflicts are directly related with the tails thickness.

To the purpose of resolve problems of conflicts, a long literature has been developed aiming to establish *sufficient* conditions on the distributions in the model in order to make the posterior distribution unaffected by conflicts (surprising events). In the pure location-parameter case, Dawid (1973) and O'Hagan (1979) proposed sufficient conditions on the data and prior distribution which allows to resolve conflict by rejecting the conflicting information in favour of the other source. Some further development of the ideas of Dawid and O'Hagan can found in O'Hagan (1988 and 1990), Pericchi *et al* (1993) and Pericchi & Sansó (1995), O'Hagan & Le (1994) and Le & O'Hagan (1998) and finally Haro-López and Smith (1999), who proposed some conditions on multivariate v-spherical family (Fernandez *et al*, 1995) involving location and scale parameters in order to bound the influence of the likelihood over the posterior distribution. However, their approach establishes conditions which are quite difficult to verify, and does not provide explicitly the limiting posterior distribution.

Andrade & O'Hagan (2006) used the theory of *regular variation* in order to resolve conflicts in Bayesian modelling of location and scale parameters structures; Andrade & O'Hagan (2011) generalised their idea to location-scale structures. The advantage of regarding heavy tails as regularly varying distributions is that regular variation provides a much easier interpretation of tails decay, since any distribution with regularly varying tails can be represented simply as a power function. Moreover, concepts created in the literature (such as Credence O'Hagan (1988)) have equivalents in the regular variation theory. See Andrade & O'Hagan (2006).

The literature developed so far has been channelled only in providing sufficient conditions to the posterior distribution become robust to atypical events. Let x be an outlier and suppose we have single parameter model, in which $x|y \overset{D}{\sim} f(x|y)$ and $y \overset{D}{\sim} p(y)$, the idea was to propose sufficient conditions on f and on p so that the posterior distribution becomes unaffected by the outlier x as it becomes large, that is $p(y|x) \to g(y)$ as $x \to \infty$, where g is a density which does not involve the outlier x.

In this work we extend the ideas proposed by Andrade & O'Hagan (2006) to wider families of heavy-tailed distributions, in which we propose *sufficient* conditions on the location and scale parameter structures. In Section 2 we provide the definitions and some properties of the classes of distributions which we use throughout the paper. In Section 3 we we find sufficient conditions on the location parameter structure in order to reject observations in the sample which are far away from the other sources of information. In Section 4 we consider the scale parameter case, in which we propose alternative conditions to those proposed by Andrade & O'Hagan (2006). We illustrate the theory in Section 5, where we provides examples involving distributions belonging to the classes studied. Finally, we make some general comments in Section 6.

2 Some classes of distribution functions

In this section we recall some important classes of functions (and some proprieties) that play an important role in the models that we discuss. The basic reference for these classes is Bingham, Goldie & Teugels (1985), which will be cited as BGT from now on.

Definition 1 *A measurable function f is regularly varying at ∞ and with index $\rho \in \Re$, written $f \in RV(\rho)$, if it satisfies:*

$$\lim_{x\to\infty} \frac{f(xy)}{f(x)} = y^{\rho}, \forall y > 0. \tag{1}$$

In particular, if $\rho = 0$, f is said to be slowly varying.

Definition 2 *A measurable function f is O-regularly varying at ∞, written $f \in ORV$, if it satisfies:*

$$\lim_{x\to\infty} \sup \frac{f(xy)}{f(x)} < \infty, \forall y > 0. \tag{2}$$

If $f \in ORV$, the upper index of f is given by

$$\alpha(f) = \lim_{y\to\infty} \frac{\log \limsup_{x\to\infty} f(xy)/f(x)}{\log(y)}, \tag{3}$$

and the lower index of f is given by $\beta(f) = \alpha(1/f)$. It can be proved (see BGT, §2.0.1) that if $f \in ORV$, then for any $\beta < \beta(f)$ and $\alpha > \alpha(f)$, there exist constants C, D and x° so that

$$Cy^\beta \leq \frac{f(xy)}{f(x)} \leq Dy^\alpha, \forall y \geq 1, \forall x \geq x^\circ. \tag{4}$$

Definition 3 *A measurable function f is in the class $\mathfrak{L}$ if it satisfies:*

$$\lim_{x\to\infty} \frac{f(x+y)}{f(x)} = 1, \forall y > 0. \tag{5}$$

Equivalently, $f \in \mathfrak{L}$ if and only if $f \circ \log \in RV(0)$. It is well known that $f \in RV(\rho)$ implies that $f \in \mathfrak{L}$. The converse statement is false in general.

Definition 4 *A density function f is a subexponential density, written $f \in SD$, if $f \in \mathfrak{L}$ and if*

$$\lim_{x\to\infty} \frac{f^{\otimes 2}(x)}{f(x)} = 2, \tag{6}$$

where $f^{\otimes 2}(x) = f \otimes f(x)$ is the 2-fold convolution of f.

It can be proved for density functions that if $f \in RV$ or $f \in \mathfrak{L} \cap ORV$ imply that $f \in SD$. See Chover *et al* (1973).

3 Location parameter models

3.1 Notation

Consider a random sample $X = (x_1, x_2, ..., x_n)$ of independent and identically distributed (*iid*) random variables with fixed sample size n. A general location parameter Bayesian model is of the form

$$\begin{aligned} x_i|y &\overset{D}{\sim} f(x_i \mid y) = f(x_i - y),\ 1 \leq i \leq n; \\ y &\overset{D}{\sim} p(y), \end{aligned}$$

where f is a fixed p.d.f. and $p(y)$ is the prior p.d.f. of y which is the parameter of interest.

Suppose that x_i, $1 \leq i \leq k < n$ are large. We set $X_L = (x_1, x_2, ..., x_k)$ and $X^U = (x_{k+1}, ..., x_n)$. We clearly have

$$\begin{aligned} f(X|y) &= \Pi_{i=1}^k f(x_i - y) \times \Pi_{j=k+1}^n f(x_j - y) \\ &= f(X_L \mid y) \times f(X^U \mid y) \\ &= L \times U. \end{aligned}$$

The posterior p.d.f. of y is given by

$$p(y \mid X) = \frac{f(X \mid y)p(y)}{\int_{\Re} f(X \mid y)p(y)dy}. \tag{7}$$

We want to investigate what happens to $p(y \mid X)$ as $x = \min(x_1, x_2, ..., x_k) \to \infty$. This is a situation in which the sample fractions X_L and X^U conflict, in the sense that they carry very diverse information, that is the likelihood of y based on X_L is settled far away from the likelihood of y based on X^U and the prior distribution $p(y)$. This kind of conflict may disturb the posterior distribution and potentially lead to wrong conclusions. In order to avoid this behaviour, the idea is to establish conditions under which

$$p(y \mid X) \to p(y \mid X^U) \propto p(y)f(X^U|y), \text{ as } x \to \infty.$$

In this case we say that the influence of the data over the posterior distribution vanishes, leaving the posterior distribution depending only on the prior distribution and the likelihood of y based on X^U. The model rejects the data X_L in favour of the prior distribution and the rest of the data. This behaviour implies that the posterior distribution is robust to atypical data, that is if x becomes too far away from the prior mode and the data X^U, X_L is rejected.

3.2 Preliminary results

If $f \in \mathfrak{L}$, then as $z \to \infty$ we have $f(z-y)/f(z) \to 1$. If $f \in \mathfrak{L}$, it follows that

$$L = \Pi_{i=1}^k f(x_i - y) \sim \Pi_{i=1}^k f(x_i) \text{ (as } x \to \infty).$$

Now we consider (cf. (7)) the integral $\int_{\Re} f(X \mid y)p(y)dy$ and write

$$\int_{\Re} f(X \mid y)p(y)dy = \int_{\Re} f(X_L \mid y)f(X^U \mid y)p(y)dy.$$

Using Fatou's lemma, we get that for $f \in \mathfrak{L}$,

$$\liminf_{x\to\infty} \frac{\int_{\Re} f(X \mid y)p(y)dy}{\Pi_{i=1}^k f(x_i)} \geq \int_{\Re} f(X^U \mid y)p(y)dy$$

and then also that

$$\limsup_{x\to\infty} p(y \mid X) \leq \frac{f(X^U \mid y)p(y)}{\int_{\Re} f(X^U \mid y)p(y)dy} = p(y \mid X^U).$$

3.3 Main results

Theorem 5 (Densities in $\mathfrak{L} \cap ORV$) *Suppose that f is a bounded density and that $f \in \mathfrak{L} \cap ORV$ with $\alpha(f) < 0$. Also assume that*

$$\int_x^\infty p(y)dy = o(\Pi_{i=1}^k f(x_i)), \text{ as } x \to \infty. \tag{8}$$

Then we have

$$p(y \mid X) \to p(y|X^U), \text{ as } x \to \infty.$$

Proof. For the integral in (7), we write

$$\begin{aligned}\int_{\Re} f(X \quad | \quad y)p(y)dy &= (\int_{-\infty}^0 + \int_0^{x/2} + \int_{x/2}^\infty)L \times U \times p(y)dy \\ &= I + II + III.\end{aligned}$$

First consider II. In II we have $0 \le y \le x/2$ and it follows that

$$x_i - x/2 \le x_i - y \le x_i, \; 1 \le i \le k$$

and then also that

$$x_i/2 \le x_i - y \le x_i, \; 1 \le i \le k.$$

First note that for $f \in \mathfrak{L}$ we have $L \sim \Pi_{i=1}^k f(x_i)$, as $x \to \infty$. Since $f \in ORV$, it follows that in II, $L/\Pi_{i=1}^k f(x_i)$ is bounded. Since f is bounded (by assumption) we have that U is bounded, then there exists a constant C such that

$$\int_{\Re} U \times p(y)dy < \infty < C \int_{\Re} p(y)dy = C,$$

hence

$$\frac{II}{\Pi_{i=1}^k f(x_i)} \to \int_0^\infty U \times p(y)dy.$$

Next we consider I. In I we have $x_i \le x_i - y$. Using (4) we get that

$$\frac{f(x_i - y)}{f(x_i)} = \frac{f(x_i(x_i - y)/x_i)}{f(x_i)} \le D \left(\frac{x_i - y}{x_i}\right)^\alpha,$$

where $\alpha(f) < \alpha$. Since $\alpha(f) < 0$, we can choose $\alpha < 0$ and then we see that in I, $L/\Pi_{i=1}^k f(x_i)$ is bounded. Again Lebesgue's theorem can be applied to obtain that

$$\frac{I}{\Pi_{i=1}^k f(x_i)} \to \int_{-\infty}^0 U \times p(y)dy.$$

For the third term III we use the assumption that f is a bounded density. In this case we obtain that

$$III \leq (\sup f(z))^n \times \int_{x/2}^{\infty} p(y)dy.$$

By our assumption on p we obtain that

$$III = o(\Pi_{i=1}^k f(x_i/2)).$$

Using $f \in ORV$,we conclude that $III = o(\Pi_{i=1}^k f(x_i))$. This proves the theorem. ■

Theorem 6 (Densities in $\mathfrak{L}$) *Suppose that $f \in \mathfrak{L}$ and that f is bounded. Also suppose that there exists $s > 0$ such that*

$$\int_{-\infty}^{0} e^{-sy} U \times p(y)dy + \int_{0}^{\infty} e^{sy} U \times p(y)dy < \infty.$$

If $\int_x^\infty p(y)dy = o(\Pi_{i=1}^k f(x_i))$, as $x \to \infty$, we have that

$$p(y \mid X) \to p(y \mid X^U), \text{ as } x \to \infty.$$

Proof. Since $f \in \mathfrak{L}$, we still have $L/\Pi_{i=1}^k f(x_i) \to 1$ as $x \to \infty$. Also we have $F := f \circ \log \in RV(0)$. Using (4), for each $\varepsilon > 0$, we can find constants A, B, z° so that

$$\begin{aligned} \frac{f(z-y)}{f(z)} &\leq Ae^{-\varepsilon y},\ y \leq 0, z \geq z^\circ, \\ \frac{f(z-y)}{f(z)} &\leq Be^{\varepsilon y},\ y \geq 0, z - y \geq z^\circ, z \geq z^\circ. \end{aligned} \tag{9}$$

Now we write the integral as follows:

$$\begin{aligned} \int_{\Re} f(X|y)p(y)dy = (\int_{-\infty}^{0} + \int_{0}^{x-z^\circ} + \int_{x-z^\circ}^{\infty})L \times U \times p(y)dy \\ = I + II + III. \end{aligned}$$

First consider I. For L we have $y \leq 0$ and $x_i \geq x$. It follows from (9) that for $x \geq z^\circ$ we have

$$L = \Pi_{i=1}^k f(x_i - y) \leq \Pi_{i=1}^k A f(x_i) e^{-\varepsilon y}.$$

Taking $s = \varepsilon k$, we can use Lebesgue's theorem to see that

$$\frac{I}{\Pi_{i=1}^k f(x_i)} \to \int_{-\infty}^{0} U \times p(y)dy.$$

Now consider II. In II we have $0 \leq y \leq x - z^\circ$ so that

$$z^\circ \leq x_i - x + z^\circ \leq x_i - y \leq x_i.$$

It follows from (9) that for $x \geq z^\circ$ we have

$$L = \Pi_{i=1}^k f(x_i - y) \leq \Pi_{i=1}^k B f(x_i) e^{\varepsilon y}.$$

Taking $s = \varepsilon k$, we can use Lebesques theorem to see that

$$\frac{II}{\Pi_{i=1}^k f(x_i)} \to \int_{0}^{\infty} U \times p(y)dy.$$

Finally consider III. Since f has a bounded density, we find that

$$III \leq (\sup f(z))^n \int_{x-z^\circ}^{\infty} p(y)dy.$$

Our assumption about p shows that

$$III = o(\Pi_{i=1}^k f(x_i - z^\circ).$$

Using $f \in \mathfrak{L}$, we obtain that $III = o(\Pi_{i=1}^k f(x_i)$. This proves the result. ■

Now, assume $k = 1$ in Section 3.1, we get just one outlier and then we have $L = f(x_1 - y)$. In this subsection we assume $f \in SD$, that is $f \in \mathfrak{L}$ and f satisfies $f \otimes f(x)/f(x) \to 2$, as $x \to \infty$.

Theorem 7 (Densities in SD) *Suppose that $f \in SD \subset \mathfrak{L}$ and that $p(|x|) = o(f(|x|))$, then $p \otimes f(x)/f(x) \to 1$.*

Proof. We write

$$\begin{aligned} p \otimes f(x) &= \left(\int_{-\infty}^{-x^\circ} + \int_{-x^\circ}^{x^\circ} + \int_{x^\circ}^{\infty} \right) p(y) f(x-y) dy \\ &= I + II + III. \end{aligned}$$

Since $p(-x) = o(1)f(-x)$ as $x \to \infty$, for $\varepsilon > 0$ we can find x° so that we have

$$I \leq \varepsilon \int_{-\infty}^{-x^\circ} f(y) f(x-y) dy \leq \varepsilon f \otimes f(x),$$

It follows that

$$\limsup \frac{I}{f(x)} \leq 2\varepsilon.$$

For III, in a similar way we find that

$$\limsup \frac{III}{f(x)} \leq 2\varepsilon.$$

Now consider II. Using $f \in \mathfrak{L}$, we get that

$$\frac{II}{f(x)} \to \int_{-x^\circ}^{x^\circ} p(y)dy.$$

By choosing x° sufficiently large, we obtain that

$$1 - \int_{-x^\circ}^{x^\circ} p(y)dy \leq \varepsilon,$$

We conclude that

$$\limsup_{x\to\infty} \left| \frac{p \otimes f(x)}{f(x)} - 1 \right| \leq 5\varepsilon.$$

Now let $\varepsilon \to 0$, to get the desired result. ∎

A following theorem provides the same result, but with conditions slightly different from Those of Theorem 7.

Theorem 8 (Densities in SD) *Assume that $f \in SD$, (this is: $f \in \mathfrak{L}$ and $f \otimes f(x)/f(x) \to 2$, as $x \to \infty$). Also assume that $p(|x|) \sim \alpha f(|x|)$ where $\alpha > 0$. Then $p \otimes f(x) \sim (\alpha + 1)f(x)$.*

Proof. We choose a in such a way that

$$(\alpha - \varepsilon)f(|x|) \leq p(|x|) \leq (\alpha + \varepsilon)f(|x|), \forall x \text{ with } |x| \geq a.$$

Now choose $x^\circ \geq a$. We reconsider I and III from the proof of Theorem 7 and get that

$$(\alpha - \varepsilon) \int_{-\infty}^{-x^\circ} f(y)f(x-y)dy \leq I \leq (\alpha + \varepsilon) \int_{-\infty}^{-x^\circ} f(y)f(x-y)dy$$

$$(\alpha - \varepsilon) \int_{x^\circ}^{\infty} f(y)f(x-y)dy \leq III \leq (\alpha + \varepsilon) \int_{x^\circ}^{\infty} f(y)f(x-y)dy$$

It follows that

$$(\alpha+\varepsilon)\left\{f\otimes f(x)-\int_{-x^\circ}^{x^\circ} f(y)f(x-y)dy\right\}\leq I+III$$
$$\leq(\alpha+\varepsilon)\left\{f\otimes f(x)-\int_{-x^\circ}^{x^\circ} f(y)f(x-y)dy\right\},$$

and using $f\in\mathfrak{L}$ we obtain that

$$(\alpha-\varepsilon)(2-\int_{-x^\circ}^{x^\circ} f(y)dy)\leq\lim_{x\to\infty}\binom{\sup}{\inf}\frac{I+III}{f(x)}\leq(\alpha+\varepsilon)(2-\int_{-x^\circ}^{x^\circ} f(y)dy).$$

For II we obtain that (use $f\in\mathfrak{L}$), we get that

$$\frac{II}{f(x)}\to\int_{-x^\circ}^{x^\circ} p(y)dy.$$

We can find x° sufficiently large such that

$$1-\varepsilon\leq\int_{-x^\circ}^{x^\circ} f(y)dy\leq 1,\ 1-\varepsilon\leq\int_{-x^\circ}^{x^\circ} p(y)dy\leq 1.$$

We get that

$$(\alpha-\varepsilon)+1-\varepsilon\leq\lim_{x\to\infty}\binom{\sup}{\inf}\frac{p\otimes f(x)}{f(x)}\leq(\alpha+\varepsilon)(1+\varepsilon)+1.$$

Now let $\varepsilon\to 0$ to get the desired result. ■

3.4 Many observations

Theorems 7 and 8 can be extended to many observations with some of them possibly being outliers. In order to show this, we use the same strategy as in Andrade & O'Hagan (2006). Consider a random sample $\boldsymbol{x}=(x_1,x_2,...,x_n)$ *iid* with a p.d.f. $f(x_i|y)=f(x_i-y)\,\forall i$, where y is a location parameter. Let also $y\overset{D}{\sim}p(y)$ (prior distribution).

Now we have the situation in which a few observations are very large (i.e. tending to infinity). In other words, we have $(x_1,...,x_k)$ $(k\leq n)$ tending to infinity. This is equivalent to think of the k observations close to each other and tending to infinity, that is $x_i=x+\xi_i$ for ξ_i fixed $(i=1,...,k)$. Thus we can write the joint distribution of $(x_1,...,x_k)$ as

$$f(x_1,...,x_k|y)=\prod_{i=1}^{k} f(x+\xi_i-y)=g(x-y),$$

which clearly keeps the location structure. In fact, the outliers behave like a single observation as $x \to \infty$.

The joint distribution of the rest of the observations is given by

$$f(x_{k+1}, ..., x_n|y) = \prod_{i=k+1}^{n} f(x_i - y) = U.$$

The posterior can be written as

$$p(y|\boldsymbol{x}) = \frac{U \times g(x-y) \times p(y)}{\int_{\mathbb{R}} U \times g(x-y) \times p(y)dy}.$$

Let $p^*(y) = U \times p(y)$, it follows that we have the same structure of Theorem 8, holds if $p^*(|x|) \sim \alpha g(|x|)$ as $x \to \infty$. In this case the posterior distribution

$$p(y|\boldsymbol{x}) \to \frac{U \times p(y)}{\int_{\mathbb{R}} U \times p(y)dy} \; (x \to \infty).$$

4 Scale parameter models

4.1 Notation

As shown by Andrade & O'Hagan (2006), regular variation provides a natural way to deal with scale parameters since the scale structure of a scale parameter model is the same as in the definition of regular variation.

Consider a sample $X = (x_1, x_2, ..., x_n)$ of independent and identically distributed (iid) with fixed sample size n. A typical scale parameter model is of the form

$$\begin{aligned} x_i|y &\overset{D}{\sim} f(x_i \mid y) = y^{-1}h(x_i/y),\ 1 \le i \le n; \\ y &\overset{D}{\sim} p(y), \end{aligned}$$

where f is the data p.d.f. and $p(y)$ is the prior p.d.f. of y which is the parameter of interest. For convenience, we assume that all random variables involved are concentrated on the positive halfline.

Suppose that x_i, $1 \le i \le k < n$ are large. As before we define $X_L = (x_1, x_2, ..., x_k)$, $x = \min(x_1, ..., x_k)$ and $X^U = (x_{k+1}, ..., x_n)$. We clearly have

$$\begin{aligned} f(X \mid y) &= y^{-n}\Pi_{i=1}^{k} h(x_i/y) \times \Pi_{j=k+1}^{n} h(x_j/y) \\ &= y^{-n} \times L \times U. \end{aligned}$$

The posterior p.d.f. of y is given by

$$p(y \mid X) = \frac{f(X \mid y)p(y)}{\int_{\Re} y^{-n} L \times U \times p(y)dy}. \tag{10}$$

4.2 Main result

Now we suppose that $h \in RV(\alpha)$. In this case it is easy to see that

$$\frac{L}{\Pi_{i=1}^{k} h(x_i)} \to y^{-\alpha k}, \text{ as } x \to \infty.$$

We need extra conditions to see what happens in (10) if $x \to \infty$. An alternative to the conditions proposed by Andrade & O'Hagan (2006) is provided by the next result.

Theorem 9 (Densities in RV) *Suppose that $h \in RV(\alpha)$ with $\alpha < 0$ and suppose that h is bounded on bounded intervals. Assume that for $\varepsilon > 0$ we have*

$$\int_0^1 y^{-(\alpha+\varepsilon)k} U \times p(y) dy + \int_1^\infty y^{-(\alpha-\varepsilon)k} U \times p(y) dy < \infty, \tag{11}$$

Then

$$\frac{1}{\Pi_{i=1}^{k} h(x_i)} \int_{\Re} L \times U \times p(y) dy \to \int_0^\infty y^{-\alpha k} U \times p(y) dy < \infty \tag{12}$$

Proof. We have

$$\int_{\Re} L \times U \times p(y) dy = (\int_0^1 + \int_1^\infty) L \times U \times p(y) dy = I + II.$$

First consider I and write

$$\frac{I}{\Pi_{i=1}^{k} h(x_i)} = \int_0^1 \Pi_{i=1}^{k} \frac{h(x_i/y)}{h(x_i)} U \times p(y) dy.$$

Since $1 \leq 1/y$ and $h \in RV(\alpha)$, for each $\varepsilon > 0$ we can find constants C and z° such that

$$\frac{h(z/y)}{h(z)} \leq C y^{-\alpha-\varepsilon}, \forall z \geq z^\circ, \forall y \leq 1.$$

It follows that

$$\Pi_{i=1}^{k} \frac{h(x_i/y)}{h(x_i)} \leq C^k y^{-\alpha k - \varepsilon k}, \forall x \geq z^\circ, \forall y \leq 1.$$

Using (11) it follows that we can apply dominated convergence and we find that

$$\frac{I}{\Pi_{i=1}^{k} h(x_i)} \to \int_0^1 y^{-\alpha k} U \times p(y) dy.$$

Now consider II. Using $h \in RV(\alpha)$, for each $\varepsilon > 0$, we can find constants D and z° such that

$$\frac{h(x_i/y)}{h(x_i)} \leq Dy^{-\alpha+\varepsilon}, \forall y \geq 1 \text{ and } x_i/y \geq z^\circ.$$

Now consider the case where $y \geq 1$, $x_i/y < z^\circ$ and $x_i \geq z^\circ$. Since we assume that h is bounded on bounded intervals, we get that

$$\begin{aligned}\frac{h(x_i/y)}{h(x_i)} &\leq \sup_{u \leq z^\circ} h(u) \frac{1}{h(x_i)} \\ &= \sup_{u \leq z^\circ} h(u) \frac{1}{x_i^{-\alpha+\varepsilon} h(x_i)} x_i^{-\alpha+\varepsilon}.\end{aligned}$$

Since $z^{-\alpha+\varepsilon}h(z) \to \infty$ as $z \to \infty$, we can find a constant D° so that

$$\frac{h(x_i/y)}{h(x_i)} \leq D^\circ x_i^{-\alpha+\varepsilon}, \forall y \geq 1, x_i/y < z^\circ, x_i \geq z^\circ.$$

Now it follows that

$$\frac{h(x_i/y)}{h(x_i)} \leq D^\circ y^{\alpha-\varepsilon} x_i^{-\alpha+\varepsilon} y^{-\alpha+\varepsilon}.$$

Since $\alpha < 0$ and $x_i/y < z^\circ$, we get that

$$\frac{h(x_i/y)}{h(x_i)} \leq D^\circ (z^\circ)^{-\alpha+\varepsilon} y^{-\alpha+\varepsilon} = Fy^{-\alpha+\varepsilon}.$$

Combining these estimates, we have proved that we can find a constant G such that

$$\frac{h(x_i/y)}{h(x_i)} \leq Gy^{-\alpha+\varepsilon}, \forall y \geq 1, \forall x_i \geq z^\circ.$$

Assumption (11) can be used and applying dominated convergence, we get that

$$\frac{II}{\Pi_{i=1}^k h(x_i)} \to \int_1^\infty y^{-\alpha k} U \times p(y) dy.$$

Combining the results for I and II, we obtain (12). ■

4.3 Remark

We briefly discuss conditions under which condition (11) holds. We prove the following result

Proposition 10 *Suppose that $h \in RV(\alpha)$ is bounded on bounded intervals. Also assume that*

$$\int_0^1 y^{\theta-(n-k)(\alpha+\varepsilon)} p(y)dy < \infty \text{ resp. } \int_1^\infty y^\theta p(y)dy < \infty.$$

Then for fixed X^U,

$$\int_0^1 y^\theta U \times p(y)dy < \infty, \text{ resp. } \int_1^\infty y^\theta U \times p(y)dy < \infty.$$

Proof. First consider an integral of form

$$\int_0^1 y^\theta U \times p(y)dy.$$

Clearly we have

$$\int_0^1 y^\theta U \times p(y)dy = (\int_0^a + \int_a^1) y^\theta U \times p(y)dy = A + B,$$

where $0 < a < 1$. First consider B. Since we assume that h is bounded on bounded intervals, we have

$$B \leq \int_a^1 y^\theta p(y)dy \leq \max(1, a^\theta) \int_a^1 p(y)dy < \infty.$$

Next consider A. Using $z^{-\alpha-\varepsilon}h(z) \to 0$, we can find z° such that $h(z) \leq \varepsilon z^{\alpha+\varepsilon}, z \geq z^\circ$. It follows that

$$h(x_i/y) \leq \varepsilon(x_i/y)^{\alpha+\varepsilon}, x_i/y \geq z^\circ.$$

Since $y \leq a$, we should choose a such that $x_i/a \geq z^\circ$, or $a \leq x_i/z^\circ$. Having done this, we find

$$\begin{aligned} A &\leq \int_a^1 y^\theta \Pi_{i=k+1}^n \varepsilon(x_i/y)^{\alpha+\varepsilon} p(y)dy \\ &= \varepsilon^{n-k}\Pi_{i=k+1}^n x_i^{\alpha+\varepsilon} \int_a^1 y^{\theta-(n-k)(\alpha+\varepsilon)} p(y)dy < \infty. \end{aligned}$$

Next consider an integral of the form

$$\int_1^\infty y^\theta U \times p(y)dy.$$

Since we assume that h is bounded on bounded intervals, we have

$$h(x_i/y) \leq \sup_{z \leq x_i} h(z) = s(x_i)$$

and then also

$$\int_1^\infty y^\theta U \times p(y)dy \leq \Pi_{i=k+1}^n s(x_i) \int_1^\infty y^\theta p(y)dy < \infty.$$

This proves the result. ■

5 Examples

We illustrate the theory with a general problem of estimating the location and scale parameters of the random sample $\boldsymbol{x} = (2, 3, 3, 4, x_5)$, where we take x_5 arbitrarily large in order to observe the behaviour of the posterior distribution of the location and the scale parameters. The general model is

$$\begin{cases} x_i|y,\sigma^2 & \overset{D}{\sim} & f(x_i|y,\sigma) = \sigma^{-1} h\left(\frac{x_i - y}{\sigma}\right) \ iid, \ i = 1, ..., 5 \\ y & \overset{D}{\sim} & p(y) \\ \sigma & \overset{D}{\sim} & \pi(\sigma) \end{cases}. \quad (13)$$

We use the `OpenBugs` software which uses the MCMC methods for sampling from the posterior distribution, in all the cases the algorithm was run until its convergence, then the posterior estimates of the location and of the scale parameters were computed.

In order to achieve rejection of the outlying observation we need to model accordingly to the theorems above. This basically means to choose suitably heavy-tailed distributions for the data and prior distributions with lighter tails for the location and the scale parameters. As our purpose is to illustrate the theory, thus we opt for quite strong prior information (small variances), which will make the MCMC algorithm to achieve convergence more quickly. Thus we expect to base the posterior estimates on the prior information and on the non-outlying observations $X^U = (2, 3, 3, 4)$. As for the data distribution, we propose four different choices for f, namely, models: (I) f light-tailed, (II) $f \in RV$, (III) $f \in \mathfrak{L}$ and (IV) $f \in SD$. Thus we assess the behaviour of the posterior estimates as we disturb the data by increasing x_5.

We need to verify if the distributions of Models (I)-(IV) satisfy the conditions of the Sections 3 and 4.

Model I) The traditional light-tailed choice for f is a normal distributions with mean (location) y and standard deviation (scale) σ. It easy to verify that the normal distributions does not belong to any of the families above. In fact,

$$\lim_{x\to\infty} f(x-y)/f(x) = \begin{cases} 1, & y=1; \\ 0, & y<0; \\ \infty, & y>0. \end{cases},$$

hence $f \notin \mathfrak{L}$, therefore $f \notin RV$ and $f \notin SD$. Notice that $f \notin ORV$, since we the limit (2) is *infinity* as $0 < y < 1$. As for the prior information, we assign $y \overset{D}{\sim} N(0, 0.05)$ and $\sigma \overset{D}{\sim} \mathrm{G}(3, 10)$.

Model II) Besides being bounded, by (1), the Student's t distribution with d degrees of freedom and is regularly varying with index $-(d+1)$. Thus we assign to $f(y_i|y, \sigma^2)$ a t distribution with $d = 4$ degrees of freedom, mean y and variance σ^2. In addition we assign $y \overset{D}{\sim} N(0, 0.05)$ for the prior distribution of y and $\sigma \overset{D}{\sim} \mathrm{G}(3, 10)$. Now we need to verify the conditions of Theorems 5 (location parameter) and 9 (scale parameter).

Location parameter: We have to show that

$$[1-\Phi(x)]/\prod_{i=1}^{n} f(x_i) \to 0, \quad \text{as } x \to \infty,$$

where Φ is the cumulative distribution of the standard normal distribution. In fact, $1 - \Phi(x) = \mathrm{erfc}(x/\sqrt{2})/2$, where erfc is the complementary error function

$$erfc(x) = 2(2\pi)^{-1/2} \int_x^{\infty} e^{-x^2/2} dx,$$

which has the asymptotic expansion

$$erfc\left(\frac{x}{\sqrt{2}}\right) = \frac{2e^{-x^2/2}}{\sqrt{2\pi}} \sum_{n=0}^{n} \frac{(-1)^n 1 \cdot 2 \cdot 3 \cdots (2n-1)}{x^{2n+1}}. \qquad (14)$$

For any k $(\leq n)$ $\prod_{i=1}^{k} f(x_i) \propto \prod_{i=1}^{k} \left(1 + x_i^2/(d-2)\right)^{-(d+1)/2}$, it follows that $[1-\Phi(x)]/\prod_{i=1}^{k} f(x_i) \to 0$ as $x \to \infty$. Hence the conditions of Theorem 5 as satisfied. Here $k = 1$ which makes the condition even easier to verify.

Scale parameter: It's straightforward by application of Proposition 10, since the prior of the scale parameter is a Gamma distribution, it follows that

$$\int_1^\infty \sigma^\theta p(\sigma) d\sigma < \infty, \text{ for all } \theta > -\alpha = -3.$$

Model III) The Exponential Power Distribution (EPD) (or generalised normal distribution) (Box & Tiao, 1973) is of the form

$$f(y_i|y,\sigma^2) \propto \frac{1}{\sigma} e^{-\left|\frac{y_i - y}{\sigma}\right|^q}.$$

This structure generalises several well know distributions. For instance, if $q = 2$ f is a normal distribution and if $q = 1$ f is the double exponential. For $0 < q \leq 1$ we have $EPD \in \mathfrak{L}$, thus we choose $q = 1/3$. From (1), the EPD is not in the RV class, hence we cannot guarantee robustness of the posterior estimate of the scale parameter, here we assign $\sigma \overset{D}{\sim} \text{G}(3, 0.01)$. As for the location parameter, let $y \overset{D}{\sim} N(0, 0.05)$, again we have to satisfy the conditions of Theorem 5. In fact, similarly to the strategy used in Model II above, $\prod_{i=1}^k f(x_i) \propto \sigma^{-k} \exp\{-\sum_{k=1}^k |x_i|^q\}$, which can be compared with (14), hence Condition (8) is verified.

Model IV) The LogNormal distribution is a well known subexponential distribution (see Goldie & Klüppelberg, 1998). The LogNormal distribution is also in $\mathfrak{L}$, but not in ORV. We consider the model

$$f(y_i|y,\sigma^2) \propto (y_i\sigma)^{-1} e^{-\frac{(\log y_i - y)^2}{2\sigma^2}}, \ i = 1, ..., 5$$

that is y_i is lognormally distributed with location parameter y and scale parameter σ. In addition we choose a Lognormal distribution for the prior distribution of μ, that is $\mu \overset{D}{\sim} \text{LogN}(0, 0.05)$, hence we satisfy the condition that $p(x) \sim f(x)$ $(x \to \infty)$ (Theorem 8). Again, as in the $\mathfrak{L}$ the SD class will not produce a robust posterior distribution for the scale parameter, thus we arbitrarily choose $\sigma \overset{D}{\sim} \text{IG}(3, 10)$.

Note that we have different models, thus we cannot compare the models estimates. In fact, we compare the behaviour of the posterior estimates in the different models.

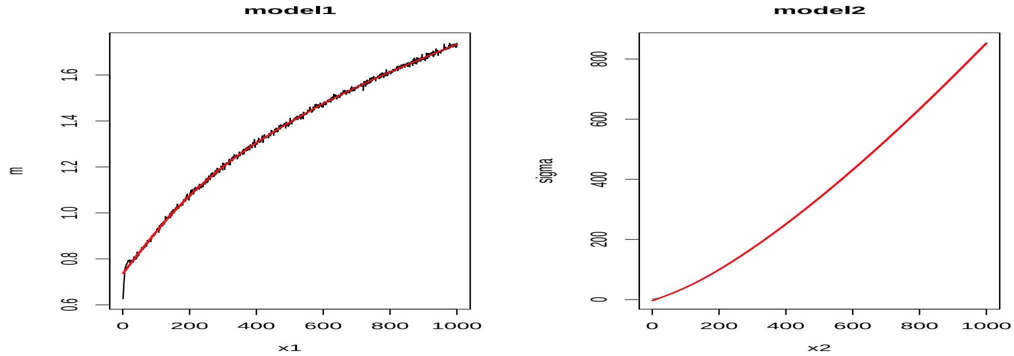
5.1 Results

Although the example looks rather simple, great part of the problems in Statistics concerns the estimation of location and scale parameters. For instance, in linear regression models the parameters of interest are the mean (which involves the regression coefficients) and the variance of the errors.

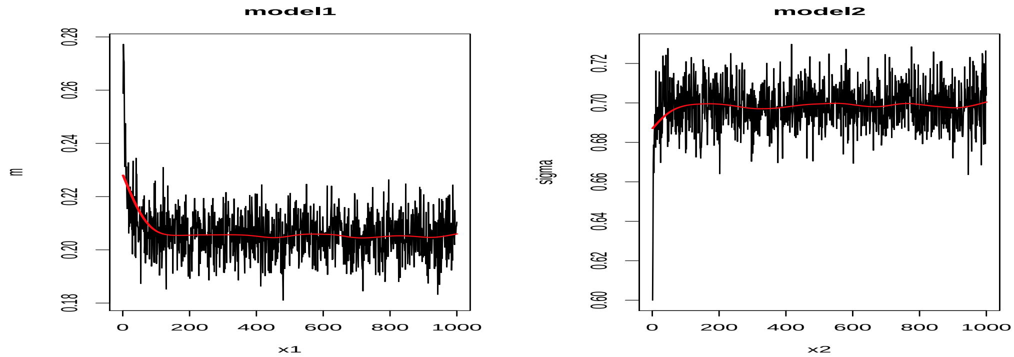
We plot the posterior estimates for each x_5 varying from 1 to 1000, we also plot the tendency line which helps to see the variation of the estimates. Figure 1(a) (Model I) shows a quite common model used in Bayesian analysis, in which both the data and the prior distributions are light-tailed. Note that in this case, when $x_5 \uparrow \infty$ the posterior estimates follow the outlying information faithfully to the infinity. This was the behaviour identified by de Finetti (1961) and described in more details by Lindley (1968). In practice this basically means that if outliers are in the data, the posterior distribution may be disturbed by it, and potentially leading wrong conclusions. As an alternative, Model II (Figure 1(b)) yields a quite robust posterior distribution. In fact, the posterior estimates for y and for σ becomes unaffected by the outlying information when it becomes too large. As pointed out by Andrade & O'Hagan (2006), the posterior estimates reject the outlying data in favor of the rest of the data and the prior information. Model III uses an $\mathfrak{L}$ distribution, which shows in Figure 1(c) that we achieve robustness only on the location parameters, whose posterior estimates tend to a constant, whereas the posterior estimates of the scale parameter tend to infinity as $x_5 \uparrow \infty$. Similarly, in Model IV we cannot control the influence of the outlier in the posterior distribution of σ, which produces estimates very sensitive to changes of x_5, in contrast the posterior estimates of the location parameter y tends to a constant, rejecting the outlier.

6 Discussion

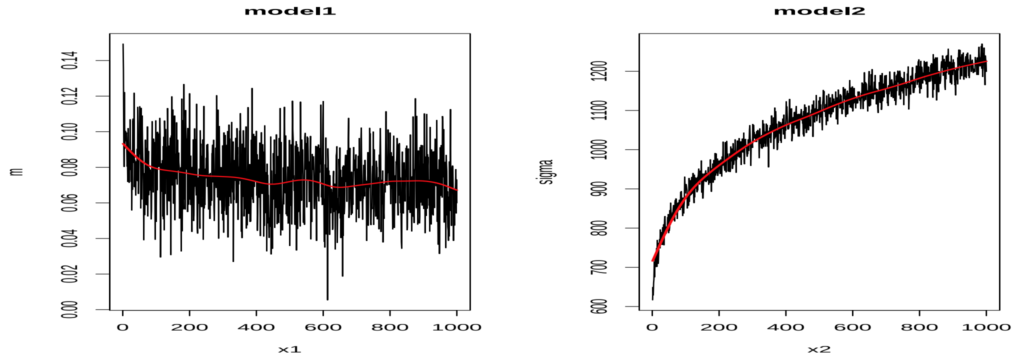
The results above concern the cases where we want to reject some observations or the whole sample in favour of the prior information, although this is the most common form of conflict (most of the problems happen in the data), in some situations one may wish to weaken the prior information in the model, perhaps for finding the prior information not so credible. In this case, the theory presented provides tools for making the the model to behave in the way the modeller wishes. For rejecting some prior information we basically need to model accordingly to the Theorems above, but how focusing in the prior information, that is assign some heavy-tailed distribution to the



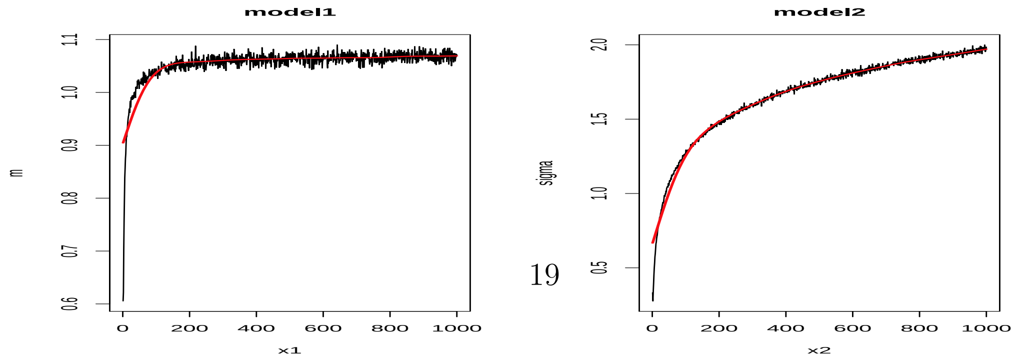
(a) Model I: Posterior estimates of y and σ.



(b) Model II: Posterior estimates of y and σ.



(c) Model III: Posterior estimates of y and σ.



(d) Model IV: Posterior estimates of y and σ.

prior information and some distribution with lighter tails to the data. For instante, in Theorem 5, in order to reject the prior distribution in favour of the data, we need to choose $p \in \mathfrak{L} \cap ORV$, with p bounded, $\alpha(p) < 0$ and $\int_x^\infty f(y)dy = o(p(x))$ as $x \to \infty$. In general we need just to exchange the prior distribution with the data distribution in the theorems presented above. Andrade & O'Hagan (2006, 2011) provides some further description of how to reject prior information of the location and the scale parameter.

Our results involve quite wide classes of heavy-tailed distribution, in particular the $\mathfrak{L}$ and SD embraces most of the distributions whose tails decay like e^{-x^q} ($q < 1$) and those with regularly varying tails, which behave like a polynomial. Distributions like the EPD, Laplace and LogNormal have been used as heavy tails in practical applications (see Pericchi & Sansó, 1995 and Pericchi *et al*, 1993), but without formal description of their classes. As shown in the examples, we cannot achieve posterior robustness on the scale parameters within the classes $\mathfrak{L}$ and SD. As Andrade & O'Hagan (2006) point out, differently from the location case in which conflicts disturb only the location of the posterior distribution, in a scale parameter structure the posterior distribution is affected both on the location as in the dispersion as some observation increases, thus we need quite heavy tails to resolve those conflicts.

The theory of Bayesian robustness modelling for resolution of conflicts has been channeled to find sufficient conditions under which the posterior distribution resolves the conflict in favour of that source of information we regard as the most credible. The literature has been focusing on the theoretical aspects rather than applications. Actually we do not know how efficient the theory is in practical applications. In particular, the verification of the conditions of the theorems might not be so straightforward in hierarchical models in which we may have many sources of prior information. Of course this is a lacuna in the area which needs more work.

References

Andrade, J.A.A. & O'Hagan, A. (2006). Bayesian robustness modelling using regularly varying distributions. *Bayesian Analysis* **1**, No. 1, 169–188.

Andrade, J.A.A. e O'Hagan, A. (2011). Bayesian robustness modelling of location and scale parameters. *To appear on The Scandinavian Journal of Statistics.*

Bingham, N. H., Goldie, C. M. & Teugels, J. L. (1987). Regular variation. *Vol. 27 of Encyclopedia of Mathematics and Its Applications.* Cambridge: Cambridge University Press.

Box, G. E. P. & Tiao, G. C. (1973). Bayesian inference in statistics analysis, New York: Wiley.

Chover, J., Ney, P. & Wainger, S. (1973). Functions of probability measures. J. Analyse Math. **26**, 255302.

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 4th Berkeley Symposium on Mathematical Statistics and Probability*, Vol. I pp. 99-210 Univ. California Press, Berkeley, California.

Fernandez, C., Osiewalski, J. & Steel, M. F. J. (1995). Modelling and inference with v-spherical distributions. *Journal of The American Statistical Association* **90**, 1331-1340.

Goldie, C. M. & Klppelberg, C. (1998). Subexponential distributions. A practical guide to heavy tails (Santa Barbara, CA, 1995), 435459, Birkhuser Boston, Boston, MA.

Haro-López, R. A. & Smith, A. F. M. (1999). On robust Bayesian analysis for location and scale parameters. *Journal of Multivariate Analysis* **70**, no. 1, 30–56.

Le, H. & O'Hagan, A. (1998). A class of bivariate heavy-tailed distributions. *Sankhya* **B 60**, 82–100.

Lindley, D. V. (1968). The choice of variables in multiple regression (with discussion). *Journal of the Royal Statistical Society* **B 30**, 31–66.

O'Hagan, A., (1979). On outliers rejection phenomena in Bayes inference. *Journal of the Royal Statistical Society* **B 41**, 358–367.

O'Hagan, A., (1988). Modelling with heavy tails. In *Bayesian Statistics* **3** (eds. J. M. Bernardo, M. H. DeGroot, D. V. Lindley & A. F. M. Smith), Oxford: Clarendon Press, 345–359.

O'Hagan, A. (1990). On outliers and credence for location parameter inference. *Journal of the American Statistical Association* **85**, 172–176.

O'Hagan, A. & Le, H., (1994). Conflicting information and a class of bivariate heavy-tailed distributions. In *Aspects of Uncertainty: a Tribute to D. V. Lindley*, (eds. P. R. Freeman & A. F. M. Smith), New York: Wiley, 311–327.

O'Hagan, A. & Forster, J. J. (2004). Bayesian inference - Kendall's Advanced Theory of Statistics, Volume 2B. London: Arnold.

Pericchi, L. R. & Sansó, B. (1995). A note on bounded influence in Bayesian analysis. *Biometrika* **82**, 1, 223–225.

Pericchi, L. R., Sansó, B. & Smith, A. F. M. (1993). Posterior cumulant relationships in Bayesian inference involving the exponential family. *Journal of the American Statistical Association* **88**, 1419–1426.