JEAN-FRANÇOIS ANGERS (*)

P-credence and outliers

CONTENTS:

1. INTRODUCTION

If x is an observation from a normal population with mean θ and variance 1, and if the prior on θ is $\mathcal{N}(0, 1)$, then its posterior is $\mathcal{N}(x/2, 1/2)$. Elementary introductions to Bayesian statistics often present this example to illustrate how the prior and the likelihood information are synthesized by Bayes' theorem. This compromise seems natural enough when both sources of information agree but it is questionable when the two sources are conflicting.

Normal theory models, and all exponential family models with conjugate prior, have this property of compromising between the various sources of information, even when they conflict. Replacing the normal prior by a prior with heavier tails, such as the Student-t distribution or the Laplace distribution causes the posterior to ignore most (or all) of the prior information when it disagrees with the data.

In the simplest case, we can consider a single observation x with mean θ and suppose that the prior mean is μ. A conflict between both sources of information occurs when $|x - \mu| \to \infty$. This situation was studied by Dawid (1973); Hill (1974). O'Hagan (1979) considered outlier rejection in a sample and O'Hagan (1988) considered more general Bayesian modeling based on Student-t distributions. The notion

(*) Département de mathématiques et de statistique, Université de Montréal, C.P. 6128, Succ. "Centre-ville", Montréal, Québec H3C 3J7 E-mail: angers@dms.umontreal.ca

of credence was first introduced in O'Hagan (1990). The use of heavy-tailed distributions is therefore a valuable tool in developing robust Bayesian procedures, limiting the influence that extreme information sources can have on posterior inferences, see for instance Meinhold and Singpurwalla (1989); Angers and Berger (1991); Carlin and Polson (1991); Angers (1992); Fan and Berger (1992); Geweke (1994); Angers (1996).

In Section 2 of this paper, we define the generalized exponential power distribution and redefine the notion of credence to account for the logarithmic, polynomial and exponential tail behavior. This new definition is called the p-credence. In Section 3, the p-credence of the marginal and the posterior distribution are studied. In Section 4, relationships between the behavior of the Bayes estimator in presence of a conflict between the two sources of information and p-credence of the likelihood and of the prior are established. It is shown that if the power of the exponential part of the prior is less than the minimum between the one of the likelihood and 1, then the posterior mean collapses to the observation. In Section 5, the case with several observations is considered. Conditions on the prior and the density of the X_i's are given in order for the posterior to ignore the outlying observations if there is any in the sample. In the last section, examples are presented to illustrate the results given in this paper.

2. Generalized exponential power density and p-credence

In Box and Tiao (1962), the exponential power density was introduced to provide a heavy-tailed alternative to the normal density. This density is given by

$$\pi(\theta) \propto \exp\left\{-\frac{1}{2}|\theta|^{\frac{2}{1+\beta}}\right\}, \text{ for } -1 < \beta \leq 1\,. \tag{1}$$

However, as shown in O'Hagan (1979), the exponential power family is outlier prone only for $\beta > 1$, that is if $2/[1+\beta] < 1$. In our generalization, we parameterize, so that, in effect we allow $-1 < \beta \leq \infty$ with the special case $\beta = \infty$ being important (see equation (3) below). In O'Hagan and Le (1994), a class of bivariate heavy-tailed distributions was introduced. Under some conditions, the marginal density of one component has tail behavior like

$$\frac{\log(2+\theta^2)}{(1+\theta^2)^\alpha}, \quad \text{for } \alpha > 0\,. \tag{2}$$

Combining equations (1) and (2), we can construct a class of distributions which encompasses all these kinds of tail behavior. A density from this class, called the generalized exponential power family, is given by

$$\begin{aligned} p(z|\gamma,\delta,\alpha,\beta) \propto{}& \max(|z|,z_0)^\alpha \log^\beta[\max(|z|,z_0)] \\ &\times \exp\{-\delta \max(|z|,z_0)^\gamma\}\ , \end{aligned} \tag{3}$$

where $\gamma \geq 0$, $\delta > 0$ if $\gamma > 0$ and by convention we define $\delta = 0$ if $\gamma = 0$. The parameters γ, δ, α, β and z_0 must satisfy the following conditions:

i) $z_0 > 1$ if $\beta \neq 0$ or $z_0 > 0$ if $\alpha < 0, \beta = 0$;

ii) $\alpha + \frac{\beta}{\log(z_0)} \leq \delta\gamma z_0^\gamma$;

iii) $\alpha \leq -1$ if $\gamma = 0$;

iv) $\beta < -1$ if $\gamma = 0$ and $\alpha = -1$.

Note that the first condition is needed in order for the density to be nonnegative when the logarithmic part is present and to avoid a singularity at $z = 0$ when $\alpha < 0$. The second condition guarantees the unimodality of $p(z|\gamma,\delta,\alpha,\beta)$ and it is always satisfied if z_0 is chosen large enough. If it is violated, the density is bimodal, each mode being symmetrical with respect to the origin. Since we are really only interested in the tails of the density, the fact that $p(z|\gamma,\delta,\alpha,\beta)$ is constant for $-z_0 < z < z_0$ is unimportant. (Note that the parameter z_0 is not listed as an argument because it is not an important parameter for this paper as will be seen in Lemma 2.) The third and fourth conditions ensure that it is a proper density. (They can be omitted if improper densities are of interest.) It should also be noted that allowing γ to be negative would provide no more generality in the the tail behavior of $p(z|\gamma,\delta,\alpha,\beta)$. In Figure 1, different forms of $p(z|\gamma,\delta,\alpha,\beta)$ are presented. In the Figure 1-a, the behavior of $p(z|\gamma,\delta,\alpha,\beta)$ as a function of z_0 is illustrated while the other parameters are set equal to $\gamma = 2$, $\delta = 1/2$, $\alpha = \beta = 1$. In this graph, the solid line corresponds to $z_0 = 1$, the dotted line, to $z_0 = 1.2$ and the dashed line, to $z_0 = 1.7$.

From this graph, we can see that z_0 affects only the middle part of the density and all three densities have the same tails. The influence of z_0 is also illustrated in Figure 1-b but for $\gamma = \delta = \beta = 0$ and $\alpha = -2$. It can be seen that if $z_0 = 0$ (solid line), then there is an infinite mode at 0, while for a non zero value of z_0 ($z_0 = 1$ for the dotted line), the density is bounded. In Figure 1-c, two different tails behavior are illustrated. The solid line represents exponential tails with $\gamma = 2$, $\delta = 1/2$, $\alpha = \beta = 0$ and the dotted line represents polynomial tails with $\gamma = \delta = \beta = 0$ and $\alpha = -2$.

In O'Hagan (1990), the notion of credence was first introduced and it was defined as the following.

Definition 1. *A density f on $\mathbb{R}$ has credence α if there exist constants $0 < k < K < \infty$ such that, for all $z \in \mathbb{R}$*

$$k \leq (1 + z^2)^{\alpha/2} f(z) \leq K \, . \tag{4}$$

This notion of credence characterizes the tail of a density by comparing it to a Student density. Essentially, equation (4) ensures that $f(z)$ is of order $|z|^{-\alpha}$ for large values of $|z|$. Note that, even if they are outlier prone, all members of the exponential power family (*cf.* equation (1)) have infinite credence. In order to take these densities into account, the notion of credence given in Definition 1 has to be generalized. Using $p(z|\gamma, \delta, \alpha, \beta)$ as a reference to classify the tails behavior of a density, we can redefine the notion of credence as follows.

Definition 2. *A density f on $\mathbb{R}$ has p-credence $(\gamma, \delta, \alpha, \beta)$, denoted by* p-cred$(f) = (\gamma, \delta, \alpha, \beta)$, *if there exist constants $0 < k \leq K \leq \infty$ such that for all $z \in \mathbb{R}$*

$$k \leq \frac{f(z)}{p(z|\gamma, \delta, \alpha, \beta)} \leq K \, ,$$

where $p(z|\gamma, \delta, \alpha, \beta)$ is defined in equation (3). *We also write* p-cred$(Z) = (\gamma, \delta, \alpha, \beta)$ *if the distribution of Z has p-credence $(\gamma, \delta, \alpha, \beta)$.*

Using Definition 2, the p-credence of the exponential power density given by equation (1) is $(2/(1 + \beta), 1, 0, 0)$ and the p-credence of the density given by equation (2) is $(0, 0, -\alpha, 1)$. Note that most of the usual symmetric densities on $\mathbb{R}$ (such as the normal, Student-t,

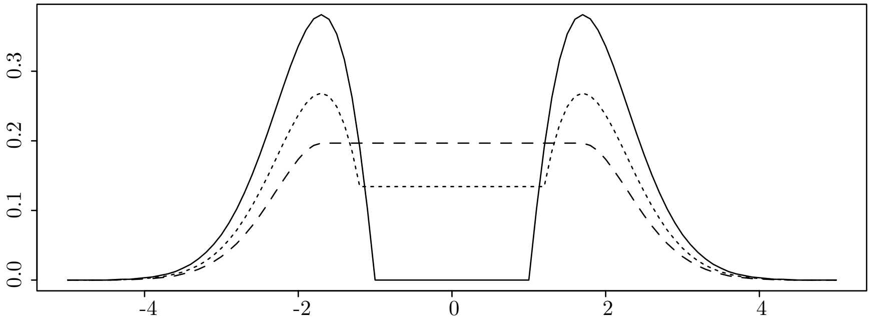

Figure 1-a

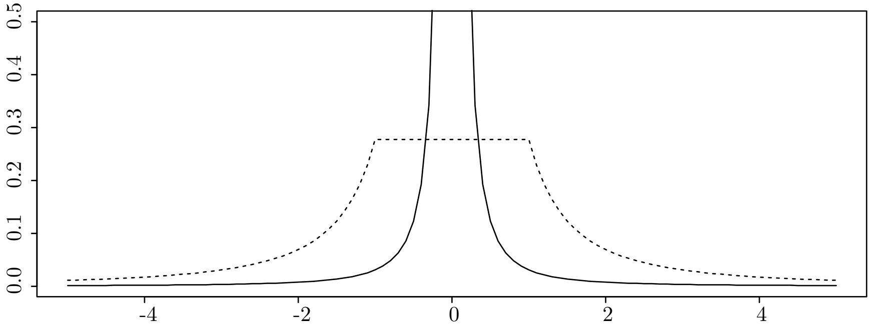

Figure 1-b

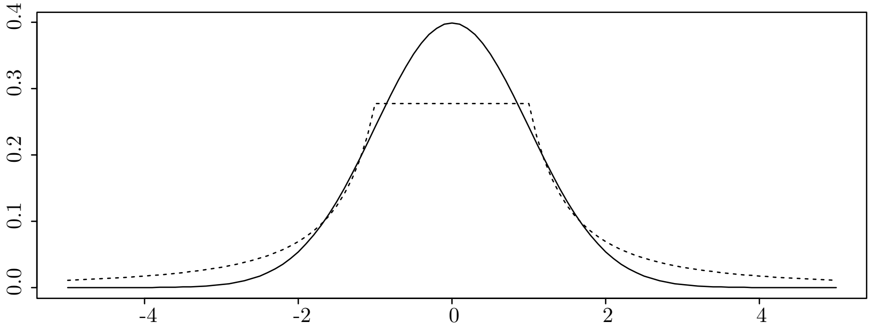

Figure 1-c

Fig. 1. Some examples of generalized exponential power densities.

Laplace, logistic densities) are cover by Definition 2. However, the p-credence for nonsymmetric densities (such as extreme value density) is not defined.

By using the notion of p-credence, a dominance relation can be established to compare two densities on $\mathbb{R}$ (*cf.* O'Hagan and Le, 1994) It should be noted that this dominance relation does not provide a complete ordering of the densities on $\mathbb{R}$ since it is possible to find two densities such that neither dominates the other. However, the following definition is sufficient to characterize the tails behavior for symmetric densities and allow us to consider the behavior of the posterior in presence of an outlier.

Definition 3. *Let f and g be any two densities on $\mathbb{R}$. We say that*

$i)$ *f dominates g, denoted by $f \succeq g$, if there exists a constant $k > 0$ such that*

$$f(z) \geq kg(z)\, \forall\, z \in \mathbb{R}\,;$$

$ii)$ *f is equivalent to g, denoted by $f \approx g$, if both $f \succeq g$ and $g \succeq f$;*

$iii)$ *f strictly dominates g, denoted by $f \succ g$, if $f \succeq g$ but $g \not\succeq f$.*

As example, let f be the Laplace density with parameter $\sqrt{2}$ and g, the standard normal density, that is

$$f(x) = \frac{1}{\sqrt{2}} \exp\{-\sqrt{2}|x|\}\,,$$
$$g(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{\frac{1}{2}x^2\right\}\,,$$

$\forall x \in \mathbb{R}$. It can be shown that

$$f(x) \geq (\sqrt{\pi}\mathrm{e}^{-1})g(x)\, \forall x \in \mathbb{R}\,.$$

Hence $f \succeq g$. However, there exist no constant k such that $g(x) \geq kf(x)\ \forall x$. Consequently, f strictly dominates g.

Note that by p-cred$(f) = (\gamma, \delta, \alpha, \beta)$, it is meant that $f \approx p(\cdot|\gamma, \delta, \alpha, \beta)$ where $p(\cdot|\gamma, \delta, \alpha, \beta)$ is defined by equation (3). These densities are ordered by the dominance relation as shown in the next proposition.

Proposition 1. *Let f and g be two densities on $\mathbb{R}$ such that* p-cred(f) $= (\gamma, \delta, \alpha, \beta)$ *and* p-cred$(g) = (\gamma', \delta', \alpha', \beta')$, *then*

$i)$ *$f \approx g$ if $\gamma' = \gamma$, $\delta' = \delta$, $\alpha' = \alpha$ and $\beta' = \beta$;*

$ii)$ *$f \succ g$ if:*

  $a)$ $\gamma' > \gamma$;

b) $\gamma' = \gamma,\ \delta' > \delta;$
c) $\gamma' = \gamma,\ \delta' = \delta,\ \alpha' < \alpha;$
d) $\gamma' = \gamma,\ \delta' = \delta,\ \alpha' = \alpha,\ \beta' < \beta.$

To prove Proposition 1, one only has to compute $p(z|\gamma, \delta, \alpha, \delta)/p(z|\gamma', \delta', \alpha', \delta')$ and to study the sign of its derivative. Being mostly algebraic manipulation, this proof is omitted.

To conclude this section and before we discuss the marginal and posterior p-credences, three lemmas and one general theorem are presented.

Lemma 1. *If* $c \neq 0, d \in \mathbb{R}$, *then* $\forall z \in \mathbb{R}$,

i) $\exists\, 0 < a \leq A < \infty$ *such that*

$$a \leq \frac{\max(|z|, z_0)}{\max(|cz + d|, z_0)} \leq A,$$

ii) $\exists\, 0 < b \leq B < \infty$ *such that*

$$b \leq \frac{\log[\max(|z|, z_0)]}{\log[\max(|cz + d|, z_0)]} \leq B\,.$$

Lemma 2. *Let* z_0 *and* z_1 *be two positive constants.* (*Note that* $\min(z_0, z_1)$ *should be larger than 1 if* $\beta \neq 0$ *and larger than 0 if* $\alpha < 0, \beta = 0$.) *Then* $\forall z \in \mathbb{R}, \exists\, 0 < a \leq A < \infty$ *such that*

$$\begin{aligned} a \leq &\left[\frac{\max(|z|, z_0)}{\max(|z|, z_1)}\right]^{\alpha} \left[\frac{\log[\max(|z|, z_0)]}{\log[\max(|z|, z_1)]}\right]^{\beta} \\ &\times \exp\{-\delta[\max(|z|, z_0)^{\gamma} - \max(|z|, z_1)^{\gamma}]\} \leq A\,. \end{aligned}$$

The proofs of these two lemmas are mostly technical and thus, they are omitted.

Lemma 3. *If* $\gamma \leq 1, d \in \mathbb{R}$, *then* $\exists\, 0 < a \leq A < \infty$ *such that*

$$a \leq \exp\{-\delta[\max(|z + d|, z_0)^{\gamma} - \max(|z|, z_0)^{\gamma}]\} \leq A,\ \forall z \in \mathbb{R}\,.$$

The proof of Lemma 3 is given in the Appendix.

Using Lemmas 1, 2 and 3, the following theorem can be shown.

THEOREM 1. *a) If* p-cred$(X) = (\gamma, \delta, \alpha, \beta)$*, then for all* $c \neq 0$, p-cred$(cX) = (\gamma, \delta/|c|^\gamma, \alpha, \beta)$.

b) In addition, if $\gamma \leq 1$*, then for all* $d \in \mathbb{R}$, p-cred$(cX + d) = (\gamma, \delta/|c|^\gamma, \alpha, \beta)$.

It should be noted that this theorem is slightly different from Theorem 1 of O'Hagan (1990) which states that the credence is invariant under linear transformations. This is due to the fact that the exponential part of equation (3) depends on the scale parameter for any positive value of γ and on the location parameter if $\gamma > 1$. When $\gamma = \delta = 0$, we obtain the same result as Theorem 1 of O'Hagan (1990).

3. Marginal and posterior p-credence

In this section, we study the p-credence of the marginal and posterior distributions. Throughout this paper, it is assumed that

$$\theta \sim \pi(\theta),\ \text{p-cred}(\theta) = \text{p-cred}(\pi) = (\gamma, \delta, \alpha, \beta),$$
$$X|\theta \sim f(x-\theta),\ \text{p-cred}(X-\theta|\theta) = \text{p-cred}(f) = \left(\gamma', \delta', \alpha', \beta'\right).$$

In the next theorem, the marginal p-credence of X is given.

THEOREM 2. *If* p-cred$(X-\theta|\theta) = (\gamma', \delta', \alpha', \beta')$ *and* p-cred$(\theta) = (\gamma, \delta, \alpha, \beta)$*, then, the marginal of* X*, denoted by* $m(x)$*, is such that*

$$m \approx \begin{cases} \pi; & \text{if } \pi \succeq f, \\ f; & \text{if } f \succ \pi. \end{cases}$$

The proof of this theorem is given in the Appendix.

Unlike O'Hagan (1990), the p-credence of the posterior cannot be expressed easily. However, there still exist members of the generalized exponential power family which bound the posterior. They are given in the next theorem. Using these bounds, bounds on the posterior expected mean and the posterior expected loss can be found.

THEOREM 3. *If* p-cred$(X-\theta|\theta) = (\gamma', \delta', \alpha', \beta')$, p-cred$(\theta) = (\gamma, \delta, \alpha, \beta)$ *then*

i) *if* $\min(\gamma, \gamma') = 0$,

$$\pi(\theta|x) \approx p(\theta - x_{**}|\gamma_{**}, \delta_{**}, \alpha+\alpha', \beta+\beta'),$$

ii) *if* $\min(\gamma, \gamma') > 0$,

$$p\left(\theta - x_{**}|\gamma_{**}, \delta_{**} + \delta_* \frac{(1+|x_{**}|/z_0)^{\gamma_*}}{z_0^{\gamma_{**}}}, \alpha+\alpha', \beta+\beta'\right) \preceq \pi(\theta|x)$$
$$\preceq p\left(\theta - x_{**}|\gamma_{**}, \delta_{**}, \alpha+\alpha', \beta+\beta'\right)$$

where

$$\gamma_* = \min(\gamma, \gamma'), \gamma_{**} = \max(\gamma, \gamma'),$$
$$\delta_* = \begin{cases} \delta'; & \text{if } f \succ \pi, \\ \delta; & \text{if } \pi \succ f, \end{cases} \delta_{**} = \begin{cases} \delta; & \text{if } f \succ \pi, \\ \delta'; & \text{if } \pi \succ f, \end{cases} x_{**} = \begin{cases} 0; & \text{if } f \succ \pi, \\ x; & \text{if } \pi \succ f. \end{cases}$$

The proof of Theorem 3 is given in the Appendix. Note that if $z_0 = c|x|^a$, where $a > \gamma_*/(\gamma_* + \gamma_{**})$, then the lower bound of part *ii*) of Theorem 3 converges toward the upper bound as $|x| \to \infty$.

4. Conflicting information

In this section, the behavior of the posterior density and the posterior expectation are investigated when $|x|$ is large, that is, when there is a conflict between the prior and the likelihood information. As mentioned in the introduction, in order to get an estimator for θ which is insensitive to prior misspecification, the prior tails have to be heavier than those of the likelihood. In this section, it is supposed that the influence of the prior should be limited in case of a conflict between the two sources of information. Hence, it is assumed that $\pi \succ f$. However, because of the symmetry between π and f, the case where one wants to limit the influence of the likelihood, that is, if $f \succ \pi$, can easily be inferred. This case is considered in the next section. In the next theorem, it is shown that, if $|x|$ is large, then under some condition on γ, the posterior and the likelihood are equivalent. Hence if some conditions are satisfied, the posterior density ignores all the information contained in the prior if there is a conflict between the two sources of information.

THEOREM 4. *If* $\pi \succeq f$ *and if* $\gamma < 1$, *then*

a) $\forall d > 0, \exists D > 0$ *such that* $\forall |x| > D, \forall |\theta - x| < d$,

$$\frac{k\underline{k}}{K_*}\left(1 - d\left|\frac{\alpha}{x}\right|\right)\left(1 - d\left|\frac{\beta}{(x-d_\beta)\log(x-d_\beta)}\right|\right)\left(1 - d\frac{\delta\gamma}{|x|^{1-\gamma}}\right)$$
$$\leq \frac{\pi(\theta|x)}{f(x-\theta)}$$
$$\leq \frac{K\overline{K}}{k_*}\left(1+d\left|\frac{\alpha}{x}\right|\right)\left(1+d\left|\frac{\beta}{(x-d_\beta)\log(x-d_\beta)}\right|\right)\left(1+d\frac{\delta\gamma}{|x|^{1-\gamma}}\right),$$

where k, k_*, $\underline{k}$, K, K_* *and* $\overline{K}$ *are appropriate constants and* $d_\beta = 0$ *if* $\beta \geq 0$ *and it is equal to* d *if* $\beta < 0$.

b) $\forall\, \theta_0 > 0$, $\lim_{|x|\to\infty} \Pr\{|\theta| < \theta_0 | X = x\} = 0$.

The proof of Theorem 4 is given in the Appendix. Note that part a) of Theorem 4 is still valid even if $\gamma \geq 1$. However, this does not imply that $\pi(\theta|x) \approx f(x-\theta)$. In Theorem 4, it is shown that if the tails of the prior density is flat enough (compare to the one of the density of $X|\theta$), then for large values of x, the posterior density is equivalent of $f(x-\theta)$ in any neighborhood of x. Hence, in case of a conflict between the prior and the likelihood information, the posterior density will ignore totally the first one.

If the likelihood has at least p moments, the following theorem is also proved in the Appendix.

THEOREM 5. *If* $\pi \succeq f$, $\gamma < \min(1, \gamma')$ *if* $\gamma' > 0$ *or* $\alpha' < \alpha - p$ *if* $\gamma' = 0$ *and if* $|x| > D$, *where* $D > 2^{\gamma'/(\gamma'-\gamma)}$ *if* $\gamma' > 0$ *and* $D > 2^{(|\alpha'|-1)/|\alpha|}$ *if* $\gamma' = 0$, *then*

$$\mathbb{E}^{\pi(\theta|x)}[\theta^p] = x^p + \begin{cases} o(|x|); & \text{if } \gamma' > 0, \\ O(|x|^{-(\alpha-\alpha'-p)}\log^{\beta'-\beta}(|x|)); & \text{if } \gamma' = 0. \end{cases}$$

Theorem 5 states that, in case of conflict between the prior and the likelihood information, the Bayes rule (under the squared error loss) will collapse to x if the tails of the prior are flat enough. From this theorem, we can obtain directly the following corollary.

Corollary 1. *If the conditions of Theorem* 5 *are satisfied and if any one of the following three additional conditions are satisfied,*

$i)$ $\gamma' > 0$,
$ii)$ $\gamma' = 0, \alpha' < \alpha - p$,
$iii)$ $\gamma' = 0, \alpha' = \alpha - p, \beta > \beta'$,
then $\lim_{|x|\to\infty}(x^p - \mathbb{E}^{\pi(\theta|x)}[\theta^p]) = 0$.

The proof of this corollary is a straightforward application of Theorem 5. Thus, it is omitted.

The notion of p-credence is applicable to a broader class of densities than the notion of credence proposed in O'Hagan (1990). Since the credence of any exponential power density (*cf.* equation (1)) is infinite, Theorems 5 and 6 of O'Hagan (1990) (which are equivalent to Theorems 4 and 5 here) cannot be used to distinguish between the outlier resistant densities ($\beta \leq 1$) and the outlier prone density ($\beta > 1$). However, this can be done using p-credence. Since the p-credence of equation (1) is $(\frac{2}{1+\beta}, \frac{1}{2}, 0, 0)$, using Theorems 4 and 5, it can be shown that the exponential power density is outlier prone if

$$\gamma = \frac{2}{1+\beta} < 1 \Leftrightarrow \beta > 1\,.$$

5. Outlier elimination

Let $x_1, x_2, \ldots, x_n$ be n independent observations with density $f(x-\theta)$. It is assumed that this density does not have any sufficient statistic beside the whole sample itself. Suppose also that p-cred$(X_i - \theta \mid \theta) = (\gamma', \delta', \alpha', \beta')$ for all $i = 1, 2, \ldots, n$ and that p-cred$(\theta) = (\gamma, \delta, \alpha, \beta)$. In this section, the behavior of the posterior density is considered when one observation, say x_n is an outlier, that is when $|x_n| \gg \max_{1\leq i\leq n-1}|x_i|$. To ease the notation, it is assume that $x_n > 0$ and $x_n > x_i$ for all $i = 1, 2, \ldots, n-1$.

In order to eliminate the influence of an outlying observation on the posterior, the density should be such that, in case of conflict between it and the prior, the posterior ignores the information contained in the observation. Hence, according to Section 4, f and π should be such that $f \succeq \pi$. Under this conditions, the posterior and the prior are equivalent when $|x|$ is large, as stated in the next theorem.

THEOREM 6. *Let $X \mid \theta \sim f(x-\theta)$ and $\theta \sim \pi(\theta)$. If $f \succeq \pi$ and if $\gamma' < 1$, then $\forall d > 0, \exists D > d + z_0$ such that $\forall |x| > D$ and $\forall |\theta| < d$,*

$$\begin{aligned}
&\frac{k\underline{k}}{K_*}\left(1 - d\left|\frac{\alpha'}{x}\right|\right)\left(1 - d\left|\frac{\beta'}{(x-d_{\beta'})\log(x-d_{\beta'})}\right|\right) \\
&\quad\times\left(1 - d\frac{\delta'\gamma'}{|x|^{1-\gamma'}}\right) \le \frac{\pi(\theta|x)}{\pi(\theta)} \\
&\le \frac{K\overline{K}}{k_*}\left(1 + d\left|\frac{\alpha'}{x}\right|\right)\left(1 + d\left|\frac{\beta'}{(x-d_{\beta'})\log(x-d_{\beta'})}\right|\right) \\
&\quad\times\left(1 + d\frac{\delta'\gamma'}{|x|^{1-\gamma'}}\right),
\end{aligned}$$

where k, k_*, $\underline{k}$, K, K_* and $\overline{K}$ are appropriate constants and $d_{\beta'} = 0$ if $\beta' \ge 0$ and it is equal to d if $\beta' < 0$.

The proof of this theorem is similar to the one of Theorem 4 and thus it is omitted. This theorem is the opposite of Theorem 4. In this setup (the likelihood tails are heavier than the prior's ones), the posterior ignores the likelihood information when it is conflicting with the prior information.

Using this theorem, bounds of the posterior of θ when the sample contained an outlier, is given in the next theorem.

THEOREM 7. *Let $x_1, x_2, \ldots, x_n$ be n independent and identically distributed random variables with density $f(x-\theta)$ and let π denote the prior density. If $|x_n| \gg \max_{1\le i\le n-1}|x_i|$, $f \succ \pi$ and if $\gamma' < 1$ then $\forall d > 0, \exists D > d + z_0$ such that $\forall |x_n| > D$ and $\forall |\theta| < d$,*

$$\frac{m(x_n)m(\underset{\sim}{x}_{(-n)})}{m(\underset{\sim}{x})}c^-_{x_n} \le \frac{\pi(\theta \mid \underset{\sim}{x})}{\pi(\theta \mid \underset{\sim}{x}_{(-n)})} \le \frac{m(x_n)m(\underset{\sim}{x}_{(-n)})}{m(\underset{\sim}{x})}c^+_{x_n},$$

where

$$\underset{\sim}{x} = (x_1, x_2, \ldots, x_n)\,,$$
$$\underset{\sim}{x}_{(-n)} = (x_1, x_2, \ldots, x_{n-1})\,,$$
$$m(x_n) = \int_\Theta \pi(\theta) f(x_n - \theta) d\theta\,,$$
$$m(\underset{\sim}{x}) = \int_\Theta \pi(\theta) \left[\prod_{i=1}^{n} f(x_i - \theta)\right] d\theta\,,$$
$$m(\underset{\sim}{x}_{(-n)}) = \int_\Theta \pi(\theta) \left[\prod_{i=1}^{n-1} f(x_i - \theta)\right] d\theta\,,$$
$$c^-_{x_n} = \frac{k\underline{k}}{K_*}\left(1 - d\left|\frac{\alpha'}{x}\right|\right)\left(1 - d\left|\frac{\beta'}{(x-d_{\beta'})\log(x-d_{\beta'})}\right|\right)\left(1 - d\frac{\delta'\gamma'}{|x|^{1-\gamma'}}\right),$$
$$c^+_{x_n} = \frac{K\overline{K}}{k_*}\left(1 + d\left|\frac{\alpha'}{x}\right|\right)\left(1 + d\left|\frac{\beta'}{(x-d_{\beta'})\log(x-d_{\beta'})}\right|\right)\left(1 + d\frac{\delta'\gamma'}{|x|^{1-\gamma'}}\right),$$

and k, k_*, $\underline{k}$, K, K_* *and* $\overline{K}$ *are the appropriate constants defined in Theorem* 6 *and* $d_{\beta'} = 0$ *if* $\beta' \geq 0$ *and it is equal to* d *if* $\beta' < 0$.

The proof of the theorem is given in the Appendix.

Theorem 7 says that if the sample contained an outlier and if the prior is chosen such that $f \succ \pi$, then the outlying observation will not influence the Bayes rule since it does not affect the posterior. This behavior is illustrated in the next section.

6. Examples

In this section, three examples are discussed to illustrate the results of Sections 4 and 5. For the first two examples, it is assumed that $X|\theta \sim N(\theta, 1)$ and that $x \gg 0$. The conditional p-credence of X is then equal to p-cred$(X - \theta|\theta) = (2, 1/2, 0, 0)$. It is often thought that if one is using a heavy-tailed prior, then the Bayes estimator is robust with respect to prior misspecification. The prior used in the first example does not satisfy the conditions of Theorem 5. Hence the resulting estimator will not discard the prior information when it is conflicting with the data. With the second example, the importance of the condition on γ given in Theorem 5 and Corollary 1 is illustrated. In the third example, we consider several observations and it is supposed that the sample contained an outlier. For this example, the prior is chosen to be a Laplace density and the density of the observations has p-credence $(\gamma', 1, 0, 0)$ where $\gamma' < 1$.

6.1. First example

For this example, the prior of θ is $p(\theta|2, 1/2, \alpha, 0)$ with $z_0 = \sqrt{\alpha}$, that is

$$\pi(\theta) \propto [\max(|\theta|, \sqrt{\alpha})]^{\alpha} \exp\left\{-\frac{1}{2}\max(\theta^2, \alpha)\right\},$$

where $\alpha > 1$ and p-cred$(\theta) = (2, 1/2, \alpha, 0)$. Using Theorem 1, it is obvious that $\pi \succ f$. Although $\pi(\theta)$ is not what it is usually called a heavy-tailed prior, its tails are still heavier than those of a normal distribution.

Under the squared error loss, the Bayes estimator of θ is given by

$$\begin{aligned}
\widehat{\theta}(x) &= \frac{\displaystyle\int_{-\infty}^{\infty} \theta p(\theta|2, 1/2, \alpha, 0) p(x-\theta|2, 1/2, 0, 0) d\theta}{\displaystyle\int_{-\infty}^{\infty} p(\theta|2, 1/2, \alpha, 0) p(x-\theta|2, 1/2, 0, 0) d\theta} \\
&= \frac{\displaystyle\int_{-\infty}^{\infty} \theta [\max(|\theta|, \sqrt{\alpha})]^{\alpha} \exp\left\{-\frac{1}{2}\left[\max(\theta^2, \alpha) + (\theta - x)^2\right]\right\} d\theta}{\displaystyle\int_{-\infty}^{\infty} [\max(|\theta|, \sqrt{\alpha})]^{\alpha} \exp\left\{-\frac{1}{2}\left[\max(\theta^2, \alpha) + (\theta - x)^2\right]\right\} d\theta} \\
&= \left[\mathrm{e}^{-x^2/4} \int_{\sqrt{\alpha}}^{\infty} \theta^{\alpha+1} \mathrm{e}^{-(\theta - x/2)^2} d\theta + \alpha^{\alpha/2} \mathrm{e}^{-\alpha/2} \int_{-\sqrt{\alpha}}^{\sqrt{\alpha}} \theta \mathrm{e}^{-\frac{1}{2}(\theta - x)^2} d\theta \right. \\
&\quad \left. - \mathrm{e}^{-x^2/4} \int_{\sqrt{\alpha}}^{\infty} \theta^{\alpha+1} \mathrm{e}^{-(\theta + x/2)^2} d\theta \right] \Big/ \left[\mathrm{e}^{-x^2/4} \int_{\sqrt{\alpha}}^{\infty} \theta^{\alpha} \mathrm{e}^{-(\theta - x/2)^2} d\theta \right. \\
&\quad \left. + \alpha^{\alpha/2} \mathrm{e}^{-\alpha/2} \int_{-\sqrt{\alpha}}^{\sqrt{\alpha}} \mathrm{e}^{-\frac{1}{2}(\theta - x)^2} d\theta + \mathrm{e}^{-x^2/4} \int_{\sqrt{\alpha}}^{\infty} \theta^{\alpha} \mathrm{e}^{-(\theta + x/2)^2} d\theta \right] \\
&\simeq \frac{\displaystyle\int_{\sqrt{\alpha}}^{\infty} \theta^{\alpha+1} \mathrm{e}^{-(\theta - x/2)^2} d\theta}{\displaystyle\int_{\sqrt{\alpha}}^{\infty} \theta^{\alpha} \mathrm{e}^{-(\theta - x/2)^2} d\theta} \qquad (5) \\
&\simeq \frac{x}{2}\left(1 + \frac{2\alpha}{\alpha(\alpha - 1) + x^2}\right) \qquad (6)
\end{aligned}$$

if $x \gg \sqrt{\alpha}$. Consequently, even if x is extremely large, $\widehat{\theta}(x)$ will never collapse to x. Hence, it is not an estimator insensitive to prior misspecification.

6.2. Second example

In this example, the prior of θ is $p(\theta|\gamma, 1, 0, 0)$ for $\gamma \leq 1$ and $z_0 = 0$. Using Theorem 1, we have that $\pi \succ f$. Under the squared error loss, and using the Taylor series expansion of $p(\theta|\gamma, 1, 0, 0)$ with respect to $\theta = 0$, the Bayes estimator for θ can be written as

$$\begin{aligned}
\widehat{\theta}(x) &= \frac{\displaystyle\int_{-\infty}^{\infty} \theta \exp\left\{-\left[|\theta|^{\gamma} + \frac{1}{2}(\theta - x)^2\right]\right\} d\theta}{\displaystyle\int_{-\infty}^{\infty} \exp\left\{-\left[|\theta|^{\gamma} + \frac{1}{2}(\theta - x)^2\right]\right\} d\theta} \\
&= \frac{\displaystyle\sum_{j=0}^{\infty} \frac{(-1)^j}{j!} \int_{-\infty}^{\infty} \theta|\theta|^{\gamma j} \mathrm{e}^{-\frac{1}{2}(\theta-x)^2} d\theta}{\displaystyle\sum_{j=0}^{\infty} \frac{(-1)^j}{j!} \int_{-\infty}^{\infty} |\theta|^{\gamma j} \mathrm{e}^{-\frac{1}{2}(\theta-x)^2} d\theta} \\
&= x + \frac{\displaystyle\sum_{j=0}^{\infty} \frac{(-1)^j}{j!} \int_{-\infty}^{\infty} |\theta|^{\gamma j} (\theta - x) \mathrm{e}^{-\frac{1}{2}(\theta-x)^2} d\theta}{\displaystyle\sum_{j=0}^{\infty} \frac{(-1)^j}{j!} \int_{-\infty}^{\infty} |\theta|^{\gamma j} \mathrm{e}^{-\frac{1}{2}(\theta-x)^2} d\theta}.
\end{aligned}$$

Using integration by parts and doing the change of variable $z = \theta - x$, we obtain

$$\widehat{\theta}(x) = x - \gamma \frac{\mathrm{sign}(x)}{|x|^{1-\gamma}} \frac{\sum_{j=0}^{\infty} \frac{(-1)^j}{j!} |x|^{\gamma j} \int_{-\infty}^{\infty} \mathrm{sign}(1 + z/x)|1 + z/x|^{\gamma j - 1 + \gamma} \phi(z) dz}{\sum_{j=0}^{\infty} \frac{(-1)^j}{j!} |x|^{\gamma j} \int_{-\infty}^{\infty} |1 + z/x|^{\gamma j} \phi(z) dz}.$$

If $|x|$ is large enough, it can be shown that the ratio of integrals in the previous equation goes to 1. Consequently,

$$\widehat{\theta}(x) \approx x - \gamma \frac{\mathrm{sign}(x)}{|x|^{1-\gamma}}.$$

Hence, $\widehat{\theta}(x)$ is insensitive to prior misspecification only if $\gamma < 1$.

6.3. Third example

In this example, let $x_1, x_2, \ldots, x_n$ be a sample of size n from a population with density given by

$$f(x-\theta) = \frac{\gamma'}{2\Gamma(1/\gamma')} \exp\{-|x-\theta|^{\gamma'}\},$$

where $\gamma' < 1$. Hence p-cred$(X-\theta \mid \theta) = (\gamma', 1, 0, 0)$. The prior on θ is chosen to be a Laplace density with scale parameter equal to 1. So p-cred$(\theta) = (1, 1, 0, 0)$. Consequently, $f \succ \pi$. It is also assume that the last observation is positive outlier, that is $x_n > 0$ and $x_n \gg x_i$ for all $i = 1, 2, \ldots, n-1$. Let

$$I_k(\underset{\sim}{x}) = \int_{-\infty}^{\infty} \theta^k \pi(\theta) \left[\prod_{i=1}^{n} f(x_i - \theta)\right] d\theta\,,$$

$$I_k(\underset{\sim}{x}_{(-n)}) = \int_{-\infty}^{\infty} \theta^k \pi(\theta) \left[\prod_{i=1}^{n-1} f(x_i - \theta)\right] d\theta\,,$$

for $k = 0, 1$. Hence, under the squared error loss, the Bayes estimator is given by

$$\widehat{\theta}(\underset{\sim}{x}) = I_1(\underset{\sim}{x})/I_0(\underset{\sim}{x})\,.$$

If x_n is large enough, it can be shown that

$$I_k(\underset{\sim}{x}) = I_k(\underset{\sim}{x}_{(-n)}) + b_k x_n^k \exp\{-a x_n\}\,,$$

where a, b_0 and b_1 are appropriate constants. Hence

$$\begin{aligned}
\widehat{\theta}(\underset{\sim}{x}) &= \frac{I_1(\underset{\sim}{x})}{I_0(\underset{\sim}{x})} \\
&= \frac{I_1(\underset{\sim}{x}_{(-n)}) + b_1 x_n \mathrm{e}^{-a x_n}}{I_0(\underset{\sim}{x}_{(-n)}) + b_0 \mathrm{e}^{-a x_n}} \\
&= \frac{\widehat{\theta}(\underset{\sim}{x}_{(-n)}) + b_1^* x_n \mathrm{e}^{-a x_n}}{1 + b_0^* \mathrm{e}^{-a x_n}}\,,
\end{aligned}$$

where $b_k^* = b_k / I_0(\underset{\sim}{x}_{(-n)})$. Consequently,

$$\widehat{\theta}(\underset{\sim}{x}) = \widehat{\theta}(\underset{\sim}{x}_{(-n)}) + b x_n \mathrm{e}^{-a x_n}\,.$$

Hence, when the sample contained an outlier, the posterior mean neglects the information provided by this outlier given that it is far enough from the other observations.

7. Conclusion

In this paper, we introduce a new family of densities, called the generalized exponential power family that captures several types of tail behavior. Using this family, we also generalize the notion of credence introduced in O' Hagan (1990). Sufficient conditions on the p-credence of the prior and the likelihood function are given in order to have an estimator which is insensitive to prior misspecification if we have only one observation (or a density with one sufficient statistic). The relationship between p-credence and outlier is also discussed.

Appendix: Proofs

Proof of Lemma 3. Without loss of generality, suppose that $z > 0$ and $d > 0$. To prove this lemma, three cases have to be considered, that is:

i) $z + d \leq z_0$,
ii) $z \leq z_0 < z + d$,
iii) $z > z_0$,

and we only have to prove that the function

$$h_d(z) = \max(|z + d|, z_0)^\gamma - \max(|z|, z_0)^\gamma$$

is bounded above and below. Obviously, $h_d(z) \geq 0$. For the first case, $h_d(z) = 0$. For the second case,

$$h_d(z) = (z + d)^\gamma - z_0^\gamma \leq (z_0 + d)^\gamma - z_0^\gamma ,$$

and for the third case,

$$\begin{aligned} h_d(z) &= (z + d)^\gamma - z^\gamma \\ &= z^\gamma \left[(1 + d/z)^\gamma - 1\right] \\ &\leq z^\gamma \left[(1 + d/z) - 1\right] \\ &\leq \frac{d}{z^{1-\gamma}} \\ &\leq \frac{d}{z_0^{1-\gamma}} . \qquad \blacksquare \end{aligned}$$

Proof of Theorem 2. Suppose that $x > 0$. Since p-cred$(X - \theta|\theta) = (\gamma', \delta', \alpha', \beta')$ and p-cred$(\theta) = (\gamma, \delta, \alpha, \beta)$, the marginal of x is bounded above as follows

$$\begin{aligned}
m(x) &= \int_{-\infty}^{\infty} f(x-\theta)\pi(\theta)d\theta \\
&\leq KK' \int_{-\infty}^{\infty} p(x-\theta|\gamma', \delta', \alpha', \beta')p(\theta|\gamma, \delta, \alpha, \beta)d\theta \\
&= KK' \left[\int_{-\infty}^{0} p(x-\theta|\gamma', \delta', \alpha', \beta')p(\theta|\gamma, \delta, \alpha, \beta)d\theta \right. \\
&\quad \left. + \int_{-\infty}^{0} p(\zeta|\gamma', \delta', \alpha', \beta')p(x-\zeta|\gamma, \delta, \alpha, \beta)d\zeta \right] \\
&\leq \frac{KK'}{2}\left[p(x|\gamma', \delta', \alpha', \beta') + p(x|\gamma, \delta, \alpha, \beta)\right] \\
&\leq KK' \max\left\{p(x|\gamma', \delta', \alpha', \beta'), p(x|\gamma, \delta, \alpha, \beta)\right\} .
\end{aligned}$$

To obtain the lower bound, we proceed as follows. For any $\epsilon > 0$, $m(x)$ can be written as

$$\begin{aligned}
m(x) &\geq kk' \left[\int_{-\infty}^{0} + \int_{0}^{\epsilon} + \int_{\epsilon}^{x-\epsilon} + \int_{x-\epsilon}^{x} + \int_{x}^{\infty}\right] p(x-\theta|\gamma', \delta', \alpha', \beta') \\
&\quad \times p(\theta|\gamma, \delta, \alpha, \beta)d\theta \\
&\geq kk' \max\left\{\int_{0}^{\epsilon} p(x-\theta|\gamma', \delta', \alpha', \beta')p(\theta|\gamma, \delta, \alpha, \beta)d\theta, \right. \\
&\quad \left. \int_{0}^{\epsilon} p(\zeta|\gamma', \delta', \alpha', \beta')p(x-\zeta|\gamma, \delta, \alpha, \beta)d\zeta \right\} \\
&\geq kk'\epsilon \max\left\{p(\epsilon|\gamma, \delta, \alpha, \beta)p(x|\gamma', \delta', \alpha', \beta'), \right. \\
&\quad \left. p(\epsilon|\gamma', \delta', \alpha', \beta')p(x|\gamma, \delta, \alpha, \beta)\right\} \\
&\geq kk'\epsilon \min\left\{p(\epsilon|\gamma, \delta, \alpha, \beta), p(\epsilon|\gamma', \delta', \alpha', \beta')\right\} \\
&\quad \times \max\left\{p(x|\gamma', \delta', \alpha', \beta'), p(x|\gamma, \delta, \alpha, \beta)\right\} . \qquad \blacksquare
\end{aligned}$$

Proof of Theorem 3. Assume that $\pi \succ f$, let $\gamma_{**} = \gamma'$ and $\delta_{**} = \delta'$. (Note that the proof for $f \succ \pi$, that is $\gamma_{**} = \gamma$, $\delta_{**} = \delta$ is similar.) The case where $\gamma_{**} = 0$ can be proved using Lemma 1 and a method similar to the one discussed in the proof of Theorem 3 of O'Hagan (1990). If $\min(\gamma, \gamma') = 0$ but either γ or γ' is greater than 0, the proof is similar to the case where $\gamma_{**} = 0$ since there is only one exponential part to deal with. Because of Lemma 2, we can assume without loss of generality that $z_0 > 0$.

Since p-cred$(X-\theta|\theta) = (\gamma', \delta', \alpha', \beta')$ and p-cred$(\theta) = (\gamma, \delta, \alpha, \beta)$ and using Lemma 1 there exist positive constants A, K and K' such that, the posterior of θ is bounded above by

$$\begin{aligned}
\pi(\theta|x) &= \frac{f(x-\theta)\pi(\theta)}{m(x)} \\
&\leq \frac{KK'}{m(x)} \max(|\theta|, z_0)^{\alpha} \max(|\theta - x|, z_0)^{\alpha'} \log^{\beta}[\max(|\theta|, z_0)] \\
&\quad \times \log^{\beta'}[\max(|\theta - x|, z_0)] \\
&\quad \times \exp\left\{-[\delta \max(|\theta|, z_0)^{\gamma} + \delta' \max(|\theta - x|, z_0)^{\gamma'}]\right\} \\
&\leq \frac{AKK'}{m(x)} \max(|\theta - x|, z_0)^{\alpha+\alpha'} \log^{\beta+\beta'}[\max(|\theta - x|, z_0)] \\
&\quad \times \exp\left\{-[\delta \max(|\theta|, z_0)^{\gamma} + \delta' \max(|\theta - x|, z_0)^{\gamma'}]\right\} .
\end{aligned} \tag{7}$$

Similarly, there exist positive constants a, k and k' such that

$$\begin{aligned}
\pi(\theta|x) \geq \frac{akk'}{m(x)} &\max(|\theta - x|, z_0)^{\alpha+\alpha'} \log^{\beta+\beta'}[\max(|\theta - x|, z_0)] \\
&\times \exp\left\{-[\delta \max(|\theta|, z_0)^{\gamma} + \delta' \max(|\theta - x|, z_0)^{\gamma'}]\right\} .
\end{aligned} \tag{8}$$

Now let $g(\theta) = \delta \max(|\theta|, z_0)^{\gamma} + \delta' \max(|\theta - x|, z_0)^{\gamma'}$ and let $\underline{g}(\theta) = \delta' \max(|\theta - x|, z_0)^{\gamma'}$. Hence $\underline{g}(\theta) \leq g(\theta)\, \forall\theta$. If we let $c(\theta) = \delta' + \delta \frac{\max(|\theta|, z_0)^{\gamma}}{\max(|\theta - x|, z_0)^{\gamma'}}$. Then $g(\theta)$ can be written as $g(\theta) = c(\theta) \max(|\theta - x|, z_0)^{\gamma'}$. Maximizing $c(\theta)$ over all values of θ, we can show that

$$c(\theta) \leq \delta' + \delta \frac{(1 + |x|/z_0)^{\gamma}}{z_0^{\gamma'}} .$$

Hence, if we let $\overline{g}(\theta) = [\delta' + \delta \frac{(1+|x|/z_0)^{\gamma}}{z_0^{\gamma'}}] \max(|\theta - x|, z_0)^{\gamma'}$, we have that $g(\theta) \leq \overline{g}(\theta)\, \forall\theta$. Consequently, using equations (7) and (8) we obtain the following bounds for the posterior, that is

$$\begin{aligned}
&\frac{akk'}{m(x)} \max(|\theta - x|, z_0)^{\alpha+\alpha'} \log^{\beta+\beta'} \\
&\qquad \times [\max(|\theta - x|, z_0)] \exp\{-\overline{g}(\theta)\} \\
&\qquad \leq \pi(\theta|x) \qquad\qquad (9)\\
&\qquad \leq \frac{AKK'}{m(x)} \max(|\theta - x|, z_0)^{\alpha+\alpha'} \log^{\beta+\beta'} \\
&\qquad \times [\max(|\theta - x|, z_0)] \exp\{-\underline{g}(\theta)\}\,. \qquad \blacksquare
\end{aligned}$$

Proof of Theorem 4. a) Assume that $x > z_0 + d$. Since $\pi \succeq f$, the ratio $\pi(\theta|x)/f(x-\theta)$ is bounded above by

$$\begin{aligned}
\frac{\pi(\theta|x)}{f(x-\theta)} &= \frac{\pi(\theta)}{m(x)} \\
&\leq \frac{K}{k_*} \left(\frac{\max(|\theta|, z_0)}{\max(|x|, z_0)}\right)^{\alpha} \left(\frac{\log(\max(|\theta|, z_0))}{\log(\max(|x|, z_0))}\right)^{\beta} \\
&\quad \times \exp\left\{-\delta\left[\max(|\theta|, z_0)^{\gamma} - \max(|x|, z_0)^{\gamma}\right]\right\} \\
&= \frac{K}{k_*} \left(\frac{\theta}{x}\right)^{\alpha} \left(\frac{\log(\theta)}{\log(x)}\right)^{\beta} \exp\left\{\delta x^{\gamma}\left[1 - \left(\frac{\theta}{x}\right)^{\gamma}\right]\right\} \\
&\leq \frac{K}{k_*} \left(1 + s_\alpha \frac{d}{x}\right)^{\alpha} \left(1 + \frac{d}{(x-d_\beta)\log(x-d_\beta)}\right)^{\beta} \\
&\quad \times \exp\left\{\delta x^{\gamma}\left[1 - \left(1 - \frac{d}{x}\right)^{\gamma}\right]\right\}, \qquad (10)
\end{aligned}$$

where

$$s_\alpha = \begin{cases} 1; & \text{if } \alpha > 0, \\ -1; & \text{if } \alpha < 0, \\ 0; & \text{if } \alpha = 0, \end{cases} \quad \text{and } d_\beta = \begin{cases} 0; & \text{if } \beta > 0, \\ d; & \text{if } \beta < 0. \end{cases}$$

If $d \ll D$ and if $x > D$, then

$$k_\alpha\left(1 + \frac{|\alpha| d}{x}\right) \leq \left(1 + s_\alpha \frac{d}{x}\right)^{\alpha} \leq K_\alpha\left(1 + \frac{|\alpha| d}{x}\right),$$

where $k_\alpha = \min(1, \frac{(1+s_\alpha d/D)^\alpha}{1+|\alpha|d/D})$ and $K_\alpha = \max(1, \frac{(1+s_\alpha d/D)^\alpha}{1+|\alpha|d/D})$. Similarly

$$\begin{aligned} k_\beta \left(1 + \frac{|\beta|d}{(x-d_\beta)\log(x-d_\beta)}\right) &\le \left(1 + \frac{d}{(x-d_\beta)\log(x-d_\beta)}\right)^\beta \\ &\le K_\beta \left(1 + \frac{|\beta|d}{(x-d_\beta)\log(x-d_\beta)}\right) \end{aligned}$$

and

$$k_\gamma \left(1 + \frac{\delta\gamma d}{x^{1-\gamma}}\right) \le \exp\left\{\delta x^\gamma [1 - \left(1 - \frac{d}{x}\right)^\gamma]\right\} \le K_\gamma \left(1 + \frac{\delta\gamma d}{x^{1-\gamma}}\right) .$$

Replacing these in equation (10) leads to the upper bound where $\overline{K} = K_\alpha K_\beta K_\gamma$. The lower bound is obtained in a similar manner.

b) Assume that $x > \theta_0 + z_0$. Using Lemmas 1 and 3, we have that

$$\begin{aligned} \Pr\{|\theta| < \theta_0 | X = x\} &= \int_{-\theta_0}^{\theta_0} \frac{\pi(\theta) f(x-\theta)}{m(x)} d\theta \\ &\le \frac{KK'}{m(x)} \int_{-\theta_0}^{\theta_0} p(\theta|\gamma,\delta,\alpha,\beta) p(x-\theta|\gamma',\delta',\alpha',\beta') d\theta \\ &\le 2\theta_0 \frac{KK'}{k_*} \frac{p(0|\gamma,\delta,\alpha,\beta) p(x-\theta_0|\gamma',\delta',\alpha',\beta')}{p(x|\gamma,\delta,\alpha,\beta)} \\ &\le 2\theta_0 \frac{AKK'}{k_*} p(0|\gamma,\delta,\alpha,\beta) x^{\alpha'-\alpha} \log^{\beta'-\beta}(x) \\ &\quad \times \exp\left\{-\delta'(x-\theta_0)^{\gamma'} + \delta x^\gamma\right\} \\ &\le 2\theta_0 \frac{AKK'}{k_*} p(0|\gamma,\delta,\alpha,\beta) x^{\alpha'-\alpha} \log^{\beta'-\beta}(x) \\ &\quad \times \exp\left\{-\delta'(x-\theta_0)^{\gamma'} \left[1 - \frac{\delta}{\delta'(x-\theta_0)^{\gamma'-\gamma}}\right]\right\} . \end{aligned} \tag{11}$$

Since $\pi \succeq f$, then either $\gamma' > \gamma$, or $\gamma' = \gamma$ and $\alpha' < \alpha$, or $\gamma' = \gamma$ and $\alpha' = \alpha$ and $\beta' < \beta$. Hence, taking the limit as $x \to \infty$, equation (11) goes to 0. A lower bound can be obtained using similar arguments. ∎

Proof of Theorem 5. Without loss of generality, assume that $x > D$. Let κ be a constant such that

$$\frac{\gamma}{\gamma'} < \kappa < 1; \text{ if } \gamma' > 0\,,$$
$$1 - \frac{|\alpha|}{|\alpha'| - 1} < \kappa < 1; \text{ if } \gamma' = 0\,,$$

Hence $\mathbb{E}^{\pi(\theta|x)}[\theta^p]$ can be written as

$$\begin{aligned} x^p - \mathbb{E}^{\pi(\theta|x)}[\theta^p] &= \sum_{l=1}^{p} \binom{p}{l} (-1)^{l+1} x^{p-l} \\ &\times \left[\int_{-\infty}^{-x^\kappa} + \int_{-x^\kappa}^{x^\kappa} + \int_{x^\kappa}^{x-x^\kappa} + \int_{x-x^\kappa}^{x+x^\kappa} + \int_{x+x^\kappa}^{\infty} \right] \zeta^l \frac{\pi(x-\zeta) f(\zeta)}{m(x)} d\zeta \\ &\le \sum_{l=1}^{p} D_l \binom{p}{l} (-1)^{l+1} x^{p-l} \left[\int_{-\infty}^{-x^\kappa} + \int_{-x^\kappa}^{x^\kappa} + \int_{x^\kappa}^{x-x^\kappa} + \int_{x-x^\kappa}^{x+x^\kappa} + \int_{x+x^\kappa}^{\infty} \right] \\ &\times \zeta^l \frac{p(x-\zeta|\gamma,\delta,\alpha,\beta) p(\zeta|\gamma',\delta',\alpha',\beta')}{p(x|\gamma,\delta,\alpha,\beta)}\,, \end{aligned}$$

where D_l are appropriate constants. It can be shown that

$$\begin{aligned} &\int_{-\infty}^{-x^\kappa} \zeta^l \frac{p(x-\zeta|\gamma,\delta,\alpha,\beta) p(\zeta|\gamma',\delta',\alpha',\beta')}{p(x|\gamma,\delta,\alpha,\beta)} d\zeta \\ &\qquad \le \frac{K' q_0}{q_1} \frac{p(x-\zeta_1|\gamma,\delta,\alpha,\beta)}{p(x|\gamma,\delta,\alpha,\beta)} [1 - p_1(x^\kappa)]\,, \end{aligned}$$

where $\zeta_1 \in (-\infty, x^\kappa)$. But, using Lemma 3,

$$\begin{aligned} &\frac{p(x-\zeta_1|\gamma,\delta,\alpha,\beta)}{p(x|\gamma,\delta,\alpha,\beta)} \le \frac{p(x-x^\kappa|\gamma,\delta,\alpha,\beta)}{p(x|\gamma,\delta,\alpha,\beta)} \\ &\le \frac{K}{k_*} \left(1 - x^{\kappa-1}\right)^\alpha \left(1 + \frac{\log(1-|x|^{\kappa-1})}{\log(x)}\right)^\beta \\ &\times \exp\left\{-\delta x^\gamma \left(\left[1 - |x|^{\kappa-1}\right]^\gamma - 1\right)\right\} \\ &\le K_1, \end{aligned}$$

since $\gamma < 1$ and x is large. It can also be shown that, for any $\epsilon > 0$,

$$1 - p_1(x^\kappa) = c_1 \int_{x^\kappa}^{\infty} u^{\alpha'+l} \log^{\beta'}(u) \exp\left\{-\delta' u^{\gamma'}\right\} du$$
$$\leq \begin{cases} c_1 x^{\kappa(\alpha'+\epsilon\beta'+l+1-\gamma')} \exp\left\{-\delta' x^{\kappa\gamma'}\right\}; & \text{if } \gamma' > 0, \\ c_1 \kappa \dfrac{\log^{\beta'}(x)}{x^{\kappa(|\alpha'|-l)}}; & \text{if } \gamma' = 0. \end{cases}$$

Consequently

$$\int_{-\infty}^{-x^\kappa} \zeta \frac{\pi(x-\zeta) f(\zeta)}{m(x)} d\zeta$$
$$\leq C_{1,\gamma'} \begin{cases} x^{\kappa(\alpha'+\epsilon\beta'+l+1-\gamma')} \exp\left\{-\delta' x^{\kappa\gamma'}\right\}; & \text{if } \gamma' > 0, \\ \dfrac{\log^{\beta'}(x)}{x^{\kappa(|\alpha'|-l)}}; & \text{if } \gamma' = 0, \end{cases} \tag{12}$$

where $C_{1,\gamma'}$ is an appropriate constant.

Since $p(\cdot|\gamma', \delta', \alpha', \beta')$ is symmetrical with respect to 0, using the mean value theorem, we have

$$\int_{-x^\kappa}^{x^\kappa} \zeta^l \frac{p(x-\zeta|\gamma,\delta,\alpha,\beta)\, p(\zeta|\gamma',\delta',\alpha',\beta')}{p(x|\gamma,\delta,\alpha,\beta)} d\zeta$$
$$= \frac{p(x-\zeta_2|\gamma,\delta,\alpha,\beta)}{p(x|\gamma,\delta,\alpha,\beta)} \times \int_{-x^\kappa}^{x^\kappa} \zeta^l p(\zeta|\gamma',\delta',\alpha',\beta') d\zeta$$
$$= \begin{cases} 0; & \text{if } l \text{ is odd}, \\ \dfrac{2p(x-\zeta_2|\gamma,\delta,\alpha,\beta)}{p(x|\gamma,\delta,\alpha,\beta)} \times \displaystyle\int_0^{x^\kappa} \zeta^l p(\zeta|\gamma',\delta',\alpha',\beta') d\zeta; & \text{if } l \text{ is even}, \end{cases} \tag{13}$$
$$\leq \begin{cases} 0; & \text{if } l \text{ is odd}, \\ \Gamma(D_{\gamma'}); & \text{if } l \text{ is even}, \end{cases}$$

where $\zeta_2 \in (-x^\kappa, x^\kappa)$ and

$$D_{\gamma'} = \begin{cases} \dfrac{\alpha' + \epsilon\beta' + l + 1}{\gamma'}; & \text{if } \gamma' > 0, \\ \beta' + 1; & \text{if } \gamma = 0. \end{cases}$$

Using similar arguments as for the first integral, it can be shown that

$$
\begin{aligned}
&\int_{x^\kappa}^{x-x^\kappa} \zeta^l \frac{p(x-\zeta|\gamma,\delta,\alpha,\beta)p(\zeta|\gamma',\delta',\alpha',\beta')}{p(x|\gamma,\delta,\alpha,\beta)} d\zeta \\
&\leq \frac{q_0}{q_1}\frac{p(x^\kappa|\gamma,\delta,\alpha,\beta)}{p(x|\gamma,\delta,\alpha,\beta)}[1-p_1(x^\kappa)] \\
&\leq \frac{q_0 K\kappa^\beta}{k_* x^{\alpha(1-\kappa)}} \exp\left\{\delta x^\gamma \left(1 - 1/x^{\gamma(1-\kappa)}\right)\right\} \\
&\quad \times \begin{cases} x^{\kappa(\alpha'+\epsilon\beta'+l+1-\gamma')}\exp\left\{-\delta' x^{\kappa\gamma'}\right\}; & \text{if } \gamma' > 0, \\ \kappa \dfrac{\log^{\beta'}(x)}{x^{\kappa(|\alpha'|-l)}}; & \text{if } \gamma' = 0, \end{cases} \\
&\leq C_{2,\gamma'} \begin{cases} x^{\kappa(\alpha'+\alpha+\epsilon\beta'+l+1-\gamma')-\alpha}\exp\Big\{-\delta' x^{\kappa\gamma'} & \\ \qquad \times \left[1-\dfrac{\delta}{\delta'}x^{\gamma-\kappa\gamma'}+\dfrac{\delta}{\delta' x^{\kappa(\gamma'-\gamma)}}\right]\Big\}; & \text{if } \gamma' > 0, \\ \kappa \dfrac{\log^{\beta'}(x)}{x^{\kappa(|\alpha|+|\alpha'|-l)-|\alpha|}}; & \text{if } \gamma' = 0. \end{cases}
\end{aligned}
\tag{14}
$$

Since $p(\cdot|\gamma,\delta,\alpha,\beta)$ is symmetrical with respect to 0,

$$
\begin{aligned}
&\int_{x-x^\kappa}^{x+x^\kappa} \zeta^l \frac{p(x-\zeta|\gamma,\delta,\alpha,\beta)p(\zeta|\gamma',\beta',\alpha',\beta')}{p(x|\gamma,\delta,\alpha,\beta)} d\zeta \\
&\qquad = \int_{-x^\kappa}^{x^\kappa} (x-\theta)^l \frac{p(\theta|\gamma,\delta,\alpha,\beta)p(x-\theta|\gamma',\beta',\alpha',\beta')}{p(x|\gamma,\delta,\alpha,\beta)} d\theta \\
&\qquad = \frac{p(x-\theta_1|\gamma',\delta',\alpha',\beta')}{p(x|\gamma,\delta,\alpha,\beta)} \sum_{j=0}^{l} (-1)^j \binom{l}{j} x^{l-j} \\
&\qquad\quad \times \int_{-x^\kappa}^{x^\kappa} \theta^j p(\theta|\gamma,\delta,\alpha,\beta) d\theta\,,
\end{aligned}
$$

where $\theta_1 \in (-x^\kappa, x^\kappa)$. However, it can be shown that

$$
\begin{aligned}
&\int_{-x^\kappa}^{x^\kappa} \theta^j p(\theta|\gamma,\delta,\alpha,\beta) d\theta \\
&\leq \begin{cases} 0; & \text{if } j \text{ is odd}, \\ 2\Gamma\left(\dfrac{\alpha+\epsilon\beta+j+1}{\gamma}\right); & \text{if } j \text{ is even and } \gamma > 0, \\ 2\Gamma(\beta+1); & \text{if } j \text{ is even and } \gamma = 0. \end{cases}
\end{aligned}
$$

Hence,

$$\begin{aligned}
&\int_{x-x^{\kappa}}^{x+x^{\kappa}} \zeta^{l} \frac{p(x-\zeta|\gamma,\delta,\alpha,\beta)p(\zeta|\gamma',\beta',\alpha',\beta')}{p(x|\gamma,\delta,\alpha,\beta)} d\zeta \\
&\leq p(x-\theta_1|\gamma',\delta',\alpha',\beta')p(x|\gamma,\delta,\alpha,\beta)\sum_{j=0}^{l}(-1)^j \binom{l}{j} x^{l-j} C_{j,\gamma} \\
&\leq C'_{\gamma} x^{l} \frac{(x-x^{\kappa})^{\alpha'}}{x^{\alpha}} \frac{\log^{\beta'}(x-x^{\kappa})}{\log^{\beta}(x)} \exp\left\{-\delta'[x-x^{\kappa}]^{\gamma'}+\delta x^{\gamma}\right\} \\
&= C'_{\gamma}(1-x^{\kappa-1})^{\alpha'}\left(1+\frac{\log(1-x^{\kappa-1})}{\log(x)}\right)^{\beta'} \frac{x^{l+\alpha'-\alpha}}{\log^{\beta-\beta'}(x)} \\
&\quad \times \exp\left\{-\delta' x^{\gamma'}\left[(1-x^{\kappa-1})-\frac{\delta}{\delta'}x^{\gamma-\gamma'}\right]\right\}, \qquad (15)
\end{aligned}$$

where $C_{j,\gamma}$ and C'_{γ} are appropriate constants.

Finally, using a similar argument used for the first and the previous integrals, the last integral can be bounded above by

$$\begin{aligned}
&\int_{x+x^{\kappa}}^{\infty} \zeta^{l} \frac{p(x-\zeta|\gamma,\delta,\alpha,\beta)p(\zeta|\gamma',\beta',\alpha',\beta')}{p(x|\gamma,\delta,\alpha,\beta)} d\zeta \\
&\leq C_{5,\gamma'} \begin{cases}
(1+x^{\kappa-1})^{\alpha'}\left(1+\dfrac{\log(1+x^{\kappa-1})}{\log(x)}\right)^{\beta'} & \\
\times x^{\alpha'-\alpha+l+\kappa(\alpha+\epsilon\beta+1-\gamma)} & \\
\times \exp\Big\{-\delta' x^{\gamma'}\Big[(1+x^{\kappa-1})^{\gamma'} & \\
+\dfrac{\delta}{\delta'}(x^{\kappa\gamma-\gamma'}-x^{\gamma-\gamma'})\Big]\Big\}; & \text{if } \gamma>0, \\
\dfrac{\log^{\beta'}(x+x^{\kappa})}{(1+x^{\kappa-1})^{|\alpha'|}} \dfrac{1}{x^{|\alpha'|-(1-\kappa)|\alpha|-l}}; & \text{if } \gamma=0
\end{cases} \qquad (16)
\end{aligned}$$

Combining equations (12) to (16), it can be shown that the dominant term is given by equation (15) if the conditions of Theorem 5 are satisfied. ■

Proof of Theorem 7. Suppose that $|x_n| \gg \max_{1\le i\le n-1} |x_i|$ and that $|x_n|$ is large enough such that the condition of Theorem 6 are satisfied. Hence, using Theorem 6, the posterior can be written as

$$
\begin{aligned}
\pi(\theta \mid \underset{\sim}{x}) &= \frac{\pi(\theta) f(x_n - \theta)\left[\prod_{i=1}^{n-1} f(x_i - \theta)\right]}{m(\underset{\sim}{x})} \\
&= \frac{m(x_n)}{m(\underset{\sim}{x})}\pi(\theta \mid x_n)\left[\prod_{i=1}^{n-1} f(x_i - \theta)\right] \\
&\le \frac{m(x_n)}{m(\underset{\sim}{x})} c_{x_n}^{+} \pi(\theta)\left[\prod_{i=1}^{n-1} f(x_i - \theta)\right] \\
&= \frac{m(x_n) m(\underset{\sim}{x}_{(-n)})}{m(\underset{\sim}{x})} c_{x_n}^{+} \pi(\theta \mid \underset{\sim}{x}_{(-n)}) \,.
\end{aligned}
$$

A lower bound for $\pi(\theta \mid \underset{\sim}{x})$ can be obtained using a similar argument. ■

ACKNOWLEDGMENTS

The author is thankful to Prof. Anthony O'Hagan for his useful comments and his helpful suggestions for improving the manuscript.

REFERENCES

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

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

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

BOX, G. and TIAO, G. (1962) A further look at robustness via Bayes's theorem, *Biometrika*, 49, 419-432.

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

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

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

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

HILL, B. (1974) *On coherence, inadmissibility and inference about many parameters in the theory of least squares*, in *Studies in Bayesian econometrics and Statistics*, eds. Fienberg, S. and Zellner, A., Amsterdam: North-Holland, pp. 555-584.

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

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

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

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

O'HAGAN, A. AND LE, H. (1994) *Conflicting Information and a Class of Bivariate Heavy-tailed Distributions*, in *Aspects of Uncertainty*, eds. Smith, A and Freeman, P., New York: John Wiley and Sons Ltd, pp. 315-331.

P-credence and outliers

SUMMARY

In a Bayesian model, a conflict between the prior and the likelihood information can occur. A way to resolve this conflict in favor of the likelihood information, is to use heavy-tailed prior. In this paper, the notion of credence for inference about a location parameter with known scale parameter is generalized to account for logarithmic, polynomial and exponential tail behavior. This generalization is called p-credence. Using the p-credence of the prior density and the likelihood function, a dominance relation between the two densities is defined. Furthermore, using the p-credence of the prior density and the likelihood function, sufficient conditions on the prior distribution are given such that the shrinkage function goes to zero when the data are conflicting with the prior belief. Finally, examples are presented to illustrate results discussed in the paper.

P-*credence* e i valori anomali

RIASSUNTO

In un modello Bayesiano, può accadere che l'informazione a priori e l'informazione contenuta nella verosimiglianza siano in conflitto. Un modo per risolvere tale conflitto in favore dell'informazione contenuta nella funzione di verosimiglianza consiste nel considerare distribuzioni a priori a code pesanti. In questo lavoro, la nozione di *credence* per l'inferenza su un parametro di posizione quando è noto il parametro di scala viene generalizzata in modo che tenga conto di un andamento della coda di tipo logaritmico, polinomiale ed esponenziale. Questa generalizzazione viene chiamata p-credence. Usando la p-credence della densità a priori e la funzione di verosimiglianza, si definisce una relazione di dominanza fra le due densità. Inoltre, usando la p-credence della densità a priori e la funzione di verosimiglianza, vengono date condizioni sufficienti sulla distribuzione a priori affinché la funzione contrazione vada a zero quando i dati sono in conflitto con le opinioni a priori. Infine, si presentano degli esempi per illustrare i risultati discussi nel lavoro.

KEY WORDS

Bayesian inference, Outlier, Heavy-tailed modeling, Generalized exponential power family, Location parameter.

[Manuscript received March 2000; final version received July 2000.]