Bayesian Robustness Modeling Using Regularly Varying Distributions

J. A. A. Andrade*   A. O'Hagan†

Abstract.   Bayesian robustness modelling using heavy-tailed distributions provides a flexible approach to resolving problems of conflicts between the data and prior distributions. See Dawid (1973) and O'Hagan (1979, 1988, 1990), who provided sufficient conditions on the distributions in the model in order to reject the conflicting data or the prior distribution in favour of the other source of information. However, the literature has almost concentrated exclusively on robustness of the posterior distribution of location parameters; little attention has been given to scale parameters. In this paper we propose a new approach for Bayesian robustness modelling, in which we use the class of regularly varying distributions. Regular variation provides a very natural description of tail thickness in heavy-tailed distributions. Using regular variation theory, we establish sufficient conditions in the pure scale parameter structure under which is possible to resolve conflicts amongst the sources of information. We also note some important differences between the scale and the location parameters cases. Finally, we obtain new conditions in the pure location parameter structure which may be easier to verify than those proposed by Dawid and O'Hagan.

Keywords: Bayesian robustness, heavy-tailed distributions, conflicting information, regular variation, credence.

1   Introduction

In Bayesian Statistics we model a particular problem by assigning distributions to the data and prior information and then we combine those distributions in order to obtain the posterior distribution. Most standard models use distributions from the exponential family which provides very useful properties (such as conjugacy) that simplify the computation of the posterior distribution. However, in some situations the use of such distributions may lead to inappropriate behaviour of the posterior distribution. In particular, sometimes the information provided by the data and the prior knowledge may disagree, in the sense that their distributions are far away from each other. When this happens, we say that there is *conflict* between the sources of information. As O'Hagan and Forster (2004 §3.35 and 8.33) show, modelling using light-tailed distributions (commonly part of the exponential family), the posterior distribution may be strongly affected by the source of information which is causing the conflict. We may have conflicts caused by the data (such as outliers) or by the prior information and, in either situations, if we use distributions with light tails such as the normal distribution, the conflict may strongly influence the posterior distribution and potentially lead to

---

*University of Sheffield, UK, http://www.shef.ac.uk/paspgr/andrade/
†University of Sheffield, UK, http://www.shef.ac.uk/pas/people/academic/ohagan.html

inappropriate conclusions.

This behaviour was first observed by de Finetti (1961) and described by Lindley (1968) who suggested that if we used heavy-tailed distributions rather than normal distributions the posterior distribution would become more robust to anomalies like outliers. Investigating this idea, Dawid (1973) established conditions in the data and the prior distributions under which it is possible to achieve outlier rejection in models with pure location parameter structure. Thus, in problems where the parameter of interest is a location parameter, modelling satisfying Dawid's conditions the posterior distribution will reject the outliers if they are sufficiently far from the rest of information. This idea was also adapted to the case in which the prior information is causing the conflict, allowing to reject the prior information in favour of the data.

A long literature followed this idea. O'Hagan (1979) proposed alternative conditions which are easier to verify than Dawid's conditions. Also O'Hagan (1988, 1990) gave some examples in the context of one-way analysis of variance as well as some asymptotic properties of heavy-tailed distributions. Le and O'Hagan (in their papers of 1994 and 1998) extended the theory to the bivariate case, focusing on the tail behaviour. Also, they proposed bivariate heavy-tailed distributions whose tails decay at different rates in different radial directions. Haro-López and Smith (1999), working within the multivariate v-spherical family (Fernandez *et al*, 1995), proposed some conditions on location and scale parameters in order to bound the influence of the likelihood over the posterior distribution, they find bounds for the difference between the posterior and the prior expectation of a function of interest as the observations tend to infinity. However, they do not compute the limiting posterior distribution as the observations become large. Also, their conditions are quite difficult to verify, mainly the evaluation of uniform convergence limits of quotients of distributions. See Haro-López and Smith (1999).

Apart from Haro-López and Smith v-spherical conditions, the theory developed until now has mostly concentrated on location-parameter structures and little attention has been given to scale parameter structures. In this work we bring the theory of *regular variation* into the context of Bayesian robustness modelling and, we propose sufficient conditions in the univariate scale-parameter structure, based on a different class of heavy-tailed distributions, namely the *regularly varying distributions*, to those considered by Dawid (1973) and O'Hagan (1979).

In the next section we will provide a formal definition of regular variation as well as some relevant properties for our theory. Also, we show how regular variation theory relates to Bayesian robustness modelling. In Section 3 we present new results on the asymptotic behaviour of the posterior distribution of a pure scale parameter. Under regular variation conditions, robust resolution of conflicts is achieved, but the rejected information source nevertheless exerts some influence even in the limit. In Section 5 we propose new conditions in the pure location parameter case based on regularly varying densities. We use a simple example of the pure scale case (Section 4) to show how the theory deals with outliers in practice. Finally, we conclude by making some general comments on the theory.

2   Regularly varying functions

2.1   Definition and properties

The theory of regular variation was initiated by Karamata (1930) who developed it to solve some problems in Tauberian theory, since when many authors have applied the theory in probability. See for instance, Feller (1971, Chapter VIII) and the extensive works of Aljančić, Bojanic and de Haan, which are deeply discussed by Bingham *et al* (1987).

Definition 2.1 (Regular variation). *We say that a measurable function f is regularly varying at ∞ with index (order) ρ, which we write $f \in R_\rho$, if $\rho \in \mathbb{R}$ and*

$$\frac{f(\lambda x)}{f(x)} \longrightarrow \lambda^\rho \ (x \to \infty) \ \forall \lambda > 0. \tag{1}$$

In particular, if $\rho = 0$ then we say that f is *slowly varying*, we write $f \in R_0$. The set of all regularly varying functions is $R = \{R_\rho : -\infty < \rho < \infty\}$. By the Characterisation Theorem (Bingham *et al*, 1987), a function f is regularly varying with index ρ if and only if it can be written as $f(x) = x^\rho \ell(x)$ for $\rho \in \mathbb{R}$ and $\ell \in R_0$.

Sometimes the behaviour of f for values of x on the left-hand tail may be of interest. In this case, if $f(\lambda x)/f(x) \to \lambda^\rho$, as $x \downarrow -\infty$, then we say that f is *regularly varying at $-\infty$ with index ρ*, we write $f \in R_\rho(-\infty)$. Furthermore, if $f : \mathbb{R}^+ \to \mathbb{R}^+$ and $f(\lambda x)/f(x) \to \lambda^\rho$, as $x \downarrow 0$, then we say that f is *regularly varying at the origin* with index ρ, we write $f \in R_\rho(0+)$. $f(x) \in R_\rho(0+)$ is equivalent to $f(1/x) \in R_{-\rho}$ and $f(x) \in R_\rho(-\infty) \Leftrightarrow f(-x) \in R_\rho$.

Some properties of regularly varying functions are quite easy to prove (see Bingham *et al*, 1987). If $f \in R_\rho$, then $f(x)^\alpha \in R_{\alpha\rho}$ $(\rho, \alpha \in \mathbb{R})$. Suppose $f_i \in R_{\rho_i}$ $(i = 1, 2)$, then $f_1 + f_2 \in R_{\max\{\rho_1, \rho_2\}}$ and $f_1 \times f_2 \in R_{\rho_1 + \rho_2}$. Also, if $\rho_2 > 0$, then $f_1(f_2(x)) \in R_{\rho_1 \rho_2}$.

Another important concept is the asymptotic equivalence of functions. We say that $\phi(x)$ and $\psi(x) \neq 0$ are *asymptotically equivalent* at infinity, written $\phi(x) \sim \psi(x)$ as $x \to \infty$, if $\phi(x)/\psi(x) \to 1$ $(x \to \infty)$.

The following results are discussed in depth by Bingham *et al* (1987).

Lemma 2.2. *If f is regularly varying of index ρ then for any chosen $A > 1$ and $\delta > 0$ there exists $X = X(A, \delta)$ such that*

$$f(w)/f(z) \leq A \times \max\{(w/z)^{\rho+\delta}, (w/z)^{\rho-\delta}\}, \ \forall (w \geq X, z \geq X).$$

Proof. See Bingham *et al* (1987) - Theorem 1.5.6, Page 25.   □

Equivalently, if $w > z$, $A^{-1}(w/z)^{\rho-\delta} \leq f(w)/f(z) \leq A(w/z)^{\rho+\delta}$, for z sufficiently large (see Resnick, 1987). In particular, if $f = \ell$ is slowly varying and ℓ is bounded

away from 0 and ∞ on every compact subset of $[0, \infty)$, then for every $\delta > 0$ there exists $A' = A'(\delta) > 1$ such that

$$\ell(w)/\ell(z) \leq A' \times \max\{(w/z)^{\delta}, (w/z)^{-\delta}\}, \text{ for all } (z > 0, w > 0).$$

Lemma 2.2 is well known as Potter's Theorem (due to Potter, 1942) and establishes global bounds for the quotient $f(w)/f(z)$ as w and z become sufficiently large.

Lemma 2.3. *If $\ell \in R_0$ is measurable and $\alpha < -1$, then $\int^{\infty} x^{\alpha} \ell(x) dx < \infty$.*

Proof. See Bingham *et al* (1987) - Proposition 1.5.10, Page 27.   □

Lemma 2.3 provides a sufficient (but not necessary) condition for the existence of integrals involving regularly varying functions.

Rapid variation

The class of *rapidly varying functions* (de Haan, 1970) contains those functions that do not satisfy the limit (1), or more simply, it embraces functions with decay or growth faster than any power function of finite order.

Definition 2.4. *A measurable function $f : [A, \infty) \to (0, \infty)$ is rapidly varying of index ∞, which we write $f \in R_{\infty}$, if for $\lambda > 1$*

$$f(\lambda x)/f(x) \to \infty \ (x \to \infty),$$

and rapidly varying of index $-\infty$, written $f \in R_{-\infty}$, if for $\lambda > 1$

$$f(\lambda x)/f(x) \to 0 \ (x \to \infty).$$

Note that Definition 2.4 can be easily adapted to the left-hand tail, i.e. considering $x \downarrow -\infty$. Also, notice that rapid variation is intimately connected with regular variation, since rapid variation is essentially the limit (1) with infinite index, but as Seneta (1976) points out, this may be misleading. For our purpose, "regular variation" will always mean finite index.

2.2   Regular variation and heavy tails

Regular variation theory is widely used in the context of stable laws, which has direct applications to domains of attraction and extreme value theory. In these contexts, the idea of heavy-tailed distributions is defined in terms of the distribution function because many results can be simplified by the monotonicity property of the distribution function (e.g. Resnick (1987), Geluk (1996), Athreya *et al* (1998)). In our context we define heavy tails directly through the probability densities in order to avoid dealing with the usual integrals yielded by the distribution function. In terms of regular variation, in

$\mathbb{R}^1$, the tail behaviour of the probability density is equivalent to the growth order of the distribution function. See Karatama's Theorems (Bingham *et al* §1.5).

Observe that Definition 2.1 is closely related with tail behaviour of f, if $f \in R_{-\rho}$ as $x \to \infty$, then the right-hand tail of f decays at an order ρ, that is the right-hand tail of $f(x)$ behaves like a power function of order $-\rho$ for x sufficiently large. In probability theory, regular and rapid variation can be used to describe the speed at which the tails of a p.d.f. decay to zero. Thus, we define heavy- and light-tailed distributions in terms of regular and rapid variation, respectively.

Precisely, we say that a density f is a *heavy-tailed* distribution if $f \in R$. Similarly, a density f is a *light-tailed* distribution if $f \notin R$ ($f \in R_{-\infty}$). In particular, if $f(x)$ is skewed, its left- and right-hand tails can decay at different rates, thus we also can use the definition of regular variation to account for the behaviour of the left- and right-hand tails of f. If $X > 0$, we say that f has *heavy left-hand tail* if $f(x) \in R$ (as $x \downarrow 0$) or that f has *heavy right-hand tail* if $f(x) \in R$ (as $x \to \infty$). Of course, if $X \in \mathbb{R}$ we assess the tail behaviour of f as $x \downarrow +\infty$ and $x \downarrow -\infty$. Furthermore, although we can think of regular variation at a finite point, here we will be interested specifically in unbounded random variables, e.g. $X \in (-\infty, \infty)$, $X \in (0, \infty)$, etc, since we will be investigating the behaviour of the posterior distribution as some observations tend to infinity.

It is easy to verify that a t distribution with d degrees of freedom is heavy-tailed (both tails), whereas the normal distribution is light-tailed.

2.3 Credence

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, i.e. we assign a probability density which we believe is reasonable to describe the phenomenon. The relevance or the weight attached to the information provided by the data or prior knowledge is implicit in the chosen probability density. In general, the variance or precision indicates how informative an information source is, and how much weight it gets in synthesis with other sources of information. But when the sources of information conflict then credence or tail weight is what matters, since conflicts are characterised by events occurring far into the tails. Thus, the speed at which the tails of the probability density decay to zero determines the importance of the source of information in the model. See discussion in O'Hagan and Forster (2004, §3.34 and §8.50-8.62).

This property of probability densities was studied by O'Hagan (1990), who introduced another measure of tail weight, namely *credence* of information, which is essentially the representation of how much we are prepared to believe in one source of information rather than another in the case of conflict. For instance, if we believe that the process that generated a sample may produce outlying information, then we represent this belief by means of credence, that is we choose a distribution for that sample which has small credence.

Credence as defined by O'Hagan (1990) concerns basically distributions whose tails

behave like $|x|^{-\rho}$ as $|x| \to \infty$ and this class of distributions is a special case of regular variation, since $f(x) \sim |x|^{-c}$ implies $f \in R_{-c}$. As a consequence, we can interpret the regular variation index as Credence. Moreover, the regular variation index provides a more general interpretation of credence in the sense that now we can think of *left- and right-credence*, that is, we can regard distributions with different left- and right-hand tails behaviour. In particular, this is quite important in scale structures, where the left- and right-hand tails always decay at different rates, hence it is natural to have different credences of observations occurring on each side of the distribution. We broaden the concept of credence for distributions with any tails behaviour, which also accounts for the left- and right-hand tails.

Definition 2.5 (RV-Credence). *If $f(x) \in R_{-c}$ $(c > 0)$ as $x \to \infty$, then we say that the right (finite) RV-Credence of f is c, likewise, if $f(x) \in R_{-d}$ $(d > 0)$ as $x \to -\infty$, the left (finite) RV-Credence of f is d.*

The notion of RV-Credence can be extended to light-tailed densities. We simply say that all light-tailed distributions have *infinite RV-Credence*, by means of Definition 2.4. We use the term "RV-Credence" alone to mean finite RV-Credence. Also, when the RV-Credence refer to the right side of the distribution we omit the term "right", writing just "RV-Credence". Obviously, if f is symmetric the left and right RV-Credences coincide. For positive random variables the left RV-Credence is taken as $x \to 0^+$. Similarly to O'Hagan (1990), RV-Credence is positive by definition and will always express how credible a piece of information is.

For instance, suppose $f(x|\theta)$ is a t distribution with mean zero and variance θ, then $f \in R_{-(d+1)}$ as $|x| \to \infty$, which means that its tails decay at an order $d + 1$. In pure location structures the likelihood $L_x(\mu) = f(y - \mu)$ has exactly the same credence as the data distribution, we just transpose the tails. However, observe that in the scale parameter case, the likelihood $L_x(\theta) = f(x|\theta)$ is a function of θ, hence the left-hand and right-hand tails will decay at different rates, precisely $L \in R_{-1}$ as $\theta \to \infty$ and $L \in R_d$ as $\theta \downarrow 0$. In this way, the RV-Credence becomes quite useful, since it allows us to account for the behaviour of both tails of the likelihood.

3 Scale parameter structure

3.1 Rejecting data information

The general idea of robustness modelling is to assign suitable distributions to the data and the prior information in such a way that those distributions express our belief of which source of information might produce conflicting distributions. The simplest and most common case of conflict is when there are outliers in the sample. Thus, for instance, if we suspect that the way some particular data were collected might produce atypical observations, then we should choose suitably the distributions in the model so that those atypical values do not affect the posterior distribution. More precisely, by assigning a regularly varying (heavy-tailed) distribution to the data and some distribution with lighter tails to the prior information, the posterior distribution will be less influenced

by outliers and more influenced by the prior distribution. In the Scale parameter case, the behaviour of the posterior distributions is rather different from that in the location case.

We begin considering a pure scale model with only one observation, where this observation is an outlier, thus there will be a conflict between the prior information and a large observation, then we propose conditions on the prior and data distributions in order to resolve the conflict in favour of the prior distribution. Subsequently, we extend this idea to many observations and when we want to resolve the conflict in favour of the data rather than the prior distribution.

Theorem 3.1 (Conditions for the scale parameter case). *Let $Y|\theta \overset{D}{\sim} f$ such that f is of the form $f(y|\theta) = (1/\theta) \times h(y/\theta)$ and $\Theta \overset{D}{\sim} p$. If for some $0 < \delta < \rho$*

(i) $h \in R_{-(\rho+1)}$, *and*

(ii) $\int_\Theta \theta^{\rho+\delta} p(\theta) d\theta < \infty$.

Then

$$p(\theta|y) \longrightarrow \frac{\theta^\rho p(\theta)}{\int_\Theta \theta^\rho p(\theta) d\theta} \quad (y \to \infty). \tag{2}$$

Proof. The posterior distribution is given by

$$p(\theta|y) = \frac{f(y|\theta)p(\theta)}{\int_\Theta f(y|\theta)p(\theta)d\theta} = \frac{(1/\theta)h(y/\theta)p(\theta)}{\int_\Theta (1/\theta)h(y/\theta)p(\theta)d\theta}.$$

Therefore, dividing the numerator and denominator by $h(y)$ and taking the limit of $p(\theta|y)$ $(y \to \infty)$, we have

$$\lim_{y\to\infty} p(\theta|y) = \lim_{y\to\infty} \frac{(1/\theta)h(y/\theta)p(\theta)/h(y)}{\int_\Theta (1/\theta)h(y/\theta)p(\theta)/h(y)d\theta} = \frac{\lim_{y\to\infty}(1/\theta)h(y/\theta)p(\theta)/h(y)}{\lim_{y\to\infty}\int_\Theta (1/\theta)h(y/\theta)p(\theta)/h(y)d\theta}, \tag{3}$$

since both numerator and denominator are continuous.

Since $h \in R_{-(\rho+1)}$, the numerator of (3) is $\theta^\rho p(\theta)$. In order to apply the same limit to the denominator of (3), we need to prove that the limit can be passed inside the integral.

Using Lemma 2.2, we have that for any chosen $c > 1$ and $\delta > 0$ and for all y sufficiently large,

$$(1/\theta)h(y/\theta)p(\theta)/h(y) \le c \times (1/\theta)p(\theta) \max\{\theta^{\rho+1-\delta}, \theta^{\rho+1+\delta}\}.$$

But notice that Condition (ii) implies that $\max\{\theta^{\rho-\delta}, \theta^{\rho+\delta}\}p(\theta)$ is integrable, since

$$
\begin{aligned}
\int_0^\infty \max\{\theta^{\rho-\delta}, \theta^{\rho+\delta}\}p(\theta)d\theta &= \int_0^1 \max\{\theta^{\rho-\delta}, \theta^{\rho+\delta}\}p(\theta)d\theta + \int_1^\infty \max\{\theta^{\rho-\delta}, \theta^{\rho+\delta}\}p(\theta)d\theta \\
&= \int_0^1 \theta^{\rho-\delta}p(\theta)d\theta + \int_1^\infty \theta^{\rho+\delta}p(\theta)d\theta \\
&\leq \int_0^1 p(\theta)d\theta + \int_1^\infty \theta^{\rho+\delta}p(\theta)d\theta,
\end{aligned}
$$

which is clearly finite, since Condition (ii) is satisfied.

Hence, by *Lebesgue's dominated convergence theorem*, we can pass the limit inside the integral in (3) and we obtain

$$
p(\theta|y) \longrightarrow \frac{\theta^\rho p(\theta)}{\int_\Theta \theta^\rho p(\theta)d\theta} \ (y \to \infty). \qquad \qquad \square
$$

Condition (i) of Theorem 3.1 basically guarantees that the distribution of y is heavy-tailed. Condition (ii) says that $p(\theta)$ must have lighter right-hand tail than those from the distribution of y and, it is also a sufficient condition for the existence of the proportionality constant in the limiting posterior distribution.

Theorem 3.1 says that, if we assign a heavy-tailed distribution to the data and some distribution to the prior information which has lighter tails than those of the data distribution, then the posterior distribution will become independent of the outlier as it become sufficiently far away the prior distribution. The prior distribution can be either light-tailed or have tails decaying slightly faster than the tails of the data distribution. However, unlike the location parameter case described by Dawid and O'Hagan, in the scale case there is not complete rejection of outliers. Notice that the limiting posterior distribution, although independent of y, is a combination of the function θ^ρ and the prior distribution, θ^ρ is the only influence that the outlier exerts over the posterior distribution. We can think of θ^ρ as an improper data distribution that carries little information about the data, or more simply, we can say that the data distribution becomes little informative. This initial influence of the outlier is due to the multiplicative structure of the scale parameter. In terms of regular variation, we defined heavy tails with multiplicative arguments, that is $f(\lambda x)/f(x) \to \lambda^{-\rho}$ $(x \to \infty)$. Clearly, multiplying any quantity by the argument of the function will affect the limit, whereas additions to the argument does not alter its limit, i.e. $f(\lambda(x+c))/f(x) \to \lambda^{-\rho}$ $(x \to \infty)$, for any c (see Proof of Theorem 5.1). Therefore, as the scale parameter by definition is multiplied by the argument, then the limiting posterior distribution will always be a combination of $\lambda^{-\rho}$ with the prior distribution. In conclusion, the outlier in the scale parameter structure will not be completely rejected, instead, there will be a *partial rejection* of the outlier.

3.2 Many observations

Following the independence structure suggested by O'Hagan (1979), Theorem 3.1 can be generalised for many observations. This is a situation where one or more outliers occur in the data and, the rest of the data agreeing with each other and with the prior distribution, in the sense that the likelihood of the outlier is far away from the likelihood of the rest of the data and the prior distribution.

Firstly, the idea of *agreeing data information* will be defined more precisely. This is basically a statement that the data are sufficiently close. For the purpose of this work, the observations x_i $(i = 1, ..., n)$ agree with each other, if the deviation of the x_i in relation to a certain constant x is fixed. This is equivalent to think of the x_i as $x_i = x + \xi_i$ $(i = 1, ..., n)$ for some fixed $|\xi_i|$. In particular, a group of outliers is a group of observations which are concentrated around a certain centre, which is far away from the other data. Thus, we extend Theorem 3.1 to many observations by taking the limit as a group outliers tend to infinity, that is as a centre x (around which the outliers are concentrated) tends to infinity. Note that the differences amongst the outliers within the group stay constant.

The most important aspect concerning this idea is that the joint distribution of a group of observations adds the RV-Credences of each observation. That is, if $h_i \in R_{-\rho_i}$, then the $\prod_{i=1}^{n} h_i(x_i/\theta) = \prod_{i=1}^{n} h_j((x + \xi_j)/\theta) \in R_{-\sum \rho_i}$ as $x \to \infty$, for some arbitrary centre x. Observe that this is exactly the nature of statistical inference, if many observations carry *similar* information about a certain parameter, so the evidence about the parameter is strengthened, and this is expressed through RV-credence.

In order to show how Theorem 3.1 applies to many observations, suppose we have a random sample $x_1, ..., x_n$ which is independently distributed according to $f_i(x_i|\theta) = (1/\theta) \times h_i(x_i/\theta)$ $(i = 1, ..., n)$, then the joint distribution of the data is given by $f(\boldsymbol{x}|\theta) = \prod_{i=1}^{n} f_i(x_i|\theta)$. Also, let $\theta \overset{D}{\sim} p(\theta)$. Notice that, since the observations are independent, $f(\boldsymbol{x}|\theta)$ can be reorganised separating the conflicting observation from the agreeing information, that is, supposing x_n is an outlier, we can write $f(\boldsymbol{x}, \theta) = f_n(x_n|\theta) \times \prod_{i=1}^{n-1} f_i(x_i|\theta) \times p(\theta)$, which is exactly the configuration of Theorem 3.1, i.e. $f_n(x_n|\theta)$ is the likelihood involving the outlier x_n and $\prod_{i=1}^{n-1} f_i(x_i|\theta) \times p(\theta)$, the agreeing information, plays the role of the prior distribution. Thus, Condition (i) and (ii) of Theorem 3.1 will be satisfied if $f_n(x_n|\theta)$ is regularly varying with some index $-(\rho + 1)$ and $\int \theta^{\rho+\delta} \times \prod_{i=1}^{n-1} f_i(x_i|\theta) \times p(\theta)d\theta < \infty$ (for some $\delta > 0$). The latest will depend on the forms of $p(\theta)$ and $f_i(x_i|\theta)$ $(i = 1, ..., n-1)$. More precisely, if $f_i(x_i|\theta)$ $(i = 1, ..., n-1)$ or $p(\theta)$ are light-tailed as $\theta \to \infty$, Condition (ii) is automatically satisfied. Alternatively, if the $f_i(x_i|\theta)$ $(i = 1, ..., n-1)$ are regularly varying $(\theta \to \infty)$ with indexes $-\beta_i \in \mathbb{R}^-$ and if $p(\theta) \in R_{-(c+1)}$, then Condition (ii) can be satisfied if $-\sum_{i=1}^{n-1} \beta_i - (c+1) + \rho + \delta < -1$, by Proposition 2.3. In other words, the existence of the limiting posterior distributions can be guaranteed if the combination of the RV-Credences of $p(\theta)$ and $\prod_{i=1}^{n-1} f_i(x_i|\theta)$ (as a function of θ) is larger than $\rho + 1$.

Under the conditions of Theorem 3.1, the posterior distribution will be given by

$$p(\theta|\boldsymbol{x}) \approx \frac{\theta^\rho \prod_{i=1}^{n-1} f_i(x_i|\theta) \times p(\theta)}{\int_\Theta \theta^\rho \prod_{i=1}^{n-1} f_i(x_i|\theta) \times p(\theta)d\theta}, \text{ for } x_n \text{ sufficiently large.}$$

Observe that following this scheme it is possible to combine any density f_i ($i = 1, ..., n$) and apply the theorem. In particular, it is also easy to show that a group of outliers is also (partially) rejected in the same way. Suppose that $X_i \overset{D}{\sim} f_i(x_i|\theta) \in R_{-(\alpha_i+1)}$ ($\alpha_i > 0$ and $i = 1, ..., n$), $p(\theta) \notin R$ and, that the observations labelled as x_1, x_2 and x_3 are outliers. Thus $f(\boldsymbol{x}|\theta) = \prod_{i=4}^n f_i(x_i|\theta) \times \prod_{i=1}^3 f_i(x_i|\theta)$. Notice that following the idea from Page 177 (that is $x_i = x' + \xi_i$ ($i = 1, 2, 3$), where x' is some arbitrary centre), we have that $f(\boldsymbol{x}^{(3)}|\theta) \in R_{-(\alpha_1+\alpha_2+\alpha_3+3)}$ as $x' \to \infty$ or, equivalently: as $x_i \to \infty$ ($i = 1, 2, 3$). In this way, in order to show that Theorem 3.1 holds, the group of outliers is made to tend to infinity by taking $x' \to \infty$ and, $P(\theta|\boldsymbol{x}^{(n-3)}) = \prod_{i=4}^n f_i(x_i|\theta) \times p(\theta) \notin R$ plays the role of the prior distribution. Consequently, the posterior distribution will be the combination of the $P(\theta|\boldsymbol{x}^{(n-3)})$ and $\theta^{-(\alpha_1+\alpha_2+\alpha_3+3)}$, that is

$$p(\theta|\boldsymbol{x}) \approx \frac{\theta^{-(\alpha_1+\alpha_2+\alpha_3+3)} \prod_{i=4}^n f_i(x_i|\theta) \times p(\theta)}{\int_\Theta \theta^{-(\alpha_1+\alpha_2+\alpha_3+3)} \prod_{i=4}^n f_i(x_i|\theta) \times p(\theta)d\theta},$$

for (x_1, x_2, x_3) sufficiently large.

In general conflict between two groups of information will be resolved automatically in favour of the group with largest total RV-Credence. In particular, if we model the n observations as i.i.d. with some regularly varying distribution, and if there is a group of k outliers in the data, as they tend to infinity, the outliers will be rejected if $k < n - k$, when the prior distribution is uninformative. More generally, the regular variation index of the prior can be considered in terms of an equivalent number of observations to be added to the non-outlying group.

3.3   Rejecting prior information

The results established in the previous sections concern the case when we want to resolve conflicts in favour of the prior distribution. In this section we consider the converse case when we want to (partially) reject the prior information in favour of the data distribution. In other words, if we somehow believe that the information available to formulate the prior distribution may produce a prior density which is far away from the data, then Theorem 3.1 can be adapted to lead to partial rejection of the prior distribution.

Consider the transformation of Θ to $\Phi = |Y|/\Theta$, thereby the posterior distribution of Θ given y is equivalent to the posterior distribution of Φ given y. Indeed, using the

Jacobian rule on the posterior distribution of $\Theta|y$, we obtain

$$
\begin{aligned}
p_\Phi(\phi|y) &= \frac{h\left(\phi y/|y|\right) \times (1/\phi) \times p(|y|/\phi)}{\int_\Phi h\left(\phi y/|y|\right) \times (1/\phi) \times p(|y|/\phi)d\phi} \\
&= \frac{h(\mathrm{sgn}(y)\phi) \times (1/\phi) \times p(|y|/\phi)}{\int_\Phi h(\mathrm{sgn}(y)\phi) \times (1/\phi) \times p(|y|/\phi)d\phi},
\end{aligned}
\tag{4}
$$

where $\mathrm{sgn}(y) = 1$ if $y \geq 0$ or $\mathrm{sgn}(y) = -1$ if $y < 0$, which we use just to indicate whether $y \downarrow -\infty$ or $y \uparrow +\infty$.

Notice that $p_\Phi(\phi|y)$ is the posterior distribution of a scale parameter, where $h(\mathrm{sgn}(y)\phi)$ and $q(y|\phi) = 1/\phi \times p(|y|/\phi)$ play the role of the prior and the likelihood, respectively. Thus, letting h have lighter tails than p and $p \in R_{-(\rho+1)}$ ($\rho > 0$), we can apply Theorem 3.1 and obtain

$$
p(\phi|y) \longrightarrow \frac{\phi^\rho h(\mathrm{sgn}(y)\phi)}{\int_\Phi \phi^\rho h(\mathrm{sgn}(y)\phi)d\phi} \ (|y| \to \infty),
$$

which is equivalent to

$$
p(\theta|y) \sim \frac{\theta^{-(\rho+1)}f(y|\theta)}{\int_\Theta \theta^{-(\rho+1)}f(y|\theta)d\theta} \ (|y| \to \infty).
$$

In other words, if we assign a heavy-tailed distribution to the prior and some distribution with lighter tails to the data, the posterior distribution will be approximately independent of the prior information if it is far away from the data, hence the posterior distribution will be based on the likelihood. However, just like in the previous section, here the prior information is not completely rejected, it becomes asymptotically equal to $\theta^{-(\rho+1)}$. Note that the posterior estimates will be based on the data information.

Consider the random sample $\boldsymbol{y} = (y_1, ..., y_n)$ and suppose the n observations agree with each other (in the sense of Section 3.2) and disagree with the prior distribution. In this way, the limiting posterior distribution naturally becomes

$$
p(\theta|\boldsymbol{y}) \approx \frac{\theta^{-(\rho+1)}f(\boldsymbol{y}|\theta)}{\int_\Theta \theta^{-(\rho+1)}f(\boldsymbol{y}|\theta)d\theta},
$$

provided that Condition (ii) is satisfied, that is $\int y^{\rho+\delta} \prod_{j=1}^n h(y_j/\theta)dy < \infty$.

3.4   Remarks

In general, in the presence of conflict, depending on how we choose the distributions in the model the posterior distribution will respond quite differently to the conflicting information. If we model assigning light-tailed distributions for both data and prior information, the conflict will not be resolved, in the sense that the posterior distribution will combine all the sources information including the conflicting information, which may influence the posterior estimates. Alternatively, modelling assigning heavy-tailed distributions to data and the prior information, the conflict will be resolved in favour of

either the data or the prior distribution, depending on which of them has lighter tails (i.e. larger RV-Credence). Thus, regardless whether the conflicting information comes from the data or the prior distribution, modelling suitably will resolve automatically problems in the model.

4   Example

We consider the simple independent random sample $\boldsymbol{y} = (1, 2, 2, 4, y_5)$, where y_5 is an outlier and the parameter of interest is the standard deviation. Then, we assign $y_j | \theta \overset{D}{\sim} f(y_j | \theta)$ $(j = 1, ..., 5)$ and $\theta \overset{D}{\sim} p(\theta)$.

One natural way of modelling is to assign the Normal Inverse-Gamma model (Model I)

$$\begin{cases} y_j | \theta & \overset{D}{\sim} & N(0, \theta^2) \\ \theta & \overset{D}{\sim} & IG(\alpha, \beta) \end{cases} , \; j = 1, ..., 5.$$

As an alternative, we will compare Model I with Model II:

$$\begin{cases} y_j | \theta & \overset{D}{\sim} & t_{(d)}(0, \theta^2) \\ \theta & \overset{D}{\sim} & IG(\alpha, \beta) \end{cases} , \; j = 1, ..., 5.$$

Notice that unlike Model I, if we choose suitably the hyperparameters d and α, Model II satisfies the conditions of Theorem 3.1. Indeed, condition (i) is directly verified by Definition 2.1, that is $t_{(d)}(y | 0, \theta) \in R_{-(d+1)}$ as $y \to \infty$, hence $f(\boldsymbol{y} | \theta) \in R_{-5(d+1)}$. Also Condition (ii) is assured since $t_{(d)}(y_j | 0, \theta) \in R_{-1}$ $(\theta \to \infty)$, which gives $f(\boldsymbol{y}_{-5} | \theta) \in R_{-4}$ $(\theta \to \infty)$, where $\boldsymbol{y}_{-5}$ is the data without the observation y_5. Thus, the combination of $p(\theta)$ and $f(\boldsymbol{y}_{-5} | \theta)$ is regularly varying with index $-(\alpha + 5)$, whence, if $\alpha + 5 > d + \delta$ (for some $\delta > 0$), we can guarantee

$$\int_{\Theta} \theta^{d+\delta} p(\theta) \times f(\boldsymbol{y}_{-5} | \theta) d\theta = \int_{\Theta} \theta^{d+\delta} \theta^{-(\alpha+5)} \ell(\theta | \boldsymbol{y}_{-5}) d\theta < \infty,$$

by Lemma 2.3, where $\ell \in R_0$. This is also related with the fact that the RV-Credences of the agreeing observations add, as a result of the combined distribution $f(\boldsymbol{y}_{-5} | \theta) \in R_{-4(d+1)}$. Clearly, this distribution has already lighter tails than $f(y_5 | \theta) \in R_{-(d+1)}$, hence the conflict between $f(\boldsymbol{y}_{-5} | \theta)$ and $f(y_5 | \theta)$ would already be resolved in favour of $f(\boldsymbol{y}_{-5} | \theta)$ regardless the RV-Credence $\alpha + 1$ of the prior distribution. In this way, when the sample size is larger than *one* a conflict can be resolved quite early, in the sense that it is not necessary to obtain y_5 very large before the posterior distribution becomes almost identical to the limiting posterior distribution.

4.1   Comparing posterior estimates

For Model II, it is necessary to choose suitably the hyperparameter α in order to satisfy the conditions of Theorem 3.1. Thus, letting $d = 4$, $\alpha = 5$ and $\beta = 8$, $\alpha + 5 > d + \delta$ (for

some $\delta > 0$) will allow to resolve the conflict caused by y_5 and it also guarantees propriety of the limiting posterior distribution. Notice that choosing any $d < 10$ the conditions would still be satisfied, since the sample of size 5 adds naturally the RV-Credences of the observations in it. The choice of β is irrelevant since the tails behaviour is fully described by α. Observe that in Model I we could choose any hyperparamaters α and β and the conditions would be satisfied, thus we use the same values as in Model II.

Figure 1 shows the behaviour of the posterior expectation and standard deviation of θ yielded by Model I. Observe that the expectation of θ increases as y_5 becomes large. In other words, y_5 is conflicting with the rest of the data and the prior information. As we used a normal distribution for the data, any large observation will affect the posterior distribution and in the extreme case where y_5 tends to infinity, the posterior estimates will also tend to infinity.
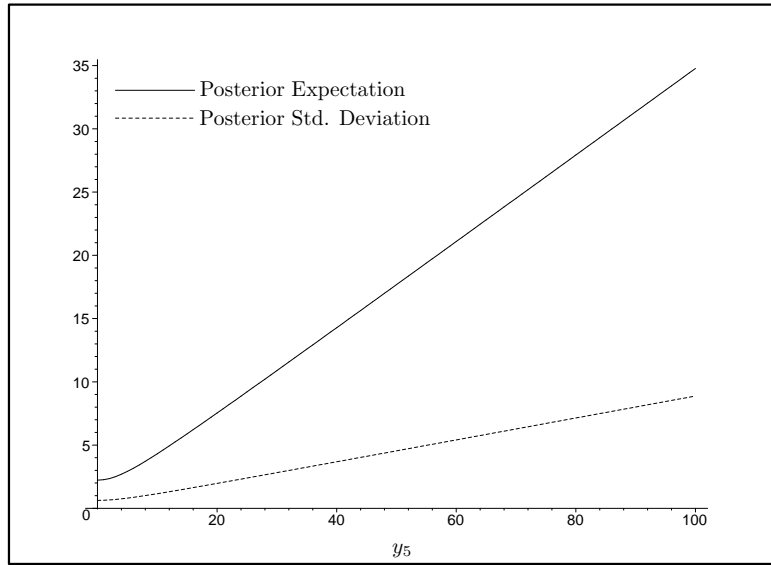

Figure 1: Model I: Posterior estimates of θ as y_5 increases

Unlike Model I, Model II yields posterior quantities which are robust to the increase of y_5. Figure 2 shows that, as y_5 becomes large, it influences the posterior quantities until a certain point and then the posterior estimates become approximately unaffected by larger values. For instance, the posterior expectation of θ is influenced by the outlying information only until y_5 is around 45, after that, $\mathrm{E}[\theta|\boldsymbol{y}]$ becomes approximately constant. Nevertheless, computing the posterior distribution of Model II *without the outlier y_5*, we find the posterior expectation $\mathrm{E}[\theta|\boldsymbol{y}_{-5}] \approx 2.69$ and the standard deviation $\mathbb{SD}[\theta|\boldsymbol{y}_{-5}] \approx 0.94$. Comparing these values with the posterior expectation of Model II (including all observations), we have $\mathrm{E}[\theta|\boldsymbol{y}] \to 4.76$ and $\mathbb{SD}[\theta|\boldsymbol{y}] \to 2.40$, as y_5 increases. That is, as we mentioned in the previous sections, y_5 is not completely rejected, it exerts some initial influence over the posterior distribution, but this influence become weaker as the outlier increases.

From Section 3.2, the limiting posterior density $p_L(\theta|y)$ yielded by Model II is given by

$$p(\theta|\boldsymbol{y}) \to p_L(\theta|\boldsymbol{y}) = \frac{\theta^d \times P(\theta|\boldsymbol{y}_{-5})}{\int_\Theta \theta^d \times P(\theta|\boldsymbol{y}_{-5})d\theta} \ (y_5 \to \infty),$$

where $P(\theta|\boldsymbol{y}_{-5}) = f(\boldsymbol{y}_{-5}|\theta) \times p(\theta)$ and $f(\boldsymbol{y}_{-5}|\theta)$ is the likelihood without y_5, thereby $P(\theta|\boldsymbol{y}_{-5}) \propto \theta^{-(\alpha+5)} \exp\left[-1/2(25 + 2\beta\theta)/\theta^2\right]$, hence

$$p(\theta|\boldsymbol{y}) \propto \theta^d \times \theta^{-(\alpha+5)} \exp\left[-\frac{1}{2}\frac{(25 + 2\beta\theta)}{\theta^2}\right], \text{ for } y_5 \text{ sufficiently large.}$$

The statistical summaries from this density coincide with those produced by the full Model II, for y_5 sufficiently large. Precisely, the posterior expectation and standard deviation of θ are approximately 4.76 and 2.40, respectively.
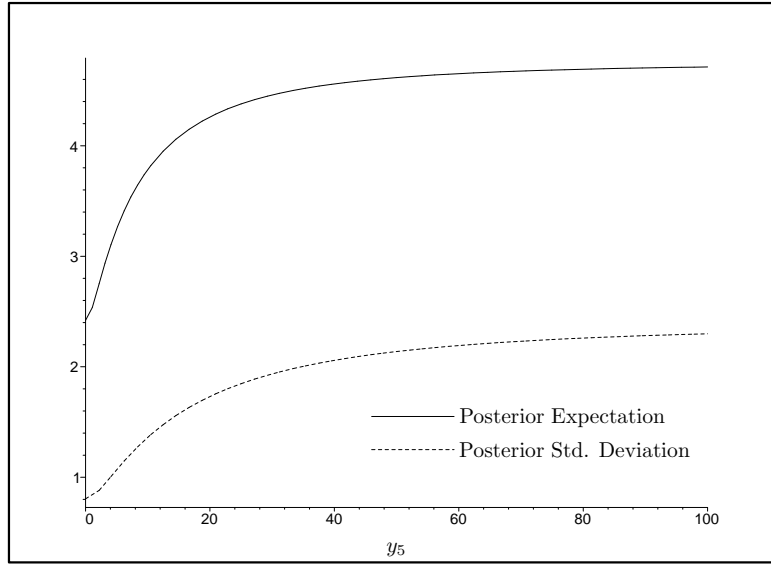

Figure 2: Model II: Posterior estimates of θ as y_5 increases

Models I and II show clearly the discrepancy between the posterior quantities produced by models involving normal and heavy-tailed distributions in the presence of conflicting information. As shown above, outliers can be partially rejected if we model satisfying the conditions of Theorem 3.1. Alternatively, if we wanted to reject the prior distribution, we would have model accordingly to Section 3.3 which would result in a posterior distribution barely influenced by the prior distributions.

5 Location parameter structure

5.1 Relation between location and scale structures

In the modelling process, we often make transformation of variables. In particular, a scale parameter structure can be transformed into a location structure by simple logarithm transformation (or vice-versa, using exponential transformation). Thus, considering a pure scale structure, one would think of taking logarithms in order to switch from multiplicative to additive arguments, finding the equivalent dual location model and then end up with the configuration of Dawid's Theorem. Unfortunately this is not possible, the theory proposed by Dawid (1973) and O'Hagan (1990) can not be applied to the scale parameter. For when we are interested in tails behaviour of the probability densities, these transformations do not preserve the tails property.

Suppose $y|\theta \overset{D}{\sim} (1/\theta) \times h(y/\theta)$ and $\theta \overset{D}{\sim} p(\theta)$, where $h \in R_{-\rho}$. Letting $X = \log Y$ and $\mu = \log \Theta$, we have the dual location model $f(x,\mu) = e^{x-\mu} h(e^{x-\mu}) \times e^{\mu} p(e^{\mu}) = f^*(x-\mu) \times p^*(\mu)$. But, observe that when we apply the logarithm transformation the regular variation properties are not transmitted to the dual model, that is $f^*(x-\mu) = e^{x-\mu} h(e^{x-\mu})$ is not necessarily regularly varying. For instance, suppose $Y|\theta \overset{D}{\sim}$ Inverse-Gamma(a,θ), where θ is an unknown scale parameter, then letting $X = \mu + \log(Y/\theta)$ we obtain $X|\mu \overset{D}{\sim} g(x|\mu)$, where μ is the location parameter and $g(x|\mu) \propto \exp\{a(x-\mu) - \exp(x-\mu)\}$. But observe that g is light-tailed, hence it does not satisfy Dawid's conditions. Thus, results such as those of Dawid and O'Hagan for location parameters cannot simply be used to derive useful outliers rejection for scale parameters, because the transformed sampling distributions do not have comparable heavy-tailed properties and are not realistic for potential modelling.

5.2 New conditions on location structure

In this section we show that the conditions in the location case established by Dawid and O'Hagan can be simplified with the use of regularly varying distributions. Note that it is not possible to apply regular variation alone to the location case in the same straightforward way as in the scale case. In fact, since the scale parameter is positive, the existence of the integral in the denominator of the limiting posterior distribution (2) is guaranteed only by assigning a prior distribution with lighter tails than the tails of data distribution, that is as long as the tails of the prior distribution decay to zero sufficiently fast in the sense of Lemma 2.3. In the location case, we integrate over $\mathbb{R}$, hence we need to have a well-behaved data distribution. Regular variation only accounts for the tails behaviour, i.e. it can only guarantee that the data distribution is well-behaved in its tails. Thus, we need some extra conditions to guarantee that the data distribution is also well-behaved in the whole real line. Therefore, we combine regular variation with some behavioral assumptions in order to establish new conditions for the location case that are relatively simpler to verify than those proposed by Dawid and O'Hagan.

Theorem 5.1 (New conditions for the location parameter case). *Let the distri-*

bution of $Y|\mu = \mu$ be $f(y - \mu)$ and $\mu \overset{D}{\sim} p$ such that:

(a) $f \in R_{-\rho}$ $(\rho > 0)$;

(b) $f > 0$ *is continuous in* $\mathbb{R}$;

(c) *There exist a C_1 and a $C_2 \geq C_1$ such that $f(y)$ is decreasing for $y \geq C_2$ and increasing for $y \leq C_1$. Also, $d \log f(y)/dy$ exists and is increasing for $y \geq C_2$;*

(d) $p \notin R$ *or* $p \in R_{-c}$ $(\mu \to \infty)$ *with* $c > \rho + 1$.

then the following conditions hold:

(i) *Given $\epsilon > 0$, $h > 0$, there exists A such that if $y > A$, then $|f(y') - f(y)| < \epsilon f(y)$ whenever $|y' - y| < h$;*

(ii) *For some B, M, $0 < f(y') < M f(y)$ whenever $y' > y > B$;*

(iii) *Let $k(\mu) = \sup_y \{f(y - \mu)/f(y)\}$, then we have $\int_\mu k(\mu) p(\mu) d\mu < \infty$.*

Consequently, given $(a) - (d)$, then

$$p(\mu|y) \longrightarrow p(\mu), \text{ as } y \to \infty.$$

Proof. Without loss of generality, let $y' = y + c$ for some $c > 0$ such that $|y' - y| = c < h$. As $f \in R_{-\rho}$, by Lemma 2.2, for any chosen $\epsilon > 0$ and $\delta > 0$ there exists $A = A(\epsilon, \delta)$ such that

$$\frac{f(y')}{f(y)} \leq (1 + \epsilon)(1 + c/y)^{-\rho + \delta}, \text{ for all } y \geq A(\epsilon, \delta).$$

Now, observe that we can find y sufficiently large so that $0 < c/y < 1$, for any fixed c. Thus, for all $y \geq A(\epsilon, \delta)$

$$\frac{f(y')}{f(y)} \leq (1 + \epsilon)(1 + c/y)^{-\rho + \delta} \leq 1 + \epsilon, \text{ for some } \epsilon > 0. \tag{5}$$

On the other hand, for some chosen $L > 1$

$$\frac{f(y)}{f(y')} \leq L \times (y/(y + c))^{-\rho - \delta} \iff L^{-1} \times (y/(y + c))^{\rho + \delta} \leq \frac{f(y')}{f(y)},$$

but notice that $L^{-1}(y/(y + c))^{\rho + \delta} < 1$, thus we can choose $\epsilon > 0$ so that $L^{-1}(y/(y + c))^{\rho + \delta} = 1 - \epsilon$ hence we can write $1 - \epsilon \leq f(y')/f(y)$. Thus, we have $1 - \epsilon \leq f(y')/f(y) \leq \epsilon + 1$, which is equivalent to condition (i), i.e. $|f(y') - f(y)| \leq \epsilon f(y)$. In particular, condition (ii) follows by letting $B = A(\epsilon, \delta)$, $M = 1 + \epsilon$ and $y' = y + c$ in (5), that is, for any chosen $c > 0$ (i.e. for any $y' > y > B$), $0 < f(y') < M f(y)$.

Condition (iii): Given (b) and (c) and condition (i), O'Hagan (1979) showed that $k(\mu) = \sup_y\{f(y-\mu)/f(y)\} \leq D/f(\mu)$ for some $D > 0$ and μ sufficiently large. But as $f \in R_{-\rho}$ we can write $k(\mu) \leq D/f(\mu) = D \times \mu^\rho \times \ell(\mu)$, where $\ell(\mu) \in R_0$. Thus

$$\int_\mu k(\mu)p(\mu)d\mu \leq D \int_\mu \mu^\rho \times \ell(\mu)p(\mu)d\mu,$$

which always exist for all $p(\mu)$ light-tailed and, alternatively, if $p \in R_{-c}$ then

$$\int_\mu \mu^{-\rho} \times \ell(\mu)p(\mu)d\mu = \int_\mu \mu^{\rho-c} \times \ell^*(\mu)d\mu < \infty, \text{ since } \rho - c < -1 \text{ and Lemma } 2.3.$$

Thus $p(\mu|y) \longrightarrow p(\mu)$, as $y \to \infty$. $\qquad\square$

In other words, regular variation of f with the regularity conditions (b) and (c) (see O'Hagan, 1979), implies Dawid's conditions, hence $p(\mu|y) \longrightarrow p(\mu)$ as $y \to \infty$. Hence, in order to satisfy Dawid's conditions, we need only satisfy Conditions $(a) - (d)$, which are much easier to verify. Indeed, regular variation can be verified simply by evaluating the limit (1), condition (b) is quite direct, also condition (c), which requires only that the tails of f should decay monotonically to zero on the extremes of the domain. Furthermore, note that Conditions (b) and (c) allow us to bypass the evaluation of $\sup_y\{f(y-\mu)/f(y)\}$ which may become relatively difficult in more complex densities.

The other case, in which we want to reject the prior distribution, also can be achieved in the same way. Also, the properties verified by Dawid and O'Hagan (such as proneness, resistance, credence, etc) also applied to the result above. In particular, credence is replaced by RV-Credence, which allows to use a broader class of distributions with different left- and right-hand tails decay.

6   Concluding remarks

The class of regularly varying distributions provides an easy way of understanding the behaviour of the posterior distribution for a scale parameter in the presence of atypical information. In particular, we have obtained explicitly the limiting posterior distribution as the conflicting information becomes sufficiently far away from the rest of the information. In addition, the conditions of Theorem 3.1 are relatively easy to verify, since, one of the conditions basically involves the evaluation of a limit and for the other condition, we can use a simple property of regular variation (Lemma 2.3) in order to guarantee the existence of the integral in the limiting posterior distribution. Furthermore, regular variation can also deal with the location parameter case (previously approached by Dawid and O'Hagan) in a simpler way, resulting in conditions which are relatively easier to verify than those proposed in the literature.

Note that, in order to illustrate the theory, we used a simple (and perhaps unrealistic) model. In practical problems where the prior information may be more complex, the verification of the conditions may become difficult. However, the properties of regular

variation may be helpful in assessing the tail behaviour of more complex distributions. For instance, the tails behaviour of a mixture of distributions may be assessed by the simple property of the sum of regularly varying functions.

Although this paper is not concerned with the detection of conflicts, there is a close connection between heavy-tailed modelling and the ideas of model criticism and measures of *surprise* discussed in, for example, O'Hagan and Forster (2004, §8.29-49).

As we showed, regular variation is an elegant representation of tails which behave like power functions and, it is sufficient to yield robust posterior distributions for either location or scale parameters. However, it is important to point out that the regularly varying class does not embrace all the distributions with robust properties. For instance, the double-exponential distribution is not regularly varying, but Pericchi and Smith (1992) noted that, in the location-parameter context, it can be used to obtain a posterior distribution robust to conflicts. However, the conflicting information is not completely rejected, but there is an initial influence which does not vanish as the conflict becomes more extreme.

More generally, the *exponential power distribution* (EPD) proposed by Box and Tiao (1973) (which embraces the double-exponential distribution) is not regularly varying but can provide robust posterior distributions for the location parameter. Furthermore, there are distributions from the exponential family which allow one to obtain robust behaviour of the posterior distribution; Pericchi *et al* (1993) provide some results (and examples).

In more complex structures where the parameters of interest are location and scale parameters, the posterior distribution can also be robust to anomalies such as outliers. However, care must be taken because the interaction amongst the data and prior distributions implies a multivariate problem which is not treatable by univariate regular variation theory. We are currently preparing another paper in which we explore the behaviour of models involving location-scale structures.

Bibliography

Aljan˘ cić, S. (1973). "Asymptotic Mercerian Theorems Involving Slowly Varying Functions." *Matematički Vesnik*, 25:331–337.

— (1982). "Regularly Varying Functions in Asymptotic Mercerian Theorems." *Académie Serbe des Sciences et des Arts, Classe des Sciences Mathmatiques et Naturelles Bulletin*, 12:23–30.

Athreya, K. B., S. N., Lahiri, and Wu, W. (1998). "Inference for Heavy Tailed Distributions." *Journal of Statistical Planning and Inference*, 66:61–75.

Bingham, N. H., Goldie, C. M., and Teugels, J. L. (1987). "Regular Variation." In *Encyclopedia of Mathematics and Its Applications*, volume 27. Cambridge: Cambridge University Press.

Bojanic, R. and Karamata, J. (1963a). "Class of Functions of Regular Asymptotic Behaviour." Technical Report 436, Mathematics Research Center, Madison, Wisconsin.

— (1963b). "Slowly Varying Functions and Asymptotic Relations." Technical Report 432, Mathematics Research Center, Madison, Wisconsin.

Box, G. E. P. and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis*. New York: Wiley.

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." In *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability*, volume I, 99–210. Berkeley, California: University of California Press.

de Haan, L. (1970). "On Regular Variation and its Application to the Weak Convergence of Sample Extremes." Mathematical Centre Tracts 32, Mathematisch Centrum, Amsterdam.

— (1974). "Equivalence Classes of Regularly Varying Functions." *Stochastic Processes and their Applications*, 2:243–259.

Feller, W. (1971). *An Introduction to Probability and its Applications*, volume II. New York: Wiley, 2nd edition.

Fernandez, C., Osiewalski, J., and Steel, M. F. J. (1995). "Modelling and Inference with V-spherical Distributions." *Journal of the American Statistical Association*, 90:1331–1340.

Geluk, J. L. (1996). "Tails of Subordinated Laws: The Regularly Varying Case." *Stochastic Processes and Their Applications*, 61:147–161.

Haro-López, R. (1999). "On Robust Bayesian Analysis for Location and Scale Parameters." *Journal of Multivariate Analysis*, 70:30–56.

Karamata, J. (1930). "Sur un Mode de Croissance Régulière des Fonctions." *Mathematica (Cluj)*, 4:38–53.

Le, H. and O'Hagan, A. (1998). "A Class of Bivariate Heavy-tailed Distributions." *Sankyā: The Indian Journal of Statistics*, B-60:82–100.

Lindley, D. V. (1968). "The Choice of Variables in Multiple Regression (With Discussion)." *Journal of the Royal Statistical Society, Series B*, 30:31–66.

Neyman, J. and Scott, E. L. (1971). "Outlier Proneness of Phenomena and of Related Distributions." In Rustagi, J. S. (ed.), *Optimizing Methods in Statistics*. New York: Academic Press.

O'Hagan, A. (1979). "On Outliers Rejection Phenomena in Bayes Inference." *Journal of the Royal Statistical Society, Series B*, 41:358–367.

— (1988). "Modelling with Heavy Tails." In *Bayesian Statistics*, volume 3, 345–359. Oxford: Claredon Press.

— (1990). "On Outliers and Credence for Location Parameter Inference." *Journal of the American Statistical Association*, 85:172–176.

O'Hagan, A. and Forster, J. J. (2004). *Bayesian Inference - Kendall's Advanced Theory of Statistics*, volume 2B. London: Arnold.

O'Hagan, A. and Le, H. (1994). "Conflicting Information and a Class of Bivariate Heavy-tailed Distributions." In Freeman, P. R. and Smith, A. F. M. (eds.), *Aspects of Uncertainty: a Tribute to D. V. Lindley*, 311–327. New York: Wiley.

Pericchi, L. R. and Sansó, B. (1995). "A Note on Bounded Influence in Bayesian Analysis." *Biometrika*, 82:223–225.

Pericchi, L. R., Sansó, B., and Smith, A. F. M. (1993). "Posterior Cumulant Relationships in Bayesian Inference Involving the Exponential Family." *Journal of the American Statistical Association*, 88:1419–1426.

Potter, H. S. A. (1942). "The Mean Value of a Dirichlet Series II." In *Proceedings of the London Mathematical Society*, volume 47, 1–19.

Resnick, S. (1987). *Extreme Values, Regular Variation and Point Process*. New York: Springer-Verlag.

Seneta, E. (1976). "Functions of Regular Variation." In *Lecture Notes in Mathematics*, volume 506. New York: Springer.

Acknowledgments