# Bayesian robustness modelling of location and scale parameters

A. O'Hagan[†] and J.A.A. Andrade[*]

[†]Dept. of Probability and Statistics, University of Sheffield, Sheffield, S3 7RH, UK.

[*]Dept. of Statistics and Applied Mathematics, Federal University of Ceara, Brazil, 60.410-670.

[*]jaa.andrade@gmail.com

**Abstract**

The modelling process in Bayesian Statistics constitutes the fundamental stage of the analysis, since depending on the chosen probability laws the inferences may vary considerably. This is particularly true when conflicts arise between two or more sources of information. For instance, inference in the presence of an outlier (which conflicts with the information provided by the other observations) can be highly dependent on the assumed sampling distribution. When heavy-tailed (e.g. t) distributions are used, outliers may be rejected whereas this kind of robust inference is not available when we use light-tailed (e.g. normal) distributions. A long literature has established sufficient conditions on location parameter models to resolve conflict in various ways. In this work we consider a location-scale parameter structure, which is more complex than the single parameter cases because conflicts can arise between three sources of information, namely the likelihood, the prior distribution for the location parameter and the prior for the scale parameter. We establish sufficient conditions on the distributions in a location-scale model in order to resolve conflicts in different ways as a single observation tends to infinity. In addition, for each case we explicitly give the limiting posterior distributions as the conflict becomes more extreme.

April 2010

*Keywords*: Bayesian robustness, conflicting information, regularly varying distributions, RV-Credence.

# 1 Introduction

## 1.1 Conflict and robustness

In Bayesian analysis we have essentially two sources of information about the parameters of interest: the data and the prior knowledge. The modelling process basically consists in choosing probability distributions to express the uncertainty about the data and the parameters so as to construct the likelihood function and the prior distribution. As with any modelling exercise, the choices we make are invariably simplifications that we believe (or at least hope) will represent our beliefs and knowledge sufficiently well for the analyses we wish to perform. The activity often referred to as model checking or model criticism is the process by which we assess whether the modelling is indeed adequate for our purposes, or whether we need to rethink.

For example, we often use normal distributions in our modelling; this is for convenience and simplicity of analysis and because we think a smooth, unimodal, reasonably symmetric density would be appropriate, not because we believe absolutely in the normal distribution. In the process of model checking we may identify one or more outliers in the data. According to our assumption of normality, such observations would be extremely unlikely, so the model checking has identified a surprising event which suggests that we should review our modelling. Models with light-tailed (e.g. normal) distributions are often said to be non-robust to the presence of outliers because inferences can be highly sensitive to outlying observations. In contrast, when heavy-tailed (e.g. t) distributions are used the influence of outliers is reduced and inferences are less sensitive to surprising data.

In the Bayesian context, a surprising event (such as outliers) may be seen as one or more sources of information about the parameters which contradict the rest of the information. For instance, outlying observations suggest quite different parameter values from those indicated by the rest of the data. We refer to this situation as a *conflict of information*. Another important kind of conflict of information in Bayesian analysis is between the prior distribution and the likelihood, where the data suggest quite different parameter values from those anticipated in the prior. O'Hagan and Forster (2004, §3.35) provide a full discussion about conflicts and the effects they can have on the posterior distribution. In particular, they discuss the relationship between conflicts and tail weight. Evans and Moshonov (2006) give some procedures to detect conflicts between the prior and the data.

Depending on the choice of distributions, conflicts may be resolved by the posterior distribution in different ways. Using light-tailed distributions typically results in an attempt to compromise between the conflicting information sources, with the result that the posterior distribution concentrates on values of the parameters that are not well supported by any of the conflicting sources. In general, such non-robust behaviour is undesirable. Heavy-tailed distributions downweight some information sources in favour of others, and if the conflict is large enough may effectively reject one of more components. However, which sources are favoured depends on the relative tail weights of their respective distributions. The response to conflict which has been identified during model checking therefore demands a choice of which of the conflicting components we believe, and so wish to base our posterior judgements on.

Robust modelling involves anticipating such choices by choosing distributions with appropriate tail thickness at the outset. But whether we wish to model robustly at the start or to rely on model checking and modifying distributions appropriately when conflicts are identified, it is important to understand how different choices of distributions lead to different conflict resolutions. The research presented here is a contribution to that understanding.

## 1.2 Existing literature

The problems of conflict in Bayesian models were first observed by de Finetti (1961) and then described by Lindley (1968). Following their ideas a long literature has been developed channelled to establish sufficient conditions on the distributions in the model in order that the posterior distribution discards some conflicting sources of information in favour of others.

The theory has mostly been developed in the context of location parameters. Dawid (1973), O'Hagan (1979) and Desgagné and Angers (2007) proposed sufficient conditions on the data and prior distribution which allows conflict to be resolved by rejecting one of the conflicting sources of information in favour of the other. Essentially, they showed that this kind of resolution of conflict can occur when we use heavy-tailed distributions for the data and/or the prior in location-parameter models. Further asymptotic results and some practical aspects of heavy-tailed modelling were studied by O'Hagan (1988 and 1990). Pericchi *et al* (1993) and Pericchi and Sansó (1995) studied distributions from the exponential family which provide robust behaviour of the posterior distribution of the location parameter. O'Hagan and Le (1994) and Le and O'Hagan (1998) extended the ideas initiated by Dawid to the bivariate location-parameter case, focusing on the tail behaviour in different radial directions. An alternative approach is given by Haro-López and Smith (1999), who proposed some conditions on a 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 does not compute the limiting posterior distribution, and their conditions are quite difficult to verify.

Most of the results obtained in the literature deal with location parameter models, which may limit their practical applications. The posterior distribution of a scale parameter, in case of conflict, is more complex than in the pure location parameter case. For instance, with standard light-tailed models, in the presence of an outlier (conflicting information) the mode of the posterior distribution moves as the outlier becomes far away from the rest of the data (similarly to the pure location parameter case), but at the same time its dispersion is affected as well. Perhaps this is the main reason why researchers have avoided working with scale parameters, since it is quite difficult to establish conditions which control both the location and the variance of the posterior distribution.

In order to address this problem, Andrade and O'Hagan (2006) brought the theory of regular variation into the context of Bayesian robustness modelling. They showed that regular variation provides an easier way of interpreting the tail behaviour of probability densities. By defining distributions with heavy tails as regularly varying functions, they established conditions on the pure scale-parameter case which lead to resolution of conflicts in favour of one of the sources of information. However, unlike the location case, in the scale-parameter case the conflicting information is not completely rejected; instead some kind of *partial rejection* is achieved. As Andrade and O'Hagan show, this is a characteristic of the scale structure. Also, they proposed alternative conditions on the pure location-parameter case which are easier to verify than those proposed by Dawid and O'Hagan.

## 1.3 Location and scale

Following the substantial literature on location parameter problems, heavy-tailed distributions have often been adopted in applied Bayesian modelling, justified on the grounds that they provide inferences with built-in robustness. However, the applications invariably have both location and scale parameters, and there is as yet almost no theory regarding how conflicts are actually resolved in such models. This use of heavy-tailed distributions rests more on a leap of faith than on a proper understanding of their implications.

In this work we consider a model with both location and scale parameters. This is a more complex structure than either pure location parameter or pure scale parameter models, since now we have three sources of information – the likelihood, the prior distribution for the location parameter and the prior distribution for the scale parameter. Depending on the three distributions we choose, one source of information may be downweighted relative to the other two, giving three different ways for the posterior to resolve any conflict. We propose sufficient conditions on the location-scale parameter structure which allow the conflict to be resolved by rejecting one of the conflicting sources of information in favour of the other sources. Our results allow the applied Bayesian to devise appropriate modelling to resolve conflicts in whatever way seems most appropriate to the context. This may be adjusted modelling in response to a conflict having been identified by model checking, but may also be prior modelling intended to provide naturally robust inference. Like almost all of the theory concerning heavy-tailed models, our results are asymptotic, showing that one source of information will be rejected as the conflict with other sources tends to *infinity*. In practice, this information is not completely rejected but is increasingly downweighted the further it becomes from the rest of the information.

We focus on the theoretical aspects rather than the applications. The theory is presented using a standard location-scale parameters model in which the prior distribution of the location parameter involves the scale parameter. In the next section we provide a brief discussion about aspects of regular variation theory that are relevant for our theory, and relate regular variation with robustness modelling. In Section 3 we introduce the location-scale structure and prove the sufficient conditions which allow us to resolve conflict as an outlying observation tends to infinity. We give some illustrative examples in Section 4, in which we show how the joint posterior distribution of the location and scale parameters responds as conflicts become more accentuated. Section 5 concludes with some general comments about the theory.

# 2 Regularly varying functions

## 2.1 Definition and properties

The theory of regular variation, initiated by Karamata (1930), has been used in many areas of probability. See for instance, Feller (1971, Chapter VIII) and Resnick (1987) for a range of applications. The basic reference for regular variation is Bingham, Goldie and Teugels (1987) (BGT), which summarises the univariate theory developed until that moment. Some other relevant references are Seneta (1976), de Haan (1970) and Resnick (1987). Also, Andrade and O'Hagan (2006) relate regular variation with other concepts of robustness modelling.

We say that a measurable function $f$ is *regularly varying* with index $\eta$, which we write as $f \in R_\eta$, if $\eta \in \mathbb{R}$ and

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

In particular, if $f \in R_0$ then we say that $f$ is *slowly varying.* The set of all regularly varying functions is $R = \{R_\eta : -\infty < \eta < \infty\}$. By the Characterisation Theorem (see BGT §1.4.1), a function $f$ is regularly varying with index $\eta$ *if and only if* it can be written as $f(x) = x^\eta \ell(x)$ for $\rho \in \mathbb{R}$ and $\ell \in R_0$. Notice that using this important result, we can reduce a regular varying function (whichever form it assumes) to the form of a simple power and a well behaved function (slowly varying).

An important concept, which will be used to prove the results in Sections 3.3 and 3.3, is the asymptotic equivalence of functions. We say that $\phi(x)$ and $\psi(x) \neq 0$ are *asymptotically equivalent* at infinity, written $\phi(x) \sim \psi(x)$, if $\phi(x)/\psi(x) \to 1$ $(x \to \infty)$. In particular, if $f(x) \sim g(x)$ and $g \in R_\eta$, then $f \in R_\eta$.

## 2.2 Regular variation and heavy tails

We shall use regular variation (RV) to characterise the tail-thickness of probability distributions. It is possible to work with quite general distributions by considering regular variation of $1-F(x)$, where $F$ is a cumulative distribution function, but for simplicity we shall suppose that distributions are continuous and will characterise tail thickness by regular variation of density functions. In this context it is worth noting that density functions typically have both right- and left-hand tails. Definition (1) considers only the right-hand tail, but we can readily define a corresponding index of weight of the left-hand tail by considering $x \to -\infty$. For the purposes of this article, where we study behaviour of the posterior distribution as an observation tends to $+\infty$, it is enough to work with the right-hand tails. Analogous results as the observation tends to $-\infty$ are easily derived but depend on left-hand tails of some densities. Note also that a proper density function with a regularly varying right-hand tail must have $\eta < -1$.

Andrade and O'Hagan (2006) defined a density function to be heavy-tailed (in the right-hand tail) if it is regularly varying. Since the RV index will typically be negative, they defined the negative of the index to be the RV *credence*. Thus a density $f$ has RV credence $\rho$ if $f \in R_{-\rho}$. This accords with the notion of credence introduced by O'Hagan (1990). Although the two definitions differ in detail, they agree for any density that satisfies both the regular variation condition and O'Hagan's condition for credence; in particular the density of a Student $t$ distribution with $d$ degrees of freedom has credence $d+1$ and also RV credence $d+1$.

As opposed to the idea of regular variation, the class of *rapidly varying functions* (de Haan, 1970) contains those functions that do not satisfy the limit (1), or in particular it embraces functions with decay or growth faster than any power function of finite order. For instance the tails of a normal density are rapidly varying. Following Andrade and O'Hagan (2006), we refer to distributions with regularly varying density as heavy-tailed (on the right).

Although the results proved here apply when all the relevant densities are regularly varying, and so heavy-tailed, analogous results for cases where one or more densities are light-tailed may be obtainable by simple extension. For instance, the normal density is a limit of the Student's $t$ density with $d$ degrees of freedom as $d$ goes to infinity. Such a case may be studied by replacing the normal density with the $t$ density, applying the results of this paper and letting $d \to \infty$. We also assume throughout that distributions are proper, and so all RV credences are greater than 1.

## 2.3 Regular variation and convolution

One important aspect of regular variation theory concerns the integrals of regularly varying functions. The so called *Abelian and Tauberian theorems* deal with the relations between regularly varying functions and their integral transforms (such as Mellin, Laplace and Fourier transforms). The Abelian theorems concern those results where one would like to verify whether an integral transform of a regularly varying function is or is not regularly varying. The Tauberian theorems concern the converse case, where a regularly varying integral transform may or may not have its integrand regularly varying. Some further discussion can be found in Pitt (1958) and BGT.

The following result concerning the convolution of two regularly varying densities was kindly provided by Professors N. Bingham and C. Goldie. Note that Proposition 2.1 is an Abelian theorem, since we want to assess an integral transform (convolution) given that its integrand is regularly varying.

**Proposition 2.1** (Convolution of regularly varying densities)**.** *If $f$ and $g$ are proper, bounded and strictly positive probability densities on $\mathbb{R}$, both regularly varying at $\infty$, then their convolution $f * g$ satisfies*

$$f * g(x) \sim f(x) + g(x),$$

*and so is also regularly varying with index the maximum of the index of $f$ and the index of $g$.*

*Proof.* See the Appendix. $\square$

In other words, the convolution tail behaves exactly like the convolving function with thicker tail.

# 3 Resolution of conflicts in location-scale structures

We now apply regular variation methods to address the asymptotic behaviour of the posterior distribution for a model with a location parameter $\mu$ and a scale parameter $\theta$, as one observation tends to infinity. All density functions will be proper (and so all RV indices are greater than 1), smooth, continuous, bounded and strictly positive over the ranges of their arguments. We use the symbol $\overset{D}{\sim}$ to denote 'is distributed with density function'. The range for $\mu$ is $(-\infty, +\infty)$ while that for $\theta$ is $(0, \infty)$, and all integrals with respect to these parameters are over those ranges unless otherwise specified.

The first subsection below reviews results from Andrade and O'Hagan (2006), relating to the cases of a single location or a single scale parameter, and the second subsection introduces our location-scale model. The next subsection addresses the three possible limiting forms of the posterior distribution when we have a single observation $y$ and $y \to \infty$. Finally we generalise to the case of multiple observations in which a single outlier goes to infinity.

## 3.1 Results for a scale parameter or a location parameter

We will use the following additional results (Theorems 3.1 and 3.3 and Corollaries 3.2 and 3.4) which are adapted from Andrade and O'Hagan (2006), who studied separately the cases of models with a scale parameter and with a location parameter. The original results use slightly different notation and are more general than we need here.

**Theorem 3.1** (Conditions on 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 $\theta \overset{D}{\sim} p$. Suppose that*

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

*(ii) $p \in R_{-\alpha}$ with $\alpha > \rho$.*

*Then*

$$p(\theta|y) \longrightarrow \frac{\theta^{\rho-1}p(\theta)}{\int_{\Theta} \theta^{\rho-1}p(\theta)d\theta}, \text{ as } y \to \infty. \tag{2}$$

**Corollary 3.2.** *Let $y|\theta \overset{D}{\sim} f$, such that $f$ is of the form $f(y|\theta) = (1/\theta) \times h(y/\theta)$ and $\theta \overset{D}{\sim} p$. Suppose that*

*(i) $p \in R_{-\alpha}$ with $\alpha > 1$, and*

*(ii) $h \in R_{-\rho}$ with $\rho > \alpha$.*

*Then*

$$p(\theta|y) \sim \frac{\theta^{-\alpha} f(y|\theta)}{\int_{\Theta} \theta^{-\alpha} f(y|\theta) d\theta}, \text{ as } y \to \infty. \tag{3}$$

Theorem 3.1 gives sufficient conditions on the data and prior distributions such that the posterior distribution rejects partially the outlier in a pure scale model. Note that we do not achieve complete rejection of the outlier because of the term $\theta^{\rho-1}$ yielded by the scale structure. Conversely, Corollary 3.2 gives conditions under which it is the prior information that is partially rejected and the posterior is based on the likelihood. Again, the rejection is partial because of the power of $\theta$ that arises from the scale structure. Observe that the asymptotic equivalence says that as $y$ increases the posterior distribution becomes the likelihood multiplied by $\theta^{-\alpha}$. Effectively, the original informative prior density is replaced by the improper prior density proportional to $\theta^{-\alpha}$. The only remnant of the original prior is in the power $-\alpha$.

**Theorem 3.3** (Conditions on the location-parameter case)**.** *Let $y|\mu \overset{D}{\sim} f(y-\mu)$ and $\mu \overset{D}{\sim} p(\mu)$ such that the following conditions hold.*

*(a)* $f \in R_{-\rho}$.

*(b)* $f > 0$ *and 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)* $p \in R_{-c}$ *with* $c > \rho$.

*Then*

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

**Corollary 3.4.** *Let $y|\mu \overset{D}{\sim} f(y-\mu)$ and $\mu \overset{D}{\sim} p(\mu)$ such that the following conditions hold.*

*(a)* $p \in R_{-c}$.

*(b)* $p > 0$ *and is continuous in* $\mathbb{R}$.

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

(*d*) $f \in R_{-\rho}$ $(\mu \to \infty)$ *with* $\rho > c$.

*Then*

$$p(\mu|y) \sim f(y|\mu), \text{ as } y \to \infty. \tag{4}$$

Theorem 3.3 and Corollary 3.4 complement the results of Dawid (1973) by proving rejection of conflicting information in a pure location problem under simpler sufficient conditions based on regular variation.

These results basically mean that assigning a prior which is lighter tailed than $f$, we obtain a posterior distribution based on the prior distribution, such that the observation $y$ is wholly rejected. Conversely, if the prior is heavier tailed than $f$, we obtain a posterior distribution based on the data distribution, so that now it is the prior information that is rejected. In the next subsections we generalise these results to a location-scale parameter structure.

### 3.2 A location-scale model

We consider the following hierarchical structure for a single observation. The generalisation for many observations follows in Subsection 3.4.

$$\begin{cases} y|\mu,\theta \overset{D}{\sim} f(y|\mu,\theta) = (1/\theta) \times h[(y-\mu)/\theta], \\ \mu|\theta \overset{D}{\sim} p(\mu|\theta) = (1/\theta) \times v(\mu/\theta), \\ \theta \overset{D}{\sim} p^*(\theta), \\ h \in R_{-\rho}\ (\rho > 1)\,, \quad v \in R_{-c}\ (c > 1)\,, \quad p^* \in R_{-\alpha}\ (\alpha > 1). \end{cases} \tag{5}$$

Notice that although hyperparameters are not explicitly stated in Model (5), it is possible to have known location and scale hyperparameters in $v$ and $p^*$ without loss of generality in the forthcoming results.

The prior structure in Model (5) is rather restrictive in the way it assumes that $\theta$ is a scale parameter for the conditional prior distribution of $\mu$, but this hierarchical structure is widely used in practice and serves our purpose of demonstrating how regular variation can be applied to examine conflicts in location-scale structures. Behaviour of the posterior distribution under different forms of prior structure, and in particular when the parameters are independent *a priori*, could be quite different and is certainly a topic for future research.

Conflict will arise if the observation is large enough so that it is inconsistent with both location and scale parameters being close to their prior means. We therefore consider behaviour of the joint posterior distribution as $y$ tends to infinity. Our approach will be based on separately examining the posterior marginal distribution of $\theta$ and the posterior conditional distribution of $\mu$ given $\theta$. The following corollary applies Proposition 2.1 to identify the regular variation index of the integrated likelihood for $\theta$.

**Corollary 3.5.** *In Model (5), define the marginal likelihood* $\varphi(y|\theta) = \int f(y|\mu,\theta) \times p(\mu|\theta)d\mu$*. Then*

*(i)* $\varphi$ *is of the form* $\varphi(y|\theta) = (1/\theta) \times g(y/\theta)$ *and*

*(ii)* $g \in R_{-\rho'}$ *as* $y \to \infty$*, where* $\rho' = \min\{\rho, c\}$.

*Proof.* See the Appendix. □

## 3.3 Posterior asymptotics

We now consider the asymptotic behaviour of the posterior joint distribution of $\mu$ and $\theta$ under Model (5) as $y \to \infty$. We find three cases, depending on which of the three RV indices $\rho$, $\alpha$ and $c$ is the smallest. In each case, the corresponding source of information is rejected, although note that in the case of rejection of the likelihood or the prior information about $\theta$ the rejection is 'partial', in the sense discussed in Subsection 3.1.

For each case of conflict we consider the limiting posterior joint distribution as a product of the marginal posterior distribution of $\theta$ and the posterior conditional of $\mu$ given $\theta$. That is provided the limits below exist we have

$$\lim_{y\to\infty} p(\mu,\theta \,|\, y) = \lim_{y\to\infty} p(\theta \,|\, y) \times \lim_{y\to\infty} p(\mu \,|\, \theta, y) \; . \tag{6}$$

These limits are then derived separately by expressing the first as a pure scale-parameter problem for $\theta$ and the second as a pure location-parameter problem for $\mu$ conditional on $\theta$.

**Theorem 3.6** (Rejection of a single observation)**.** *Consider an observation $y$ following Model (5) and suppose that the following conditions hold.*

*(i)* *$h > 0$ and is continuous. There exist a $C_1$ and a $C_2 \geq C_1$ such that $h(y)$ is decreasing for $y \geq C_2$ and increasing for $y \leq C_1$. Also, $d \log h(y)/dy$ exists and is increasing for $y \geq C_2$.*

*(ii)* *$\rho < \min\{\alpha, c\}$.*

*Then as $y \to \infty$,*

$$p(\mu, \theta|y) \longrightarrow \frac{\theta^{\rho-1} p(\mu|\theta) p^*(\theta)}{\int_\Theta \theta^{\rho-1} p^*(\theta) d\theta}.$$

*Proof.* See the Appendix. □

In simple terms, Theorem 3.6 says that if a large observation occurs in the data, modelling that observation with some suitably heavy-tailed distribution will produce a posterior distribution which is robust to that observation, in the sense that the joint posterior distribution of $\theta$ and $\mu$ equals their joint prior distribution, except for the scale parameter term. Condition $(i)$ ensures that the likelihood is well behaved in the tails. Condition $(ii)$ simply requires the priors for $\theta$ and $\mu$ to be lighter-tailed than $h$.

**Theorem 3.7** (Partial rejection of $p^*(\theta)$)**.** *Consider Model (5) and assume $\alpha < \min\{\rho, c\}$, then,*

$$p(\mu, \theta|y) \sim \frac{\theta^{-\alpha} \times f(y|\mu, \theta) p(\mu|\theta)}{\int_\Theta \int_\mu \theta^{-\alpha} \times f(y|\mu, \theta) p(\mu|\theta) d\mu d\theta}, \text{ as } y \to \infty. \tag{7}$$

*Proof.* See the Appendix. □

It is important to note the difference between the asymptotics here and in the case of Theorem 3.6. In both cases, we first consider the conflict in the marginal distribution of $\theta$ between the prior distribution $p^*$ and the marginal likelihood represented by $g$. In Theorem 3.6, this conflict is resolved by rejecting the data information, so that the posterior tends to the prior. In this case, $\theta$ is *a posteriori* finite as $y \to \infty$, and hence conflict arises also in the conditional distribution of $\mu$ given $\theta$. In Theorem 3.6 this conflict is resolved by again rejecting the data information. In the next result, we show how this conflict may be resolved by rejecting the conditional prior distribution.

**Theorem 3.8** (Partial rejection of $p(\mu|\theta)$)**.** *Consider an observation $y$ following Model (5) and suppose that the following conditions hold.*

*(i)* $v > 0$ *and continuous. There exist a $C_1$ and a $C_2 \geq C_1$ such that $v(\mu)$ is decreasing for $\mu \geq C_2$ and increasing for $\mu \leq C_1$. Also, $d \log v(\mu)/d\mu$ exists and is increasing for $\mu \geq C_2$;*

*(ii)* $c < \min\{\alpha, \rho\}$.

*Then,*

$$p(\mu, \theta|y) \sim \frac{\theta^{c-1} f(y|\mu, \theta) p^*(\theta)}{\int_\mu \int_\Theta \theta^{c-1} f(y|\mu, \theta) p^*(\theta) d\mu d\theta}, \text{ as } y \to \infty.$$

*Proof.* See the Appendix. □

The above three theorems cover all the possible cases as long as one of the three RV credences $\rho$, $\alpha$ and $c$ is strictly less than the other two (and given regular behaviour of the tails of $h$ or$v$ as required). If two RV credences tie for smallest then none of the sources of information will be rejected. The situation will be similar to that of a single location parameter, where if the prior and likelihood have equal RV credences the limiting posterior continues to accommodate both the prior and likelihood by having two diverging modes, one centred on each of the conflicting information sources.

## 3.4 Multiple observations

The preceding results have shown how in the case of a single observation we may, as the observation tends to infinity, reject the likelihood, the prior conditional distribution of $\mu$ given $\theta$, or the prior marginal distribution of $\theta$. It is clear that we cannot reject two of these components, since the result would be an improper joint distribution for $\mu$ and $\theta$.

We now consider multiple observations following the hierarchical structure

$$\begin{cases} y_i|\mu,\theta & \overset{D}{\sim} \quad f(y_i|\mu,\theta) = 1/\theta \times h_i[(y_i-\mu)/\theta] \text{ ind. } (i=1,...,n), \\ \mu|\theta & \overset{D}{\sim} \quad p(\mu|\theta) = (1/\theta) \times v(\mu/\theta), \\ \theta & \overset{D}{\sim} \quad p^*(\theta), \\ h_i \in R_{-\rho_i}, \ \rho_i > 1, \ (i=1,...,n), & \\ v \in R_{-c}, \ c > 1, \qquad p^* \in R_{-\alpha}, \ \alpha > 1. & \end{cases} \tag{8}$$

Note that the model allows a different density to each observation and every density can have a different RV credence, although in many applications they will all be the same. Again all densities are bounded and positive.

**Lemma 3.9.** *Consider data $\boldsymbol{y} = (y_1, ..., y_n)$ following Model (8). Then the posterior distribution has the form*

$$\begin{cases} \mu|\boldsymbol{y},\theta & \overset{D}{\sim} \quad P(\mu|\boldsymbol{y},\theta) = (1/\theta) \times V[\mu/\theta|\boldsymbol{y}], \\ \theta|\boldsymbol{y} & \overset{D}{\sim} \quad P^*(\theta|\boldsymbol{y}), \end{cases} \tag{9}$$

*where*

(*i*) $V[\mu/\theta|\boldsymbol{y}] \propto v\,(\mu/\theta) \times \prod_{i=1}^n h_i\,((y_i-\mu)/\theta)$ *and* $V \in R_{-(c+\sum_{i=1}^n \rho_i')}$, *and*

(*ii*) $P^*(\theta|\boldsymbol{y}) \propto p^*(\theta) \times \int_\mu f(\boldsymbol{y}|\mu,\theta)p(\mu|\theta)d\mu$ *and* $P^* \in R_{-(\alpha+n)}$.

*Proof.* See the Appendix. □

Lemma 3.9 basically characterises the posterior distribution as having the same form as the prior distribution in (8), but with changed regular variation indices. In particular, suppose we have data $\boldsymbol{y} = (y_1, ..., y_n)$ and let $y_n$ be taken sufficiently large so that it becomes an outlier. Lemma 3.9 allows the multiple observations problem to be transformed into the single observation case, by conditioning all the densities on the non-outlying data information $\boldsymbol{y}^{(n-1)} = (y_1, ..., y_{n-1})$.

The following corollary specifies how the conditions originally established through Theorem 3.6 apply to the case of several independent observations with a single outlier, so that the outlier is rejected. It follows directly from Theorem 3.6 with $y_n$ as the single observation $y$ and with the densities $v$ and $p^*$ replaced by the distributions $V$ and $P^*$ from Lemma 3.9, but using only the first $n-1$ observations $\boldsymbol{y}^{(n-1)} = (y_1, ..., y_{n-1})$.

**Corollary 3.10** (Rejection of a single outlier in $n$ observations)**.** *Consider data* $\boldsymbol{y} = (y_1, ..., y_n)$ *following Model (8), and suppose the following conditions are satisfied:*

*(i)* $h_n > 0$. *There exist a* $C_1$ *and a* $C_2 \geq C_1$ *such that* $h_n(y)$ *is decreasing for* $y \geq C_2$ *and increasing for* $y \leq C_1$. *Also,* $d \log h_n(y)/dy$ *exists and is increasing for* $y \geq C_2$*;*

*(ii)* $\rho_n < \min\{\alpha + n - 1, c + \sum_{i=1}^{n-1} \rho_i\}$.

*Then*

$$p(\mu, \theta|\boldsymbol{y}) \longrightarrow \frac{\theta^{\rho_n - 1} f(\boldsymbol{y}^{(n-1)}|\mu, \theta) p(\mu|\theta) p^*(\theta)}{\int_\mu \int_\Theta \theta^{\rho_n - 1} f(\boldsymbol{y}^{(n-1)}|\mu, \theta) p(\mu|\theta) p^*(\theta) d\mu d\theta}, \text{ as } y_n \rightarrow \infty. \tag{10}$$

It is straightforward also to obtain versions of Theorems 3.7 and 3.8 for the case of a single outlier in a sample of $n$ observations. Notice in these cases that the partial rejection of prior information involves rejecting not just the original prior $p^*(\theta)$ or $p(\mu \,|\, \theta)$ but also that information in the non-outlying observations which relates to $\theta$ or $\mu$, respectively.

When the observations are a random sample of size $n > 1$ from a common distribution having density $h$ with RV credence $\rho$, then condition *(ii)* of Corollary 3.10 becomes $\rho < \min\{\alpha + n - 1, c + (n-1)\rho\}$. Then $\rho$ is certainly less than $c + (n-1)\rho$, so the condition reduces to $\rho < \alpha + n - 1$. Therefore the outlier will be rejected provided $\alpha$ is sufficiently large or the sample is large enough. Otherwise the information about $\theta$ coming from the prior distribution and from $\boldsymbol{y}^{(n-1)}$ will be (partially) rejected and the posterior distribution for $\theta$ will be based on the outlier alone (except for a power of $\theta$ deriving from the scale structure). In particular, note that if $\alpha < \rho - 1$ we can have the situation where the asymptotic behaviour of the posterior depends on the sample size, and can change dramatically simply by adding or removing an observation.

By allowing the observations to have different RV credences, however, we admit the more exotic possibility that a single outlier can have more credence than the combined information in the rest of the sample, so that the outlier is retained and the information about $\mu$ in the prior and the rest of the data is rejected.

We could have a group of outlying observations all tending to infinity together. This situation was addressed in the context of a location parameter by O'Hagan (1990), who let groups of observations tend to infinity in such a way that the distances between them were fixed. In the location parameter case it was enough that the centre of the group was tending to infinity, in order to study the conflict, and the case of a group of outliers was reduced to that of a single outlier. In the case of both location and scale parameters the situation is more complex. In particular, we would expect to be able to reject both parts of the prior distribution, because the likelihood (multiplied by a negative power of $\theta$) will provide a proper joint posterior distribution. We cannot address this situation using the tools developed here, because the marginal likelihood for $\theta$ will not just be of the form $\theta^{-1}g(y'/\theta)$, where $y'$ represents the centre of the group. The marginal likelihood will also provide information about $\theta$ from the distances between the members of the group, and asymptotically this information may not be rejected.

# 4 Examples

We now present numerical examples of each of the three kinds of conflict resolution. For $h$ and $v$ we use Student $t$ densities, remembering that the density $t_d$ with $d$ degrees of freedom has RV credence $d + 1$. The $t$ distributions satisfy the conditions $(i)$ of Theorems 3.6 and 3.8. For $p^*$ we use inverse-gamma distributions. The distribution $IG(a, b)$ has mean $b/a$ and RV credence $a + 1$.

In each example we will observe the posterior joint density converging to the limiting form predicted by the theory.

## 4.1 Partial rejection of an outlier

Consider the general model (8) and suppose that we have a random sample $\boldsymbol{y} = (1, 2, 4, 4, y_5)$, where we we will allow $y_5$ to increase towards infinity, becoming an outlier. We let $h_i$ be the $t_3$ density for $i = 1, 2, 3, 4, 5$, $v$ be the $t_{19}$ density and $p^*$ be the $IG(7, 8)$ density.

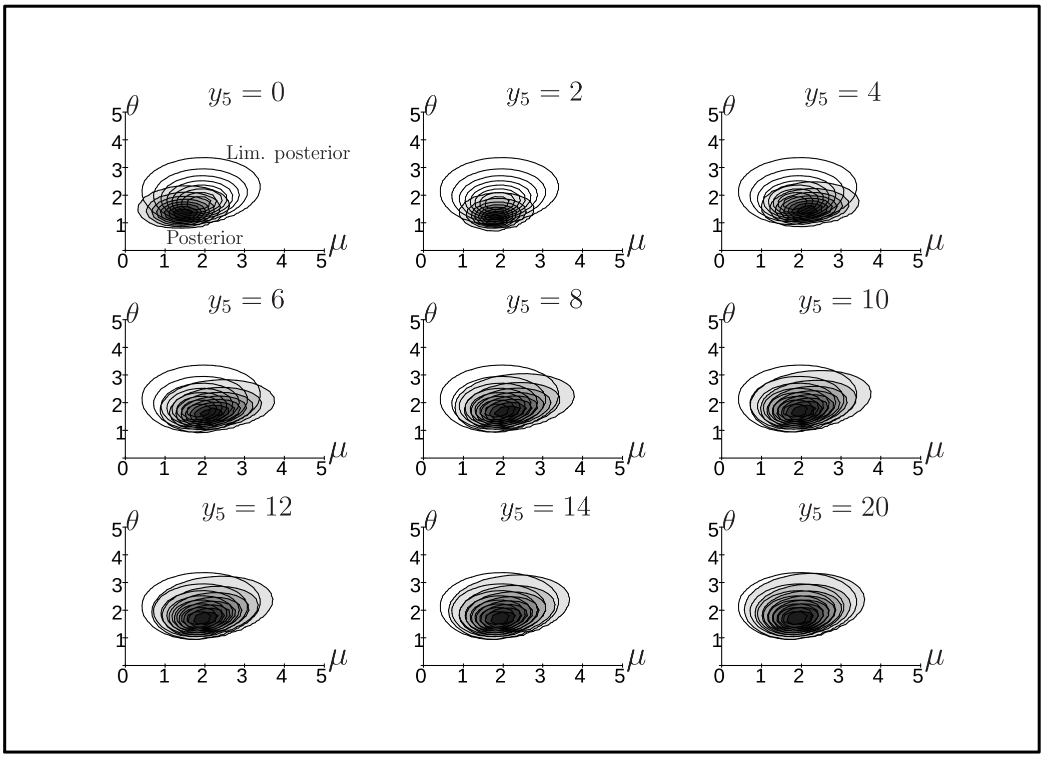


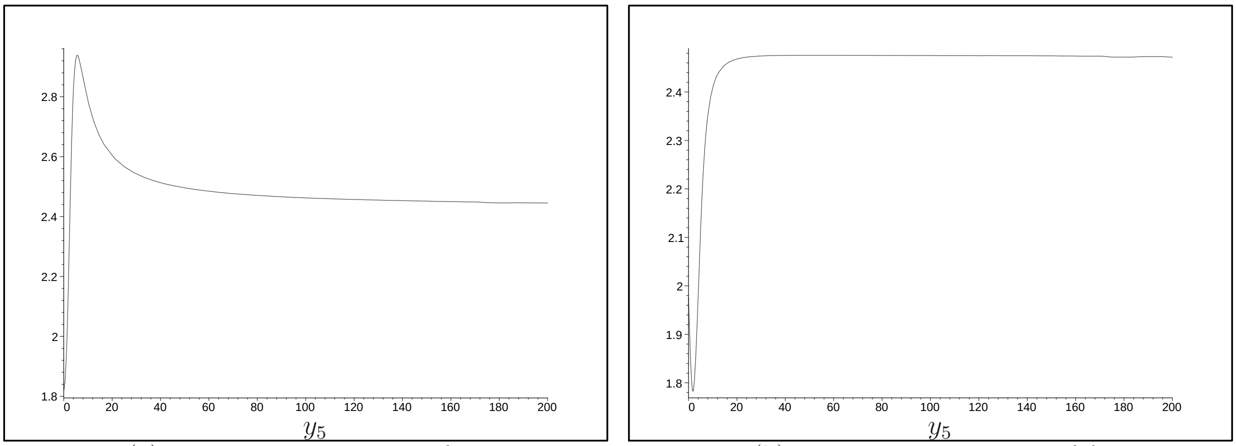

Figure 1: Partial rejection of the outlier $y_5$

Thus the conditions of Corollary 3.10 are found to hold with $n = 5$, $\rho = 4$, $c = 20$ and $\alpha = 8$. In the limit as $y_5 \to \infty$ we should find that the posterior distribution has the form

$$p(\mu, \theta|\boldsymbol{y}) \propto \theta^{\rho} f(\boldsymbol{y}^{(4)}|\mu, \theta)p(\mu|\theta)p^*(\theta), \tag{11}$$

where $\boldsymbol{y}^{(4)} = (1, 2, 4, 4)$ is the data without the outlier $y_5$.

Figure 1 shows the result of numerical evaluation of the actual posterior for some increasing values of $y_5$. In each panel of Figure 1 both the actual posterior (for that value of $y_5$) and the limiting posterior distribution given in (11) are plotted. Clearly, the posterior distribution tends to the limiting posterior distribution as $y_5$ becomes large. As for the posterior expectation of $\mu$ and $\theta$, Figure 2 shows that the posterior expectation of $\mu$ follows the outlier until around $y_5 = 15$, then it begins to reject $y_5$ in favour of the limiting posterior expectation of $\mu$, which is approximately 2.42. Likewise, the posterior expectation of $\theta$ increases with the outlier until around $y_5 = 40$, then it becomes unaffected by further values of $y_5$, converging to the limiting posterior expectation of $\theta$, which is approximately




(a) Posterior expectation of $\mu$

(b) Posterior expectation of $\theta$

Figure 2: Posterior expectations of $\mu$ and $\theta$ as $y_5$ increases

2.47. It is important to notice that the initial influence of the outlier does not fade away completely like in the pure location case. There is an initial influence over the estimates which is not recovered, just like in the pure scale case: it is possible only to achieve partial rejection of the outlier in the posterior quantities concerning the scale parameter. If we achieved total rejection of the outlier we would expect to have the posterior estimates for $y$ sufficiently large equal to those of the posterior distribution without the outlier $p(\mu, \theta|\boldsymbol{y}^{(4)}) \propto f(\boldsymbol{y}^{(4)}|\mu, \theta)p(\mu|\theta)p^*(\theta)$. However, the posterior expectations of $\mu$ and $\theta$ produced by this distribution are approximately 2.33 and 1.90, respectively, which do not match with those values obtained from limiting posterior distribution (11).

## 4.2 Partial rejection of the prior distribution of $\theta$

In this example we consider the simpler Model (5) for a single observation $y$, which we let increase towards infinity. We let $h$ be the $t_7$ density, $v$ be the $t_9$ density and $p^*$ be the $IG(3, 50)$ density.

We now find that the conditions of Theorem 3.7 hold with $\rho = 8$, $c = 10$ and $\alpha = 4$. Therefore, we should find that in the limit as $y \to \infty$ the prior distribution of $\theta$ is partially rejected in favour of the data and the prior distribution of $\mu$ given $\theta$. For large $y$ the posterior

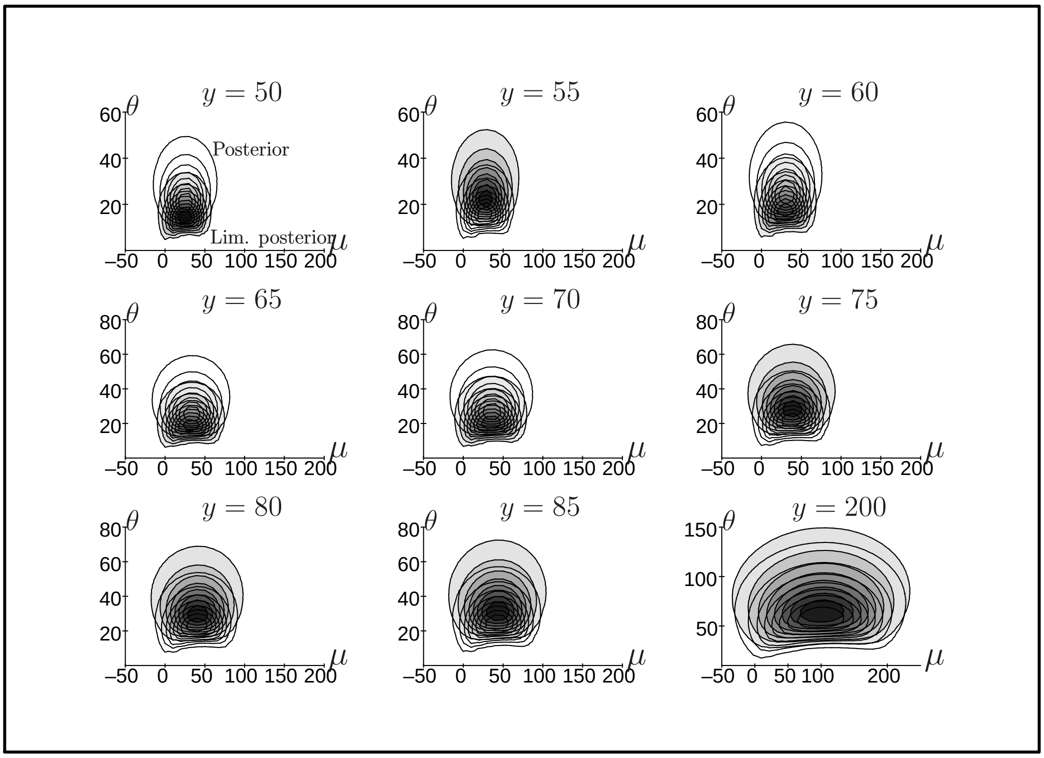



Figure 3: Partial rejection of the prior distribution of $\theta$

distribution will be approximately

$$p(\mu, \theta|y) \propto \theta^{-\alpha} f(y|\mu, \theta) p(\mu|\theta).$$

Indeed, increasing $y$ so that $p^*$ conflict with the data and $p(\mu|\theta)$, the contour plots in Figure 3 shows that the posterior distribution tends to the limiting posterior distribution.

## 4.3 Partial rejection of the prior distribution of $\mu$ given $\theta$

Our third example again uses the simpler model 5 for a single observation $y$, which we let increase towards infinity. We again let $h$ be the $t_7$ density, but now $v$ is the $t_6$ density and $p^*$ is the $IG(9, 50)$ density.

We now find that the conditions of Theorem 3.8 hold with $\rho = 8$, $c = 7$ and $\alpha = 10$. Therefore, we should find that in the limit as $y \to \infty$ the prior distribution of $\mu$ given $\theta$ is partially rejected in favour of the data and the prior distribution of $\theta$. For large $y$ the posterior distribution will be approximately

$$p(\mu, \theta|y) \propto \theta^{c-1} f(y|\mu, \theta) p^*(\theta).$$

This behaviour is shown in Figure 4, in which the posterior and the limiting posterior distribution again become close as $y$ becomes large.

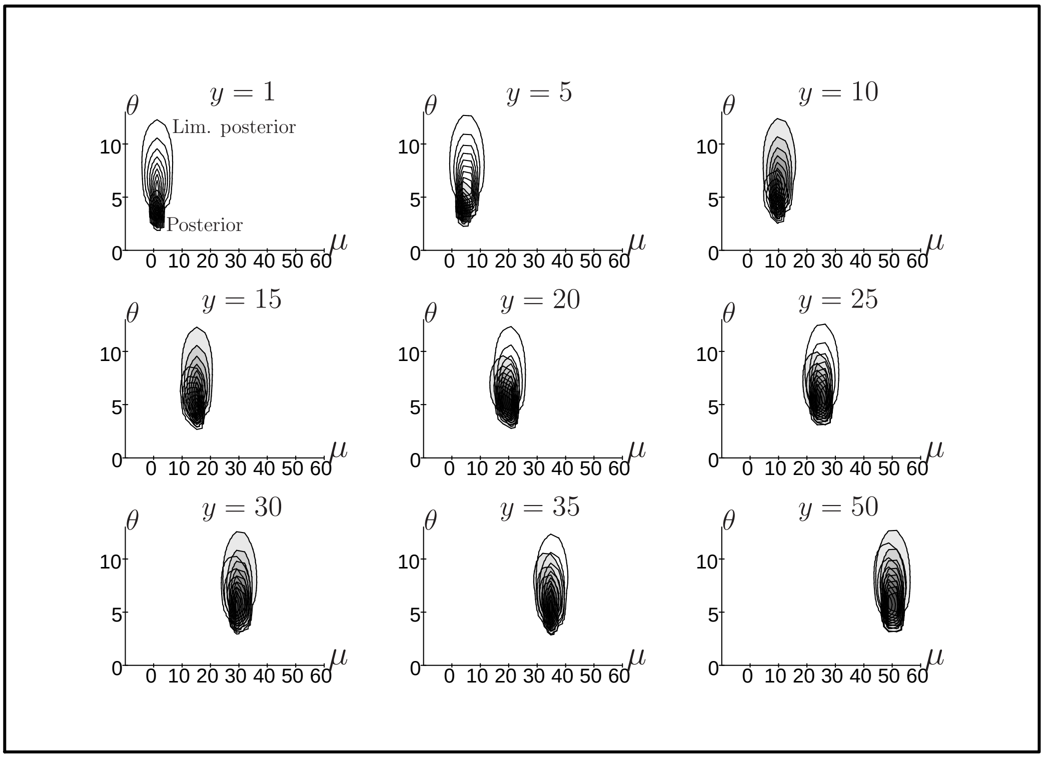



Figure 4: Partial rejection of the prior distribution of $\mu$

# 5 Other classes of heavy-tailed distributions

We have considered here a specific class of heavy-tailed distributions, those with regularly varying densities. There are a number of important alternative classes, some of which have been studied in the context of Bayesian modelling, as discussed in Section 1.2.

We have mentioned the class of distributions with credence values as defined by O'Hagan (1990) and extended to two-dimensional location-parameter models by O'Hagan and Le (1994) and Le and O'Hagan (1998). In O'Hagan's definition a density has credence

$d+1$ if it can be bounded above and below by multiples of the Student $t$ density with $d$ degrees of freedom. This is similar to RV credence, but with some differences. A distribution could have credence $c$ but not be regularly varying, because its density may fluctuate between the upper and lower bounds in such a way that the limit (1) does not exist. On the other hand, a density proportional to $\ln(x)$ times a $t$ density with $d$ degrees of freedom has RV credence $d+1$ but does not have credence according to O'Hagan's definition.

The Lévy class of *$\alpha$-stable distributions* (Feller, 1971 §VI) involves all random variables $X$ for which the linear combination of two *iid* copies of $X$ has the same distribution as $X$. We write $X \overset{D}{\sim} \boldsymbol{S}(\alpha, \beta, \gamma, \delta)$, where $\alpha$ is the *index of stability*, $\beta$ is the *symmetry coefficient*, $\gamma > 0$ is the *scale parameter* and $\delta \in \mathbb{R}$ is the *location parameter*. See Nolan (2010) for a review on stable distributions. The class embraces both heavy- and light-tailed distributions. For instance, for $0 < \alpha < 2$ at least one of the tails is asymptotically power law. If $-1 < \beta < 1$ both tails are asymptotically power law, if $\beta = -1$ only the left tail is asymptotically power law, if $\beta = 1$ only the right tail is asymptotically power law. For instance, if $0 < \alpha < 2$ and $-1 < \beta \leq 1$ then $f(x) \sim k(1-\beta)x^{-(\alpha+1)}$, where $k$ is a constant involving $\alpha$ and the location and the scale parameters. Note that is easy to show that if $f$ has $\alpha-$stable tail with $0 < \alpha < 2$ then $f \in R_{-(\alpha+1)}$ at the same tail, therefore the present theory also embraces $\alpha-$stable distributions such as the Cauchy and the Lévy distributions.

Another important class of heavy-tailed distributions is the *subexponential class* which involves the distributions whose tails decay is slower than any exponential function. We say a random variable $X \overset{D}{\sim} f(x)$ is subexponential, denoted by $f \in SD$, if $f$ satisfies: (1) $f(x-\mu)/f(x) \to 1\,(x \to \infty)$, $\forall \mu \in \mathbb{R}$ and (2) $f * f(x)/f(x) \to 2\,(x \to \infty)$. Condition (1) implies that $f$ belongs to the class $L$ and Condition (2) requires that the $2-$fold convolution of $f$ is asymptotically equivalent to $f$. $L$ is a heavy tail class itself and $SD \subset L$. The regularly varying distributions are subsets of the $SD$ and $L$ classes, that is $R \subset SD \cap L$ (See Feller (1971) §VIII.8). Nonetheless, note that the two conditions for subexponentiallity considers only location shifts in the density, that is Condition (1) compares the tails of $f$ when its argument is shifted by $\mu$ and Condition (2) compares the 2-fold location convolution of $f$ with $f$. Therefore, there is no guarantee that under scale structure this property will be satisfied. For instance, the *power exponential distribution* (Box and Tiao, 1973) $f(x) \propto e^{-|x|^q}$

$(0 < q \leq 1)$, which is in $L$, is well known as a heavy-tailed density for location parameters but it is not regularly varying.

There is certainly scope for more research based on Bayesian inference using these and other classes of heavy-tailed distributions. In particular, none of these classes has yet been examined in the context of scale parameters. However, the property of regular variation implies a regularity of the density under arbitrary scaling which we believe is particularly attractive in studying asymptotic posterior behaviour with scale parameters.

# 6 Discussion

## 6.1 Credence and regular variation

Regular variation gives a simple and natural interpretation of tail thickness. In particular, RV–credence of a source of information is related to the concept of credence introduced in O'Hagan (1990). An advantage of using regular variation to define tail thickness is that conditions of the resulting theorems are simpler and easier to verify than under other approaches. Indeed the regular variation of a distribution is obtained by simply evaluating the limit (1).

The theory allows a broader interpretation of (RV) credence, which varies within $(1, \infty)$. RV credence close to 1 indicates that we believe that a source of information is likely to produce conflicting information, whereas large RV credence means we believe that the information is very reliable and will not cause conflicts. When an outlier becomes large, it conflicts with the rest of the data and the prior distribution. Depending on the RV credence of the outlier and of the other sources of information, we have seen three distinct ways in which this conflict can be resolved. Andrade and O'Hagan (2006) showed, in the pure location and pure scale parameter cases, that the posterior distribution will be based on the source of information with largest total RV credence. In particular, they showed that a group of outliers could be rejected if their total RV credence was smaller than that of the prior and the remaining data. As shown, here the outcome is similar, that is the sources of information with largest credences will dominate the posterior distribution. However, considering a group of outliers, the situation with location and scale parameters is more complex. Although we would expect the total RV credence still to be the key factor regarding the location parameter $\mu$, the outlying group can still provide information about the scale parameter $\theta$. We need further investigation on this matter.

## 6.2 Implications of our work

The presence of conflicting information has always been a concern in Statistics. The most common form of conflict is the outlier, which usually is treated as an aberration; many of the traditional approaches to deal with outliers try to identify and delete them from the data. For instances of these techniques, see Barnett and Lewis (1978). The essence of the Bayesian perspective is to use all available information, including prior information, thus the idea of simply removing outlying observations is not natural in a Bayesian framework. As Neyman and Scott (1971) point out, in many situations outliers should not be treated as an anomaly (and removed from the data), but as a natural characteristic of the phenomenon. On the other hand, conflicts such as outliers can strongly influence the results and lead to unstable inferences. The present work is a contribution to the growing Bayesian literature on modelling with heavy tails, in which conflicts are resolved asymptotically in appropriate ways. Furthermore, the theory may be useful not only in the model check task, but also to express our initial assumptions and beliefs concerning the phenomenon of interest.

The general idea is to assign a heavy-tailed distribution to each source of information, with a credence value that reflects the extent to which we will believe that information to be useful if it conflicts with other sources. Note that this subjective aspect of modelling is inevitable, since at some point we have to choose densities from a range of distributions, hence the theory may give an important tool to avoid problems of misspecification of the model.

There has been very little study of the behaviour of heavy-tailed models outside the very simple class of location-parameter models. The present work begins a study of location and scale parameter models. This is a much more complex undertaking, and our analysis has required careful asymptotics to achieve the results presented here. We have considered just a single outlier and a particular location-scale model structure; nevertheless, we already see important differences from the familiar theory of location parameter conflicts. In particular, we can reject either the prior for $\mu$ (given $\theta$) or that of $\theta$, and in respect of $\theta$ we also have the kind of partial rejection that was found in Andrade and O'Hagan (2006) for pure scale parameter models.

## 6.3 Limitations and future work

We have already remarked on some of the limitations in our results. We have assumed that all distributions have regularly varying tails and have only considered asymptotics as an observation tends to $+\infty$. Although we have indicated how the case of an observation tending to $-\infty$ could be handled and also how some results could be obtained for rapidly varying tails, the details are not explored here. Similarly, all of our asymptotics involve convergence in distribution. It does not necessarily follow, for instance, that posterior moments will converge to the moments of the limiting distribution. It would not be difficult to obtain analogous results for moments, and experience in location parameter models suggests that conditions like $\rho < \min\{c, \alpha\}$ would need to be replaced by $\rho + k < \min\{c, \alpha\}$ in order to have convergence of order-$k$ moments.

More seriously, future work is needed to extend this article in various directions — to the case of multiple outliers and groups; to other location-scale prior structures; to more complex models and to hierarchical models; and to consider alternative classes of heavy-tailed distributions.

# Appendix - Proofs

**Proof of Proposition 2.1.**

$$\begin{aligned}
f * g(x) &= \int_{-\infty}^{\infty} f(u)g(x-u)du \\
&= \int_{-\infty}^{x/2} f(u)g(x-u)du + \int_{x/2}^{\infty} f(u)g(x-u)du \\
&= \int_{-\infty}^{x/2} f(u)g(x-u)du + \int_{-\infty}^{x/2} g(u)f(x-u)du \\
&=: B(x) + C(x).
\end{aligned}$$

We can write

$$\frac{C(x)}{f(x)} = \int_{-\infty}^{\infty} h_x(u)du, \tag{12}$$

where

$$h_x(u) = g(u)\frac{f(x-u)}{f(x)}\mathbf{1}_{u<x/2}.$$

Now $f \in R_{-\rho}$, where $\rho \geq 1$ since $f \in L^1$. BGT (Theorem 1.5.6), with $A = 2$, $\delta = \rho/2$, gives the existence of $X$ such that

$$\frac{f(y)}{f(x)} \leq 2\max\{(y/x)^{-\rho/2}, (y/x)^{-3\rho/2}\} \ (x \geq X, y \geq X).$$

So for $x \geq 2X$ we find that

- for $0 < u < x/2$, $1/2 < \frac{x-u}{x} < 1$, so $\frac{f(x-u)}{f(x)} \leq 2\left(\frac{x-u}{x}\right)^{-3\rho/2} < 2^{1+3\rho/2}$, and
- for $-\infty < u \leq 0$, $1 \leq \frac{x-u}{x} < \infty$, so $\frac{f(x-u)}{f(x)} \leq 2\left(\frac{x-u}{x}\right)^{-\rho/2} < 2$.

Therefore

$$0 < h_x(u) \leq 2^{1+3\rho/2}g(u) \text{ for all } u \text{ and sufficiently large } x.$$

As $f$, $g$ are assumed bounded and positive, and as wi$g \in L^1$, we can use the Dominated Convergence Theorem to obtain the limit of the integral in (12). For each fixed $u$, $h_x(u) \to g(u)$ as $x \to \infty$, so

$$\frac{C(x)}{f(x)} \longrightarrow \int_{-\infty}^{\infty} g(u)du = 1.$$

We have also $B(x)/g(x) \to 1$ by interchange of notation. In summary $f*g(x) = B(x)+C(x)$, but we have proved that $C(x) \sim f(x)$ and $B(x) \sim g(x)$, hence $f * g(x) = B(x) + C(x) \sim$

$f(x)+g(x)$. By BGT §1.5.7$(iii)$, it follows that $f*g(x)$ is also regularly varying with index given by the maximum of the index of $f$ and $g$. $\square$

**Proof of Corollary 3.5.** *Part* $(i)$. Applying the transformation $\theta \to t = \mu/\theta$ on $\varphi$,

$$\varphi(y|\theta) = \frac{1}{\theta}\int_{\mathbb{R}} h\left(\frac{y}{\theta} - t\right) \times v(t)dt,$$

but this latest integral is clearly a convolution of the densities $h$ and $v$, thus letting

$$g\left(\frac{y}{\theta}\right) = \int_{\mathbb{R}} h\left(\frac{y}{\theta} - t\right) \times v(t)dt, \tag{13}$$

$\varphi(y|\theta) = (1/\theta)\times g(y/\theta)$. *Part* $(ii)$. Notice that, according to Proposition 2.1, the convolution $g(y/\theta) \sim h(y/\theta) + v(y/\theta)$. Hence, $g \in R_{-\rho'}$, where $\rho' = \min\{\rho, c\}$. $\square$

**Proof of Theorem 3.6.** First, the marginal posterior distribution of $\theta$ is given by

$$p(\theta|y) \propto p^*(\theta) \times \int_{\mu} \frac{1}{\theta^2} \times h\left(\frac{y-\mu}{\theta}\right) v\left(\frac{\mu}{\theta}\right) d\mu.$$

Letting $\varphi(y|\theta) = (1/\theta)\times g(y/\theta)$, where $g(y/\theta) = \int_\mu h\left((y-\mu)/\theta\right) v\left(\mu/\theta\right) d\mu/\theta$, then $g \in R_{-\rho}$ by Corollary 3.5. Thus, this becomes a pure scale model, where $\varphi(y|\theta)$ plays the role of the likelihood and, if combined with the prior distribution $p^*(\theta)$, gives

$$p(\theta|y) \longrightarrow \frac{\theta^{\rho-1}p^*(\theta)}{\int_\Theta \theta^{\rho-1}p^*(\theta)d\theta}\ (y\to\infty),$$

due to Theorem 3.1. The limit of $p(\mu|\theta, y)$ as $y \to \infty$ is $p(\mu|\theta)$ by direct application of Theorem 3.3. As a consequence

$$p(\mu,\theta|y) \longrightarrow \frac{\theta^{\rho-1}p^*(\theta)}{\int_\Theta \theta^{\rho-1}p^*(\theta)d\theta} \times p(\mu|\theta). \qquad \square$$

**Proof of Theorem 3.7.** Consider again (6) and $\varphi(y|\theta) = (1/\theta) \times g(y/\theta)$, where $g \in R_{-\min\{\rho,c\}}$ by Corollary 3.5. Thus the marginal posterior of $\theta$ is given by

$$p(\theta|y) = \frac{p^*(\theta)\times\varphi(y|\theta)}{\int_\Theta p^*(\theta)\times\varphi(y|\theta)d\theta} = \frac{p^*(\theta)\times(1/\theta)\times g(y/\theta)}{\int_\Theta p^*(\theta)\times(1/\theta)\times g(y/\theta)d\theta}. \tag{14}$$

Since $g$ has lighter tails than $p^*$, then by Corollary 3.2,

$$p(\theta|y) \sim \frac{\theta^{-(\alpha+1)} g(y/\theta)}{\int_\Theta \theta^{-(\alpha+1)} g(y/\theta) d\theta} \; (y \to \infty).$$

Therefore the asymptotic posterior distribution of $\theta/y$ has density $g^*$, where

$$g^*(t) = \frac{t^{\alpha-1} g(t^{-1})}{\int_0^\infty t^{\alpha-1} g(t^{-1}) dt} . \tag{15}$$

Therefore $\theta$ is *a posteriori* of order $y$, in the sense that with posterior probability that converges to 0.98, say, $\theta$ is between $g^*_{0.01} y$ and $g^*_{0.99} y$, where $g^*_q$ is the $q$-quantile of the distribution with density $g^*$. Now consider the conditional posterior distribution of $\mu$ given $\theta$. Note that $\theta$ is a scale parameter in both the prior conditional distribution of $\mu$ and the likelihood. There is only conflict between the prior and the data, in respect of $\mu$, if the distance between $y$ and the prior mean becomes large $r$elative to $\theta$. However, we have just seen that as $y \to \infty$ the posterior distribution of $\theta$ makes it proportional to $y$. Therefore, this conflict does not arise. As $y \to \infty$ the posterior conditional distribution of $\mu$ given $\theta$ (for all $\theta$ having non-negligible posterior probability) continues to combine both the likelihood and the prior, with neither component being rejected.

This rather heuristic explanation can be made mathematically rigorous by noting that the conditional posterior of $\mu$ given $\theta$ has the form

$$p(\mu|\theta, y) \propto p(\mu|\theta) f(y|\mu, \theta) \propto v\left(\frac{\mu}{\theta}\right) h\left(\frac{y}{\theta} - \frac{\mu}{\theta}\right) .$$

So letting $\tilde{\mu} = \mu/\theta$ and $\tilde{y} = y/\theta$ we have

$$p(\tilde{\mu}|\theta, y) \propto v(\tilde{\mu}) h(\tilde{y} - \tilde{\mu}) . \tag{16}$$

For any positive $\epsilon$ and $\delta$ we can find $Y$ such that $\forall y > Y$, $\theta \in (g^*_\epsilon, g^*_{1-\epsilon})$ with probability at least $1 - 2\epsilon - \delta$. Hence as $y \to \infty$, $\tilde{y}$ remains bounded between finite bounds with arbitrarily high posterior probability. This is the sense in which no conflict arises, and as $y \to \infty$ the distribution (16) is the asymptotic posterior distribution of $\tilde{\mu}$ given $\theta$ almost surely in $\theta$.

As a result, the asymptotic joint posterior distribution is

$$\begin{aligned}
p(\mu, \theta \,|\, y) &\sim \frac{\theta^{-(\alpha+1)} g(y/\theta)}{\int_\Theta \theta^{-(\alpha+1)} g(y/\theta) d\theta} \times \frac{f(y \,|\, \mu, \theta) p(\mu \,|\, \theta)}{\int_\mu f(y \,|\, \mu, \theta) p(\mu \,|\, \theta) d\mu} \\
&\sim \frac{\theta^{-\alpha} f(y \,|\, \mu, \theta) p(\mu \,|\, \theta)}{\int_\Theta \theta^{-(\alpha+1)} g(y/\theta) d\theta} \\
&\sim \frac{\theta^{-\alpha} f(y \,|\, \mu, \theta) p(\mu \,|\, \theta)}{\int_\Theta \int_\mu \theta^{-\alpha} f(y \,|\, \mu, \theta) p(\mu \,|\, \theta) d\mu d\theta},
\end{aligned}$$

where we used the fact that $\theta^{-1} g(y/\theta) = \psi(y|\theta) = \int_\mu f(y \,|\, \mu, \theta) p(\mu \,|\, \theta) d\mu$. □

**Proof of Theorem 3.8.** The limiting marginal distribution of $\theta$ follows as in Theorem 3.6, except that now condition $(ii)$ means that $g \in R_{-c}$ and we have

$$p(\theta|y) \longrightarrow \frac{\theta^{c-1} p^*(\theta)}{\int_\Theta \theta^{c-1} p^*(\theta) d\theta} \quad (y \to \infty).$$

The conditional posterior distribution of $\mu$ given $\theta$ now follows using Corollary 3.4, and we have

$$p(\mu|\theta, y) \sim \frac{f(y|\mu, \theta)}{\int_\mu f(y|\mu, \theta) d\mu} \quad (y \to \infty).$$

Now, combining the two limits we have

$$\begin{aligned}
p(\mu, \theta|y) &\sim \frac{\theta^{c-1} p^*(\theta)}{\int_\Theta \theta^{c-1} p^*(\theta) d\theta} \times \frac{f(y|\mu, \theta)}{\int_\mu f(y|\mu, \theta) d\mu} \\
&\sim \frac{\theta^{c-1} f(y|\mu, \theta) p^*(\theta)}{\int_\mu \int_\Theta \theta^{c-1} f(y|\mu, \theta) p^*(\theta) d\mu d\theta}, \text{ as } y \to \infty. \quad \square
\end{aligned}$$

**Proof of Lemma 3.9.** $(i)$. The density of $\mu$ given $(\boldsymbol{y}, \theta)$ is given by

$$P(\mu|\boldsymbol{y}, \theta) = \frac{f(\boldsymbol{y}|\mu, \theta) \times p(\mu|\theta)}{\int_\mu f(\boldsymbol{y}|\mu, \theta) \times p(\mu|\theta) d\mu} \tag{17}$$

$$\propto \prod_{i=1}^{n} \frac{1}{\theta} \times h_i\left(\frac{y_i - \mu}{\theta}\right) \times \frac{1}{\theta} v\left(\frac{\mu}{\theta}\right)$$

$$\propto \prod_{i=1}^{n} h_i\left(\frac{y_i - \mu}{\theta}\right) \times v\left(\frac{\mu}{\theta}\right), \tag{18}$$

but notice that $y_i$ and $\theta$ are fixed, and we think of the above expression as a function of $\mu/\theta$, hence (18) is clearly of the form (9) with $V$ as given in ($i$). To prove that $V$ has the regular variation index given in ($i$), notice the product of regularly varying functions is also regular varying with index given by the sum of the regular variation indexes of the $h_i s$ and $v$.

($ii$). The distribution of $\theta$ given $\boldsymbol{y}$ is given by

$$\begin{aligned} P^*(\theta|\boldsymbol{y}) &= \frac{f(\boldsymbol{y}|\theta) \times p^*(\theta)}{\int_\mu f(\boldsymbol{y}|\theta) \times p^*(\theta) d\theta} \quad (19) \\ &\propto p^*(\theta) \times \int_\mu f(\boldsymbol{y}|\mu, \theta) p(\mu|\theta) d\mu \\ &\propto p^*(\theta) \times \theta^{-n} \int_\mu \prod_{i=1}^n h_i \left( \frac{y_i - \mu}{\theta} \right) \times v \left( \frac{\mu}{\theta} \right) d\mu/\theta. \quad (20) \end{aligned}$$

Then, in order to see that $\int \prod_{i=1}^n h_i \left( y_i/\theta - m \right) \times v(m) dm$ is $O(1)$ in $\theta$, notice that the assumption that the $h_i$ are continuous also implies that they are bounded (since they are density functions), and hence $\prod_{i=1}^n h_i \left( (y_i - \mu)/\theta \right)$ is bounded in $\theta$. Thus there exists $M > 0$, such that $\prod_{i=1}^n h_i \left( (y_i - \mu)/\theta \right) \leq M$. But, as $v$ is integrable, by the dominated convergence theorem,

$$\int \prod_{i=1}^n h_i \left( \frac{y_i}{\theta} - m \right) \times v(m) dm \longrightarrow \int \prod_{i=1}^n h_i \left( -m \right) \times v(m) dm \ (\theta \to \infty).$$

Hence, we can replace this integral by a slowly varying function $\ell(\theta)$ so that for all $\boldsymbol{y}$

$$\begin{aligned} P^*(\theta|\boldsymbol{y}) &\propto p^*(\theta) \times \theta^{-n} \times \ell(\theta) \\ &\propto \theta^{-(\alpha+n)} \times \ell_{p^*}(\theta) \times \ell(\theta), \end{aligned}$$

where $\ell_{p^*}(\theta)$ is the slowly varying function yielded by $p^*$. Thereby, $P^*(\theta|\boldsymbol{y}) \in R_{-(\alpha+n)}$. $\square$

# References


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

Barnett, V. and Lewis, T. (1978). Outliers in statistical data. Wiley: New York.

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

Box, G. E. P. and Tiao, G. C. (1973). Bayesian inference in statistics 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. *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability*, Vol. I pp. 99-210 Univ. California Press, Berkeley, California.

Desgagné, A. and Angers, J-F (2007) Conflicting information and location parameter inference, *Metron* **65**, 67–97.

de Haan, L. (1970). On regular variation and its application to the weak convergence of sample extremes. *Mathematical Centre Tracts* **32**, Mathematisch Centrum, Amsterdam.

Evans, M. and Moshonov, H. (2006). Checking for Prior-Data Conflict. *Bayesian Analysis* **1**, No. 4, 893–914.

Feller, W. (1971). An introduction to probability and its applications. Vol. II, 2nd Ed. New York: Wiley.

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.

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

Nolan, J. P. (2010) Stable Distributions - Models for Heavy Tailed Data, *Note: In progress, Chapter 1 online at academic2.american.edu/∼jpnolan*, Boston: Birkhäuser.

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 and 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. (2003). HSSS model criticism (with discussion). In *Highly Structured Stochastic Systems*, P. J. Green, N. L. Hjort and S. T. Richardson (eds), 423-453 Oxford University Press.

O'Hagan, A. and 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 and A. F. M. Smith), New York: Wiley, 311–327.

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

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

Pitt, H. R. (1958). Tauberian theorems. London: Oxford University Press.

Resnick, S. (1987). Extreme values, regular variation and point process. New York: Springer-Verlag.

Seneta, E. (1976). Functions of regular variation. *Lecture Notes in Mathematics Vol.* **506**, New York: Springer.