Bayesian Analysis (2013) **8**, Number 1, pp. 187–220

Full Robustness in Bayesian Modelling of a Scale Parameter

Alain Desgagné *

Abstract. Conflicting information, arising from prior misspecification or outlying observations, may contaminate the posterior inference in Bayesian modelling. The use of densities with sufficiently heavy tails usually leads to robust posterior inference, as the influence of the conflicting information decreases with the importance of the conflict. In this paper, we study full robustness in Bayesian modelling of a scale parameter. The log-slowly, log-regularly and log-exponentially varying functions as well as log-exponential credence (LE-credence) are introduced in order to characterize the tail behaviour of a density. The asymptotic behaviour of the marginal and the posterior is described and we find that the scale parameter given the complete information converges in distribution to the scale given the non-conflicting information, as the conflicting values (outliers and/or prior's scale) tend to 0 or ∞, at any given rate. We propose a new family of densities defined on $\mathbb{R}$ with a large spectrum of tail behaviours, called generalized exponential power of the second form (GEP_2), and its exponential transformation defined on $(0, \infty)$, called log-GEP_2, which proves to be helpful for robust modelling. Practical considerations are addressed through a case of combination of experts' opinions, where non-robust and robust models are compared.

Keywords: Bayesian robustness, conflicting information, log-exponentially varying functions, log-regularly varying functions, log-slowly varying functions, LE-credence, Log-GEP_2

1 Introduction

In Bayesian analysis, conflicting information may contaminate the posterior inference. Conflict can arise from outlying observations as well as prior misspecification. The outcome depends on the relative tail behaviour of the involved densities. The conflict is usually resolved by modelling the densities with sufficiently heavy tails.

Outlier rejection in Bayesian analysis was first described by De Finetti (1961), where the simplest case with a single observation having mean θ was considered. The theory has mostly evolved for location parameter inference, see for instance Dawid (1973), O'Hagan (1979, 1988, 1990), Angers (2000) and Desgagné and Angers (2007).

Robustness for scale parameter inference was first considered by Andrade and O'Hagan (2006). They show that modelling with regularly varying densities leads to partial robust inference, in the sense that the influence of conflicting information is limited. They first consider the simple case of one observation combined with the prior. It is

*Département de mathématiques, Université du Québec à Montréal, Montréal, Québec, Canada desgagne.alain@uqam.ca

 DOI:10.1214/13-BA808

then generalized to multiple observations when one group of outliers moves to infinity as a block. Andrade and O'Hagan (2011) also consider partial robustness for the location-scale parameter inference. A good description of the existing literature is given in this paper.

In our paper, we study robustness in Bayesian modelling of a scale parameter. In contrast to Andrade and O'Hagan (2006), we give some conditions to obtain full robustness, in the sense that the information provided by the conflicting densities is completely rejected in the posterior inference. Our context is also more general, as we consider multiple observations with many conflicting sources (prior and/or observations) that move to 0 or ∞ at any given rate.

In Section 2, we define three classes of functions called log-slowly, log-regularly and log-exponentially varying functions. Analogously to the regularly varying functions used in Andrade and O'Hagan (2006), these classes describe the tail behaviour of a function. We find that modelling densities with log-exponentially varying functions may lead to full Bayesian robustness in a scale parameter inference. In connection with these classes, we also define log-exponential credence (LE-credence) as a measure of tail behaviour, which proves to be useful to characterize and order different tails.

In Section 3, the resolution of conflicts in a scale parameter model is analysed. We start with the description of the mathematical context in Section 3.1. The Bayesian model consists of positive conditionally independent random variables $X_i \mid \sigma$ sharing the same scale parameter σ. We also detail the meaning of robustness in our context, especially its asymptotic nature.

In Section 3.2, the main result of this paper is given. The asymptotic behaviour of the marginal and the posterior is described and we find that the random variable σ given the complete information converges in distribution to the random variable σ given the non-conflicting information, as the conflicting values (outliers and/or prior) tend to 0 or ∞, at any given rate. Conditions for the achievement of this full robustness are given. They concern the tail behaviour of the prior and the likelihood.

In Section 3.3, we derive two special cases. We analyse the classical case of a single observation combined with the prior. We see that a conflict can be resolved either in favour of the prior or the observation, depending on their relative tail behaviour. A useful case in practice, where the tail behaviour is the same for all densities, is also studied. We see that full robustness is achieved if the non-conflicting values exceed the conflicting values.

In Section 4.1, we propose the exponential transformation of densities defined on the real line as a method to devise appropriate densities defined on $(0, \infty)$ to achieve robustness in our context. Using this approach, we also propose in Section 4.2 a new family of densities defined on $\mathbb{R}$, called generalized exponential power of the second form (GEP_2), and its exponential transformation defined on $(0, \infty)$, called log-GEP_2. The large spectrum of tail behaviours of the log-GEP_2 density is useful for robust modelling. Special cases of this density are given in Section 4.3, such as the well-known log-normal.

An example is given in Section 5, where non-robust and robust models are compared. Practical considerations are addressed through a case of combination of experts' opinions. Finally we conclude in Section 6 and the proofs are given in Section 7.

2 Log-exponentially varying functions

2.1 Definitions

Tail behaviour is a key component for robust Bayesian modelling. Therefore, we introduce three classes of functions called log-slowly, log-regularly and log-exponentially varying functions, following the idea of regularly varying functions developed by Karamata (1930).

We assume that every function is defined on $(0, \infty)$, continuous and strictly positive. Definitions and results are given for the asymptotic behaviour at ∞, with the differences for the asymptotic behaviour at 0 put in square brackets.

We say that the functions $f(z)$ and $g(z)$ are *asymptotically equivalent* at $\infty[0]$, written $f(z) \sim g(z)$ as $z \to \infty[0]$, if

$$f(z)/g(z) \to 1 \text{ as } z \to \infty[0].$$

Definition 1 (Log-slowly varying function)**.** *We say that a measurable function g is log-slowly varying at $\infty[0]$, written $g \in L_0(\infty)[L_0(0)]$, if for all $\nu > 0$, we have*

$$g(z^\nu) \sim g(z) \text{ as } z \to \infty[0].$$

For instance, the functions 1 and $\log|\log z|$ are both log-slowly varying at 0 and ∞.

Definition 2 (Log-regularly varying function)**.** *We say that a measurable function g is log-regularly varying at $\infty[0]$ with index $\rho \in \mathbb{R}$, written $g \in L_\rho(\infty)[L_\rho(0)]$, if for all $\nu > 0$, we have*

$$g(z^\nu)/g(z) \to \nu^{-\rho} \text{ as } z \to \infty[0],$$

or equivalently (for $\rho \neq 0$), if there exists a constant $A > 1$ such that for $z \geqslant A[z \leqslant 1/A]$, g can be written as

$$g(z) = |\log z|^{-\rho} S(z), \text{ where } S \in L_0(\infty)[L_0(0)].$$

Definition 3 (Log-exponentially varying function)**.** *We say that a measurable function g is log-exponentially varying at $\infty[0]$ with index (γ, δ, α), written $g \in L_{\gamma,\delta,\alpha}(\infty)$ $[L_{\gamma,\delta,\alpha}(0)]$, if there exists a constant $A > 1$ such that for $z > A[z < 1/A]$, g can be written as*

$$g(z) = \exp(-\delta|\log z|^\gamma)|\log z|^{-\alpha} S(z),$$

where $S \in L_0(\infty)[L_0(0)]$, $\gamma \geqslant 0, \delta \geqslant 0, \alpha \in \mathbb{R}$.

By convention, we set $\gamma = 0$ if and only if $\delta = 0$. The class of log-exponentially varying functions includes the class of log-regularly varying functions (when $\gamma = \delta = 0$ and $\alpha = \rho$), which includes itself the class of log-slowly varying functions (when $\rho = 0$). Notice that $g(z) \in L_{\gamma,\delta,\alpha}(0)$ if and only if $g(1/z) \in L_{\gamma,\delta,\alpha}(\infty)$.

Definition 4 (LE-credence)**.** *We say that the right [left] LE-credence of a measurable function g is (γ, δ, α) if $g \in L_{\gamma,\delta,\alpha}(\infty)[L_{\gamma,\delta,\alpha}(0)]$.*

The concept of LE-credence has interesting properties and interpretation regarding the tails of a function, as described in Section 2.2. We can also define the ordering of LE-credence as follows.

Definition 5 (Ordering of LE-credence)**.** *For two given LE-credences denoted by $(\gamma_1, \delta_1, \alpha_1)$ and $(\gamma_2, \delta_2, \alpha_2)$, we write*

i) $(\gamma_1, \delta_1, \alpha_1) = (\gamma_2, \delta_2, \alpha_2)$ *if* $\gamma_1 = \gamma_2$, $\delta_1 = \delta_2$, $\alpha_1 = \alpha_2$ *and we say that the LE-credences are equal,*

ii) $(\gamma_1, \delta_1, \alpha_1) > (\gamma_2, \delta_2, \alpha_2)$ *if* $\gamma_1 > \gamma_2$ *or* $\gamma_1 = \gamma_2$, $\delta_1 > \delta_2$ *or* $\gamma_1 = \gamma_2$, $\delta_1 = \delta_2$, $\alpha_1 > \alpha_2$ *and we say that* $(\gamma_1, \delta_1, \alpha_1)$ *is larger than* $(\gamma_2, \delta_2, \alpha_2)$.

2.2 Properties

For two functions g_1 and g_2 that are log-exponentially varying at $\infty[0]$, we can order the asymptotic behaviour of their tails, using LE-credence, as follows.

Proposition 1 (Ordering of tails)**.** *If $g_1 \in L_{\gamma_1,\delta_1,\alpha_1}(\infty[0])$ and $g_2 \in L_{\gamma_2,\delta_2,\alpha_2}(\infty[0])$, then*

$$(\gamma_1, \delta_1, \alpha_1) > (\gamma_2, \delta_2, \alpha_2) \text{ implies that } g_1(z)/g_2(z) \to 0 \text{ as } z \to \infty[0].$$

If $(\gamma_1, \delta_1, \alpha_1) = (\gamma_2, \delta_2, \alpha_2)$, then the ratio of the log-slowly varying part of g_1 and g_2 determines the tails dominance.

Proof. It is omitted since it is mostly algebraic.

The terminology LE-credence follows the idea of credence defined in O'Hagan (1990), which describes essentially the tails that behave like $|z|^{-\beta}$. Analogously, LE-credence describes the tails that behave like the function $\exp(-\delta|\log z|^{\gamma})|\log z|^{-\alpha}S(z)$. LE-credence can be interpreted as a measure of the tail's thickness. According to Proposition 1, the function with the smallest LE-credence has the heaviest tail. For instance, if $f(z)$ is the density of a log-normal distribution with parameter $\tau^2 = 1/2$ (see Section 4.3), it can be verified that the left and right LE-credences of $zf(z)$ are given by $(2, 1, 0)$. If $f(z)$ is the density of a log-double-Pareto distribution with parameter $\lambda = 1$ (see Section 4.3), it can be verified that the left and right LE-credences of $zf(z)$ are given by $(0, 0, 2)$. According to Definition 5, we have $(2, 1, 0) > (0, 0, 2)$, which means that the tails of the log-double-Pareto are heavier than those of the log-normal. As it will be seen in Section 3, in case of conflict between two sources of information, the source with the largest LE-credence will be preferred to the other one (with additional conditions). LE-credence is therefore a measure of confidence in a source of information in case of conflict. In our example, the source modelled using the log-normal would be preferred in case of conflict.

LE-credence is also useful to determine if a log-exponentially varying function is integrable, as described in the next proposition.

Proposition 2 (Integrability)**.** *For a function $zf(z) \in L_{\gamma,\delta,\alpha}(\infty[0])$, there exists a constant $A > 1$ such that $f(z)$ is integrable on $z > A[z < 1/A]$, if*

i) $(\gamma, \delta, \alpha) > (0, 0, 1)$*,* ***ii)*** $(\gamma, \delta, \alpha) = (0, 0, 1)$*, with the log-slowly varying part of $zf(z)$ having a decay sufficiently fast (e.g. $(\log|\log z|)^{-\beta}$, with $\beta > 1$).*

Proof. It is omitted since it is mostly algebraic.

In particular, if f is a probability density function, we know from Proposition 2 that LE-credences larger than $(0, 0, 1)$ in both tails of $zf(z)$ are sufficient to guarantee that f is proper.

An important characteristic of a log-exponentially varying function with LE-credence sufficiently small is its asymptotic scale invariance, as shown in the next proposition.

Proposition 3 (Scale invariance)**.** *If $g \in L_{\gamma,\delta,\alpha}(\infty[0])$ and $\gamma < 1$, then we have, for $\sigma > 0$,*

$$g(\sigma z) \sim g(z) \text{ as } z \to \infty[0].$$

Or equivalently, if $zf(z) \in L_{\gamma,\delta,\alpha}(\infty[0])$ and $\gamma < 1$, then we have, for $\sigma > 0$,

$$(1/\sigma)f(z/\sigma) \sim f(z) \text{ as } z \to \infty[0].$$

Proof. See Section 7.1.

In particular, if f is a probability density function, we know from Proposition 3 that $f(z)$ is scale invariant as $z \to \infty[0]$ if the first term (γ) of the right [left] LE-credence of $zf(z)$ is smaller than 1.

In the next proposition, we give the asymptotic behaviour, as $z \to \infty[0]$, of the density (evaluated at z) of the product of two independent and identically distributed random variables with density f. Though the connection is not apparent yet, this result is essential for the proof of our results of robustness given in Section 3.

Proposition 4 (Product of random variables)**.** *If f is a proper density, such that $zf(z)$ has a right [left] monotonic tail and $zf(z) \in L_{\gamma,\delta,\alpha}(\infty[0])$ with $\gamma < 1$, then*

$$\textbf{\textit{i)}} \int_0^\infty (1/\sigma)f(z/\sigma)f(\sigma)\,d\sigma \sim 2f(z) \text{ as } z \to \infty[0],$$

and

$$\textbf{\textit{ii)}} \sup_{\sigma>0} f(z/\sigma)f(\sigma)/f(z) \to \sup_{\sigma>0} \sigma f(\sigma) \text{ as } z \to \infty[0].$$

Proof. See Section 7.2.

Note that ii) is still valid if we relax the condition that f needs to be a proper density to the condition that $\sigma f(\sigma)$ is bounded on $\sigma > 0$, to allow for the cases $\gamma = \delta = 0$ and $0 \leqslant \alpha \leqslant 1$. Only a minor and trivial change has to be done in the proof.

Notice that for a proper and continuous density f defined on $(0, \infty)$, it is easy to show that the function $\sigma f(\sigma)$ is bounded (which is not necessarily the case for $f(\sigma)$ as $\sigma \to 0$), with $\sigma f(\sigma) \to 0$ as $\sigma \to 0$ or $\sigma \to \infty$.

By monotonicity of the right [left] tail of $zf(z)$, it is meant that

$$y \geqslant z\,[y \leqslant z] \text{ implies that } yf(y) \leqslant zf(z) \text{ as } z \to \infty[0]. \tag{1}$$

3 Resolution of conflicts in a scale parameter model

3.1 The scale parameter model and notations

The Bayesian structure is given as follows.

i) Let $X_1, \ldots, X_n$ be n random variables conditionally independent given σ with their conditional densities given by $X_i \mid \sigma \overset{\mathcal{D}}{\sim} (1/\sigma)f_i(x_i/\sigma)$,

ii) the prior density of σ is given by $\sigma \mid \phi \overset{\mathcal{D}}{\sim} (1/\phi)\pi(\sigma/\phi)$,

where $x_1, \ldots, x_n$, σ, $\phi > 0$.

As in Section 2, we assume that $f_1, \ldots, f_n$ and π are defined on $(0, \infty)$, continuous and strictly positive. In addition, we assume that each density is proper. The only exception is the possibility for the prior to be modelled by a non-informative (hence considered as non-conflicting) improper density (such as $\pi(\sigma) \propto 1/\sigma$), as long as $\sigma\pi(\sigma)$ is bounded on $\sigma > 0$.

It necessarily means that the functions $zf_1(z), \ldots, zf_n(z)$ and $z\pi(z)$ are bounded on $z > 0$, with a limit of 0 in their tails as $z \to 0$ and $z \to \infty$ (except for an improper prior). Notice that the limit of the densities $f_1(z), \ldots, f_n(z), \pi(z)$ at 0 can be anything ranging from 0 to infinity. We also assume that both tails of $zf_1(z), \ldots, zf_n(z)$ and $z\pi(z)$ are monotonic, as defined in (1).

The scale parameter of the prior, ϕ, is considered known. The parameter ϕ represents the information provided by the prior, and in this sense plays in our inference the same role as the observations $x_1, \ldots, x_n$. Any other hyperparameters are assumed to be known and are implicitly included in the densities.

We study robustness of the inference on σ in the presence of extreme observations x_i and/or misspecification of the prior's scale ϕ. The nature of the results is asymptotic, in the sense that some x_i and/or ϕ are going to 0 or ∞. We find conditions on the densities to obtain full robustness, given by the complete rejection of the information provided by the conflicting values.

Among the n observations, denoted by $\mathbf{x_n} = (x_1, \ldots, x_n)$, we assume that k of them, denoted by the vector $\mathbf{x_k}$, form a group of non-outlying observations or fixed values. We assume that l of them, denoted by the vector $\mathbf{x_l}$, are considered as left outliers (smaller than the fixed values) and r of them, denoted by the vector $\mathbf{x_r}$, are considered as right outliers (larger than the fixed values), with $k+l+r=n$. Similarly, the prior scale ϕ can be considered as a fixed value, left outlier or right outlier.

For $i=0,1,\ldots,n$, we define three binary functions k_i, l_i and r_i as follows. If x_i is a fixed value, we set $k_i=1$; if it is a left outlier, we set $l_i=1$; and if it is a right outlier, we set $r_i=1$. Similarly, if ϕ is a fixed value, we set $k_0=1$; if it is a left outlier, we set $l_0=1$; and if it is a right outlier, we set $r_0=1$. These functions are set to 0 otherwise. We have $k_i+l_i+r_i=1$ for $i=0,1,\ldots,n$, with $\sum_{i=1}^n k_i=k$, $\sum_{i=1}^n l_i=l$ and $\sum_{i=1}^n r_i=r$. If the prior is a non-informative improper density, ϕ is treated as a fixed value, with $k_0=1$.

We assume that each conflicting value (outlier/prior) is going to 0 or ∞ at any given rate. They can move as a block, as assumed in Andrade and O'Hagan (2006), but they are not restricted to any patterns. The conflicting values can therefore be combined in one variable ω, defined as follows:

i) $\omega=\min(1/\phi,1/\mathbf{x_l},\mathbf{x_r})$ if $l_0=1$, **ii)** $\omega=\min(1/\mathbf{x_l},\mathbf{x_r})$ if $k_0=1$,
iii) $\omega=\min(\phi,1/\mathbf{x_l},\mathbf{x_r})$ if $r_0=1$.

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

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

$$\pi(\sigma\mid\mathbf{x_n},\phi)=m(\mathbf{x_n}\mid\phi)^{-1}(1/\phi)\pi(\sigma/\phi)\prod_{i=1}^n(1/\sigma)f_i(x_i/\sigma).$$

Let the posterior density of σ considering only the fixed values $\mathbf{x_k}$ and ϕ^{k_0} ($\phi^{k_0}=\phi$ if $k_0=1$; $\phi^{k_0}=1$ if $k_0=0$) be denoted by $\pi(\sigma\mid\mathbf{x_k},\phi^{k_0})$ and its corresponding marginal density be denoted by $m(\mathbf{x_k}\mid\phi^{k_0})$, with

$$\pi(\sigma\mid\mathbf{x_k},\phi^{k_0})=m(\mathbf{x_k}\mid\phi^{k_0})^{-1}(1/\sigma)((\sigma/\phi)\pi(\sigma/\phi))^{k_0}\prod_{i=1}^n((1/\sigma)f_i(x_i/\sigma))^{k_i}. \quad (2)$$

In this case, the prior density is $(1/\sigma)\times((\sigma/\phi)\pi(\sigma/\phi))^{k_0}$, or equivalently $(1/\phi)\pi(\sigma/\phi)$ if $k_0=1$ or $1/\sigma$ if $k_0=0$.

Similarly, we define the posterior density of σ considering only the left conflicting values $\mathbf{x_l}$ and ϕ^{l_0} and that considering only the right conflicting values $\mathbf{x_r}$ and ϕ^{r_0} as follows:

$$\pi(\sigma\mid\mathbf{x_l},\phi^{l_0})\propto(1/\sigma)((\sigma/\phi)\pi(\sigma/\phi))^{l_0}\prod_{i=1}^n((1/\sigma)f_i(x_i/\sigma))^{l_i} \text{ and} \quad (3)$$

$$\pi(\sigma\mid\mathbf{x_r},\phi^{r_0})\propto(1/\sigma)((\sigma/\phi)\pi(\sigma/\phi))^{r_0}\prod_{i=1}^n((1/\sigma)f_i(x_i/\sigma))^{r_i}. \quad (4)$$

Furthermore, we have

$$\sigma\pi(\sigma\mid\mathbf{x_n},\phi)\propto\sigma\pi(\sigma\mid\mathbf{x_k},\phi^{k_0})\times\sigma\pi(\sigma\mid\mathbf{x_l},\phi^{l_0})\times\sigma\pi(\sigma\mid\mathbf{x_r},\phi^{r_0}).$$

It is interesting to notice that the posterior considering the whole information can be seen as a combination of three posterior densities considering the fixed values, the left and the right conflicting values.

3.2 Resolution of conflicts

Using the Bayesian context described in Section 3.1, the main theorem of this paper is now presented.

Theorem 1 (Robustness)**.** *If the following conditions are satisfied:*

i)

$$z\pi(z) \in \begin{cases} L_{\gamma_0,\delta_0,\alpha_0}(\infty) \text{ with } \gamma_0 < 1; & \text{if } l_0 = 1; \\ L_{\gamma_0',\delta_0',\alpha_0'}(0) \text{ with } \gamma_0' < 1; & \text{if } r_0 = 1, \end{cases}$$

ii) *for* $i = 1, \ldots, n$,

$$zf_i(z) \in \begin{cases} L_{\gamma_i,\delta_i,\alpha_i}(\infty) \text{ with } \gamma_i < 1; & \text{if } r_i = 1; \\ L_{\gamma_i',\delta_i',\alpha_i'}(0) \text{ with } \gamma_i' < 1; & \text{if } l_i = 1, \end{cases}$$

iii)

$$\frac{(z\pi(z))^{k_0} \prod_{i=1}^n ((1/z)f_i(1/z))^{k_i}}{((1/z)\pi(1/z))^{l_0} \prod_{i=1}^n (zf_i(z))^{l_i}} \to 0 \text{ as } z \to 0,$$

iv)

$$\frac{(z\pi(z))^{k_0} \prod_{i=1}^n ((1/z)f_i(1/z))^{k_i}}{((1/z)\pi(1/z))^{r_0} \prod_{i=1}^n (zf_i(z))^{r_i}} \to 0 \text{ as } z \to \infty,$$

then we have the following results:

a) $m(\mathbf{x_n} \mid \phi) \sim m(\mathbf{x_k} \mid \phi^{k_0}) \left((1/\phi)\pi(1/\phi)\right)^{l_0+r_0} \prod_{i=1}^n f_i(x_i)^{l_i+r_i}$ *as* $\omega \to \infty$,

b) $\pi(\sigma \mid \mathbf{x_n}, \phi) \to \pi(\sigma \mid \mathbf{x_k}, \phi^{k_0})$, $\sigma > 0$, *as* $\omega \to \infty$,

c) $\sigma\pi(\sigma \mid \mathbf{x_n}, \phi) \to 0$, *as* $|\log \sigma| \to \infty$, $\omega \to \infty$,

d) $\sigma \mid \mathbf{x_n}, \phi \xrightarrow{\mathcal{D}} \sigma \mid \mathbf{x_k}, \phi^{k_0}$ *as* $\omega \to \infty$.

Proof. See Section 7.3.

The four conditions involve only the tails of the densities. More precisely, the right conflicting values $(\mathbf{x_r}, \phi^{r_0})$ imply the right tail of $z\pi(z)$ if ϕ is fixed and its left tail if $\phi \to \infty$, and the left tail of $zf_i(z)$ if x_i is fixed or its right tail if $x_i \to \infty$, $i = 1 \ldots, n$. The left conflicting values $(\mathbf{x_l}, \phi^{l_0})$ imply the opposite tails.

In conditions i) and ii) we require, for the concerned tails of the conflicting densities, that $z\pi(z)$ and $zf_i(z)$ are log-exponentially varying with $0 \leqslant \gamma < 1$, which ensures that these tails are sufficiently heavy. Notice that there are no such contraints on the densities of the fixed values; we could for instance choose densities with lighter tails.

Using equations (2), (3) and (4), conditions iii) and iv) can be written respectively as

$$\frac{\sigma\pi(\sigma \mid \mathbf{x_k}, \phi^{k_0} = \mathbf{1})}{(1/\sigma)\pi(1/\sigma \mid \mathbf{x_l}, \phi^{l_0} = \mathbf{1})} \to 0 \text{ as } \sigma \to 0, \text{ and}$$
$$\frac{\sigma\pi(\sigma \mid \mathbf{x_k}, \phi^{k_0} = \mathbf{1})}{(1/\sigma)\pi(1/\sigma \mid \mathbf{x_r}, \phi^{r_0} = \mathbf{1})} \to 0 \text{ as } \sigma \to \infty,$$

where $\mathbf{1}$ is a vector of 1. The interpretation of conditions iii) and iv) is easier with this writing. The left tail of $\sigma\pi(\sigma \mid \mathbf{x_k}, \phi^{k_0} = \mathbf{1})$ must be lighter than the right tail of $\sigma\pi(\sigma \mid \mathbf{x_l}, \phi^{l_0} = \mathbf{1})$ and the right tail of $\sigma\pi(\sigma \mid \mathbf{x_k}, \phi^{k_0} = \mathbf{1})$ must be lighter than the left tail of $\sigma\pi(\sigma \mid \mathbf{x_r}, \phi^{r_0} = \mathbf{1})$.

Notice that this interpretation may suggest that the left or right conflicting values are treated as a block, even if we specified earlier that they go to 0 or ∞ at any given rate. However there is no contradiction since rejecting conflicting values when they move as a block requires the strongest conditions.

The asymptotic behaviour of the marginal and the posterior are given respectively in results a) and b). For any fixed value $\sigma > 0$, the posterior considering the entire information behaves as the posterior considering only the non-conflicting values. The conflicting information is then completely rejected. The convergence of $\sigma\pi(\sigma \mid \mathbf{x_n}, \phi)$ to 0 is established in result c), as $\sigma, \omega \to \infty$ or $1/\sigma, \omega \to \infty$ at any given rate. It means in particular that $\sigma\pi(\sigma \mid \mathbf{x_n}, \phi)$ converges to 0 for any area around the conflicting values, and that an eventual mode at these values will also decrease to 0. Notice that $\sigma\pi(\sigma \mid \mathbf{x_n}, \phi)/(\sigma\pi(\sigma \mid \mathbf{x_k}, \phi^{k_0}))$, as $\omega \to \infty$ and $|\log\sigma| \to \infty$, has a form of 0/0 and its limit can be anywhere between 0 and ∞, depending on the relation between σ and ω. In result d), the convergence in distribution is understood as

$$\Pr[\sigma \leqslant d \mid \mathbf{x_n}, \phi] \to \Pr[\sigma \leqslant d \mid \mathbf{x_k}, \phi^{k_0}], \text{ for any } d > 0, \text{ as } \omega \to \infty.$$

Therefore any estimators of σ based on quantiles of the posterior density are robust to outliers and misspecified priors.

3.3 Special cases

A single observation

Two interesting special cases are described in this section. We first study the inference with a sample of a single observation. While not realistic in practice, it gives some intuition of how the robustness works. The Bayesian structure is simplified as follows:

i) let X_1 be a random variable with its conditional density given by
$X_1 \mid \sigma \overset{\mathcal{D}}{\sim} (1/\sigma)f_1(x_1/\sigma)$,

ii) the prior density of σ is given by $\sigma \mid \phi \overset{\mathcal{D}}{\sim} (1/\phi)\pi(\sigma/\phi)$,

where x_1, σ, $\phi > 0$. We assume that the densities π and f_1 are proper. We must decide which source of information we trust in case of conflict, between the prior and the observation.

Consider first that we choose to resolve an eventual conflict in favour of the prior, or stated another way, the conflicting value would be the observation x_1. Starting from the general context, we set $n = 1$, $k = 0$, $l + r = 1$, $k_0 = 1$, $l_0 = r_0 = 0$, $k_1 = 0$, $l_1 + r_1 = 1$, $\mathbf{x_n} = x_1$, $\mathbf{x_k} = \varnothing$, $\omega = 1/x_1$ if $l_1 = 1$ and $\omega = x_1$ if $r_1 = 1$. We can verify that

$$m(\mathbf{x_k} \mid \phi^{k_0}) = 1 \text{ and } \pi(\sigma \mid \mathbf{x_k}, \phi^{k_0}) = (1/\phi)\pi(\sigma/\phi).$$

Theorem 1 is then simplified as follows:

If the following conditions are satisfied:

$$zf_1(z) \in \begin{cases} L_{\gamma_1,\delta_1,\alpha_1}(\infty) \text{ with } \gamma_1 < 1; & \text{if } r_1 = 1; \\ L_{\gamma_1',\delta_1',\alpha_1'}(0) \text{ with } \gamma_1' < 1; & \text{if } l_1 = 1, \end{cases}$$

and

$$\frac{z\pi(z)}{(zf_1(z))^{l_1}} \to 0 \text{ as } z \to 0 \text{ and } \frac{z\pi(z)}{(zf_1(z))^{r_1}} \to 0 \text{ as } z \to \infty,$$

then we have the following results:

a) $m(x_1 \mid \phi) \sim f_1(x_1)$ as $\omega \to \infty$,

b) $\pi(\sigma \mid x_1, \phi) \to (1/\phi)\pi(\sigma/\phi)$, $\sigma > 0$, as $\omega \to \infty$,

c) $\sigma\pi(\sigma \mid x_1, \phi) \to 0$, as $|\log \sigma| \to \infty$, $\omega \to \infty$,

d) $\sigma \mid x_1, \phi \xrightarrow{\mathcal{D}} \sigma \mid \phi$ as $\omega \to \infty$.

In practice, we may wish to be protected against a potential outlier in any direction, that is too small or too large. It suffices then to satisfy the conditions for both $l_1 = 1$ and $r_1 = 1$. We choose f such that $zf_1(z)$ is log-exponentially varying at 0 and ∞ (not necessarily with the same LE-credence) with $\gamma_1, \gamma_1' < 1$. And we choose π such that $\pi(z)/f_1(z) \to 0$, as $z \to 0$ and $z \to \infty$, that is the left and right tails of π are respectively lighter than the left and right tails of f_1.

It follows that the marginal of X_1 behaves asymptotically as $f_1(x_1)$ and the posterior behaves as the prior, the source that we decided to trust in case of conflict. The outlier is completely rejected.

Consider now that we choose to resolve an eventual conflict in favour of the observation, or stated another way, the conflicting value would be the prior's scale ϕ.

In this case, we set $n = 1$, $k = 1$, $l + r = 0$, $k_0 = 0$, $l_0 + r_0 = 1$, $k_1 = 1$, $l_1 = r_1 = 0$, $\mathbf{x_n} = \mathbf{x_k} = x_1$, $\omega = 1/\phi$ if $l_0 = 1$ and $\omega = \phi$ if $r_0 = 1$. We can verify that

$$m(\mathbf{x_k} \mid \phi^{k_0}) = 1/x_1 \text{ and } \pi(\sigma \mid \mathbf{x_k}, \phi^{k_0}) = (1/\sigma)(x_1/\sigma) f_1(x_1/\sigma).$$

Theorem 1 is then simplified as follows:

If the following conditions are satisfied:

$$z\pi(z) \in \begin{cases} L_{\gamma_0,\delta_0,\alpha_0}(\infty) \text{ with } \gamma_0 < 1; & \text{if } l_0 = 1; \\ L_{\gamma_0',\delta_0',\alpha_0'}(0) \text{ with } \gamma_0' < 1; & \text{if } r_0 = 1, \end{cases}$$

and

$$\frac{z f_1(z)}{(z\pi(z))^{l_0}} \to 0 \text{ as } z \to \infty \text{ and } \frac{z f_1(z)}{(z\pi(z))^{r_0}} \to 0 \text{ as } z \to 0,$$

then we have the following results:

a) $m(x_1 \mid \phi) \sim (1/x_1)(1/\phi)\pi(1/\phi)$ as $\omega \to \infty$,

b) $\pi(\sigma \mid x_1, \phi) \to (1/\sigma)(x_1/\sigma) f_1(x_1/\sigma)$, $\sigma > 0$, as $\omega \to \infty$,

c) $\sigma\pi(\sigma \mid x_1, \phi) \to 0$, as $|\log \sigma| \to \infty$, $\omega \to \infty$,

d) $\sigma \mid x_1, \phi \xrightarrow{\mathcal{D}} \sigma \mid x_1$ as $\omega \to \infty$.

In practice, we may wish to be protected against a potential misspecification of the prior's scale in any direction, that is too small or too large. It suffices then to satisfy the conditions for both $l_0 = 1$ and $r_0 = 1$. We choose π such that $z\pi(z)$ is log-exponentially varying at 0 and ∞ (not necessarily with the same LE-credence) with $\gamma_0, \gamma_0' < 1$. And we choose f_1 such that $f_1(z)/\pi(z) \to 0$, as $z \to 0$ and $z \to \infty$, that is the left and right tails of f_1 are respectively lighter than the left and right tails of π.

It follows that the product of x_1 with the marginal of X_1 behaves asymptotically as $(1/\phi)\pi(1/\phi)$ and the posterior behaves as $(1/\sigma)(x_1/\sigma)f_1(x_1/\sigma)$. The latter is based exclusively on the information provided by the likelihood, that is the observation x_1 and the density f_1, which is the source that we decided to trust in case of conflict. The information provided by the prior $(1/\phi)\pi(\sigma/\phi)$, that is the scale parameter ϕ and the density π, is completely ignored. Notice that the limiting posterior can be interpreted as the posterior density given x_1, where $1/\sigma$ plays the role of the prior. It is interesting to observe that even if the chosen prior in the model is completely rejected, it still remains the non-informative prior $1/\sigma$, inherent in the Bayesian structure. In result d), the random variable $\sigma \mid x_1$ has the density given in result b).

The same tail behaviour for each source of information

A simple and practical model consists in choosing the same tail behaviour for each tail of each density, that is

$$z\pi(z) \sim zf_1(z) \sim \cdots \sim zf_n(z), \text{ as } z \to 0 \text{ or } z \to \infty.$$

Notice that the shape of each density may be distinct since the conditions involve only the asymptotic tail behaviour. In case of conflict, we choose to give the same weight to each source of information.

Since $z\pi(z) \sim (1/z)\pi(1/z)$ and $zf_i(z) \sim (1/z)f_i(1/z)$, conditions iii) and iv) are simplified as follows:

$$(zf_1(z))^{(k_0+k)-(l_0+l)} \to 0 \text{ and } (zf_1(z))^{(k_0+k)-(r_0+r)} \to 0 \text{ as } z \to \infty,$$

which is equivalent to $k_0+k > \max(l_0+l, r_0+r)$ since $zf_1(z) \to 0$ as $z \to \infty$. Theorem 1 is then simplified as follows:

If for $i = 1, \ldots, n$, we have

$$z\pi(z), zf_i(z) \in L_{\gamma,\delta,\alpha}(0) \text{ and } z\pi(z), zf_i(z) \in L_{\gamma,\delta,\alpha}(\infty), \text{ with } \gamma < 1,$$

and

$$k_0 + k > \max(l_0 + l, r_0 + r),$$

then results a) to d) follow.

For instance, there are up to 10 conflicting values at left and 10 others at right that could be rejected with a group of 11 fixed or non-conflicting values.

If all the conflicting values are on the same side, we have

$$\max(l_0 + l, r_0 + r) = (n + 1) - (k_0 + k),$$

and

$$k_0 + k > \max(l_0 + l, r_0 + r) \Leftrightarrow k_0 + k > (n + 1)/2,$$

that is the conflicting values will be completely rejected if they are less numerous than the fixed values, in other words, if the non-conflicting values represent more than half of the sources of information.

4 Log-exponentially varying densities

An important condition of robustness, given in Theorem 1, requires choosing log-exponentially varying densities in our modelling, with tails sufficiently heavy. However, for most of the known densities f defined on $(0, \infty)$, the function $zf(z)$ is not log-exponentially varying, and if it is the case, its tails are not sufficiently heavy for robustness.

For instance, if a gamma density is given by

$$f(z) = e^{-z} z^{\delta-1} \Gamma(\delta)^{-1}, \tag{5}$$

where $\delta > 0$ and $\Gamma(\delta) = \int_0^\infty u^{\delta-1}\, e^{-u} du$ is the gamma function, then we have $zf(z) \sim e^{-z} z^{\delta}\, \Gamma(\delta)^{-1}$ as $z \to \infty$, and

$$zf(z) \sim z^{\delta}\, \Gamma(\delta)^{-1} = \exp(-\delta |\log z|)\Gamma(\delta)^{-1} \text{ as } z \to 0.$$

We observe that $zf(z) \in L_{1,\delta,0}(0)$, with the log-slowly varying part equal to $\Gamma(\delta)^{-1}$, but it is not sufficiently heavy. Furthermore, $zf(z)$ is not log-exponentially varying at ∞.

A special attention should be paid to the left tail of a density defined on $(0, \infty)$. For a density f such that $zf(z) \in L_{\gamma,\delta,\alpha}(0)$, that is

$$f(z) = (1/z) \exp(-\delta |\log z|^{\gamma}) |\log z|^{-\alpha} S(z), \text{when } z < 1/A \text{ for some } A > 1,$$

we can show that the limit of $f(z)$ as $z \to 0$ is

i) a constant $t > 0$ if $(\gamma, \delta, \alpha) = (1, 1, 0)$ and the log-slowly varying function $S(z)$ converges to t as $z \to 0$,

ii) 0 if $(\gamma, \delta, \alpha) > (1, 1, 0)$ or $(\gamma, \delta, \alpha) = (1, 1, 0)$ and $S(z) \to 0$ as $z \to 0$,

iii) ∞ if $(\gamma, \delta, \alpha) < (1, 1, 0)$ or $(\gamma, \delta, \alpha) = (1, 1, 0)$ and $S(z) \to \infty$ as $z \to 0$.

Notice that $f(z) \sim S(z)$ as $z \to 0$ if $(\gamma, \delta, \alpha) = (1, 1, 0)$. For instance, for the gamma density given by (5) with an LE-credence of $(1, \delta, 0)$, we can verify that $f(z) \to 1$ if $\delta = 1$; $f(z) \to 0$ if $\delta > 1$; $f(z) \to \infty$ if $\delta < 1$. It means that the condition of robustness $\gamma < 1$ implies that $f(z) \to \infty$ as $z \to 0$.

To help the search of valid densities for robustness purposes, we propose in Section 4.1 a method of transformation of densities defined on $\mathbb{R}$ in order to devise a density f such that $zf(z)$ is log-exponentially varying in both tails. In Section 4.2, we propose a family of densities resulting from this method of transformation.

4.1 Exponential transformation of densities defined on the real line

A simple method to devise densities defined on $(0, \infty)$ consists of the exponential transformation of densities defined on $\mathbb{R}$. For most of the known densities on $\mathbb{R}$, the transformation results in densities f such that $zf(z)$ is log-exponentially varying at 0 and ∞.

A well-known case is the exponential transformation of the standardized normal density in a log-normal density f. The function $zf(z)$ is then log-exponentially varying at 0 and ∞ and its LE-credence is (2,1/2,0) in both tails, which is however too large for robustness purposes.

In general, we start with a random variable Y having a density $g(y)$ defined on $\mathbb{R}$ that is unimodal, symmetric with respect to the origin ($g(y) = g(-y)$), with a scale parameter set to any value (1 by default). Let the cumulative distribution function of a random variable Y be denoted by $F_Y(\cdot)$ and its a-quantile be denoted by $Q_Y(a)$, such that

$$1 - F_Y(Q_Y(a)) = \Pr[Y > Q_Y(a)] = a.$$

Notice that by symmetry we have $F(-y) = 1 - F(y)$ and $Q_Y(a) = -Q_Y(1-a)$, for $0 < a < 1$, with $Q_Y(1/2) = 0$.

We proceed now to the exponential transformation $Z = \exp(\tau Y)$. The density of Z is then given by

$$f(z) = (\tau z)^{-1} g(\tau^{-1} \log z),$$

where $z, \tau > 0$. Notice the symmetry given by $zf(z) = (1/z)f(1/z)$ resulting from $g(y) = g(-y)$, which means that the tail behaviour of $zf(z)$ is the same at 0 and ∞. If we add a scale parameter $s > 0$, we have

$$Z \mid s \overset{\mathcal{D}}{\sim} (1/s)f(z/s) = (\tau z)^{-1} g(\tau^{-1} \log(z/s)).$$

It follows that $F_{Z|s}(z) = F_Y(\tau^{-1} \log(z/s))$ and we obtain the relation of symmetry given by $F_{Z|s}(s/z) = 1 - F_{Z|s}(sz)$. It also can be verified, for $0 < a < 1$, that

$$Q_{Z|s}(a) = s \exp(\tau Q_Y(a)) \text{ and } Q_{Z|s}(1-a) = s/\exp(\tau Q_Y(a)) = s^2/Q_{Z|s}(a).$$

In particular, the median is $Q_{Z|s}(1/2) = s$. The geometric mean of $Q_{Z|s}(1-a)$ and $Q_{Z|s}(a)$ is hence the median s.

The parameter τ can be used to control the dispersion of the density f. To simplify the choice of τ, we can proceed as follows. For a given probability $1-a$ (for instance $1-a = 0.95$), we choose an interval $(1/b, b)$, where $b > 1$, that includes $1-a$ of the mass of the density $f(z)$, that is such that

$$1 - a = \Pr[1/b \leqslant Z \leqslant b] = \int_{1/b}^{b} f(z)dz.$$

It necessarily follows that $b = \exp(\tau Q_Y(a/2))$, and solving for τ, we find

$$\tau = (\log b)/Q_Y(a/2). \tag{6}$$

Notice that the choice of τ in (6) does not depend on the median s.

It is equivalent and possibly easier to choose an interval $(s/b, sb)$ around a given median s that includes $1-a$ of the mass of the density $(1/s)f(z/s)$, since

$$\int_{1/b}^{b} f(z)dz = \int_{s/b}^{sb} (1/s)f(z/s)dz = \Pr[s/b \leqslant Z \leqslant sb \mid s].$$

It it also equivalent, in a Bayesian context, to choose a $(1-a)$-credibility interval for s given by $(z/b, zb)$, since

$$\int_{1/b}^{b} f(z)dz = \int_{1/b}^{b} f(s)ds = \int_{z/b}^{zb} (1/s)(z/s)f(z/s)ds = \Pr[z/b \leqslant s \leqslant zb \mid z],$$

where $(1/s)(z/s)f(z/s)$ represents the posterior density of s given the observation z, with $1/s$ as the non-informative prior and $(1/s)f(z/s)$ as the model.

For example, we can choose $g(y) = C(\nu)(\nu + y^2)^{-(\nu+1)/2}$, a Student density with $\nu > 0$ degrees of freedom, where $C(\nu)$ is the appropriate normalizing constant. Then the "log-Student" density with parameters ν and τ is given by

$$f(z) = C(\nu)(\tau z)^{-1}(\nu + (\log z)^2/\tau^2)^{-(\nu+1)/2}, \tag{7}$$

where $z, \tau > 0$. The function $zf(z)$ is log-exponentially varying at 0 and ∞ with its LE-credence given by $(0, 0, \nu + 1)$ in both tails, for any values of τ. Therefore, the log-Student is a good candidate for robust inference.

Two other examples are the exponential transformation of the symmetric Laplace and logistic densities, that behave as $g(y) \sim \exp(-|y|)$ as $|y| \to \infty$. It can be verified that the "log-Laplace" and "log-logistic" densities behave as $zf(z) \sim \exp(-(1/\tau)|\log z|)$ as $|z| \to 0$ or $|z| \to \infty$, with their LE-credence given by $(1, \tau^{-1}, 0)$. However, their tails are not sufficiently heavy for robustness purposes.

To facilitate the search of appropriate densities for robustness's sake, we introduce in the next section a new family of densities.

4.2 The log-GEP$_2$ distribution

The generalized exponential power (GEP) density was first introduced by Angers (2000) and then by Desgagné and Angers (2005) with a minor modification about the sign of some parameters. It is a symmetric density around the origin, defined on the real line, with a constant part in the center. Its interest lies in the large spectrum of its tail behaviour. We propose in this section a modified GEP density with no constant part, called generalized exponential power of the second form (GEP$_2$), and its exponential transformation defined on $(0, \infty)$, called log-GEP$_2$.

The GEP$_2$ density is built with the right tail of the GEP density (proportional to $\exp(-\delta y^\gamma) y^{-\alpha} (\log y)^{-\beta}$), translated to the origin and doubled.

Definition 6. *A random variable Y has a GEP$_2$ distribution, written $Y \sim GEP_2(\gamma, \delta, \alpha, \beta, \theta)$, if its density is given by*

$$g(y) = (1/2)K(\gamma, \delta, \alpha, \beta, \theta)\exp(-\delta(|y| + \theta)^\gamma)(|y| + \theta)^{-\alpha}(\log(|y| + \theta))^{-\beta},$$

where $y \in \mathbb{R}$, $\gamma \geqslant 0, \delta \geqslant 0, \alpha \in \mathbb{R}, \beta \in \mathbb{R}, \theta \geqslant 0$, and

$$(K(\gamma, \delta, \alpha, \beta, \theta))^{-1} = \int_{\theta}^{\infty} \exp(-\delta y^\gamma) y^{-\alpha} (\log y)^{-\beta} dy. \tag{8}$$

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

The GEP_2 density is symmetric with respect to the origin and is generally unimodal, except possibly when $\alpha < 0$ and/or $\beta < 0$. In this case, it suffices to choose θ large enough to guarantee unimodality, such that

$$\gamma\delta\theta^{\gamma} + \alpha + \beta/\log\theta \geqslant 0. \tag{9}$$

In particular, if $\beta = 0$, equation (9) is simplified as $\theta \geqslant |\alpha|^{1/\gamma}(\gamma\delta)^{-1/\gamma}$ if $\alpha < 0$.

The log-GEP_2 density is devised using the method of transformation described in Section 4.1.

Definition 7. *A random variable Z has a log-GEP_2 distribution, written $Z \sim$ log-GEP_2 $(\gamma, \delta, \alpha, \beta, \theta, \tau)$, if its density is given by*

$$\begin{aligned} f(z) = {} & (1/2)K(\gamma, \delta, \alpha, \beta, \theta)(\tau z)^{-1}\exp(-\delta(\tau^{-1}|\log z| + \theta)^{\gamma}) \\ & \times (\tau^{-1}|\log z| + \theta)^{-\alpha}(\log(\tau^{-1}|\log z| + \theta))^{-\beta}, \end{aligned}$$

where $z > 0$, $\tau > 0$. The domain and constraints on the other parameters are given in the definition of the GEP_2 distribution and the constant $K(\gamma, \delta, \alpha, \beta, \theta)$ is given by equation (8).

If $\gamma \leqslant 1$ or $\gamma > 1, \alpha = \beta = \theta = 0$, it can be verified that $zf(z)$ is log-exponentially varying at 0 and ∞ with LE-credences given by $(\gamma, \delta/\tau^{\gamma}, \alpha)$. Notice that the parameter θ has no impact on the tails if $\gamma \leqslant 1$. The term $(\log(\tau^{-1}|\log z| + \theta))^{-\beta}$ is log-slowly varying at 0 and ∞ since it behaves as $(\log|\log z|)^{-\beta}$ in both tails. A scale parameter can be added, and as noticed in Section 4.1, it corresponds to the median.

4.3 Special cases of the log-GEP_2 distribution

We now describe in more detail some special cases of the log-GEP_2 density for which the normalizing constant is analytically tractable.

Log-normal distribution

If $\gamma = 2, \delta = 1/2, \alpha = \beta = \theta = 0$, then Z has a log-normal density with parameter τ, given by

$$f(z) = (2\pi)^{-1/2}(\tau z)^{-1}\exp(-(\log z)^2/(2\tau^2)). \tag{10}$$

The function $zf(z)$ is log-exponentially varying at 0 and ∞ with LE-credence given by $(2, (2\tau^2)^{-1}, 0)$. We have

$$F_Z(z) = \Phi(\tau^{-1}\log z) \text{ and } Q_Z(a) = \exp(\tau\Phi^{-1}(1-a)),$$

where Φ is the usual cumulative distribution function of a standardized normal.

Log-Laplace distribution

If $\gamma = \delta = 1, \alpha = \beta = \theta = 0$, then Z has a log-Laplace density with parameter τ, given by

$$f(z) = (2\tau z)^{-1} \exp(-(1/\tau)|\log z|) = (2\tau z)^{-1} z^{-(1/\tau)\ \mathrm{sign}(\log z)}.$$

The function $zf(z)$ is log-exponentially varying at 0 and ∞ with LE-credence given by $(1, \tau^{-1}, 0)$. Furthermore, we have

$$F_Z(z) = (1/2) \exp(-(1/\tau)|\log z|) \text{ if } z \leqslant 1$$

and

$$Q_Z(a) = \exp(-\tau \log(2a)) \text{ if } a \leqslant 1/2.$$

Notice that $F_Z(1/z) = 1 - F_Z(z)$ and $Q_Z(1-a) = 1/Q_Z(a)$. Furthermore, we can simulate Z using $Z = Q_Z(a)$ if a is generated from a uniform distribution.

Log-double-Pareto distribution

If $\gamma = \delta = \beta = 0, \alpha = \lambda + 1$, $\theta = 1$, where $\lambda > 0$, then Z has a log-double-Pareto density with parameters λ and τ, given by

$$f(z) = \lambda(2\tau z)^{-1}(\tau^{-1}|\log z| + 1)^{-(\lambda+1)}. \tag{11}$$

The function $zf(z)$ is log-exponentially varying at 0 and ∞ with LE-credence given by $(0, 0, \lambda + 1)$. Furthermore, we have

$$F_Z(z) = (1/2)(\tau^{-1}|\log z| + 1)^{-\lambda} \text{ if } z \leqslant 1 \text{ and}$$

$$Q_Z(a) = \exp(\tau((2a)^{-1/\lambda} - 1)) \text{ if } a \leqslant 1/2.$$

Log-exponential-power distribution

If $\gamma > 0, \delta > 0, \alpha = 0, \beta = \theta = 0$, then Z has a log-exponential-power density with parameters γ, δ and τ, given by

$$f(z) = \gamma\delta^{1/\gamma}(2\Gamma(1/\gamma))^{-1}(\tau z)^{-1} \exp(-\delta\tau^{-\gamma}|\log z|^{\gamma}).$$

The function $zf(z)$ is log-exponentially varying at 0 and ∞ with LE-credence given by $(\gamma, \delta\tau^{-\gamma}, 0)$. Furthermore, we have

$$F_Z(z) = \frac{\Gamma(1/\gamma, \delta\tau^{-\gamma}|\log z|^{\gamma})}{2\Gamma(1/\gamma)} \text{ if } z \leqslant 1,$$

where $\Gamma(a, b) = \int_b^{\infty} u^{a-1}\, e^{-u} du$ is the incomplete gamma function, for $a > 0$ and $b \geqslant 0$. In particular, $\Gamma(a, 0) = \Gamma(a)$ is the gamma function.

Notice that the log-exponential-power distribution includes the log-normal if $\gamma = 2$, $\delta = 1/2$ and the log-Laplace if $\gamma = \delta = 1$.

Log-double-generalized-gamma distribution

If $\gamma > 0$, $\delta = 1$, $\alpha = 1-\lambda$ ($\lambda > 0$), $\beta = 0$, $\theta \geqslant 0$, with $\theta > 0$ if $\lambda \neq 1$ and $\theta \geqslant ((\lambda-1)/\gamma)^{1/\gamma}$ if $\lambda > 1$, then Z has a log-double-generalized-gamma density with parameters γ, λ, θ and τ, given by

$$f(z) = \gamma(2\Gamma(\lambda/\gamma, \theta^\gamma))^{-1}(\tau z)^{-1}\exp(-(\tau^{-1}|\log z| + \theta)^\gamma)(\tau^{-1}|\log z| + \theta)^{\lambda-1}.$$

The function $zf(z)$ is log-exponentially varying at 0 and ∞ with LE-credence given by $(\gamma, \tau^{-\gamma}, 1-\lambda)$ if $\gamma \leqslant 1$ or $\gamma > 1$, $\lambda = 1$, $\theta = 0$. Otherwise if $\gamma > 1$, the parameter θ has an impact on the exponential term and $zf(z)$ is no longer log-exponentially varying. However it is not a problem for robustness, since $\gamma < 1$ must be satisfied for conflicting densities. Furthermore, we have

$$F_Z(z) = \frac{\Gamma(\lambda/\gamma, (\tau^{-1}|\log z| + \theta)^\gamma)}{2\Gamma(\lambda/\gamma, \theta^\gamma)} \text{ if } z \leqslant 1.$$

Notice that the log-double-generalized-gamma$(\gamma, \lambda, \theta, \tau)$ distribution includes the log-exponential-power$(\gamma, \delta = 1, \tau)$ if $\lambda = 1$ and $\theta = 0$.

5 Example

Table 1: Prior and experts' predictions, with their 95% credibility intervals.

	Prior	**Experts**				
		1	**2**	**3**	**4**	**5**
Predictions $\phi, x_1, \ldots, x_5$	4.5	2.25	4	4.5	5.2	12
Left bound ϕ/b_0 or x_i/b_i	1.25	1.5	2.5	3	3.2	10
Right bound ϕb_0 or $x_i b_i$	16.2	3.375	6.4	6.75	8.45	14.4
b_i	3.6	1.5	1.6	1.5	1.625	1.2
τ_i (log-normal)	.6535	.2069	.2398	.2069	.2477	.0930
τ_i (log-double-Pareto)	1.5610	.4941	.5728	.4941	.5917	.2222
τ_i (log-Student)	.4983	.1577	.1828	.1577	.1889	.0709

We consider a simple example of a portfolio manager that needs a prediction on the volatility of a stock index represented by σ, where the volatility is measured say by the standard deviation of the next twelve monthly returns. He asks 5 experts (quantitative models for instance) for their prediction on the volatility σ, represented by the observations $x_1, \ldots, x_5$, as well as a measure of confidence on their prediction, denoted by $b_i > 1$, where $(x_i/b_i, x_i b_i)$ represents a 95% credibility interval for σ, for $i = 1, \ldots, 5$. The manager wants to combine this information with his prior beliefs using the Bayesian model described in Section 3.1. The prediction given by the prior is based on the historical median of the volatility, represented by ϕ, and the measure of confidence is denoted by $b_0 > 1$, where 95% of the past volatilities lie in the interval $(\phi/b_0, \phi b_0)$.

Data are given in Table 1. The predictions and intervals are expressed in percentages. Each of the densities $\pi, f_1, \ldots, f_5$ are modelled using three different distributions, namely the log-normal(τ_i) as given by (10) with LE-credence given by $(2, (2\tau_i^2)^{-1}, 0)$, the log-double-Pareto($\lambda = 5, \tau_i$) as given by (11) with LE-credence given by $(0, 0, 6)$, and the log-Student($\nu = 5, \tau_i$) as given by (7) with LE-credence given by $(0, 0, 6)$, $i = 0, 1, \ldots, 5$. Notice that τ_0 is associated to π and τ_i to f_i. For each distribution, the parameter τ_i is set according to (6), where $a = 0.05$. The values of τ_i are given in Table 1.

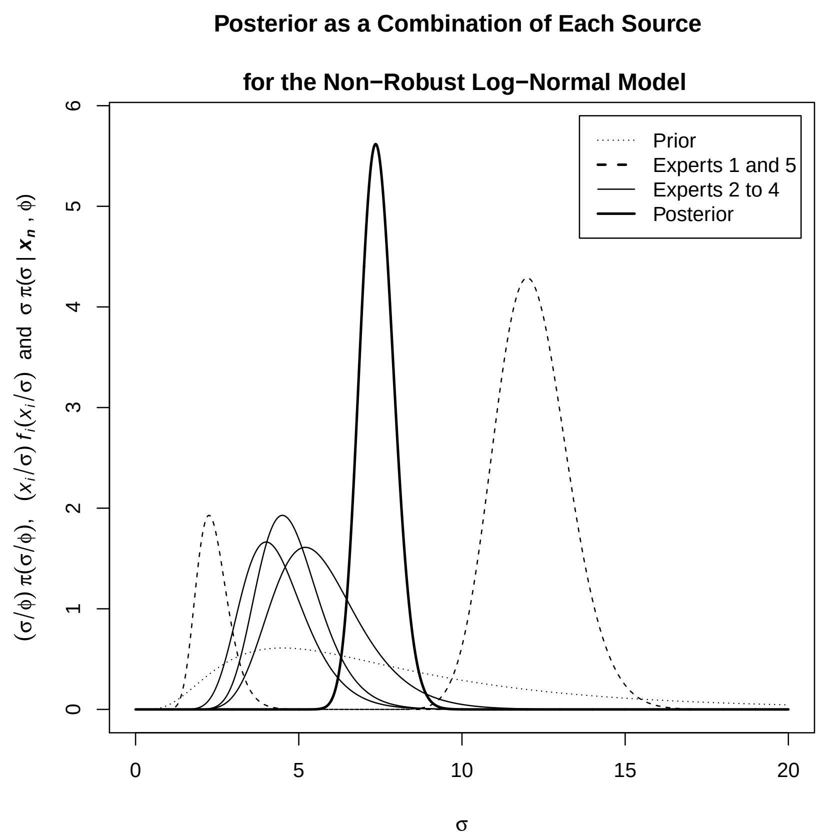

Figure 1: Example of non-robust modelling.

The inference on σ is performed using its posterior density given by

$$\sigma\pi(\sigma|\mathbf{x_5}, \phi) \propto (\sigma/\phi)\pi(\sigma/\phi) \prod_{i=1}^{5} (x_i/\sigma) f_i(x_i/\sigma).$$

The functions $(\sigma/\phi)\pi(\sigma/\phi)$ and $(x_i/\sigma)f_i(x_i/\sigma)$, as well as their product $\sigma\pi(\sigma|\mathbf{x_5}, \phi)$, are plotted in Figure 1 when the densities are log-normal and in Figure 2 when the densities are log-Student. The graph for the log-double-Pareto densities is similar to Figure 2 and

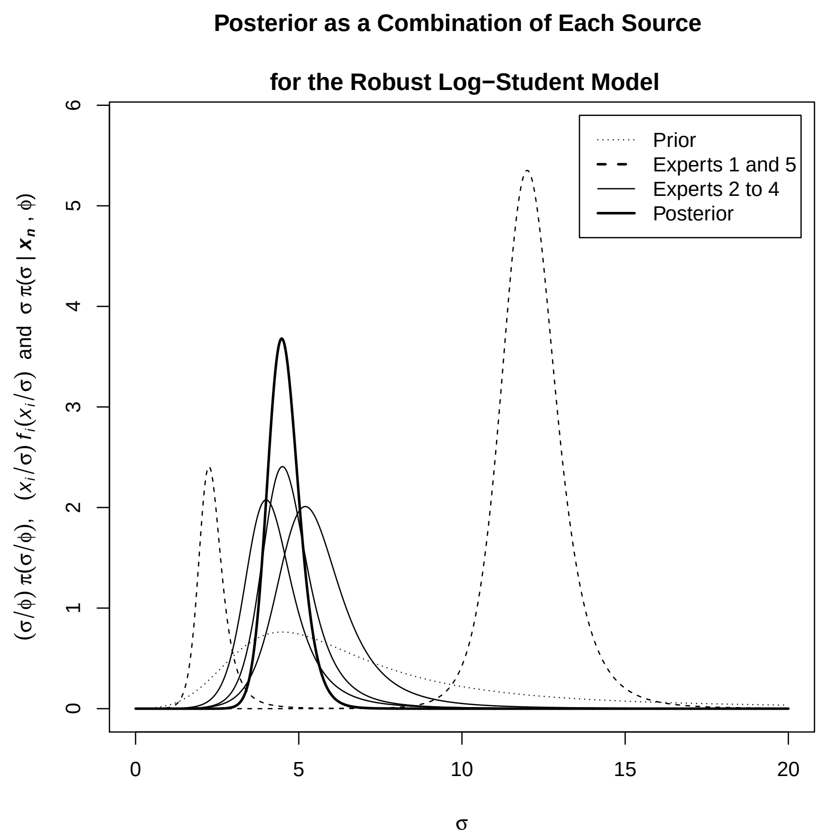

Figure 2: Example of robust modelling.

hence is not shown. We see that experts 2, 3 and 4 (thin solid lines) provided similar information. The prior information (dotted line) is also similar, but much more diffuse. However, the information provided by experts 1 and 5 (dashed lines) seems in conflict with the other sources, the common area shared with the others being small.

For the non-robust log-normal model (Figure 1), the posterior information (thick solid line) lies in an area largely ignored by most of the other sources of information, except for the quasi non-informative prior and for expert 3. For the robust log-Student model (Figure 2), the posterior information (thick solid line) agrees with the prior and experts 2, 3 and 4, but shares only a small area with the information given by experts 1 and 5, suggesting that the outliers are mostly rejected.

Posterior estimation of σ is done using either the expectation $E(\sigma \mid \mathbf{x_5}, \phi)$ or the median $Q_{\sigma|\mathbf{x_5},\phi}(1/2)$. Results are given in Table 2. Different scenarios are considered: i) experts 1 to 5 are included in the model; ii) experts 1 and 5 are excluded; iii) only

Table 2: Posterior estimation of σ.

		Experts			
Model	**Estimator**	**1,2,3,4,5**	**2,3,4**	**2,3,4,5**	**1,2,3,4**
1) Log-normal	$E(\sigma \mid \mathbf{x_5}, \phi)$	7.37	4.56	8.64	3.74
	$Q_{\sigma\mid\mathbf{x_5},\phi}(1/2)$	7.36	4.52	8.62	3.71
2) Log-Student	$E(\sigma \mid \mathbf{x_5}, \phi)$	4.51	4.55	4.95	4.17
(log-normal prior)	$Q_{\sigma\mid\mathbf{x_5},\phi}(1/2)$	4.48	4.52	4.87	4.17
3) Log-d.-Pareto	$E(\sigma \mid \mathbf{x_5}, \phi)$	4.53	4.53	4.74	4.35
(log-normal prior)	$Q_{\sigma\mid\mathbf{x_5},\phi}(1/2)$	4.50	4.50	4.67	4.35
4) Log-Student	$E(\sigma \mid \mathbf{x_5}, \phi)$	4.51	4.55	4.92	4.18
(truncated prior)	$Q_{\sigma\mid\mathbf{x_5},\phi}(1/2)$	4.48	4.52	4.86	4.18
5) Log-d.-Pareto	$E(\sigma \mid \mathbf{x_5}, \phi)$	4.52	4.52	4.68	4.38
(truncated prior)	$Q_{\sigma\mid\mathbf{x_5},\phi}(1/2)$	4.50	4.50	4.61	4.39

expert 1 is excluded; iv) only expert 5 is excluded.

For the model where each density is log-normal (model 1 in Table 2), the posterior expectation and the moments exist. However, if each density is either log-Student or log-double-Pareto, the moments do not exist even if the posterior is proper, its tails being too heavy. In this case, the inference can be done using quantiles such as the median. Credibility intervals can also be calculated. If one prefers working with expectation and moments, we propose two solutions.

A first solution is to use a log-normal density for one source of information. Ideally, we choose a source that we absolutely trust in case of conflict (since it will never be rejected) or that is so diffuse that a conflict with other sources is practically impossible, as is the case in our example with the prior. This way, the moments exist and we can estimate σ with the posterior expectation. In our example, we use a log-normal for the prior in models 2 and 3 (see Table 2).

A second solution is to use a truncated prior. In our example, we use a truncated log-Student in model 4 and a truncated log-double-Pareto in model 5. We can choose points of truncation, say $1/t$ and t, with $t > 1$ as large as we want, in accordance with our context. For the calculation of the median with truncation, it would suffice to choose t large enough to approach the median without truncation as close as we want, since the posterior is proper. Theoretically, the (absolute) moments being infinite, their value calculated with truncated prior increases with t. However, in practice it happens that it increases at such a slow rate that it is not noticeable. It is the case in our example, where the calculation of moments seems totally insensible to the points of truncation beyond a certain threshold. Whether we choose $t = 30$ or $t = 1000$, it makes no difference in the calculation of the expectation and variance, at a precision of at least 6 decimals. In our context, we are quite sure that the volatility will fall between 0.001% and 1000%. This solution is interesting only if the moments are practically insensitive to the chosen points of truncation for t beyond a certain threshold. Further analysis would be necessary to better understand when this method works. For instance, we noticed in our example that this insensitivity increases with the number of observations in the model.

Using Theorem 1, we know how the posterior behaves in the presence of conflicting information. Robustness with the log-normal model is not guaranteed, since its LE-credence is too large. For the log-Student or log-double-Pareto models, where every tail has the same behaviour, robustness is guaranteed if the number of non-outlying values is larger than the maximum between the number of left and right conflicting values. In our example, it means that information given by experts 1 and 5 would be rejected as they move away in each direction. Even if the results of Theorem 1 are asymptotic, we see in Table 2 that rejection occurs efficiently with finite observations.

Notice first that every model gives essentially the same results when the conflicting values (experts 1 and 5) are excluded, see the second column of Table 2. That was expected (and desirable) with the way we chose the parameters τ_i. We can also observe that estimation using either the posterior expectation or median gives similar results. In the same way, estimation using a truncated or a log-normal prior for the robust models gives similar results. Even the robust log-Student and log-double-Pareto models give comparable results, which can be explained by their identical LE-credence.

We see that the log-normal model is largely influenced by expert 1 ($E(\sigma \mid \mathbf{x_n}, \phi) = 3.74$) and by expert 5 ($E(\sigma \mid \mathbf{x_n}, \phi) = 8.64$) when they are added separately in the model. The log-normal model is still contaminated when they are both added in the model ($E(\sigma \mid \mathbf{x_n}, \phi) = 7.37$). For the robust models, the influence of experts 1 and 5 is already quite small and theory tells us that it would decrease to nothing if the conflict would increase.

6 Conclusion

Full robustness has been investigated in Bayesian modelling of a scale parameter. The log-exponentially varying functions have been introduced to provide a framework for the characterization of the eligible densities that lead to robust inference. LE-credence has been defined as the vector of parameters associated with a log-exponentially varying function and has proved to be useful to characterize the thickness of a tail and to order different tails. The log-regularly and log-slowly varying functions have also been defined as a subclass of the the log-exponentially varying functions.

The main results are given in Theorem 1. Their nature is asymptotic, not in the classical way where the sample size $n \to \infty$, but in the sense that the conflicting values (some observations and/or the prior's scale) move to 0 or ∞. Nevertheless, the results are still useful with finite information as it is shown in our example of combination of experts' opinions.

Essentially, robustness is guaranteed if: 1) the appropriate tail of a conflicting density, say $f(z)$, is sufficiently heavy, more precisely if $zf(z)$ is log-exponentially varying;

2) if the right tail of the posterior density considering only the non-conflicting values is lighter than the left tail of the posterior density considering only the large conflicting values; 3) if the left tail of the posterior density considering only the non-conflicting values is lighter than the right tail of the posterior density considering only the small conflicting values. These conditions are intuitive and easy to verify. They are based only on the densities of the model and on some limits; there are no integrals, derivatives or cumulative distribution functions involved.

The principal result of robustness is given by the convergence in distribution of the random variable σ given the complete information to the random variable σ given the non-conflicting information, as the conflicting values (outliers and/or prior) tend to 0 or ∞, at any given rate. Full robustness is achieved asymptotically, as the influence of the conflicting values disappears completely as they move apart. We also found that if the tail behaviour is the same for all densities, full robustness is guaranteed if the non-conflicting values exceeds the conflicting values.

Practical concerns have also been addressed. The log-GEP$_2$ density has been introduced to compensate for the rarity of densities appropriate for full robustness in a scale parameter structure. A log-GEP$_2$ density, say $f(z)$, has the property that $zf(z)$ has the same tail behaviour at 0 and ∞, which is useful if we want to be equally protected against conflicting values in all directions. Its large tail behaviour can be helpful for a user: it includes the log-normal density, the log-Laplace and a diversity of log-exponentially and log-regularly varying functions.

Practical considerations have also been addressed through an example of combination of experts' opinions. Prediction of a scale parameter of interest is given by different experts, as well as a measure of confidence on their prediction. It is shown how to reflect this confidence in the modelling of the densities. The non-robust log-normal model is compared with the robust log-Student and log-double-Pareto models. Even though the theoretical results are asymptotic, the phenomenon of rejection occurred quite well in this example with only five observations.

We also proposed solutions for the cases where modelling leads to posterior inference with no existence of the moments, for instance when all densities are log-Student. One can simply use quantiles, median and credibility intervals. If working with moments is preferred, one solution is to model a non-informative or diffuse source of information (usually the prior) with the lighter-tailed log-normal distribution. A second solution is to use a truncated prior. We found that in practice (at least in our example), the choice of the points of truncation beyond a certain threshold has no perceptible impact on the posterior moments, at a precision of at least 6 decimals. While not conventional, we think it is worthwhile to further investigate this approach.

This paper can be generalized in different ways. While the class of log-exponentially varying functions is quite large, it still can be widened. We think the class of slowly varying functions could be a good starting point, as suggested by Proposition 3. Furthermore, our results of robustness could be extended to include convergence of the posterior expectation of functions. Finally, a thorough investigation on how the robustness performs in practice with different modelling would be interesting.

7 Proofs

Notice that, as mentioned in Section 2, the square brackets distinguish asymptotic behaviour at 0 from that at ∞.

7.1 Proof of Proposition 3

Since $g \in L_{\gamma,\delta,\alpha}(\infty[0])$, we can write $g(z) \sim \exp(-\delta|\log z|^{\gamma})|\log z|^{-\alpha}S(z)$, with $S \in L_0(\infty[0])$, as $z \to \infty[0]$. Then, considering $\sigma > 0$ and using the Taylor series development of $|\log z + \log \sigma|^{\gamma}$, we have, as $z \to \infty[0]$,

$$\begin{aligned}
\frac{g(z\sigma)}{g(z)} &\sim \frac{\exp(-\delta|\log z + \log\sigma|^{\gamma})|\log z + \log\sigma|^{-\alpha}S(z\sigma)}{\exp(-\delta|\log z|^{\gamma})|\log z|^{-\alpha}S(z)} \\
&\sim \exp\left(-\delta\sum_{k=1}^{\infty}\frac{(\operatorname{sign}(\log z)\log\sigma)^{k}\gamma\cdot(\gamma-1)\cdots(\gamma-(k-1))}{k!|\log z|^{k-\gamma}}\right)\frac{S(z\sigma)}{S(z)} \\
&\sim \frac{S(z\sigma)}{S(z)}\exp\Big(-\delta\Big(\operatorname{sign}(\log z)(\log\sigma)\gamma|\log z|^{-(1-\gamma)} \\
&\qquad\qquad + (\log\sigma)^{2}\gamma(\gamma-1)(1/2)|\log z|^{-(2-\gamma)} + \ldots\Big)\Big) \\
&\sim \frac{S(z\sigma)}{S(z)},
\end{aligned}$$

as long as $0 \leqslant \gamma < 1$. It suffices now to show that $S(z\sigma) \sim S(z)$, for any $\sigma > 0$. The power invariance of $S(z)$ can be written as follows: $\forall\lambda > 1$, we have, as $z \to \infty[0]$, $1/\lambda \leqslant \nu \leqslant \lambda \Rightarrow S(z^{\nu})/S(z) \to 1$, or equivalently

$$\min(z^{1/\lambda}, z^{\lambda}) \leqslant a \leqslant \max(z^{1/\lambda}, z^{\lambda}) \Rightarrow S(a)/S(z) \to 1.$$

If $\max(z, 1/z)$ is large enough, specifically if $\max(z, 1/z) \geqslant \lambda^{\lambda/(\lambda-1)}$, then it can be verified that

$$\min(z^{1/\lambda}, z^{\lambda}) \leqslant z/\lambda < z\lambda \leqslant \max(z^{1/\lambda}, z^{\lambda}).$$

It follows that, as $z \to \infty[0]$,

$$z/\lambda \leqslant a \leqslant z\lambda \Rightarrow S(a)/S(z) \to 1 \text{ or } 1/\lambda \leqslant \sigma \leqslant \lambda \Rightarrow S(z\sigma)/S(z) \to 1.$$

7.2 Proof of Proposition 4

The proof for the case $z \to 0$ is omitted since it is similar to the case $z \to \infty$. Using the symmetry of $f(z/\sigma)f(\sigma)$ around $\sigma = \sqrt{z}$, it can be verified that

$$\begin{aligned}
\int_0^{\infty}(1/\sigma)f(z/\sigma)f(\sigma)/f(z)\,d\sigma &= 2\int_0^{\sqrt{z}}(1/\sigma)f(z/\sigma)f(\sigma)/f(z)\,d\sigma \\
&= 2\int_0^{\sqrt{z}}(1/\sigma)(z/\sigma)f(z/\sigma)\sigma f(\sigma)/(zf(z))\,d\sigma.
\end{aligned}$$

Consider an intermediate variable $\tau \to \infty$ as well as $z \to \infty$. We first choose a value of $\tau > 1$ as large as we want, and once τ chosen, we can choose a value of z as large as we want. We split the integral in three parts between $0 < 1/\tau < \tau < \sqrt{z}$.

Firstly, if $0 < \sigma \leqslant 1/\tau$,

$$(z/\sigma)f(z/\sigma)\sigma f(\sigma)/(zf(z)) \leqslant \sigma f(\sigma) \to 0 \text{ as } \sigma \leqslant 1/\tau \to 0,$$

using the monotonicity of the right tail of $zf(z)$ for any z larger than a certain constant, since $z/\sigma \geqslant z\tau \geqslant z$ for any $\tau > 1$. Similarly we have

$$2\int_0^{1/\tau} (1/\sigma)f(z/\sigma)f(\sigma)/f(z)\,d\sigma \leqslant 2\int_0^{1/\tau} f(\sigma)\,d\sigma \to 0 \text{ as } \tau \to \infty.$$

Secondly, if $1/\tau \leqslant \sigma \leqslant \tau$,

$$\lim_{z\to\infty} f(z/\sigma)f(\sigma)/f(z) = \sigma f(\sigma), \tag{12}$$

using Proposition 3 since $1/\tau \leqslant \sigma \leqslant \tau$ and $z \to \infty$. We have

$$\sigma f(\sigma) \leqslant \sup_{1/\tau\leqslant\sigma\leqslant\tau} \sigma f(\sigma) \to \sup_{\sigma>0} \sigma f(\sigma) \text{ as } \tau \to \infty.$$

Notice that for a chosen τ, if z is large enough, equation (12) means that $f(z/\sigma)f(\sigma)/f(z)$ is bounded by say $2\sigma f(\sigma)$ for $1/\tau \leqslant \sigma \leqslant \tau$. Therefore, we can use Lebesgue's dominated convergence theorem to pass the limit $z \to \infty$ inside the integral and we have

$$\lim_{z\to\infty} 2\int_{1/\tau}^{\tau} (1/\sigma)f(z/\sigma)f(\sigma)/f(z)\,d\sigma = 2\int_{1/\tau}^{\tau} f(\sigma)\,d\sigma \to 2 \text{ as } \tau \to \infty.$$

For the third part of the integral, consider first $\gamma = \delta = 0$, that is $zf(z) \in L_\alpha(\infty)$, where (γ, δ, α) is the LE-credence of $zf(z)$. If $\tau \leqslant \sigma \leqslant \sqrt{z}$,

$$\begin{aligned}(z/\sigma)f(z/\sigma)\sigma f(\sigma)/(zf(z)) &\leqslant \frac{\sqrt{z}f(\sqrt{z})}{zf(z)}\sigma f(\sigma)\\ &\sim 2^\alpha \sigma f(\sigma) \to 0 \text{ as } \sigma \geqslant \tau \to \infty,\end{aligned}$$

using the monotonicity of the right tail of $zf(z)$ for any z larger than a certain constant since $z/\sigma \geqslant \sqrt{z}$, and using the definition of $zf(z) \in L_\alpha(\infty)$. Similarly we have

$$2\int_{\tau}^{\sqrt{z}} (1/\sigma)f(z/\sigma)f(\sigma)/f(z)\,d\sigma \leqslant 2^{1+\alpha}\int_{\tau}^{\infty} f(\sigma)\,d\sigma \to 0 \text{ as } \tau \to \infty.$$

Consider now $0 < \gamma < 1$, $\delta > 0$ and $\alpha \in \mathbb{R}$. If $\tau \leqslant \sigma \leqslant \sqrt{z}$, we have

$$\begin{aligned}
\frac{(z/\sigma)f(z/\sigma)\sigma f(\sigma)}{zf(z)} &= \frac{\exp(-\delta(\log(z/\sigma))^\gamma)\exp(-\delta(\log\sigma)^\gamma)}{\exp(-\delta(\log z)^\gamma)}\frac{g(z/\sigma)g(\sigma)}{g(z)} \\
&= \frac{\exp(-\delta(\log(z/\sigma))^\gamma)\exp(-\delta(\log\sigma)^\gamma)}{\exp(-\delta(\log z)^\gamma)\exp(-\delta(2-2^\gamma)(\log\sigma)^\gamma)} \\
&\quad \times \exp(-\delta(2-2^\gamma)(\log\sigma)^\gamma)\frac{g(z^{1-\nu})g(\sigma)}{g(z)} \\
&\leqslant 2^{|\alpha|}\exp(-\delta(\log z)^\gamma((1-\nu)^\gamma + (2^\gamma-1)\nu^\gamma - 1)) \\
&\quad \times \exp(-\delta(2-2^\gamma)(\log\sigma)^\gamma)\exp(\psi(\log\sigma)^\gamma) \\
&\leqslant 2^{|\alpha|}\exp(-\delta(2-2^\gamma-\psi)(\log\sigma)^\gamma) \to 0 \text{ as } \sigma \geqslant \tau \to \infty.
\end{aligned}$$

In the first equality, since $\tau \leqslant \sigma \leqslant \sqrt{z} \leqslant z/\sigma \leqslant z/\tau \leqslant z$, it suffices to choose τ large enough to write $af(a) = \exp(-\delta(\log a)^\gamma)g(a)$ with $g(a) \in L_\alpha(\infty)$, where $a \in \{\sigma, z/\sigma, z\}$. In the second equality, we defined $\nu = (\log\sigma)/\log z$, so we can write $\sigma = z^\nu$ and $z/\sigma = z^{1-\nu}$. Furthermore, we can verify that $0 < \nu \leqslant 1/2$ if $1 < \tau \leqslant \sigma \leqslant \sqrt{z}$. In the first inequality, we used the definition of $g(z) \in L_\alpha(\infty)$, that is $g(z^{1-\nu}) \sim g(z)(1-\nu)^{-\alpha}$, and since $1/2 \leqslant 1-\nu < 1$, $g(z^{1-\nu}) \leqslant g(z)2^{|\alpha|}$. We also defined ψ such that $0 < \psi < 2 - 2^\gamma$. This means that $\exp(-\psi(\log\sigma)^\gamma)g(\sigma) \leqslant 1$ for $\sigma \geqslant \tau$, if τ is chosen large enough since the exponential term dominates the log-regularly varying $g(\sigma)$. In the second inequality, it can be verified that the function $(1-\nu)^\gamma + (2^\gamma-1)\nu^\gamma - 1$ is non-negative for $\nu \in (0, 1/2]$, since it is concave and finds its minimum at $\nu = 0$ and $\nu = 1/2$ for a value of 0.

Similarly we have

$$\begin{aligned}
&2\int_\tau^{\sqrt{z}} (1/\sigma)f(z/\sigma)f(\sigma)/f(z)\, d\sigma \\
&\quad \leqslant 2^{1+|\alpha|}\int_\tau^\infty \frac{1}{\sigma}\exp(-\delta(2-2^\gamma-\psi)(\log\sigma)^\gamma)\, d\sigma \\
&\quad = 2^{1+|\alpha|}\int_{\log\tau}^\infty \exp(-\delta(2-2^\gamma-\psi)\theta^\gamma)\, d\theta \to 0 \text{ as } \tau \to \infty.
\end{aligned}$$

7.3 Proof of Theorem 1

The proof of results a) to d) of Theorem 1 are given in this section. We first need to introduce intermediate functions and results. Let the function $H(\sigma, \phi, \mathbf{x_n})$ be defined as

$$H(\sigma, \phi, \mathbf{x_n}) = \pi(\sigma \mid \mathbf{x_k}, \phi^{k_0}) \left(\frac{(\sigma/\phi)\pi(\sigma/\phi)}{(1/\phi)\pi(1/\phi)}\right)^{l_0+r_0} \prod_{i=1}^n \left(\frac{(1/\sigma)f_i(x_i/\sigma)}{f_i(x_i)}\right)^{l_i+r_i}. \tag{13}$$

We can verify that

$$H(\sigma, \phi, \mathbf{x_n}) = \frac{m(\mathbf{x_n} \mid \phi)\pi(\sigma \mid \mathbf{x_n}, \phi)}{m(\mathbf{x_k} \mid \phi^{k_0})((1/\phi)\pi(1/\phi))^{l_0+r_0}\prod_{i=1}^n f_i(x_i)^{l_i+r_i}}. \tag{14}$$

Note that our assumptions on the densities $\pi, f_1, \ldots, f_n$ imply that the posterior $\pi(\sigma \mid \mathbf{x_k}, \phi^{k_0})$ and $\pi(\sigma \mid \mathbf{x_n}, \phi)$ are proper densities. Considering that, and using (14), we obtain

$$\int_0^\infty H(\sigma, \phi, \mathbf{x_n})\, d\sigma = \frac{m(\mathbf{x_n} \mid \phi)}{m(\mathbf{x_k} \mid \phi^{k_0})((1/\phi)\pi(1/\phi))^{l_0+r_0} \prod_{i=1}^n f_i(x_i)^{l_i+r_i}}. \tag{15}$$

From (15), we see that result a) can be written as follows:

$$\int_0^\infty H(\sigma, \phi, \mathbf{x_n})\, d\sigma \to 1 \text{ as } \omega \to \infty,$$

where $H(\sigma, \phi, \mathbf{x_n})$ is given by (13). Furthermore, dividing (14) by (15), we find

$$\pi(\sigma \mid \mathbf{x_n}, \phi) = H(\sigma, \phi, \mathbf{x_n}) \Big/ \int_0^\infty H(\sigma, \phi, \mathbf{x_n})\, d\sigma. \tag{16}$$

Equation (16) is useful for the proof of result c). Finally, from (13) and (16), we have

$$\begin{aligned} \frac{\pi(\sigma \mid \mathbf{x_n}, \phi)}{\pi(\sigma \mid \mathbf{x_k}, \phi^{k_0})} &= \frac{1}{\int_0^\infty H(\sigma, \phi, \mathbf{x_n})\, d\sigma} \left(\frac{(\sigma/\phi)\pi(\sigma/\phi)}{(1/\phi)\pi(1/\phi)}\right)^{l_0+r_0} \\ &\quad \times \prod_{i=1}^n \left(\frac{(1/\sigma) f_i(x_i/\sigma)}{f_i(x_i)}\right)^{l_i+r_i}. \end{aligned} \tag{17}$$

Equation (17) is useful for the proof of result b). We now present some results.

Lemma 1. *Conditions iii) and iv) of Theorem 1 are respectively equivalent to the following equations:*

$$\begin{aligned} &\frac{\sigma\pi(\sigma \mid \mathbf{x_k}, \phi^{k_0})}{((1/\sigma)\pi(1/\sigma))^{l_0} \prod_{i=1}^n (\sigma f_i(\sigma))^{l_i}} \to 0 \textit{ as } \sigma \to 0, \textit{ and} \\ &\frac{\sigma\pi(\sigma \mid \mathbf{x_k}, \phi^{k_0})}{((1/\sigma)\pi(1/\sigma))^{r_0} \prod_{i=1}^n (\sigma f_i(\sigma))^{r_i}} \to 0 \textit{ as } \sigma \to \infty. \end{aligned}$$

Proof. The proofs for conditions iii) and iv) being similar, we only present them for the latter. We show both directions of the equivalence. Let $p = \min(\phi, \mathbf{x_k})$ and $q = \max(\phi, \mathbf{x_k})$ if $k_0 = 1$, or let $p = \min(\mathbf{x_k})$ and $q = \max(\mathbf{x_k})$ if $k_0 = 0$. We have, as $\sigma \to \infty$,

$$\begin{aligned} \frac{\sigma\pi(\sigma \mid \mathbf{x_k}, \phi^{k_0})}{((1/\sigma)\pi(1/\sigma))^{r_0} \prod_{i=1}^n (\sigma f_i(\sigma))^{r_i}} &\propto \frac{((\sigma/\phi)\pi(\sigma/\phi))^{k_0} \prod_{i=1}^n ((x_i/\sigma) f_i(x_i/\sigma))^{k_i}}{((1/\sigma)\pi(1/\sigma))^{r_0} \prod_{i=1}^n (\sigma f_i(\sigma))^{r_i}} \\ &\sim \frac{((\sigma/\phi)\pi(\sigma/\phi))^{k_0} \prod_{i=1}^n ((x_i/\sigma) f_i(x_i/\sigma))^{k_i}}{((q/\sigma)\pi(q/\sigma))^{r_0} \prod_{i=1}^n ((\sigma/q) f_i(\sigma/q))^{r_i}} \\ &\leqslant \frac{((\sigma/q)\pi(\sigma/q))^{k_0} \prod_{i=1}^n ((q/\sigma) f_i(q/\sigma))^{k_i}}{((q/\sigma)\pi(q/\sigma))^{r_0} \prod_{i=1}^n ((\sigma/q) f_i(\sigma/q))^{r_i}} \to 0 \text{ as } \sigma/q \to \infty \Leftrightarrow \sigma \to \infty. \end{aligned}$$

If we consider the opposite direction of the equivalence, we have

$$\frac{(\sigma\pi(\sigma))^{k_0}\prod_{i=1}^{n}((1/\sigma)f_i(1/\sigma))^{k_i}}{((1/\sigma)\pi(1/\sigma))^{r_0}\prod_{i=1}^{n}(\sigma f_i(\sigma))^{r_i}} \sim \frac{(\sigma\pi(\sigma))^{k_0}\prod_{i=1}^{n}((1/\sigma)f_i(1/\sigma))^{k_i}}{((1/(\sigma p)\pi(1/(\sigma p)))^{r_0}\prod_{i=1}^{n}((\sigma p)f_i(\sigma p))^{r_i}}$$
$$\leqslant \frac{((\sigma p/\phi)\pi(\sigma p/\phi))^{k_0}\prod_{i=1}^{n}((x_i/(\sigma p))f_i(x_i/(\sigma p)))^{k_i}}{((1/(\sigma p)\pi(1/(\sigma p)))^{r_0}\prod_{i=1}^{n}((\sigma p)f_i(\sigma p))^{r_i}}$$
$$\propto \frac{(\sigma p)\pi(\sigma p \mid \mathbf{x_k})}{((1/(\sigma p)\pi(1/(\sigma p)))^{r_0}\prod_{i=1}^{n}((\sigma p)f_i(\sigma p))^{r_i}} \to 0 \text{ as } \sigma p \to \infty \Leftrightarrow \sigma \to \infty.$$

Scale invariance for conflicting densities and monotonicity of the tails of $z\pi(z)$ and $zf_i(z)$ are used. □

Corollary 1. *There exists a non-decreasing step function $h_1(\sigma)$ defined on $(0,1)$ such that $h_1(\sigma) \to 0$ as $\sigma \to 0$ and*

$$\frac{\sigma\pi(\sigma \mid \mathbf{x_k}, \phi^{k_0})}{((1/\sigma)\pi(1/\sigma))^{l_0}\prod_{i=1}^{n}(\sigma f_i(\sigma))^{l_i}} \leqslant h_1(\sigma), \text{ for } \sigma < 1,$$

and there exists a non-increasing step function $h_2(\sigma)$ defined on $(1,\infty)$ such that $h_2(\sigma) \to 0$ as $\sigma \to \infty$ and

$$\frac{\sigma\pi(\sigma \mid \mathbf{x_k}, \phi^{k_0})}{((1/\sigma)\pi(1/\sigma))^{r_0}\prod_{i=1}^{n}(\sigma f_i(\sigma))^{r_i}} \leqslant h_2(\sigma), \text{ for } \sigma > 1.$$

The existence of the functions $h_1(\sigma)$ and $h_2(\sigma)$ simply arises from the definition of limit.

Proposition 3 (scale invariance) and Proposition 4 (product of random variables) can be used on the tails involved in conditions i) and ii). A corollary of Proposition 4 is given as follows.

There exists a constant $K > 1$ such that

$$\int_0^\infty (1/\sigma)f(z/\sigma)f(\sigma)/f(z)\,d\sigma < K, \text{ as } z \to \infty[0],$$

and

$$\sup_{\sigma>0} f(z/\sigma)f(\sigma)/f(z) < K, \text{ as } z \to \infty[0].$$

And by a change of variable $\sigma' = 1/\sigma$, we also have

$$\int_0^\infty (1/\sigma)f(z\sigma)f(1/\sigma)/f(z)\,d\sigma < K, \text{ as } z \to \infty[0],$$

and

$$\sup_{\sigma>0} f(z\sigma)f(1/\sigma)/f(z) < K, \text{ as } z \to \infty[0].$$

Notice that K can be chosen large enough to be valid for all involved densities.

Proof of result a) of Theorem 1

Consider an intermediate variable $\tau \to \infty$ as well as $\omega \to \infty$. We first choose a value of $\tau > 1$ as large as we want, and once τ chosen, we choose a value of ω as large as we want. The integral of result a) is divided into three parts between $0 < 1/\tau < \tau < \infty$.

If $1/\tau \leqslant \sigma \leqslant \tau$,

$$\lim_{\omega\to\infty} H(\sigma, \phi, \mathbf{x_n}) = \pi(\sigma \mid \mathbf{x_k}, \phi^{k_0}),$$

using Proposition 3 since $1/\tau \leqslant \sigma \leqslant \tau$ and $\omega \to \infty$. Notice that for a chosen τ, if ω is large enough, it means that $H(\sigma, \phi, \mathbf{x_n})$ is bounded by say $2\pi(\sigma \mid \mathbf{x_k}, \phi^{k_0})$ for $1/\tau \leqslant \sigma \leqslant \tau$, which is integrable. Therefore, we can use Lebesgue's dominated convergence theorem to pass the limit $\omega \to \infty$ inside the integral and we have

$$\lim_{\omega\to\infty} \int_{1/\tau}^{\tau} H(\sigma, \phi, \mathbf{x_n})\, d\sigma = \int_{1/\tau}^{\tau} \pi(\sigma \mid \mathbf{x_k}, \phi^{k_0})\, d\sigma \to 1 \text{ as } \tau \to \infty.$$

If $\sigma \geqslant \tau$ and $r_0 + r > 0$, then

$$\begin{aligned}
&\sigma H(\sigma, \phi, \mathbf{x_n})\\
&\leqslant \sigma\pi(\sigma \mid \mathbf{x_k}, \phi^{k_0}) \left(\frac{(\sigma/\phi)\pi(\sigma/\phi)}{(1/\phi)\pi(1/\phi)}\right)^{r_0} \prod_{i=1}^{n} \left(\frac{(x_i/\sigma) f_i(x_i/\sigma)}{x_i f_i(x_i)}\right)^{r_i}\\
&= \frac{\sigma\pi(\sigma \mid \mathbf{x_k}, \phi^{k_0})}{((1/\sigma)\pi(1/\sigma))^{r_0} \prod_{i=1}^{n} (\sigma f_i(\sigma))^{r_i}} \left(\frac{\pi(\sigma/\phi)\pi(1/\sigma)}{\pi(1/\phi)}\right)^{r_0} \prod_{i=1}^{n} \left(\frac{f_i(x_i/\sigma) f_i(\sigma)}{f_i(x_i)}\right)^{r_i}\\
&\leqslant h_2(\tau) \left(\frac{\pi(\sigma/\phi)\pi(1/\sigma)}{\pi(1/\phi)}\right)^{r_0} \prod_{i=1}^{n} \left(\frac{f_i(x_i/\sigma) f_i(\sigma)}{f_i(x_i)}\right)^{r_i}\\
&\leqslant h_2(\tau) K^{r+1} \to 0 \text{ as } \tau \to \infty.
\end{aligned}$$

Monotonicity of the tails of $z\pi(z)$ and $zf_i(z)$ is used in the first equality for any ω larger than a certain constant since $\sigma \geqslant \tau > 1$. In the second inequality, we use $h_2(\sigma) \leqslant h_2(\tau)$ for $\sigma \geqslant \tau > 1$ from condition iv). Proposition 4 is used in the last inequality. Similarly, we have

$$\begin{aligned}
&\int_{\tau}^{\infty} H(\sigma, \phi, \mathbf{x_n})\, d\sigma\\
&\qquad \leqslant h_2(\tau) \int_0^{\infty} \frac{1}{\sigma} \left(\frac{\pi(\sigma/\phi)\pi(1/\sigma)}{\pi(1/\phi)}\right)^{r_0} \prod_{i=1}^{n} \left(\frac{f_i(x_i/\sigma) f_i(\sigma)}{f_i(x_i)}\right)^{r_i} d\sigma\\
&\qquad \leqslant h_2(\tau) K^{r+1} \to 0 \text{ as } \tau \to \infty.
\end{aligned}$$

If $\sigma \geqslant \tau$ and $r_0 + r = 0$, we have

$$\sigma H(\sigma, \phi, \mathbf{x_n}) \leqslant \sigma\pi(\sigma \mid \mathbf{x_k}, \phi^{k_0}) \to 0 \text{ as } \sigma \geqslant \tau \to \infty,$$

and

$$\int_{\tau}^{\infty} H(\sigma, \phi, \mathbf{x_n})\, d\sigma \leqslant \int_{\tau}^{\infty} \pi(\sigma \mid \mathbf{x_k}, \phi^{k_0}) d\sigma \to 0 \text{ as } \tau \to \infty.$$

If $\sigma \leqslant 1/\tau$ and $l_0 + l > 0$, then

$$\begin{aligned}
&\sigma H(\sigma, \phi, \mathbf{x_n}) \\
&\leqslant \sigma\pi(\sigma \mid \mathbf{x_k}, \phi^{k_0}) \left(\frac{(\sigma/\phi)\pi(\sigma/\phi)}{(1/\phi)\pi(1/\phi)}\right)^{l_0} \prod_{i=1}^{n} \left(\frac{(x_i/\sigma) f_i(x_i/\sigma)}{x_i f_i(x_i)}\right)^{l_i} \\
&= \frac{\sigma\pi(\sigma \mid \mathbf{x_k}, \phi^{k_0})}{((1/\sigma)\pi(1/\sigma))^{l_0} \prod_{i=1}^{n} (\sigma f_i(\sigma))^{l_i}} \left(\frac{\pi(\sigma/\phi)\pi(1/\sigma)}{\pi(1/\phi)}\right)^{l_0} \prod_{i=1}^{n} \left(\frac{f_i(x_i/\sigma) f_i(\sigma)}{f_i(x_i)}\right)^{l_i} \\
&\leqslant h_1(1/\tau) \left(\frac{\pi(\sigma/\phi)\pi(1/\sigma)}{\pi(1/\phi)}\right)^{l_0} \prod_{i=1}^{n} \left(\frac{f_i(x_i/\sigma) f_i(\sigma)}{f_i(x_i)}\right)^{l_i} \\
&\leqslant h_1(1/\tau) K^{l+1} \to 0 \text{ as } \tau \to \infty.
\end{aligned}$$

Monotonicity of the tails of $z\pi(z)$ and $zf_i(z)$ is used in the first equality for any ω larger than a certain constant since $\sigma \leqslant 1/\tau < 1$. In the second inequality, we use $h_1(\sigma) \leqslant h_1(1/\tau)$ for $\sigma \leqslant 1/\tau < 1$ from condition iii). Proposition 4 is used in the last inequality. Similarly, we have

$$\begin{aligned}
&\int_{0}^{1/\tau} H(\sigma, \phi, \mathbf{x_n})\, d\sigma \\
&\qquad \leqslant h_1(1/\tau) \int_{0}^{\infty} \frac{1}{\sigma} \left(\frac{\pi(\sigma/\phi)\pi(1/\sigma)}{\pi(1/\phi)}\right)^{l_0} \prod_{i=1}^{n} \left(\frac{f_i(x_i/\sigma) f_i(\sigma)}{f_i(x_i)}\right)^{l_i} d\sigma \\
&\qquad \leqslant h_1(1/\tau) K^{l+1} \to 0 \text{ as } \tau \to \infty.
\end{aligned}$$

If $\sigma \leqslant 1/\tau$ and $l_0 + l = 0$, we have

$$\sigma H(\sigma, \phi, \mathbf{x_n}) \leqslant \sigma\pi(\sigma \mid \mathbf{x_k}, \phi^{k_0}) \to 0 \text{ as } \sigma \leqslant 1/\tau \to 0,$$

and

$$\int_{0}^{1/\tau} H(\sigma, \phi, \mathbf{x_n})\, d\sigma \leqslant \int_{0}^{1/\tau} \pi(\sigma \mid \mathbf{x_k}, \phi^{k_0}) d\sigma \to 0 \text{ as } \tau \to \infty.$$

Proof of result b) of Theorem 1

From equation (17), we find that

$$\frac{\pi(\sigma \mid \mathbf{x_n}, \phi)}{\pi(\sigma \mid \mathbf{x_k}, \phi^{k_0})} \to 1 \text{ as } \omega \to \infty,$$

for any $\sigma > 0$, using result a) of Theorem 1 and Proposition 3.

Proof of result c) of Theorem 1

From equation (16) and result a), we have

$$\sigma\pi(\sigma \mid \mathbf{x_n}, \phi) = \sigma H(\sigma, \phi, \mathbf{x_n}) \Big/ \int_0^\infty H(\sigma, \phi, \mathbf{x_n})\, d\sigma$$
$$\sim \sigma H(\sigma, \phi, \mathbf{x_n}) \text{ as } \omega \to \infty.$$

And as shown in the proof of result a),

$$\sigma H(\sigma, \phi, \mathbf{x_n}) \to 0 \text{ as } \omega \to \infty \text{ and } \sigma \to 0 \text{ or } \sigma \to \infty,$$

at any given rate.

Proof of result d) of Theorem 1

We first write result d) as follows. $\forall d > 0$, we have, as $\omega \to \infty$,

$$\left| \int_d^\infty \pi(\sigma \mid \mathbf{x_n}, \phi)\, d\sigma - \int_d^\infty \pi(\sigma \mid \mathbf{x_k}, \phi^{k_0})\, d\sigma \right| \to 0.$$

We also write result b) as follows. $\forall \tau > 1$, we have, as $\omega \to \infty$,

$$1/\tau \leqslant \sigma \leqslant \tau \Rightarrow \pi(\sigma \mid \mathbf{x_n}, \phi)/\pi(\sigma \mid \mathbf{x_k}, \phi^{k_0}) \to 1.$$

We choose any fixed $d > 0$. Then consider an intermediate variable $\tau \to \infty$ as well as $\omega \to \infty$. We choose a value of τ as large as we want, and once τ is chosen, we choose a value of ω as large as we want, to make the difference in absolute values as close as we want to 0. In particular, we choose $\tau \geqslant \max(1/d, d) \Leftrightarrow 1/\tau \leqslant d \leqslant \tau$, which means that $(d, \tau) \in (1/\tau, \tau)$.

Firstly, since $\pi(\sigma \mid \mathbf{x_k}, \phi^{k_0})$ is a proper density, we have,

$$\int_\tau^\infty \pi(\sigma \mid \mathbf{x_k}, \phi^{k_0})\, d\sigma \to 0 \text{ as } \tau \to \infty.$$

Secondly, considering that $\pi(\sigma \mid \mathbf{x_k}, \phi^{k_0})$ and $\pi(\sigma \mid \mathbf{x_n}, \phi)$ are proper densities and using result b), we have, as $\omega \to \infty$,

$$\int_\tau^\infty \pi(\sigma \mid \mathbf{x_n}, \phi)\, d\sigma \leqslant 1 - \int_{1/\tau}^\tau \pi(\sigma \mid \mathbf{x_n}, \phi)\, d\sigma \sim 1 - \int_{1/\tau}^\tau \pi(\sigma \mid \mathbf{x_k}, \phi^{k_0})\, d\sigma$$
$$\to 0, \text{ as } \tau \to \infty.$$

Thirdly, using result b), we have, as $\omega \to \infty$,

$$\begin{aligned}
\left| \int_d^\tau \pi(\sigma \mid \mathbf{x_n}, \phi)\, d\sigma - \int_d^\tau \pi(\sigma \mid \mathbf{x_k}, \phi^{k_0})\, d\sigma \right| \\
&\leqslant \int_d^\tau \left| \pi(\sigma \mid \mathbf{x_n}, \phi) - \pi(\sigma \mid \mathbf{x_k}, \phi^{k_0}) \right| \, d\sigma \\
&= \int_d^\tau \pi(\sigma \mid \mathbf{x_k}, \phi^{k_0}) \left| \pi(\sigma \mid \mathbf{x_n}, \phi)/\pi(\sigma \mid \mathbf{x_k}, \phi^{k_0}) - 1 \right| \, d\sigma \\
&\leqslant \left| \pi(\sigma^* \mid \mathbf{x_n}, \phi)/\pi(\sigma^* \mid \mathbf{x_k}, \phi^{k_0}) - 1 \right| \int_d^\tau \pi(\sigma \mid \mathbf{x_k}, \phi^{k_0})\, d\sigma \\
&\leqslant \left| \pi(\sigma^* \mid \mathbf{x_n}, \phi)/\pi(\sigma^* \mid \mathbf{x_k}, \phi^{k_0}) - 1 \right| \to 0, \text{ as } \omega \to \infty,
\end{aligned}$$

for a $\sigma^* \in (d, \tau) \in (1/\tau, \tau)$.

Combining these three results, we have

$$\begin{aligned}
\left| \int_d^\infty \pi(\sigma \mid \mathbf{x_n}, \phi)\, d\sigma - \int_d^\infty \pi(\sigma \mid \mathbf{x_k}, \phi^{k_0})\, d\sigma \right| \\
&\leqslant \left| \int_d^\tau \pi(\sigma \mid \mathbf{x_n}, \phi)\, d\sigma - \int_d^\tau \pi(\sigma \mid \mathbf{x_k}, \phi^{k_0})\, d\sigma \right| \\
&+ \int_\tau^\infty \pi(\sigma \mid \mathbf{x_n}, \phi)\, d\sigma + \int_\tau^\infty \pi(\sigma \mid \mathbf{x_k}, \phi^{k_0})\, d\sigma \\
&\to 0 \text{ as } \tau, \omega \to \infty.
\end{aligned}$$

References

Andrade, J. A. A. and O'Hagan, A. (2006). "Bayesian robustness modeling using regularly varying distributions." *Bayesian Analysis*, 1: 169–188. 187, 188, 193

Andrade, J. A. A. and O'Hagan, A. (2011). "Bayesian robustness modelling of location and scale parameters." *Scandinavian Journal of Statistics*, 38: 691–711. 188

Angers, J.-F. (2000). "P-credence and outliers." *Metron*, 58: 81–108. 187, 201

Dawid, A. P. (1973). "Posterior expectations for large observations." *Biometrika*, 60: 664–667. 187

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 1, 199–210. Berkeley: University of California Press. 187

Desgagné, A. and Angers, J.-F. (2005). "Importance sampling with the generalized exponential power density." *Statistics and Computing*, 15: 189–195. 201

— (2007). "Conflicting information and location parameter inference." *Metron*, 65: 67–97. 187

Karamata, J. (1930). "Sur un mode de croissance régulière des fonctions." *Mathematica (Cluj)*, 4: 38–53. 189

O'Hagan, A. (1979). "On outlier rejection phenomena in Bayes inference." *Journal of the Royal Statistical Society, Series B*, 41: 358–367. 187

— (1988). "Modelling with heavy tails." In Bernardo, J. M., DeGroot, M. H., Lindley, D. V., and Smith, A. F. M. (eds.), *Bayesian Statistics 3*, 345–359. Oxford: Clarendon Press. 187

— (1990). "Outliers and credence for location parameter inference." *Journal of the American Statistical Association*, 85: 172–176. 187, 190