Université de Montréal

Quelques utilisations de la densité GEP en analyse bayésienne sur les familles de position-échelle

par

Alain Desgagné

Département de mathématiques et de statistique

Faculté des arts et des sciences

Thèse présentée à la Faculté des études supérieures

en vue de l'obtention du grade de

Philosophiæ Doctor (Ph.D.)
en statistique

mai 2005

Université de Montréal

Faculté des études supérieures

Cette thèse intitulée

Quelques utilisations de la densité GEP en analyse bayésienne sur les familles de position-échelle

présentée par

Alain Desgagné

a été évaluée par un jury composé des personnes suivantes :

Roch Roy
(président-rapporteur)

Jean-François Angers
(directeur de recherche)

Louis Doray
(membre du jury)

Liqun Wang
(examinateur externe)

François Bellavance
(représentant du doyen de la FES)

Thèse acceptée le:
30 mai 2005

RÉSUMÉ

L'utilisation des distributions à ailes relevées est un outil précieux dans le développement de méthodes bayésiennes robustes, limitant l'influence des valeurs aberrantes sur l'inférence *a posteriori*. Dans un premier temps, le comportement de la densité *a posteriori* du paramètre de position est étudié, lorsque l'échantillon contient des valeurs aberrantes. La notion de p-crédence à gauche et à droite est introduite afin de caractériser et ordonner les ailes gauche et droite d'une grande classe de densités, en comparant leurs ailes à celles d'une densité de puissance d'exponentielles généralisée (GEP).

Dans le premier article, la densité GEP est proposée comme fonction d'importance dans les simulations Monte Carlo dans le contexte d'estimation des moments *a posteriori* du paramètre de position. Cela permet d'obtenir des résultats fiables et efficaces, même s'il y a des sources d'information conflictuelles. La simulation d'observations provenant d'une densité GEP est aussi discutée.

Dans le deuxième article, des conditions sur les ailes de la densité *a priori* et de la vraisemblance, basées sur la p-crédence à gauche et à droite, sont établies afin de déterminer la proportion d'observations pouvant être rejetées lorsque celles-ci sont extrêmes. Il est démontré que la distribution *a posteriori* converge en loi vers la distribution *a posteriori* obtenue à partir de l'échantillon excluant les valeurs aberrantes, lorsque ces dernières tendent vers plus ou moins l'infini, à n'importe quel taux. Un exemple de combinaison de prévisions du rendement de l'indice S&P 500 est présenté.

Finalement, dans le troisième article, le comportement de la densité *a posteriori* du paramètre d'échelle est étudié lorsque l'échantillon contient des valeurs aberrantes et que les observations sont positives. La notion de log-crédence à gauche et à droite est introduite afin de caractériser les ailes gauches et droites d'une densité définie sur $\mathbb{R}^+$. Des conditions sur les ailes de la densité *a priori* et de la vraisemblance, basées sur la log-crédence à gauche et à droite, sont établies afin de déterminer la proportion d'observations pouvant être rejetées lorsque celles-ci sont extrêmes (observations très petites ou grandes par rapport aux autres). Un exemple de combinaison de prévisions de la volatilité des rendements de l'indice S&P 500 est présenté.

MOTS CLÉS :

Inférence bayésienne, Modèle à ailes relevées, Famille de puissance d'exponentielles généralisée, Crédence, Valeurs aberrantes, Paramètre de position, Paramètre d'échelle, Convergence en loi, Simulations Monte Carlo, Fonction d'importance

SUMMARY

The use of heavy-tailed distributions is a valuable tool in developing robust Bayesian procedures, limiting the influence of outliers on posterior inference. The behavior of the posterior density, when the sample contains outliers, is first investigated for the location parameter. The notion of left and right p-credence is introduced to characterize and to order the left and right tails of a large class of densities by comparing their tails to those of the generalized exponential power (GEP) density.

In the first paper, the GEP density is proposed as an importance function in Monte Carlo simulations in the context of estimation of posterior moments of a location parameter. It allows us to obtain reliable and effective results, even if there are conflicting sources of information. Simulation of observations from the GEP density is also addressed.

In the second paper, conditions on the tails of the prior and the likelihood, using left and right p-credence, are established to determine the proportion of observations that can be rejected when they are considered extreme. It is shown that the posterior distribution converges in law to the posterior that would be obtained from the reduced sample, excluding the outliers, as they tend to plus or minus infinity, at any given rate. An example of combination of predictions of the S&P 500 index return is presented.

Finally, in the third paper, the behavior of the posterior density of the scale parameter is investigated when the sample contains outliers and only positive observations. The notion of left and right log-credence is introduced to characterize respectively the left and right tails of a density defined on $\mathbb{R}^+$. Conditions on the tails of the prior and the likelihood, using left and right log-credence, are established to determine the proportion of observations that can be rejected when they are considered extreme (observations very small or large relatively to the other ones). An example of combination of predictions of the volatility of the S&P 500 index return is presented.

KEY WORDS :

Bayesian inference, Heavy-tailed modeling, Generalized exponential power family, Outlier, Credence, Location parameter, Scale parameter, Convergence in law, Monte Carlo simulations, Importance sampling

REMERCIEMENTS

Je tiens d'abord à remercier chaleureusement mon directeur de recherche Jean-François Angers. Par ses qualités humaines, il a contribué grandement à créer un climat de travail dans lequel il fut très agréable d'évoluer. Il m'a proposé un projet de recherche que j'ai adoré et il m'a accompagné tout au long de ce projet en partageant généreusement son expertise et son expérience. Aussi, j'ai apprécié grandement son leadership, qui m'a permis de bénéficier à la fois de toute l'autonomie nécessaire pour me développer comme chercheur, de sa grande disponibilité et de son appui dans les moments plus difficiles.

Je remercie également les membres du jury pour la lecture de ma thèse ainsi que pour les améliorations qu'ils ont apportées.

Je remercie les professeurs, les étudiants et le personnel du département que j'ai côtoyés durant ces années, pour toutes les conversations échangées et les bons moments passés ensemble qui ont permis de rendre mon doctorat très agréable.

Plus particulièrement, je veux remercier les professeurs Robert Cléroux pour le support qu'il m'a donné au début de mon doctorat et Louis Doray pour les conversations sur le monde de l'actuariat et son appui dans ma recherche fructueuse d'un poste de professeur.

Je tiens aussi à remercier plus particulièrement quelques collègues étudiants pour leur amitié et les très bons moments partagés ensemble : Alexandre Leblanc, Pierre Lafaye de Micheaux, Alice Dragomir, Pascal Croteau et Jean-François Boudreau.

Je veux souligner l'apport financier important que j'ai obtenu de la part du CRSNG (Conseil de recherches en sciences naturelles et en génie du Canada), du FQRNT (Le Fonds québécois de la recherche sur la nature et les technologies), de la SOA (Society of Actuaries), du Département de mathématiques et de statistique de l'Université de Montréal ainsi que de mon directeur de recherche.

Finalement, un grand merci à ces personnes qui ont m'ont aidé indirectement dans ce projet : mes parents Yves et Pauline, mes deux soeurs Chantal et Manon et la grande famille que j'aime beaucoup, Ingrid de Lafontaine, Anne-Marie Castilloux, Marie Flagollet, Louis Beaudoin, Jean-Guy St-Jean et Jean-Christophe Boze.

TABLE DES MATIÈRES

SIGLES ET ABRÉVIATIONS

cdf : Fonction cumulative de distribution

GEP : Puissance d'exponentielles généralisée

S&P 500 : Indice boursier Standard & Poors 500

$|\cdot|$: Valeur absolue

$\mathbb{R}$: Nombres réels

$\mathbb{R}^+$: Nombres réels positifs

$\mathbb{N}$: Nombres naturels

$\xrightarrow{\mathcal{L}}$: Convergence en loi

$\mathbb{E}[\cdot]$: Espérance

$\mathrm{Var}[\cdot]$: Variance

$\mathbb{I}_{[a]}$: Function indicatrice de l'ensemble $\{a\}$

p-cred(f) : P-crédence de la densité f

p-cred$^+(f)$: P-crédence de l'aile de droite de la densité f

p-cred$^-(f)$: P-crédence de l'aile de gauche de la densité f

log-cred$^+(f)$: Log-crédence de l'aile de droite de la densité f

log-cred$^-(f)$: Log-crédence de l'aile de gauche de la densité f

LISTE DES FIGURES

LISTE DES TABLEAUX

INTRODUCTION

Avec la progression de l'informatisation dans les entreprises et les institutions gouvernementales, de plus en plus de bases de données sont maintenant disponibles. Une mine d'information souvent se cache dans ces données et le développement d'outils statistiques permettant de les analyser devient nécessaire. Par exemple, l'efficacité d'un médicament peut être démontrée en utilisant l'information contenue dans les bases de données de la Régie de l'assurance-maladie du Québec, ou le risque d'accident automobile d'un assuré peut être mieux évalué en utilisant l'information contenue dans la base de données d'une compagnie d'assurance, ce qui permet de mieux évaluer la portion de la prime d'assurance due au risque.

Une autre conséquence de l'avènement d'ordinateurs de plus en plus puissants est la possibilité accrue de développer des outils statistiques de plus en plus performants. Plusieurs approches statistiques, utilisant par exemple les simulations ou le calcul numérique, sont maintenant utilisables via l'ordinateur afin d'analyser ces bases de données.

Une de ces approches statistiques est l'inférence bayésienne. Elle consiste essentiellement à combiner l'information provenant des données avec de l'information *a priori* réputée indépendante des données, pour en tirer de l'information *a posteriori*. L'approche bayésienne nécessite souvent des calculs numériques intensifs et le développement des ordinateurs a sans doute contribué à son essor.

Un des critères de performance recherché en statistique est la robustesse des modèles face aux valeurs aberrantes. Pour la plupart des échantillons, un modèle statistique peut être très efficace pour estimer un paramètre inconnu, mais peut ne plus être adéquat dès qu'une ou plusieurs valeurs extrêmes ou aberrantes apparaissent dans les données.

Le but de cette thèse est de développer des outils statistiques bayésiens robustes, qui demeurent efficaces même si des valeurs aberrantes viennent contaminer les données. Le contexte bayésien est décrit plus en détail dans la prochaine section.

CONTEXTE BAYÉSIEN

Soient n variables aléatoires $X_1, ..., X_n$ conditionnellement indépendantes étant donné le paramètre θ. La densité conditionnelle de $X_i|\theta$ est donnée par $f_i(x_i|\theta)$, avec $\theta \in \Theta$ et $X_i \in \mathbb{D}$, $i = 1, ..., n$, où Θ et $\mathbb{D}$ sont des sous-ensembles de $\mathbb{R}$. Notez que l'espace paramétrique Θ peut être multidimensionnel, mais seul le cas unidimensionnel est considéré dans cette thèse. L'inférence est faite sur le paramètre θ à partir des données observées $x_1, ..., x_n$. Notez que dans l'approche classique (fréquentiste), une nouvelle expérimentation produirait un nouvel échantillon $x_1, ..., x_n$ et la valeur (inconnue) de θ resterait la même, tandis que dans l'approche bayésienne, une nouvelle expérimentation produirait une nouvelle valeur de θ. L'aspect aléatoire dans l'approche bayésienne est donc mis sur le paramètre θ plutôt que sur les données.

L'information *a priori* sur θ est donc intégrée dans le modèle par l'entremise d'une loi *a priori* sur θ, dénotée par $\pi(\theta)$ (voir DeGroot, 1970). Il est donc possible d'inférer sur le paramètre θ sans même observer de données, à l'aide de la loi *a priori*. Le paradigme bayésien consiste à mettre à jour cette densité *a priori* en y intégrant les observations $x_1, ..., x_n$ par le théorème de Bayes. Rappelons que par le théorème de Bayes, $\Pr[A|B] = \Pr[B|A]\Pr[A]/\Pr[B]$, où A et B sont deux sous-ensembles d'un espace échantillonal S, c'est-à-dire $A \subseteq S$ et $B \subseteq S$. Nous obtenons alors la densité *a posteriori* de θ, dénotée par $\pi(\theta|x_1, ..., x_n)$ et donnée par

$$\pi(\theta|x_1, ..., x_n) = \frac{\pi(\theta)\prod_{i=1}^{n} f_i(x_i|\theta)}{m(x_1, ..., x_n)},$$

où $m(x_1, ..., x_n) = \int_{\Theta} \pi(\theta)\prod_{i=1}^{n} f_i(x_i|\theta)d\theta$ est la densité marginale conjointe de $X_1, ..., X_n$. L'inférence sur θ est effectuée à l'aide de sa densité *a posteriori*, tenant compte à la fois de l'information *a priori* et des données. Notez que la densité *a posteriori* est proportionnelle au produit des densités de toutes les sources d'information. Selon le critère d'optimisation (fonction de coûts) choisi, le paramètre θ peut être estimé par l'espérance, la médiane ou le mode *a posteriori* de θ.

Dans cette thèse, nous nous intéressons aux densités des observations qui font partie d'une famille de position-échelle. Soit une variable aléatoire Y ayant comme densité $f(y)$, alors la densité de $X = \sigma Y + \mu$, étant donné μ et σ connus, est donnée par $\sigma^{-1} f([x - \mu]/\sigma)$, où $\mu \in \mathbb{R}$ est le paramètre de position et $\sigma > 0$ est le paramètre d'échelle. Nous disons que la densité de X est une famille de position-échelle, ou simplement une famille de position si σ est fixé ou encore une famille d'échelle si μ est fixé. Une variation du paramètre de position entraîne une translation de la densité et une variation du paramètre d'échelle entraîne un étirement ou un rétricissement de la densité. Nous aborderons dans cette thèse la famille de position dans les deux premiers articles et la famille d'échelle dans le troisième article. Le paramètre θ, donné dans le cadre bayésien, sera donc un paramètre de position ou un paramètre d'échelle.

Robustesse

Un exemple souvent présenté comme introduction à la statistique bayésienne consiste à faire l'hypothèse que la densité *a priori* et la fonction de vraisemblance (densité des observations) sont représentées par des densités normales. Si x est une observation provenant d'une population normale avec une moyenne donnée par θ et une variance de 1, ce qui est dénoté par $N(\theta, 1)$, et que la densité *a priori* de θ est $N(0, 1)$, alors on peut démontrer que la densité *a posteriori* de θ est $N(x/2, 1/2)$. Notez que la densité de l'observation appartient à une famille de position. Selon l'information *a priori*, la distribution du paramètre θ est centrée en 0, tandis que l'observation est donnée par x. La densité *a posteriori* est centrée en $x/2$, ce qui représente un compromis entre les deux sources d'information. La densité *a posteriori* de θ est proportionnelle au produit de la densité *a priori* et de la vraisemblance considérée comme une fonction de θ, soit le produit d'une densité $N(0, 1)$ et d'une densité $N(x, 1)$. Rappelons que la densité normale est unimodale et symétrique par rapport à sa moyenne. Une petite variance indique une grande certitude de la source d'information et vice-versa. Par exemple, selon l'information *a priori*, il y a moins d'une chance sur 1000 que $|\theta|$ soit supérieur à 3,3, puisque la densité de θ est $N(0, 1)$. De la même façon, selon la vraisemblance, il y a moins d'une chance sur 1000 que $|\theta - x|$ soit supérieur à 3,3. Le compromis de $x/2$ semble naturel si les deux sources d'information sont compatibles. Par exemple si $x = 3$, la densité *a posteriori* fera un compromis avec une moyenne de $x/2 = 1,5$, ce qui est en accord à la fois avec l'information *a priori* et avec la vraisemblance. Toutefois, si $x = 10$ et que la densité *a posteriori* fait un compromis avec une moyenne de $x/2 = 5$, ce n'est pas souhaitable, puisqu'il est en désaccord à la fois avec l'information *a priori* (moins d'une chance sur 1000 d'être inférieur à -3,3 ou supérieur à 3,3) et avec l'observation (moins d'une chance sur 1000 d'être inférieur à 6,7 ou supérieur à 13.3). Il serait souhaitable dans ce cas d'écarter la source d'information jugée la moins fiable et que la densité *a posteriori* se rapproche de la source d'information ayant la plus grande crédibilité en cas de conflit. Si toutefois nous faisons l'hypothèse que la densité *a priori* est représentée par une densité ayant des ailes suffisamment relevées, telle la densité Student-t, et que la vraisemblance est toujours représentée par une densité normale, le conflit sera réglé en faveur de la fonction de vraisemblance.

L'utilisation des densités à ailes relevées est un outil important dans le développement de méthodes bayésiennes robustes, limitant l'influence des valeurs aberrantes sur l'inférence *a posteriori*. Le rejet de valeurs aberrantes a d'abord été décrit par De Finetti (1961), dans le cas le plus simple où il n'y a qu'une observation avec une moyenne de θ. Des résultats théoriques ont été donnés par Dawid (1973) et Hill (1974). O'Hagan (1979) a considéré le rejet de valeurs aberrantes dans un échantillon et O'Hagan (1988) a considéré une modélisation bayésienne plus générale basée sur les densités Student-t. O'Hagan (1990) a introduit la notion de crédence pour caractériser et ordonner des ailes de densités symétriques ayant un comportement de type polynomial, telle la densité Student-t. Il a également donné des résultats de rejet de valeurs aberrantes basés sur la crédence. Cette notion a été généralisée à la p-crédence par Angers (2000) afin d'englober une plus grande classe de densités. D'autres auteurs ont également abordé le rejet de valeurs aberrantes, par exemple Meinhold et Singpurwalla (1989), Angers et Berger (1991), Carlin et Polson (1991), Angers (1992), Fan et Berger (1992), Geweke (1994) et Angers (1996).

Densité GEP

La p-crédence caractérise les ailes de densités symétriques en comparant ses ailes à celles d'une densité de référence appelée famille de puissance d'exponentielles généralisée (GEP). La forme générale de la densité GEP, telle qu'introduite par Angers (2000), est donnée par

$$\begin{aligned} p(z|\gamma, \delta, \alpha, \beta, z_0) &= K(\gamma, \delta, \alpha, \beta, z_0) \exp\{-\delta \max(|z|, z_0)^\gamma\} \\ &\quad \times \max(|z|, z_0)^{-\alpha} \log^{-\beta}[\max(|z|, z_0)] \\ &\propto \begin{cases} e^{-\delta|z|^\gamma} |z|^{-\alpha} \log^{-\beta}|z|; & \text{si } |z| > z_0, \\ e^{-\delta z_0^\gamma} z_0^{-\alpha} \log^{-\beta} z_0; & \text{si } |z| \leq z_0, \end{cases} \end{aligned}$$

où $z \in \mathbb{R}$, $\gamma \geq 0$, $\delta \geq 0$ (nous posons $\delta = 0$ lorsque $\gamma = 0$), $\alpha \in \mathbb{R}, \beta \in \mathbb{R}$, $z_0 \geq 0$ et $K(\gamma, \delta, \alpha, \beta, z_0)$ est la constante de normalisation.

La densité GEP comprend un terme exponentiel, polynomial et logarithmique. Elle est symétrique par rapport à 0 et constante entre $-z_0$ et z_0. Les quatre autres paramètres $\gamma, \delta, \alpha, \beta$ déterminent le comportement des ailes. La plupart des densités connues définies sur $\mathbb{R}$ ou sur $\mathbb{R}^+$ ont le même comportement dans les ailes que la densité GEP, ce qui en fait une densité de référence utile. Aussi, comme la robustesse est liée principalement à l'épaisseur des ailes des densités, la densité GEP s'avère un outil puissant pour développer des méthodes statistiques bayésiennes robustes.

Dans cette thèse, trois utilisations de la densité GEP en robustesse bayésienne sont présentées sous forme de trois articles.

Premier article

Dans le premier article, intitulé *Importance Sampling with the Generalized Exponential Power Density*, la densité GEP est proposée comme fonction d'importance dans les simulations Monte Carlo, dans le contexte de l'estimation des moments *a posteriori* du paramètre de position. Il est possible de caractériser les ailes de la densité *a posteriori* par la p-crédence (voir Angers, 2000), ce qui permet de choisir les paramètres de la densité GEP de façon à ce que les ailes de la fonction d'importance soit légèrement plus relevées ou équivalentes à celles de la densité *a posteriori*. Aussi il est possible de simuler des observations à partir de la densité GEP, ce qui est abordé dans l'article. Souvent le choix d'une fonction d'importance se fait sur la base du cas par cas. Il est toutefois possible, avec la densité GEP, de choisir une fonction d'importance selon une méthode qui est la même peu importe les choix de la densité *a priori* et de la vraisemblance, en autant que leur p-crédence soit définie. Aussi la présence de valeurs aberrantes ne compromet pas l'efficacité de cette méthode, ce qui n'est souvent pas le cas pour des choix de fonctions d'importance *ad hoc*. D'autres approches permettant de choisir une fonction d'importance, basées sur la discrétisation aléatoire par exemple (voir Fu et Wang, 2002), peuvent aussi être efficaces. Il s'agit souvent de faire un compromis judicieux entre l'efficacité et la simplicité, et c'est dans cet esprit que la méthode proposée dans le premier article a été construite.

Dans les deuxième et troisième articles, nous nous intéressons aux conditions portant sur la densité *a priori* et sur la vraisemblance qui sont suffisantes pour que la densité *a posteriori* rejette les valeurs aberrantes. Dans le deuxième article, la vraisemblance est une famille de position, tandis que dans le troisième, elle est une famille d'échelle. Les conditions portent essentiellement sur l'épaisseur des ailes.

Deuxième article

Dans le deuxième article, intitulé *Outliers and choice of the prior for location parameter inference*, la p-crédence est d'abord généralisée afin de caractériser les ailes de gauche et de droite d'une densité définie sur les réels, ce qui permet de considérer la robustesse en présence de densités asymétriques. Dans un tel cas, une observation trop petite par rapport aux autres pourrait par exemple être rejetée, mais ne le serait pas si elle était trop grande, ou vice-versa. Dans un premier temps, les conditions de robustesse sont exprimées à l'aide de la p-crédence. Ces conditions sont relativement faciles à vérifier, contrairement à celles énoncées dans Dawid (1973) par exemple. Comme la p-crédence est définie pour la plupart des densités connues, ces conditions sont d'un grand intérêt en pratique. Rappelons que les résultats de O'Hagan (1990) ne s'appliquent que pour les densités ayant un comportement de type polynomial dans les ailes, tandis que les résultats du deuxième article s'appliquent en plus pour les densités ayant un comportement de type exponentiel et logarithmique dans les ailes.

Les résultats de robustesse sont donnés sous forme de convergence en loi. Si les conditions sont respectées, il est démontré que la densité *a posteriori* converge en loi vers la densité *a posteriori* que nous aurions obtenue à partir d'un échantillon excluant les valeurs aberrantes, à mesure que celles-ci tendent vers plus ou moins l'infini. D'une part ces résultats sont plus forts que ceux donnés dans O'Hagan (1990) et Angers (2000), qui ont démontré que le ratio des densités *a posteriori* avec l'échantillon complet et avec celui excluant les valeurs aberrantes est borné par des constantes positives, pour n'importe quelle valeur finie du paramètre de position. Le prix pour obtenir la convergence en loi a été d'ajouter une condition sur la régularité des ailes de la densité *a priori* et de la vraisemblance. Toutefois, cette condition est satisfaite pour la plupart des lois connues. D'autre part, les résultats de convergence sont donnés lorsqu'on est en présence d'un échantillon comportant possiblement une ou plusieurs valeurs aberrantes qui tendent vers plus ou moins l'infini, et ce à n'importe quel taux donné.

Dans un deuxième temps, les conditions sont présentées d'une façon plus générale, sans utiliser la p-crédence. Même si leur utilisation en pratique devient moins intéressante, elles demeurent relativement simples. L'interprétation des conditions devient toutefois plus aisée, puisque l'influence de chaque aile de la densité *a priori* et de la vraisemblance sur le rejet des valeurs aberrantes peut être observée. Essentiellement, un groupe de valeurs aberrantes tendant vers l'infini sera rejeté si l'aile de gauche de la densité proportionelle au produit de leurs densités est suffisamment relevée et plus relevée que l'aile de droite de la densité proportionnelle au produit des densités des observations non aberrantes et de la densité *a priori*. Le rejet des valeurs aberrantes tendant vers moins l'infini est similaire, sauf que le rôle des ailes de gauche et de droite est inversé. Le paramètre de position est la variable aléatoire, les observations étant considérées fixes. Les résultats de convergence sont valides quelque soit le taux auquel les valeurs aberrantes tendent vers plus ou moins l'infini. Quoiqu'ayant peu d'intérêt en pratique, il serait possible d'ajouter d'autres conditions de convergence si le taux auquel chaque valeur aberrante tend vers plus ou moins l'infini était spécifié.

Troisième article

Dans le troisième article, intitulé *Outliers for scale parameter inference and positive observations*, la robustesse est étudiée lorsque la densité des observations est une famille d'échelle et que l'inférence est faite sur un paramètre d'échelle. Jusqu'à présent, aucun article n'a été publié sur ce sujet. Lorsque les observations sont considérées positives, il s'agit essentiellement de transposer les résultats du deuxième article par une transformation exponentielle (ou logarithmique si on fait le chemin inverse). Le paramètre de position devient un paramètre d'échelle et le domaine réel devient un domaine réel positif. Les conditions, résultats et preuves de robustesse sur le paramètre d'échelle sont toutefois présentés indépendemment de leur contrepartie du monde du paramètre de position. En effet, nous aurions pu être tentés de ne pas présenter le détail des preuves en référant simplement au deuxième article et en précisant qu'une transformation exponentielle doit être effectuée. Quoique facile à faire pour certaines parties des preuves, il s'est avéré d'une part qu'il n'est pas du tout trivial en général de transposer les preuves, et d'autre part que des éléments nouveaux doivent souvent être apportés.

D'abord la notion de log-crédence est introduite, qui consiste à comparer les ailes d'une densité définie sur les réels positifs à celle d'une densité log GEP. La densité log GEP est définie simplement comme une transformation exponentielle de la densité GEP, c'est-à-dire si $X > 0$ a une densité log GEP, alors $\log X$ a une densité GEP (un cas particulier correspond aux lois normale et log-normale). Tout comme dans le cas de la p-crédence, le comportement des ailes de densités de type polynomial et logarithmique est considéré dans la log-crédence, mais le comportement exponentiel est modifié et un terme logarithme de logarithme y est ajouté.

L'interprétation de l'aile de gauche est un peu différente car le domaine est borné à gauche par 0. L'aile de gauche peut monter vers l'infini, descendre vers 0 ou se diriger vers une constante positive. On dira par exemple que l'aile de gauche sera plus relevée si elle monte vers l'infini que si elle descend à 0. La log-crédence peut mesurer l'aile de gauche d'une densité définie sur les réels positifs, ce que la p-crédence ne peut faire. L'interprétation des conditions et des résultats de convergence pour le cas du paramètre d'échelle est similaire au cas du paramètre de position. Les ailes des densités des valeurs aberrantes doivent être aussi suffisamment relevées, mais encore plus relevées que dans le cas du paramètre de position.

Le cas du paramètre d'échelle quand les observations sont réelles n'est pas considéré dans cet article, mais il est facile de le faire en généralisant les résultats en utilisant la symétrie par rapport à l'origine. Dans le cas d'observations réelles, il n'y a toutefois pas de valeurs aberrantes près de 0, et il en résulte que l'aile de gauche de la densité *a priori* du paramètre d'échelle n'a plus d'influence sur la robustesse. Une suite à ces travaux de thèse est en cours, où le cas de la famille de position-échelle est considéré, lorsque les données sont réelles.

Bibliographie

ANGERS, J.-F. (1992) Use of Student-t prior for the estimation of normal means : A computational approach, *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, *IMS Lecture Notes - Monograph Series*, Vol. **29**, pp. 273-283.

ANGERS, J.-F. (2000) P-Credence and outliers, *Metron*, **58**, 81-108.

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

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

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

DE FINETTI, B. (1961) The Bayesian approach to the rejection of outliers, *Proceedings of the Fourth Berkeley Symposium on Probability and Statistics*, **(Vol. 1)**, Berkeley : University of California Press, pp. 199-210.

DEGROOT, M.H. (1970) *Optimal statistical decisions*, McGraw-Hill, New York.

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

FU, J.C. et WANG, L. (2002) A random-discretization based Monte Carlo sampling method and its applications, *Methodology and Computing in Applied Probability*, **4**, 5-25.

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

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

MEINHOLD, R. et 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, *Bayesian Statistic III*, eds.

Bernardo, J.M., DeGroot M.H., Lindley, D.V., 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.

Chapitre 1

IMPORTANCE SAMPLING WITH THE GENERALIZED EXPONENTIAL POWER DENSITY

Cet article a été publié et sa référence est la suivante : DESGAGNÉ, A. et ANGERS, J.-F. (2005) Importance Sampling with the Generalized Exponential Power Density, *Statistics and Computing*, **15**, 189-195.

Abstract : In this paper, the generalized exponential power (GEP) density is proposed as an importance function in Monte Carlo simulations in the context of estimation of posterior moments of a location parameter. This density is divided in five classes according to its tail behaviour which may be exponential, polynomial or logarithmic. The notion of p-credence is also defined to characterize and to order the tails of a large class of symmetric densities by comparing their tails to those of the GEP density. The choice of the GEP density as an importance function allows us to obtain reliable and effective results when p-credences of the prior and the likelihood are defined, even if there are conflicting sources of information. Characterization of the posterior tails using p-credence can be done. Hence, it is possible to choose parameters of the GEP density in order to have an importance function with slightly heavier tails than the posterior. Simulation of observations from the GEP density is also addressed.

Key words : Importance sampling, Credence, Heavy tail density, Numerical integration, Monte Carlo.

1.1. Introduction

The use of heavy-tailed distributions is a valuable tool in developing robust Bayesian procedures, limiting the influence of extremes on posterior inference. (See for instance Meinhold and Singpurwalla, 1989; O'Hagan, 1990; Angers and Berger, 1991; Carlin and Polson, 1991; Angers, 1992; Fan and Berger, 1992; Geweke, 1994; Angers, 1996). O'Hagan (1990) introduced the notion of credence to characterize the tails of a symmetric density on the real line. This notion has been generalized to p-credence by Angers (2000) to accommodate a wider class of densities. P-credence of a density is determined by comparing its tail to a reference density introduced in Angers (2000), called the generalized exponential power (GEP) density.

An application of this density in Monte Carlo simulations with importance sampling is proposed in this paper, in the context of estimation of the posterior moments of a location parameter. The GEP density may be a good candidate for the importance function in Monte Carlo simulations because simulation of observations from this density is possible. Furthermore, characterization of the posterior tails using p-credence (see Angers, 2000) makes it possible to choose the parameters of the GEP density such that the tails of the importance function are slightly heavier than those of the posterior.

In Section 1.2, the GEP density is introduced and the notion of p-credence is defined to characterize and to order tails of densities. In Section 1.3, the importance sampling using the GEP density as an importance function is addressed. The setting is given in Section 1.3.1. The selection of parameters of the GEP density is addressed in Sections 1.3.2 and 1.3.3 and simulation of observations from this density is considered in Section 1.3.4. Finally, an example is provided in Section 1.4.

1.2. Generalized exponential power density and dominance relation using p-credence

The generalized exponential power density is introduced in Section 1.2.1.

1.2.1. Generalized exponential power density

The general form of the density of the generalized exponential power family as introduced by Angers (2000) is given by

$$p(z|\gamma,\delta,\alpha,\beta,z_0) = K(\gamma,\delta,\alpha,\beta,z_0)\exp\{-\delta\max(|z|,z_0)^\gamma\}$$
$$\times\max(|z|,z_0)^{-\alpha}\log^{-\beta}[\max(|z|,z_0)]$$
$$\propto\begin{cases} e^{-\delta|z|^\gamma}\,|z|^{-\alpha}\log^{-\beta}|z|\,; & \text{if } |z|>z_0,\\ e^{-\delta z_0^\gamma}\,z_0^{-\alpha}\log^{-\beta}z_0; & \text{if } |z|\le z_0,\end{cases} \tag{1.2.1}$$

where $z\in\mathbb{R}$, $\gamma\geq 0$, $\delta\geq 0$ (we set $\delta=0$ when $\gamma=0$), $\alpha\in\mathbb{R},\beta\in\mathbb{R}$, $z_0\geq 0$ and $K(\gamma,\delta,\alpha,\beta,z_0)$ is the normalizing constant. (Note that the parameters α and β of the GEP density defined in Angers (2000) have been changed respectively to $-\alpha$ and $-\beta$ in order to ease the comparison of p-credences.) In addition, the parameters γ, δ, α, β and z_0 must satisfy the following conditions :

$$\textbf{C1}\ :\ z_0>\begin{cases}1; & \text{if } \beta\neq 0,\\ 0; & \text{if } \alpha\neq 0,\beta=0\,;\end{cases}$$

$$\textbf{C2}\ :\ \alpha+\tfrac{\beta}{\log z_0}+\delta\gamma z_0^\gamma\geq 0\,;$$

$$\textbf{C3}\ :\ \alpha\geq 1 \text{ if } \gamma=0\,;$$

$$\textbf{C4}\ :\ \beta>1 \text{ if } \gamma=0,\alpha=1.$$

The first condition is needed in order for the density to be strictly positive and bounded. The second condition guarantees the unimodality of the density and it is always satisfied if z_0 is chosen to be large enough. The third and fourth conditions ensure that it is a proper density. The density is symmetric with respect to the origin and is constant for $-z_0\leq z\leq z_0$, which puts emphasis on the tails.

The family of GEP densities which satisfy conditions C1 to C4 can be divided in five subsets as shown in Table 1.1. Each subset is determined by the tail behaviour of the density. The right tail of a GEP density is equivalent to a generalized gamma density for type II, it is equivalent to a log-gamma density for type III, a Pareto density for type IV and a log-Pareto density for type V. Type I corresponds to a general case. Note that types III and IV are heavy-tailed distributions and type V is a super heavy-tailed distribution (see Reiss and Thomas, 1997).

TAB. 1.1. Five types of the GEP density

Type	Parameters				
I	$\gamma > 0,$	$\delta > 0,$	$\alpha \in \mathbb{R},$	$\beta \neq 0,$	$z_0 > 1$ and $\alpha + \frac{\beta}{\log z_0} + \delta\gamma z_0^\gamma \geq 0$
II	$\gamma > 0,$	$\delta > 0,$	$\alpha \in \mathbb{R},$	$\beta = 0,$	$z_0 > \max\left(0, \frac{-\alpha}{\delta\gamma}\right)^{1/\gamma}$ $(z_0 \geq 0$ if $\alpha = 0)$
III	$\gamma = 0,$	$\delta = 0,$	$\alpha > 1,$	$\beta \neq 0,$	$z_0 > \max\left(1, e^{-\beta/\alpha}\right)$
IV	$\gamma = 0,$	$\delta = 0,$	$\alpha > 1,$	$\beta = 0,$	$z_0 > 0$
V	$\gamma = 0,$	$\delta = 0,$	$\alpha = 1,$	$\beta > 1,$	$z_0 > 1$

Many known distributions (see Johnson, Kotz and Balakrishnan, 1994) are special cases of the GEP density. If the parameters α and z_0 of the GEP density of type II are set to 0, it gives the exponential power density (see Box and Tiao, 1962). In addition, if the parameter γ is set to 2, it gives the normal density and if the parameter γ is set to 1, it gives the Laplace density. Furthermore, the right tail of the GEP density of type II is equivalent to a Weibull density if $\alpha = 1 - \gamma$, a gamma density if $\gamma = 1$ and $\alpha < 1$, a Rayleigh density if $\gamma = 2$ and $\alpha = -1$ and a Maxwell-Boltzmann density if $\gamma = 2$ and $\alpha = -2$.

The normalizing constant $K(\gamma, \delta, \alpha, \beta, z_0)$ and the j^{th} moment of $|Z|$ are given in Table 1.2, except for the GEP density of type I, which may be evaluated using Monte Carlo simulations with importance sampling.

Note that in Table 1.2, $\Gamma(\lambda, a)$ is the incomplete gamma function defined by

$$\Gamma(\lambda, a) = \int_a^\infty e^{-u} u^{\lambda-1} du,$$

$\lambda \in \mathbb{R}, a > 0\,(a \geq 0$ if $\lambda > 0)$. In particular, when $a = 0$ and $\lambda > 0$, $\Gamma(\lambda, 0)$ is the gamma function and it is denoted by $\Gamma(\lambda)$.

1.2.2. Dominance relation using p-credence

The GEP density was introduced to provide a benchmark for the characterization of the tail behaviour of a density. Such a characterization is addressed by the notion of p-credence, defined in Angers (2000) as follows :

TAB. 1.2. Normalizing constant and moments of the GEP density

Type	Normalizing constant $K(\gamma,\delta,\alpha,\beta,z_0)$	Moments $\mathbb{E}(\lvert Z\rvert^j),\ j>0$
II	$\frac{1}{2}\left[\mathrm{e}^{-\delta z_0^\gamma}\, z_0^{1-\alpha} + \frac{\Gamma\left(\frac{1-\alpha}{\gamma},\delta z_0^\gamma\right)}{\gamma\delta^{\frac{1-\alpha}{\gamma}}}\right]^{-1}$	$\dfrac{\frac{1}{j+1}\,\mathrm{e}^{-\delta z_0^\gamma}\, z_0^{1-\alpha+j} + \frac{\Gamma\left(\frac{1-\alpha+j}{\gamma},\delta z_0^\gamma\right)}{\gamma\delta^{\frac{1-\alpha+j}{\gamma}}}}{\mathrm{e}^{-\delta z_0^\gamma}\, z_0^{1-\alpha} + \frac{\Gamma\left(\frac{1-\alpha}{\gamma},\delta z_0^\gamma\right)}{\gamma\delta^{\frac{1-\alpha}{\gamma}}}}$
III	$\frac{1}{2}\left[z_0^{1-\alpha}\log^{-\beta} z_0 + \frac{\Gamma(1-\beta,(\alpha-1)\log z_0)}{(\alpha-1)^{1-\beta}}\right]^{-1}$	$\dfrac{\frac{1}{j+1} z_0^{1-\alpha+j}\log^{-\beta} z_0 + \frac{\Gamma(1-\beta,(\alpha-j-1)\log z_0)}{(\alpha-j-1)^{1-\beta}}}{z_0^{1-\alpha}\log^{-\beta} z_0 + \frac{\Gamma(1-\beta,(\alpha-1)\log z_0)}{(\alpha-1)^{1-\beta}}}$ $(j<\alpha-1)$
IV	$\frac{(\alpha-1)z_0^{\alpha-1}}{2\alpha}$	$\frac{(\alpha-1)z_0^j}{(j+1)(\alpha-j-1)},\ (j<\alpha-1)$
V	$\frac{(\beta-1)(\log z_0)^\beta}{2(\beta-1+\log z_0)}$	∞

Definition 1. *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 k, K $(0<k\le K<\infty)$ such that for all $z\in\mathbb{R}$*

$$k \le \frac{f(z)}{p(z|\gamma,\delta,\alpha,\beta,z_0)} \le K,$$

where $p(z|\gamma,\delta,\alpha,\beta,z_0)$ is given by equation (1.2.1). We also write p-cred$(Z)=$ $(\gamma,\delta,\alpha,\beta)$ *if the density of Z has p-credence $(\gamma,\delta,\alpha,\beta)$.*

The notion of p-credence characterizes the tail behaviour of a density by comparing it to a GEP density. Essentially, this definition ensures that $f(z)$ is of order $\mathrm{e}^{-\delta|z|^\gamma}|z|^{-\alpha}\log^{-\beta}|z|$ for large values of $|z|$. P-credence is defined for densities having the same behaviour in the left and right tails, like the symmetric densities for example. Note that the parameter z_0 is not listed as an argument in p-credence since it has no influence on the tail behaviour (see Angers, 2000). By Definition 1, it is trivial to see that p-credence of $p(z|\gamma,\delta,\alpha,\beta,z_0)$ is $(\gamma,\delta,\alpha,\beta)$. It should also be noted that allowing γ to be negative would provide no more generality in the tail behaviour of a density. Furthermore, most of the usual symmetric densities on $\mathbb{R}$ (such as the normal, Student's t, Laplace and logistic) are covered by this definition of p-credence.

Once the tail behaviour of densities has been characterized by p-credence, a dominance relation can be established to compare them.

Definition 2. *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 $g \not\succeq f$.

Note that if p-cred$(f) = (\gamma, \delta, \alpha, \beta)$ then $f \approx p(\cdot|\gamma, \delta, \alpha, \beta, z_0)$, where $p(\cdot|\gamma, \delta, \alpha, \beta, z_0)$ is given by equation (1.2.1). The densities are ordered by the dominance relation as shown in Proposition 1.

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) $\gamma' = \gamma$, $\delta' = \delta$, $\alpha' = \alpha$, $\beta' < \beta$.

If $f \approx g$, we say that f and g have the same p-credence and we write $(\gamma', \delta', \alpha', \beta') = (\gamma, \delta, \alpha, \beta)$. If $f \succ g$, we say that p-credence of f is lower than p-credence of g and we write $(\gamma', \delta', \alpha', \beta') < (\gamma, \delta, \alpha, \beta)$. Finally we write $(\gamma', \delta', \alpha', \beta') \leq (\gamma, \delta, \alpha, \beta)$ if $f \succeq g$. Note that p-credence of the GEP densities of types I and II is larger than that of types III and IV, these latter ones having themselves p-credence larger than that of type V.

Here α and β have been defined differently from Angers (2000) in order to ease the comparison of p-credences. In fact, when p-cred$(f) = (\gamma', \delta', \alpha', \beta')$ and p-cred$(g) = (\gamma, \delta, \alpha, \beta)$ are compared using Proposition 1, the parameters are compared from left to right. As soon as an inequality between two parameters occurs, we say that the density with the largest parameter has the largest p-credence.

1.3. Importance sampling

In this section, the estimation of the posterior moments in Bayesian inference with location parameter is studied. It is assumed that p-credence of the prior and the likelihood are defined. The GEP density is proposed as an importance function when the estimation is performed using Monte Carlo simulations with importance sampling.

1.3.1. Setting

Consider $n+1$ densities $f_i(x_i - \theta)$, $i = 0, ..., n$, defined on $\mathbb{R}$ with

i) p-cred$(f_i) = (\gamma_i, \delta_i, \alpha_i, \beta_i)$, $i = 0, ..., n$,

ii) the density of the data $X_i|\theta$ is $f_i(x_i - \theta)$, $i = 1, ..., n$,

iii) the prior density of θ is $f_0(x_0 - \theta)$, where x_0 is a known prior location parameter.

If the vector composed of the prior location and the observations is denoted by $\mathbf{x} = (x_0, x_1, ..., x_n)$, then for $j \in \mathbb{N}$, the j^{th} posterior moment is given by $\mathbb{E}(\theta^j|\mathbf{x}) = I_j(\mathbf{x})/I_0(\mathbf{x})$, where

$$I_j(\mathbf{x}) = \int_{-\infty}^{\infty} \theta^j \prod_{i=0}^{n} f_i(x_i - \theta)\, d\theta.$$

The algorithm of Monte Carlo for the estimation of $\mathbb{E}(\theta^j|\mathbf{x})$ consists in generating $\theta_1, ..., \theta_m$ from an importance function $g(\theta)$ and estimating $I_j(\mathbf{x})$ by

$$\widehat{I}_j(\mathbf{x}) = \frac{1}{m} \sum_{k=1}^{m} \theta_k^j w(\theta_k),$$

where the weight function $w(\theta_k)$ is given by

$$w(\theta_k) = \frac{\prod_{i=0}^{n} f_i(x_i - \theta_k)}{g(\theta_k)}.$$

The j^{th} posterior moment is then estimated by $\widehat{\mathbb{E}}(\theta^j|\mathbf{x}) = \widehat{I}_j(\mathbf{x})/\widehat{I}_0(\mathbf{x})$. Note that any constant can be multiplied to the weight function since it cancels out in the evaluation of $\widehat{\mathbb{E}}(\theta^j|\mathbf{x})$.

The choice of $g(\theta)$ is the main issue of this section. The GEP density with a location parameter, given by $p(\theta - \mu^*|\gamma^*, \delta^*, \alpha^*, \beta^*, z_0^*)$, is proposed as an importance function. The selection of parameters is addressed in the next subsections.

1.3.2. The uniform part of the importance function

The first criterion for the choice of the parameters of the GEP density is that the importance function should be close to the posterior. The posterior density can be multimodal, with possible modes around the prior location and around each observation. It is thus difficult to choose an importance function close to the posterior. At least, their mass should be in the same area. This will be addressed by the uniform part of the GEP density.

The parameters μ^* and z_0^* are chosen to ensure that the uniform part of the GEP density is covering most of the prior and the likelihood (expressed as a function of θ). Let us first define the $(100p)^{th}$ percentile of f_i for $i = 0, ..., n$, denoted by $q_{p,i}$, such that $\int_{-\infty}^{q_{p,i}} f_i(z)\, dz = p$. The location parameter of the importance function is then given by

$$\mu^* = \frac{m_1 + m_2}{2}, \tag{1.3.1}$$

where

$$m_1 = \min_{i=0,\dots,n} [x_i + q_{p,i}] \text{ and } m_2 = \max_{i=0,\dots,n} [x_i + q_{1-p,i}],$$

and $0 < p < 0.5$. Furthermore, to ensure that the uniform part is covering at least the area $[m_1, m_2]$, z_0^* must satisfy the condition

$$z_0^* \geq \frac{m_2 - m_1}{2}. \tag{1.3.2}$$

In practice, the choice of $p = 0.05$ seems appropriate to cover a sufficient part of the prior and the likelihood. The choice of p being arbitrary, a rough approximation of the percentiles is sufficient.

1.3.3. Characterization of the posterior using p-credence

The second criterion for the choice of the parameters of the GEP density is that the importance function dominates the posterior. This is addressed with p-credence of the importance function, given by $(\gamma^*, \delta^*, \alpha^*, \beta^*)$. These parameters are chosen to make p-credence of the importance function lower than p-credence of the posterior.

P-credence of the posterior density with one observation is given in Angers (2000). This result can be generalized to $n \geq 1$ as follows :

$$\pi(\theta|\mathbf{x}) \preceq p(\theta - \mu'|\gamma', \delta', \alpha', \beta', z_0')$$

where

$$\gamma' = \max_{i=0,\ldots,n} \gamma_i, \quad \delta' = \sum_{i=0}^{n} \delta_i \mathbb{I}_{[\gamma_i=\gamma']} - \epsilon_1 \ \mathbb{I}_{[\gamma'>1]},$$

$$\alpha' = \sum_{i=0}^{n} \alpha_i, \quad \beta' = \sum_{i=0}^{n} \beta_i,$$

$\mu' \in \mathbb{R}$, $0 < \epsilon_1 < \sum_{i=0}^{n} \delta_i \mathbb{I}_{[\gamma_i=\gamma']}$, z_0' satisfies conditions C1, $\mathbb{I}_{[a]}$ is the indicator function of the set $\{a\}$ and $\pi(\theta|\mathbf{x})$ is the posterior density under the setup given in Section 1.3.1.

Choosing the parameters of the importance function such that

$$(\gamma^*, \delta^*, \alpha^*, \beta^*) = (\gamma', \delta', \alpha', \beta') \tag{1.3.3}$$

ensures that it dominates the posterior. Then, z_0^* can be chosen in order to satisfy conditions C1 and C2 in addition to the condition given by equation (1.3.2), which gives

$$z_0^* = \operatorname{argmin}_{z_0' \geq \max[\frac{m_2 - m_1}{2}, \mathbb{I}_{[\beta' \neq 0]} + \epsilon_2]} \left(\alpha' + \frac{\beta'}{\log z_0'} + \delta'\gamma'(z_0')^{\gamma'} \geq 0 \right), \tag{1.3.4}$$

where $\epsilon_2 > 0$. Recall that condition C2 is always satisfied if z_0' is large enough.

Note that ϵ_1 and ϵ_2 must be specified. In practice, it seems appropriate to choose $\epsilon_1 = \min\left(0.01, \frac{1}{2}\sum_{i=0}^{n} \delta_i \mathbb{I}_{[\gamma_i=\gamma']}\right)$ and $\epsilon_2 = 0.01$. A larger value of ϵ_1 would give an importance function with much heavier tails, which is not necessary.

1.3.4. Simulation of observations from the GEP density

The third and last desired criterion for the choice of the parameters of the GEP density consists in being able to simulate observations from the importance function. The GEP density, given by $p(z|\gamma, \delta, \alpha, \beta, z_0)$, is symmetric with respect to the origin and is uniform between $-z_0$ and z_0. Hence, if the mass of the uniform part is denoted by q_0 and given by

$$q_0 = \Pr[-z_0 < Z \leq z_0] = K(\gamma, \delta, \alpha, \beta, z_0) 2 \, e^{-\delta z_0^{\gamma}} \, z_0^{1-\alpha} \log^{-\beta} z_0,$$

TAB. 1.3. Simulation of an observation z from a GEP density on (z_0, ∞), (note : $w \sim U[0,1]$)

Type	**z**
II ($\alpha < 1$)	$\left[\frac{1}{\delta} F^{-1}_{\left(\frac{1-\alpha}{\gamma}\right)} \left(w + (1-w) F_{\left(\frac{1-\alpha}{\gamma}\right)}(\delta z_0^{\gamma})\right)\right]^{1/\gamma}$
III ($\beta < 1$)	$\exp\left\{\frac{F^{-1}_{(1-\beta)}\left(w + (1-w) F_{(1-\beta)}((\alpha-1)\log z_0)\right)}{\alpha-1}\right\}$
IV	$\frac{z_0}{w^{\frac{1}{\alpha-1}}}$
V	$\exp\left\{\frac{\log z_0}{w^{\frac{1}{\beta-1}}}\right\}$

where $K(\gamma, \delta, \alpha, \beta, z_0)$ is the normalizing constant given in Table 1.2, then an observation must be simulated from $(-\infty, -z_0]$ with probability $\frac{1-q_0}{2}$, from a uniform $[-z_0, z_0]$ with probability q_0 and from (z_0, ∞) with probability $\frac{1-q_0}{2}$.

An observation z is generated from (z_0, ∞) with the inverse transformation method, depending on the type of the GEP density as shown in Table 1.3. Note that $F_{(\lambda)}(\cdot)$ is the cdf of a gamma distribution with shape and scale parameters respectively equal to $\lambda > 0$ and 1, and $F^{-1}_{(\lambda)}(\cdot)$ is its inverse cdf. An observation z from $(-\infty, -z_0]$ is generated in the same way, except for a change of sign.

There are three cases for which direct simulation with the inverse transformation method is not possible, that is the GEP density of type I, type II with $\alpha \geq 1$ and type III with $\beta \geq 1$.

However, it is possible to simulate observations with the rejection method (see Ross, 1997). This algorithm generates a value from a proposed distribution, which is accepted or rejected according to a probability based on the ratio of the density of interest and the proposed density. For more details, see Desgagné and Angers (2003).

A proposal for each one of these three cases is suggested in Table 1.4. The densities have been chosen for their balance between simplicity and efficiency. They are GEP densities, like the densities of interest, and differ from them only by one or two parameters. Direct simulation of observations from the proposed densities is done using Table 1.3.

TAB. 1.4. Proposals when direct simulation from $p(z|\gamma,\delta,\alpha,\beta,z_0)$ is not possible

Type	Proposal :	a GEP density of
I or II ($\alpha \geq 1$)	type II ($\alpha < 1$) :	$p(z\vert\gamma,\delta,\alpha^{**}_{(\gamma,\delta,\alpha,\beta,z_0)},0,z_0)$
III ($\beta \geq 1$)	type III ($\beta < 1$) :	$p(z\vert 0,0,\alpha,1-\epsilon_3,z_0)$

For the case of the GEP density of type III when $\beta \geq 1$, β is simply replaced by $1-\epsilon_3$, where $\epsilon_3 > 0$. For type II when $\alpha \geq 1$ and for type I, β is set to 0 and α is replaced by $\alpha^{**}_{(\gamma,\delta,\alpha,\beta,z_0)}$. The objective was to choose a density as close as possible to the density of interest, but with heavier tails. A criterion which respects this objective consists in choosing

$$\alpha^{**}_{(\gamma,\delta,\alpha,\beta,z_0)} = \arg\min_{\alpha_0<1}\left[\sup_z \frac{p(z|\gamma,\delta,\alpha,\beta,z_0)}{p(z|\gamma,\delta,\alpha_0,0,z_0)}\right].$$

This criterion ensures that the probability of acceptance in the rejection method is maximized (see Robert, 1996). Explicitly, we have

$$\alpha^{**}{}_{(\gamma,\delta,\alpha,\beta,z_0)} = \begin{cases} \arg\min_{\alpha_0\in\left[\alpha+\frac{\beta}{\log z_0},\min(1,\alpha)\right)} K^{-1}(\gamma,\delta,\alpha_0,0,z_0)(\alpha-\alpha_0)^{\beta}; & \text{if } \alpha+\frac{\beta}{\log z_0} < \min(1,\alpha), \\ \min(1-\epsilon_4,\alpha); & \text{otherwise}, \end{cases} \tag{1.3.5}$$

where $\epsilon_4 > 0$ and $K(\gamma,\delta,\alpha_0,0,z_0)$ is the normalizing constant for type II given in Table 1.2.

In practice, $\epsilon_3 = 0.01$ and $\epsilon_4 = 0.01$ seems appropriate. Furthermore, the minimization of $K^{-1}(\gamma,\delta,\alpha_0,0,z_0)(\alpha-\alpha_0)^{\beta}$ with respect to α_0 has to be done numerically. However it can be shown that this function is strictly convex.

Note that the proposed density can also be used as an importance function in Monte Carlo simulations with importance sampling to evaluate the normalizing constant or the moments of the GEP density of type I, where no analytic formulae are available.

1.3.5. Selection of parameters of the importance function

As mentioned in Section 1.3.1, the proposed importance function is the GEP density given by $p(\theta - \mu^*|\gamma^*, \delta^*, \alpha^*, \beta^*, z_0^*)$, as defined by equation (1.2.1). The parameters are determined by equations (1.3.1), (1.3.3) and (1.3.4).

If the importance function is a GEP density for which direct simulation is not possible, there are two methods to handle it. Firstly, observations can be simulated with the rejection method, as seen in Section 1.3.4. Secondly, the importance function can be replaced by the appropriate density as given in Table 1.4. In both cases observations are generated from the same density, but all the observations are kept in the second method while some are rejected in the first one. It is simpler and more effective to use the second method. More explicitly, the modifications of the importance function are as follows :

i) if $\gamma^* > 0, \delta^* > 0, \beta^* \neq 0$ or $\gamma^* > 0, \delta^* > 0, \alpha^* \geq 1$,
then replace α^* by $\alpha^{**}_{(\gamma^*,\delta^*,\alpha^*,\beta^*,z_0^*)}$ and β^* by 0,

ii) if $\gamma^* = 0, \delta^* = 0, \alpha^* > 1, \beta^* \geq 1$, then replace β^* by $1 - \epsilon_3$,

where $\alpha^{**}_{(\gamma^*,\delta^*,\alpha^*,\beta^*,z_0^*)}$ is given by equation (1.3.5) and $\epsilon_3 > 0$. Note that it can be shown that this change affects neither the unimodality of the importance function, nor its dominance on the posterior.

1.4. Example

1.4.1. Setting

Suppose that a portfolio manager needs a prediction on the return of the S&P 500 index for the next day. He asks five experts for their prediction on the return as well as a 95% confidence interval on this prediction. The manager wants to combine this information with his prior beliefs using the Bayesian model described in Section 1.3.1. According to this setting, the manager chooses

$$f_i(x_i - \theta) = \frac{1}{\sigma_i} T_5 \left(\frac{x_i - \theta}{\sigma_i} \right)$$

for $i = 0, 1, ..., 5$, where $T_5(\cdot)$ is a Student density with 5 degrees of freedom, which means that $f_i(\cdot)$ is a Student density with a scale parameter σ_i. If the standard deviation of f_i is denoted by s_i, then $\sigma_i = \sqrt{0.6}\, s_i$. The Student density is chosen to ensure a robust inference (see Angers, 2000). With standard deviations equal, the opinion of each source of information has the same weight.

1.4.2. Data

The collected information for the predicted return is $\mathbf{x} = (0, -0.6, 0.3, 0.5, 0.7, 1.0)$. Note that all numbers in this example are expressed in percentages. The standard deviations of the predictions, extracted from the confidence intervals given by the managers and the prior, are vectorized as $\mathbf{s} = (1, 0.5, 1, 0.25, 0.5, 0.5)$. For example, the prior beliefs on the predicted return consist in a mean of 0 and a standard deviation of 1. Note that the vector of the scale parameters of f_i is then given by $\sigma = (0.775, 0.387, 0.775, 0.194, 0.387, 0.387)$.

The moments of the posterior distribution of θ are estimated using Monte Carlo simulations with importance sampling. Three importance functions, as described below, are compared.

1.4.3. The importance functions

The first importance function is the GEP density given by $p(\theta - \mu^*|\gamma^*, \delta^*, \alpha^*, \beta^*, z_0^*)$. Its parameters are chosen according to formulae given in Sections 1.3.2 to 1.3.5. It is easy to show that the $(100p)^{th}$ and $(100(1-p))^{th}$ percentiles of f_i are evaluated respectively as $q_{p,i} = -2.015\sigma_i$ and $q_{1-p,i} = 2.015\sigma_i$ if p is set to 0.05. Then it is possible to evaluate $m_1 = \min_{i=0,\ldots,n}[x_i + q_{p,i}] = -1.561$, $m_2 = \max_{i=0,\ldots,n}[x_i + q_{1-p,i}] = 1.861$ and the location parameter $\mu^* = \frac{m_1+m_2}{2} = 0.15$.

It can be shown that p-credence of each source of information is $(0, 0, 6, 0)$. It is then easy to show that $\gamma^* = \delta^* = 0$, $\alpha^* = 36$ and $\beta^* = 0$. Finally it can be verified that $z_0^* = \frac{m_2-m_1}{2} = 1.711$ satisfies equation (1.3.4). The density of the importance function is then given by $p(\theta - 0.15|0, 0, 36, 0, 1.711)$. This is a GEP density of type IV (the right tail being a Pareto density) for which direct simulation is possible using the method described in Section 1.3.4.

TAB. 1.5. Standard error and 95% confidence interval for the estimate $\widehat{\mathbb{E}}(\theta|\mathbf{x})$ after 10,000 simulations, for the first example

Importance function	Standard error	95% C.I.
$p(\theta - 0.15 \mid 0, 0, 36, 0, 1.711)$	0.003	0.481 to 0.493
T(df=35, μ=0.423, σ=0.177)	0.002	0.483 to 0.491
T(df=5, μ=0.423, σ=0.141)	0.002	0.483 to 0.491

The two other importance functions are, respectively, a Student distribution with 35 and 5 degrees of freedom, both centered at 0.423 with a standard deviation of 0.183 (that is a scale parameter of respectively 0.177 and 0.141). The degrees of freedom are chosen to match p-credence of the importance function with respectively that of the posterior and the prior density. The mean and standard deviation are chosen to match those of the posterior if

$$f_i(x_i - \theta) = \frac{1}{\sigma_i} N\left(\frac{x_i - \theta}{\sigma_i}\right),$$

where $N(\cdot)$ is the standard normal density. In this case, it can be shown that

$$\mathbb{E}(\theta|\mathbf{x}) = \frac{\sum_{i=0}^{n}(x_i/s_i^2)}{\sum_{i=0}^{n}(1/s_i^2)} \text{ and } \operatorname{Var}(\theta|\mathbf{x}) = \frac{1}{\sum_{i=0}^{n}(1/s_i^2)}.$$

1.4.4. Results

The posterior mean and standard deviation are respectively evaluated to $\mathbb{E}(\theta|\mathbf{x}) = 0.487$ and $\sqrt{\operatorname{Var}(\theta|\mathbf{x})} = 0.171$. The prediction on the return of the S&P 500 index for the next day is then estimated to 0.487% with an approximative 95% confidence interval of (0.145%, 0.829%). The standard error and the approximative 95% confidence interval for $\widehat{\mathbb{E}}(\theta|\mathbf{x})$ estimated with 10,000 Monte Carlo simulations are given for each importance function in Table 1.5. This is a case where no apparent conflict exists between the sources of information, which explains that the results are similar, the two concurrent importance functions having a slightly better precision. Next, we consider the case by permuting the second and fourth elements ($s_1 = 0.25$ and $s_3 = 0.5$) of the vector of the standard deviations of the predictions. Then a conflict between the prediction of the

TAB. 1.6. Standard error and 95% confidence interval for the estimate $\widehat{\mathbb{E}}(\theta|\mathbf{x})$ after 10,000 simulations, for the second example

Importance function	Standard error	95% C.I.
$p(\theta - 0.15\|0, 0, 36, 0, 1.711)$	0.004	0.336 to 0.352
T(df=35, μ=-0.017, σ=0.177)	0.050	0.244 to 0.444
T(df=5, μ=-0.017, σ=0.141)	0.013	0.318 to 0.370

first expert (-0.6) and the other sources of information occurs. The importance functions remain the same ones, except for the Student densities which are now centered at -0.017 instead of 0.423.

The posterior mean is then evaluated to $\mathbb{E}(\theta|\mathbf{x}) = 0.344$ and the posterior standard deviation to $\sqrt{\mathrm{Var}(\theta|\mathbf{x})} = 0.283$. The standard error and the approximative 95% confidence interval for $\widehat{\mathbb{E}}(\theta|\mathbf{x})$ estimated with 10,000 Monte Carlo simulations are given for each importance functions in Table 1.6. (Based on normality assumption, the interval corresponds to the mean plus or minus two standard errors.) This is a case where a conflict exists between the sources of information. The proposed method is not affected by the conflicting information when the GEP density is the importance function, the standard error of the estimate being similar to that of the first example. However the precision of the estimate is seriously affected with the two concurrent importance functions.

The posterior density as well as the three importance functions are shown in Figures 1(a) and 1(b) for the two examples. The location of the Student densities in the second example underestimates the location of the posterior, which is explained by the influence of the conflicting information of the first expert (-0.6). The densities $f_i(x_i - \theta) = T_5((x_i - \theta)/\sigma_i)$ are shown for $i = 0, ..., 5$ in Figures 1(c) and 1(d) for the two examples.

In Figures 2(a) and 2(b), the weights (divided by their maximum) are plotted over their corresponding variates generated in Monte Carlo simulations, for each importance function and for the two examples. The weights are expected to be larger and more concentrated in the main area of the posterior. Also the weights

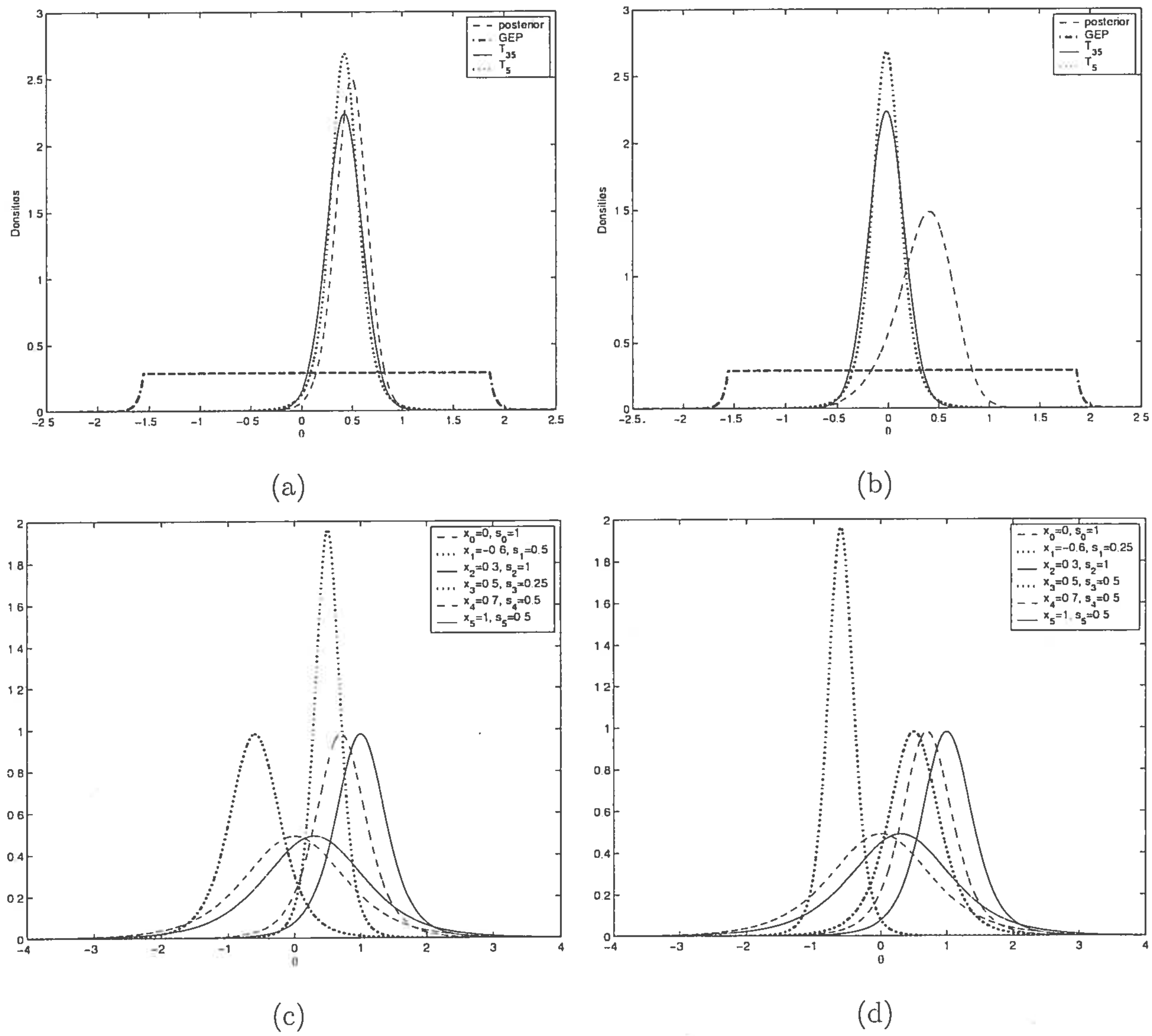
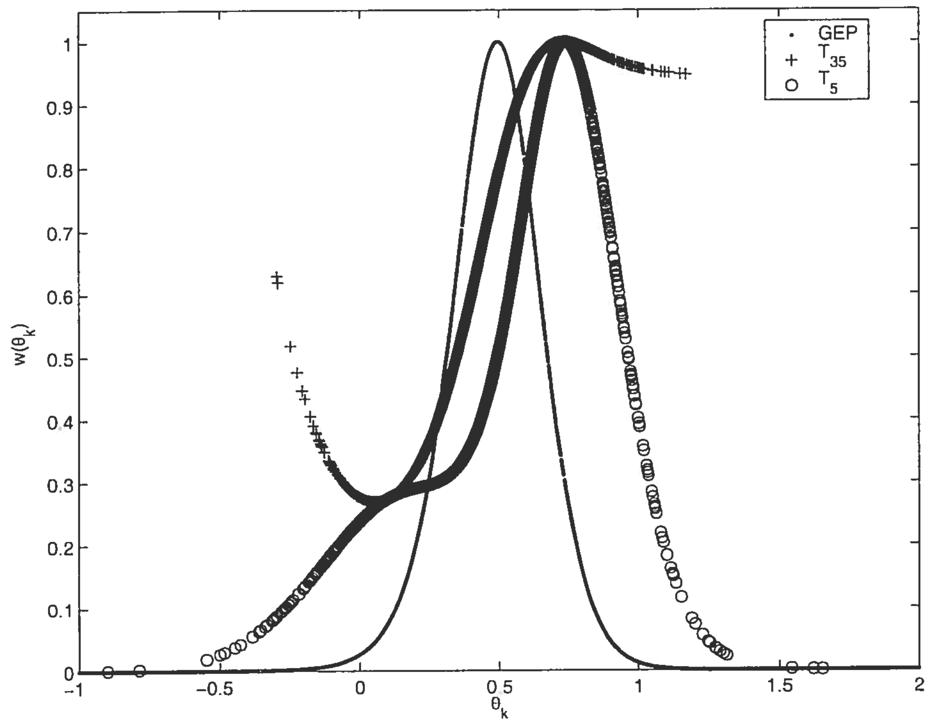
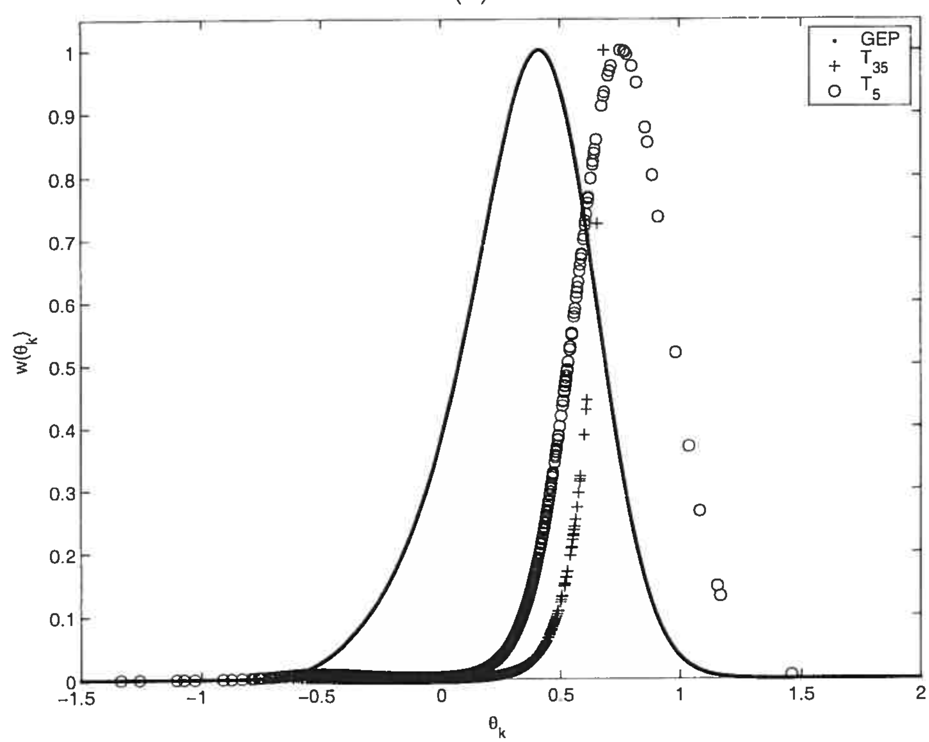


FIG. 1.1. The posterior and the importance functions (a) for the first example, (b) for the second example; the densities $f_i(x_i - \theta)$ for $i = 0, ..., 5$ (c) for the first example, (d) for the second example.

should decrease towards 0 when the observations move away from the posterior. These features are satisfied when the importance function is the GEP density, and also for the Student density with 5 degrees of freedom except for the mass of the weights located on the right of the posterior in the second example. However, the Student density with 35 degrees of freedom fails to incorporate these even for the first example.



(a)



(b)

FIG. 1.2. The weights (divided by their maximum) over their corresponding observation for each importance function (a) for the first example (b) for the second example.

1.5. Conclusion

The generalized exponential power density has been proposed as an importance function in Monte Carlo simulations in the context of the estimation of posterior moments of a location parameter. It can be difficult to choose an appropriate importance function and it must often be done for each case. If p-credences of the prior and the likelihood are defined, the parameters of the GEP density are obtained by the equations given in this paper, in an automatic way whatever the model and the data are. Note that p-credence is defined for most of the usual symmetric distributions defined on the real line with an exponential, polynomial or logarithmic behaviour in their tails.

The choice of the GEP density allows us to obtain reliable results, even if there are conflicting sources of information. Furthermore, since p-credence of the GEP density is slightly lower than that of the posterior, the Monte Carlo simulations remain effective as illustrated in Section 1.4. The simulation of observations from the GEP density has been addressed with the inverse transformation method.

1.6. Acknowledgments

The authors are thankful to the NSERC (Natural Sciences and Engineering Research Council of Canada), the FQRNT (Le Fonds québécois de la recherche sur la nature et les technologies) and the SOA (Society of Actuaries) for their financial support. They also would like to thank the referees, Prof. Yogendra Chaubey and James Merleau for their useful comments which have brought substantial improvements on a previous draft.

1.7. References

ANGERS, J.-F. (1992) *Use of Student-t prior for the estimation of normal means : A computational approach, 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, *IMS Lecture Notes - Monograph Series*, Vol. **29**, pp. 273-283.

ANGERS, J.-F. (2000) P-Credence and outliers, *Metron*, **58**, 81-108.

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.

DESGAGNÉ, A. and ANGERS, J.-F. (2003) *Computational aspect of the generalized exponential power density*, Technical report CRM-2918, Université de Montréal (http ://www.crm.umontreal.ca/pub/Rapports/2900-2999/2918.pdf).

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.

JOHNSON, N.L., KOTZ, S. AND BALAKRISHNAN, N. (1994) *Continuous univariate distributions*, Vol. 1 (second edition), Wiley, New York.

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

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

REISS, R.-D. and THOMAS, M. (1997) *Statistical analysis of extreme values*, Birkhauser Verlag, Basel.

ROBERT, C. (1996) *Méthodes de Monte Carlo par chaînes de Markov*, Economica, Paris.

ROSS, S.M. (1997) *Simulation* (second edition), Academic Press, San Diego.

Chapitre 2

OUTLIERS AND CHOICE OF THE PRIOR FOR LOCATION PARAMETER INFERENCE

Cet article a été soumis pour publication en février 2005 dans la revue *Metron*. Le premier auteur est Alain Desgagné et le coauteur est le directeur de recherche Jean-François Angers. La contribution de Alain Desgagné à cet article consiste en la conception, recherche, développement, programmation informatique et rédaction de toutes les parties de l'article, sous la supervision du directeur de recherche.

Abstract

The use of heavy-tailed distributions is a valuable tool in developing robust Bayesian procedures, limiting the influence of outliers on posterior inference. In this paper, the behavior of the posterior density of the location parameter is investigated when the sample contains outliers. The notion of left and right p-credence is introduced to characterize respectively the left and right tail of a density. Simple conditions on the tails of the prior and the likelihood, using left and right p-credence, are established to determine the proportion of observations that can be rejected as outliers. It is shown that the posterior distribution converges in law to the posterior that would be obtained from the reduced sample, excluding the outliers, as they tend to plus or minus infinity, at any given rate. An example of combination of predictions of the S&P 500 index return is presented.

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

2.1. Introduction

The use of heavy-tailed distributions is a valuable tool in developing robust Bayesian procedures, limiting the influence of outliers on posterior inference. 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. Theorical results were given by Dawid (1973) and 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.

Outliers rejection based on the notion of credence was first introduced by O'Hagan (1990). Credence is a measure of tails for symmetric densities with Student-type tails. This notion was generalized to p-credence by Angers (2000) to accommodate a wider class of densities. Other authors approached outliers rejection, see for instance Meinhold and Singpurwalla (1989), Angers and Berger (1991), Carlin and Polson (1991), Angers (1992), Fan and Berger (1992), Geweke (1994) and Angers (1996).

In Section 2.2, the relation between outliers rejection and p-credence is considered. In Section 2.2.1, the notion of p-credence is generalized to left and right p-credences in order to characterize each tail of a density distinctly. These measures are defined for a large class of densities with exponential, polynomial and logarithmic tails behavior, which now makes possible to order the right tail of most of the known densities defined on $\mathbb{R}^+$ using right p-credence or the densities defined on $\mathbb{R}$ using both left and right p-credences.

In Section 2.2.2, the behavior of the posterior density of the location parameter is investigated when the sample contains outliers. Simple conditions on the tails of the prior and the likelihood, using left and right p-credences, are established to determine the proportion of observations that can be rejected as outliers. We show that the posterior distribution converges in law to the posterior that would be obtained from the reduced sample, excluding the outliers, as they tend to $+\infty$ or $-\infty$, at any given rate. In Section 2.2.3, a special case with one observation is given.

In Section 2.3, the same convergence results are given when we specify which one of the observations are outliers and if they are positive or negative outliers. The conditions on the tails of the prior and the likelihood are also generalized to encompass densities excluded by left and right p-credences. The influence of the left and right tails of each observation's density and the prior is established clearly in these conditions. This also provides a good framework to generalize these results to the case where the scale parameter is unknown.

In Section 2.4, an example of combination of predictions of the S&P 500 index return is given. While the results given in this paper have a strong theorical component, it is shown in this example that they also have useful and easy applications in real context.

2.2. Outliers rejection using p-credence

In this section, conditions on prior and likelihood are established using p-credence to obtain robust Bayesian inference on the location parameter. The influence of the outliers on the posterior density is expected to decrease when the outliers become extreme.

In Section 2.2.1, p-credence is defined to characterize the tails of a density. In Section 2.2.2, conditions are presented to obtain convergence of the posterior density based on all observations to the posterior excluding the outliers, as the absolute values of the outliers tend to ∞. In Section 2.2.3, a special case with one observation is presented.

2.2.1. A measure of the tails : left and right p-credences

Conditions of robustness concern mainly the tails of the prior and observations' densities. In consequence, measures for the tails of a density, called left and right p-credences, are introduced. These measures are analog to the p-credence (see Angers, 2000). The p-credence is a measure for densities with the same tails behavior (symmetric densities for example) and was defined as follows.

Definition 3. *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 $0 < k \leq K < \infty$ such that for all*

$z \in \mathbb{R}$

$$k \leq \frac{f(z)}{p(z|\gamma,\delta,\alpha,\beta,z_0)} \leq K,$$

where $p(z|\gamma,\delta,\alpha,\beta,z_0) \propto \mathrm{e}^{-\delta\tilde{z}^{\gamma}}\, \tilde{z}^{-\alpha} \log^{-\beta} \tilde{z}$, that is the generalized exponential power density (GEP), up to a normalizing constant, where $\tilde{z} = \max(|z|, z_0)$ (see Desgagné and Angers, 2005).

Other measures, called left and right p-credences, are proposed in this paper in order to characterize each tail distinctly. If $f(z)$ is the density of a random variable Z, then right p-credence is denoted by p-cred$^+(f)$ or p-cred$^+(Z)$ and is defined as follows.

Definition 4. *A density f has right p-credence $(\gamma,\delta,\alpha,\beta)$ if there exists a constant $K > 0$ such that*

$$\lim_{z\to\infty} \frac{f(z)}{\mathrm{e}^{-\delta|z|^{\gamma}}\, |z|^{-\alpha} \log^{-\beta}|z|} = K.$$

Since the GEP density is symmetric, the definition of left p-credence is identical, except $\lim_{z\to\infty}$ is replaced by $\lim_{z\to-\infty}$. It is denoted by p-cred$^-(f)$ or p-cred$^-(Z)$. Note that p-cred$^-(Z)$ = p-cred$^+(-Z)$.

The definition of left and right p-credences concerns only the tails of a density, while the definition of p-credence concerns the entire real line. In the definition of left and right p-credences, a tail of the density f is proportional to the corresponding tail of the GEP density with parameters $(\gamma,\delta,\alpha,\beta)$, which ensures a certain smoothness in the tail. The domain of the parameters of the left and right p-credences is $\gamma \geq 0, \delta \geq 0$ (by convention $\delta = 0$ if $\gamma = 0$), $\alpha \in \mathbb{R}$ and $\beta \in \mathbb{R}$. Left and right p-credences are defined for most of the known densities on $\mathbb{R}$ and right p-credence is defined for most of the known densities on $\mathbb{R}^+$ (see Desgagné and Angers, 2003).

Once right (or left) p-credence of two densities have been determined, a dominance relation can be established to compare and order their tails, as described in Proposition 2.

Proposition 2. *Let f and g be two densities such that*
p-cred$^+(f) = (\gamma,\delta,\alpha,\beta)$ *and* p-cred$^+(g) = (\gamma',\delta',\alpha',\beta')$.

i) *If* $\gamma' = \gamma, \delta' = \delta, \alpha' = \alpha, \beta' = \beta$, *then the right p-credences of* f *and* g *are equal, which is denoted by* $(\gamma', \delta', \alpha', \beta') = (\gamma, \delta, \alpha, \beta)$. *Their right tails are equivalent, which means that* $\lim_{z \to \infty} \frac{f(z)}{g(z)} = k$ *for a positive constant* k.

ii) *The right p-credence of* g *is smaller than that of* f, *which is denoted by* $(\gamma', \delta', \alpha', \beta') < (\gamma, \delta, \alpha, \beta)$, *if*

a) $\gamma' < \gamma$,

b) $\gamma' = \gamma, \delta' < \delta$,

c) $\gamma' = \gamma, \delta' = \delta, \alpha' < \alpha$,

d) or $\gamma' = \gamma, \delta' = \delta, \alpha' = \alpha, \beta' < \beta$.

The right tail of g *strictly dominates the right tail of* f, *which means that* $\lim_{z \to \infty} \frac{f(z)}{g(z)} = 0$.

The proof of Proposition 2 is given in Angers (2000). The left tails of two densities are compared and ordered in a similar way using left p-credence. The left tail of the density with the smallest left p-credence dominates the left tail of the other density.

2.2.2. Outliers rejection using left and right p-credences

In this section, the behavior of the posterior density and the posterior moments of the location parameter are investigated when the sample contains outliers, that is when there is a conflict between some extreme observations and the information provided by the prior and the other observations. Using left and right p-credences, conditions on the tails of the prior density and the likelihood are established in order to obtain robust posterior inference. The influence of the outliers decreases as they become more extreme and eventually the outliers are rejected. Consider the following Bayesian context.

i) Let $X_1, ..., X_n$ be n random variables conditionally independent given θ with the conditional densities of $X_i|\theta$ given by $f_i(x_i - \theta)$, where $X_i \in \mathbb{R}, \theta \in \mathbb{R}$, $i = 1, ..., n$.

ii) The prior density of θ is $\pi_\theta(\theta - x_0)$, where $x_0 \in \mathbb{R}$ is a known location parameter.

The densities $\pi_\theta, f_1, ..., f_n$ are assumed to be proper, positive everywhere and bounded above. We assume that the sample consists of a block of k observations and of $n-k$ outliers, $0 \leq k \leq n$. Assume without loss of generality that the outliers are the observations denoted by $x_{k+1}, ..., x_n$. The differences $x_1 - x_0, ..., x_k - x_0$ are assumed fixed. The distances between each outlier $x_{k+1}, ..., x_n$ and the block composed of $x_0, x_1, ..., x_k$ tend to ∞, i.e. $|x_{k+1} - x_0| \to \infty, ..., |x_n - x_0| \to \infty$. Generally, we consider that $x_0, x_1, ..., x_k$ are fixed and that the outliers tend to plus or minus infinity. However, we could consider that $x_0, x_1, ..., x_k$ are not fixed, as long as they are moving as a block and as long as their distances with the outliers tend to infinity.

Let the posterior density of θ be denoted by $\pi(\theta|\underset{\sim}{x}_n)$ if all n observations are considered, and denoted by $\pi(\theta|\underset{\sim}{x}_k)$ if only non-outlier observations denoted by $x_1, ..., x_k$ are considered. Let also the marginal density of $\underset{\sim}{x}_n = (x_1, ..., x_n)$ be denoted by $m(\underset{\sim}{x}_n)$ and the marginal density of $\underset{\sim}{x}_k = (x_1, ..., x_k)$ be denoted by $m(\underset{\sim}{x}_k)$. Finally, let the vector of the distances between the outliers and x_0 be denoted by $\underset{\sim}{\phi}_1 = (|x_{k+1} - x_0|, ..., |x_n - x_0|)$. The notation $\underset{\sim}{\phi}_1 \to \infty$ means that each term of the vector tends to ∞ at any given rate.

Theorem 1. *Suppose that* p-cred$^-(f_i)$ = p-cred$^+(f_i) = (\gamma', \delta', \alpha', \beta')$, $i = 1, ..., n$ *and* p-cred$^-(\pi_\theta)$ = p-cred$^+(\pi_\theta) = (\gamma, \delta, \alpha, \beta)$. *For any integer k such that $0 \leq k \leq n$ and for any $x_0, x_1, ..., x_k$ such that $x_1 - x_0, ..., x_k - x_0$ are fixed, if*

i) $\gamma' < 1$, $k \geq n/2$ *or*

ii) $\gamma' < 1$, $k < n/2$, $(\gamma, \delta, \alpha, \beta) > (\gamma', \delta'(n-2k), \alpha'(n-2k), \beta'(n-2k))$

then

a) $\lim_{\underset{\sim}{\phi}_1 \to \infty} \frac{\pi(x_0|\underset{\sim}{x}_n)}{\pi(x_0|\underset{\sim}{x}_k)} = 1$,

b) for any $h > 0$ and for all θ such that $|\theta - x_0| \leq h$, $\lim_{\underset{\sim}{\phi}_1 \to \infty} \frac{\pi(\theta|\underset{\sim}{x}_n)}{\pi(\theta|\underset{\sim}{x}_k)} = 1$,

c) for any $h > 0$ and $j \in (k+1, ..., n)$, $\lim_{\underset{\sim}{\phi}_1 \to \infty} \Pr[|\theta - x_j| \leq h|\underset{\sim}{x}_n] = 0$,

d) $(\theta - x_0)|\underset{\sim}{x}_n \xrightarrow{\mathcal{L}} (\theta - x_0)|\underset{\sim}{x}_k$ *as $\underset{\sim}{\phi}_1 \to \infty$, where the density of the random variables $(\theta - x_0)|\underset{\sim}{x}_n$ and $(\theta - x_0)|\underset{\sim}{x}_k$ evaluated at the point y are given by* $\pi(y + x_0|\underset{\sim}{x}_n)$ *and* $\pi(y + x_0|\underset{\sim}{x}_k)$.

In addition, for any positive integer p, *if* $\mathbb{E}^{\pi_\theta(\theta)}[|\theta|^p] < \infty$ *and* $(\gamma, \delta, \alpha, \beta) > (\gamma', \delta'(n-2k), \alpha'(n-2k)+p, \beta'(n-2k))$ *when* $k < n/2$, *then*

e) $\lim_{\phi_1 \to \infty} \mathbb{E}^{\pi(\theta|\underline{x}_n)}[(\theta - x_0)^p] = \mathbb{E}^{\pi(\theta|\underline{x}_k)}[(\theta - x_0)^p]$.

Proof. See the Appendix, Section 2.6.8.

Note that for each density $\pi_\theta, f_1, ..., f_n$, the left and right tails have the same behavior. This condition is needed to obtain the same robustness, whether the outliers are on the left or on the right of x_0. Furthermore, p-credences of $f_1, ..., f_n$ are assumed to be identical. This condition ensures the same robustness against any extreme observations among $x_1, x_2, ..., x_n$.

The condition $\gamma' < 1$ ensures that left and right p-credences of $f_1, ..., f_n$ are sufficiently small, or equivalently that the density's tails of each potential outlier are sufficiently heavy. This also ensures that the posterior can reject up to $[n/2]$ outliers, where $[a]$ stands for the integer part of a. The conditions tell us that the posterior can reject more than $[n/2]$ outliers if the left and right p-credences of the prior are sufficiently large relatively to those of the likelihood. If one is not interested into putting too much confidence on the prior, then choosing $(\gamma, \delta, \alpha, \beta) \leq (\gamma', \delta', \alpha', \beta')$ ensures that the posterior can reject up to $[n/2]$ outliers, but not necessarily $[n/2] + 1$, if for example all the outliers are on the right of x_0. At the other extreme, if one wants to put a large confidence in the prior, then choosing $(\gamma, \delta, \alpha, \beta) > (\gamma', n\delta', n\alpha', n\beta')$ ensures rejection of up to n outliers, even for the extreme cases where all the outliers are on the right of x_0 or all are on the left of x_0. An intermediate choice of $(\gamma, \delta, \alpha, \beta)$ such that $(\gamma', \delta', \alpha', \beta') < (\gamma, \delta, \alpha, \beta) \leq (\gamma', n\delta', n\alpha', n\beta')$ gives intermediate results. Note that for specific directions of the outliers, it could be possible that the number of outliers rejected by the posterior is larger than the number suggested by Theorem 1. This case is addressed in Section 2.3 with Theorem 2.

Asymptotic behavior of the posterior density is established through results a) to d), as $|x_i - x_0| \to \infty$ at any given rate, for $i = k+1, ..., n$. Note that if x_0 is fixed, the limit can be rewritten as $|x_i| \to \infty$, $i = k+1, ..., n$.

Result a) says that the influence of the outliers on the posterior density evaluated at x_0 is asymptotically null. Note that result a) can be rewritten as

$\lim_{\phi_1 \to \infty} \frac{m(\underset{\sim}{x}_k) \prod_{i=k+1}^{n} f_i(x_i - x_0)}{m(\underset{\sim}{x}_n)} = 1$ since $\pi(x_0 | \underset{\sim}{x}_n) = m^{-1}(\underset{\sim}{x}_n) \pi_\theta(0) \prod_{i=1}^{n} f_i(x_i - x_0)$ and $\pi(x_0 | \underset{\sim}{x}_k) = m^{-1}(\underset{\sim}{x}_k) \pi_\theta(0) \prod_{i=1}^{k} f_i(x_i - x_0)$. It says that the asymptotic behavior of the marginal $m(\underset{\sim}{x}_n)$ and $\prod_{i=k+1}^{n} f_i(x_i - x_0)$ is equivalent, since $m(\underset{\sim}{x}_k)$ do not depend on $\underset{\sim}{\phi}_1$. Result b) says that the influence of the outliers on the posterior density in any finite neighborhood of x_0 (or $x_1, ..., x_k$) is asymptotically null. Result c) says that the posterior density tends to 0 in any finite neighborhood of the outliers. Finally, the convergence in distribution of the random variable $(\theta - x_0) | \underset{\sim}{x}_n$ to the random variable $(\theta - x_0) | \underset{\sim}{x}_k$ is established in result d), as $\underset{\sim}{\phi}_1 \to \infty$, where the densities of $(\theta - x_0) | \underset{\sim}{x}_n$ and $(\theta - x_0) | \underset{\sim}{x}_k$ evaluated at the point y are respectively given by $\pi(y + x_0 | \underset{\sim}{x}_n)$ and $\pi(y + x_0 | \underset{\sim}{x}_k)$. An equivalent result is given by $\lim_{\underset{\sim}{\phi}_1 \to \infty} \Pr[\theta - x_0 \leq d | \underset{\sim}{x}_n] = \Pr[\theta - x_0 \leq d | \underset{\sim}{x}_k]$, for any $d \in \mathbb{R}$ (see the Appendix, Section 2.6.5). Note that $\pi(y + x_0 | \underset{\sim}{x}_k)$ depends only on the fixed differences $\underset{\sim}{x}_k - x_0$, while $\pi(y + x_0 | \underset{\sim}{x}_n)$ depends also on $x_{k+1} - x_0, ..., x_n - x_0$.

If the distance between a given observation x_j $(j > k)$ and the center of $x_0, x_1, ..., x_k$ increases but remains smaller than a certain threshold, the influence of this observation on the posterior density usually increases. However, if this distance increases beyond the threshold, the influence of the observation begins to decrease to eventually be null.

Result e) says that the influence of the outliers on the p^{th} posterior moment centered at x_0 is asymptotically null, as long as the p^{th} absolute prior moment exists and as long as the number of outliers is less than a maximum number specified by left and right p-credences of the prior and likehood. The result is also true for the p^{th} moment centered at any value with a fixed distance from x_0. For example, if x_0 is fixed, result e) can be rewritten as $\lim_{|x_{k+1}|, ..., |x_n| \to \infty} \mathbb{E}^{\pi(\theta | \underset{\sim}{x}_n)}[\theta^p] = \mathbb{E}^{\pi(\theta | \underset{\sim}{x}_k)}[\theta^p]$.

2.2.3. Conflicting information with one observation

An interesting special case of Theorem 1 is given in Corollary 1, when the information is provided only by the prior and one observation. The prior centered at x_0 and the likelihood (when considered as a function of θ) centered at x_1 behave in a symmetric way. The posterior density converges in distribution to the source of information (prior or likelihood) with the largest left and right p-credences, if the first parameter (γ or γ') of the smallest left and right p-credences is less than one. The case where the posterior converge to the likelihood is given in Corollary 2. Using the symmetry between the prior and the likelihood, Corollary 2 follows from Corollary 1 by interchanging $\pi_\theta(a)$ with $f_1(-a)$ for any $a \in \mathbb{R}$ and x_0 with x_1.

Corollary 1. *If $\gamma' < 1$ and $(\gamma, \delta, \alpha, \beta) > (\gamma', \delta', \alpha', \beta')$ then*

a) $\lim_{|x_1-x_0|\to\infty} \frac{\pi(x_0|x_1)}{\pi_\theta(0)} = \lim_{|x_1-x_0|\to\infty} \frac{f_1(x_1-x_0)}{m(x_1)} = 1$,

b) for any $h > 0$ and θ such that $|\theta - x_0| \leq h$, $\lim_{|x_1-x_0|\to\infty} \frac{\pi(\theta|x_1)}{\pi_\theta(\theta-x_0)} = 1$,

c) for any $h > 0$, $\lim_{|x_1-x_0|\to\infty} \Pr[|\theta - x_1| \leq h|x_1] = 0$,

d) $(\theta - x_0)|x_1 \xrightarrow{\mathcal{L}} (\theta - x_0)$ *as $|x_1 - x_0| \to \infty$, where the density of the random variables $(\theta - x_0)|x_1$ and $(\theta - x_0)$ evaluated at the point y are given by $\pi(y + x_0|x_1)$ and $\pi_\theta(y)$.*

In addition, if for any positive integer p, $\mathbb{E}^{\pi_\theta(\theta)}[|\theta|^p] < \infty$ and $(\gamma, \delta, \alpha, \beta) > (\gamma', \delta', \alpha' + p, \beta')$, then

e) $\lim_{|x_1-x_0|\to\infty} \mathbb{E}^{\pi(\theta|x_1)}[(\theta - x_0)^p] = \mathbb{E}^{\pi_\theta(\theta-x_0)}[(\theta - x_0)^p]$.

Corollary 2. *If $\gamma < 1$ and $(\gamma', \delta', \alpha', \beta') > (\gamma, \delta, \alpha, \beta)$ then*

a) $\lim_{|x_1-x_0|\to\infty} \frac{\pi(x_1|x_1)}{f_1(0)} = \lim_{|x_1-x_0|\to\infty} \frac{\pi_\theta(x_1-x_0)}{m(x_1)} = 1$,

b) for any $h > 0$ and θ such that $|\theta - x_1| \leq h$, $\lim_{|x_1-x_0|\to\infty} \frac{\pi(\theta|x_1)}{f_1(x_1-\theta)} = 1$,

c) for any $h > 0$, $\lim_{|x_1-x_0|\to\infty} \Pr[|\theta - x_0| \leq h|x_1] = 0$,

d) $(\theta - x_1)|x_1 \xrightarrow{\mathcal{L}} (\theta - x_1)$ *as $|x_1 - x_0| \to \infty$, where the density of the random variables $(\theta - x_1)|x_1$ and $(\theta - x_1)$ evaluated at the point y are given by $\pi(y + x_1|x_1)$ and $f_1(y)$.*

In addition, if for any positive integer p, $\mathbb{E}^{f_1(\theta)}[|\theta|^p] < \infty$ and $(\gamma', \delta', \alpha', \beta') > (\gamma, \delta, \alpha + p, \beta)$, then

e) $\lim_{|x_1-x_0|\to\infty} \mathbb{E}^{\pi(\theta|x_1)}[(\theta - x_1)^p] = \mathbb{E}^{f_1(x_1-\theta)}[(\theta - x_1)^p]$.

2.3. Outliers rejection with general conditions

While conditions of Theorem 1 can be satisfied for most of the known symmetric densities using left and right p-credences, it is still possible to relax and generalize these conditions such that they can be satisfied for any distributions. In this section, conditions in Theorem 1 are given without using left and right p-credences. Conditions are also relaxed such that the left and right tails of each density may have their own asymptotic behavior. It is now possible to consider any distributions, such that the Gumbel distribution for instance, where the left p-credence is not defined and its right tail is heavier than its left tail. Conditions are also given such that it is possible to specify which observation is an outlier along with its direction. The new theorem with relaxed conditions is given in Theorem 2 in Section 2.3.2.

Even if conditions in Theorem 1 are a special case of the conditions in Theorem 2, they are still very general and useful in practice since it is easier to determine if the conditions are satisfied with the help of left and right p-credences.

However, Theorem 2 is interesting from a theorical point of view. For instance, it is possible to see the influence of each density's tail in the rejection of outliers. Since Theorem 1 is a special case of Theorem 2, it could be possible to extend the definition of left and right p-credences to include a larger class of densities (log-normal for instance), as long as the conditions of Theorem 2 are satisfied.

Some conditions of Theorem 2 are introduced in Section 2.3.1. These conditions concern mainly the thickness and regularity of the tails of a density.

2.3.1. Conditions of thickness and regularity for the tails of a density

The tails of the likelihood must satisfy certain conditions of thickness and regularity when robust inference is expected. In Theorem 1, these conditions are given using left and right p-credences. They are more general in Theorem 2. (Note that the conditions are the same for the left and right tails, except for the support of the density which is given in parentheses for the left tail.)

Three conditions of thickness and regularity for the tails of a density f are given by conditions C1 to C3 as follows. The density f is assumed to be proper, positive everywhere and bounded above.

C1 : $\forall \epsilon > 0, \forall h > 0$, there exists a constant $A_1(\epsilon, h)$ such that $z > A_1(\epsilon, h)$ ($z < -A_1(\epsilon, h)$ for the left tail) and $|\theta| \leq h \Rightarrow 1 - \epsilon \leq \frac{f(z+\theta)}{f(z)} \leq 1 + \epsilon$.

For conditions C2 and C3, there exist constants A_2 and $M_2 > 1$ and a proper density g such that for all $z > A_2$ ($z < -A_2$ for the left tail),

C2 : $\frac{f^2(z/2)}{f(z)g(z/2)} \leq M_2$,

C3 : $\frac{d^2}{dz^2} \log f^*(z) \geq \frac{d^2}{dz^2} \log g(z) \geq 0$,

where f^* is f or any other proper densities which satisfy $\frac{1}{M_2} \leq \frac{f(z)}{f^*(z)} \leq M_2$ for all $z > A_2$ ($z < -A_2$ for the left tail).

In condition C1, the ratio of the density f measured in two points with any fixed distance approaches 1 when the two points increase in the right tail. This ensures that the tail is sufficiently heavy. (Note that the interpretation of the conditions is done only for the right tail to ease the text, but it is similar for the left tail.) For example, if $f(z)$ is the density of a normal distribution, $\lim_{z \to \infty} \frac{f(z+1)}{f(z)} = 0$ and condition C1 is not satisfied. If $f(z)$ is the density of a Student distribution, $\lim_{z \to \infty} \frac{f(z+\theta)}{f(z)} = 1$, for any fixed $\theta \in \mathbb{R}$ and condition C1 is satisfied. For conditions C2 and C3 on the right tail, the density

$$g(z) = \begin{cases} \frac{\epsilon}{2}(1 + |z|)^{-(1+\epsilon)}; & \text{if } \lim_{z \to \infty} f(z)|z|^{1+\epsilon} = 0, \\ f^*(z); & \text{otherwise,} \end{cases}$$

is usually appropriate, for any choice of $\epsilon > 0$. The same density $g(z)$ is also usually appropriate when the left tail is considered, except $\lim_{z \to \infty}$ is replaced by $\lim_{z \to -\infty}$ in the first row.

The density f^* may be chosen as f or any other proper densities with a right tail of the same order. Condition C2 ensures that the decreasing of the density measured in a point $z/2$ and in a point z is bounded as z increases in the right tail, which also ensures that the tail is sufficiently heavy. Condition C3 ensures that the logarithm of the right tail of the densities f^* and g are convex, with the convexity of $\log f^*(z)$ more pronounced than that of $\log g(z)$. It can be shown that a logarithmically convex function is also convex, therefore the right tails of f^* and g are also convex.

2.3.2. Outlier rejection

Consider the Bayesian context given in Section 2.2.2.

i) Let $X_1, ..., X_n$ be n random variables conditionally independent given θ with the conditional densities of $X_i|\theta$ given by $f_i(x_i - \theta)$, where $X_i \in \mathbb{R}, \theta \in \mathbb{R}$, $i = 1, ..., n$.

ii) The prior density of θ is $\pi_\theta(\theta - x_0)$, where $x_0 \in \mathbb{R}$ is a known location parameter.

The densities $\pi_\theta, f_1, ..., f_n$ are assumed to be proper, positive everywhere and bounded above. We assume that the sample consists of a block of k observations around x_0, $m - k$ outliers on the left of x_0 and $n - m$ outliers on the right of x_0 $(0 \leq k \leq m \leq n)$. Assume without loss of generality that the left outliers are the observations denoted by $x_{k+1}, ..., x_m$ and the right outliers are the observations denoted by $x_{m+1}, ..., x_n$. The differences $x_1 - x_0$,...,$x_k - x_0$ are assumed fixed. The distances between each outlier and the block composed of $x_0, x_1, ..., x_k$ tend to ∞. Let the vector of the distances between the outliers and x_0 be denoted by $\underset{\sim}{\phi_2} = (-(x_{k+1} - x_0), ..., -(x_m - x_0), x_{m+1} - x_0, ..., x_n - x_0)$. Then $\underset{\sim}{\phi_2} \to \infty$, which means that each term of the vector tends to ∞ at any given rate.

Theorem 2. *For any integer k and m such that $0 \leq k \leq m \leq n$ and for any $x_0, x_1, ..., x_k$ such that $x_1 - x_0$,...,$x_k - x_0$ are fixed, if conditions C1 to C3 are satisfied on the left tails of $f_{k+1}, ..., f_m$ and on the right tails of $f_{m+1}, ..., f_n$, and if*

$$\lim_{\theta \to -\infty} \frac{\pi_\theta(\theta) \prod_{i=1}^{k} f_i(x_i - x_0 - \theta)}{\prod_{i=k+1}^{m} f_i(\theta)} = 0 \text{ when } k < m, \tag{2.3.1}$$

and

$$\lim_{\theta \to \infty} \frac{\pi_\theta(\theta) \prod_{i=1}^{k} f_i(x_i - x_0 - \theta)}{\prod_{i=m+1}^{n} f_i(\theta)} = 0 \text{ when } m < n, \tag{2.3.2}$$

then

a) $\lim_{\underset{\sim}{\phi_2} \to \infty} \frac{\pi(x_0|\underset{\sim}{x}_n)}{\pi(x_0|\underset{\sim}{x}_k)} = 1$,

b) for any $h > 0$ and for all θ such that $|\theta - x_0| \leq h$, $\lim_{\underset{\sim}{\phi_2} \to \infty} \frac{\pi(\theta|\underset{\sim}{x}_n)}{\pi(\theta|\underset{\sim}{x}_k)} = 1$, *c) for any $h > 0$ and $j \in (k+1, ..., n)$,* $\lim_{\underset{\sim}{\phi}_2 \to \infty} \Pr[|\theta - x_j| \leq h | \underset{\sim}{x}_n] = 0$,

d) $(\theta - x_0)|\underset{\sim}{x}_n \xrightarrow{\mathcal{L}} (\theta - x_0)|\underset{\sim}{x}_k$ *as* $\underset{\sim}{\phi}_1 \to \infty$*, where the density of the random variables* $(\theta - x_0)|\underset{\sim}{x}_n$ *and* $(\theta - x_0)|\underset{\sim}{x}_k$ *evaluated at the point y are given by* $\pi(y + x_0|\underset{\sim}{x}_n)$ *and* $\pi(y + x_0|\underset{\sim}{x}_k)$.

In addition, for any function $w(\cdot)$ on $\mathbb{R}$ such that $\mathbb{E}^{\pi_\theta(\theta)}[|w(\theta)|] < \infty$ *and* $|w(\theta)|\,\pi_\theta(\theta)$ *is bounded above, if*

$$\lim_{\theta \to -\infty} \frac{w(\theta)\pi_\theta(\theta) \prod_{i=1}^{k} f_i(x_i - x_0 - \theta)}{\prod_{i=k+1}^{m} f_i(\theta)} = 0 \text{ when } k < m, \tag{2.3.3}$$

and

$$\lim_{\theta \to \infty} \frac{w(\theta)\pi_\theta(\theta) \prod_{i=1}^{k} f_i(x_i - x_0 - \theta)}{\prod_{i=m+1}^{n} f_i(\theta)} = 0 \text{ when } m < n, \tag{2.3.4}$$

then

e) $\lim_{\underset{\sim}{\phi}_2 \to \infty} \mathbb{E}^{\pi(\theta|\underset{\sim}{x}_n)}[w(\theta - x_0)] = \mathbb{E}^{\pi(\theta|\underset{\sim}{x}_k)}[w(\theta - x_0)]$.

Proof. See the Appendix, Sections 2.6.1 to 2.6.7.

Note that if $k = 0$, $\prod_{i=1}^{k} f_i(x_i - x_0 - \theta)$ is set to 1. Conditions C1 to C3 ensure that a density's tail is logarithmically convex and sufficiently heavy. Since the numerators in the conditions given by equations (2.3.1) and (2.3.2) are proportional to $\pi(\theta + x_0|\underset{\sim}{x}_k)$, these two conditions can be interpreted as follows : the left tail of the density proportional to $\prod_{i=k+1}^{m} f_i(\theta)$ is heavier than the left tail of $\pi(\theta + x_0|\underset{\sim}{x}_k)$ and the right tail of the density proportional to $\prod_{i=m+1}^{n} f_i(\theta)$ is heavier than the right tail of $\pi(\theta + x_0|\underset{\sim}{x}_k)$. Note that there is no conditions on the right tail of the left outliers' densities or on the left tail of the right outliers' densities.

The results a) to d) are identical to those given in Theorem 1, except that $\lim_{\underset{\sim}{\phi}_1 \to \infty}$ is replaced by $\lim_{\underset{\sim}{\phi}_2 \to \infty}$. The results are valid for a specific direction of the outliers. Note that if x_0 is fixed, the limit can be rewritten as $x_{k+1} \to -\infty, ..., x_m \to -\infty, x_{m+1} \to \infty, ..., x_n \to \infty$.

Finally result e) establishes the convergence of the posterior expectation of any function such that the prior expectation of the absolute value of the function exists, if equations (2.3.3) and (2.3.4) are satisfied. The interpretation of these two equations is identical to that of equations (2.3.1) and (2.3.2), where the prior $\pi_\theta(\theta)$ has been replaced by a density proportional to $|w(\theta)|\,\pi_\theta(\theta)$, which is proper since $\mathbb{E}^{\pi_\theta(\theta)}[|w(\theta)|] < \infty$.

2.4. Example

Suppose that a portfolio manager needs a prediction on the return of the S&P 500 index for the next day. He asks 10 experts for their prediction on the return as well as a standard deviation on this prediction. The manager wants to combine this information with his prior beliefs using the Bayesian model described in Section 2.2.2. According to this setting, the manager chooses $\pi_\theta(\theta - x_0) = \frac{1}{\sigma_0} T_{10}\left(\frac{\theta - x_0}{\sigma_0}\right)$ and $f_i(x_i - \theta) = \frac{1}{\sigma_i} T_{10}\left(\frac{x_i - \theta}{\sigma_i}\right)$ for $i = 1, ..., 10$, where $T_{10}(\cdot)$ is a Student density with 10 degrees of freedom and σ_i is a scale parameter, $i = 0, 1, ..., 10$. If the standard deviations of these densities are denoted by s_i, then $\sigma_i = \sqrt{0.8}\, s_i, i = 0, 1, ..., 10$. Equal standard deviations of 1 are assumed for the prior and all observations, which means that each source of information has the same weight. In particular, the prior has the weight of one observation. It follows that $\sigma_i = \sqrt{0.8}$, $i = 0, ..., 10$. The prior belief on the predicted return is given by $x_0 = 0$.

2.4.1. First case

In the first case, the collected information from the experts is $(x_1, ..., x_{10}) = (-2, -2, -1, -1, 0, 1, 1, 2, 2, 0)$. Note that all numbers in this example are expressed in percentages. For example, the first expert's beliefs on the predicted return consists in a mean of -2 and a standard deviation of 1. Note that the posterior mean of θ is estimated using Monte Carlo simulations with importance sampling, see Desgagné and Angers (2005). For these data, the posterior mean of θ is $\mathbb{E}(\theta|\mathbf{x}) = 0$, where $\mathbf{x} = (x_1, ..., x_{10})$.

Consider now that the last observation $x_{10} = 0$ increases to 1, 3, 10 and 100, in order that x_{10} becomes eventually a positive outlier or consider that x_{10} decreases to -1, -3, -10 and -100, in order that x_{10} becomes eventually a negative outlier.

TAB. 2.1. The posterior mean $\mathbb{E}(\theta|\mathbf{x})$ for different values of x_{10}, when $(x_1, ..., x_9) = (-2, -2, -1, -1, 0, 1, 1, 2, 2)$.

x_{10}	0	1	3	10	100
$\mathbb{E}(\theta\|\mathbf{x})$	0	0.137	0.251	0.134	0.014
x_{10}	0	-1	-3	-10	-100
$\mathbb{E}(\theta\|\mathbf{x})$	0	-0.137	-0.251	-0.134	-0.014

Theorem 1 specifies how the posterior will behave in presence of this outlier. It can be shown that p-cred$^-(f_i)$ = p-cred$^+(f_i) = (0, 0, 11, 0)$, $i = 1, ..., 10$ and p-cred$^-(\pi_\theta)$ = p-cred$^+(\pi_\theta) = (0, 0, 11, 0)$. With these p-credences, only condition i) of Theorem 1 can be satisfied, i.e. $\gamma' < 1$ and $k \geq n/2$, which means that results a) to e) of Theorem 1 hold as long as the number of outliers is less or equal than 5.

The posterior expectations of θ are given in Table 2.1 for $x_{10} = 0, 1, 3, 10, 100$ and $x_{10} = 0, -1, -3, -10, -100$. When x_{10} increases from 0 to a certain threshold (around 3 for this case), the posterior expectation also increases. Beyond this threshold, the influence of x_{10} decreases to eventually be nul, as $x_{10} \longrightarrow \infty$. In this limit case, the posterior mean considering all observations tends to that considering only $x_1, ..., x_9$, which is 0 in this example. The interpretation is the same for the negative outlier.

2.4.2. Second case

In the second case, the collected information from the experts is $(x_1, ..., x_{10}) = (-1, -1, 0, 1, 1, 0, 0, 0, 0, 0)$. For these data, the posterior mean of θ is also $\mathbb{E}(\theta|\mathbf{x}) = 0$. Consider now that the last five observations $x_6, ..., x_{10}$ are equal and that they increase from 0 to 2, 5, 10 and 100, in order that they become eventually some positive conflicting information. According to Theorem 1, the posterior will eventually reject completely x_6 to x_{10} as they become more and more extreme.

The posterior expectations of θ are given in Table 2.2 for $x_6 = 0, 2, 5, 10, 100$, $x_6 = x_7 = x_8 = x_9 = x_{10}$. When the last five observations increase from 0 to a certain threshold (around 5 for this case), the posterior mean also increases.

TAB. 2.2. The posterior mean $\mathbb{E}(\theta|\mathbf{x})$ for different values of $x_6 = x_7 = x_8 = x_9 = x_{10}$, when $(x_1, ..., x_5) = (-1, -1, 0, 1, 1)$.

$x_6 = ... = x_{10}$	0	2	5	10	100
$\mathbb{E}(\theta\|\mathbf{x})$	0	0.995	2.003	0.939	0.085

TAB. 2.3. The posterior mean $\mathbb{E}(\theta|\mathbf{x})$ for different values of $x_6 = x_7 = x_8 = x_9 = x_{10} = x_{11}$, when $(x_1, ..., x_5) = (-1, -1, 0, 1, 1)$.

$x_6 = ... = x_{10} = x_{11}$	0	2	5	10	100	1000
$\mathbb{E}(\theta\|\mathbf{x})$	0	1.11	2.81	8.06	92.17	922.67

Beyond this threshold, their influence decreases to eventually be nul, as they tend to infinity. In this limit case, the posterior mean considering all observations tends to that considering only $x_1, ..., x_5$, which is 0 in this example. The interpretation is the same if negative values of x_5 to x_{10} are considered.

If another extreme observation denoted as x_{11} is added, where $x_6 = x_7 = x_8 = x_9 = x_{10} = x_{11}$, it can be seen in Table 2.3 that the posterior cannot reject anymore the extremes, even when they go to plus or minus infinity.

2.4.3. Third case

If one is ready to put much confidence in the prior, in the extent that all observations would be rejected if they were in conflict with the prior, it is possible to do it according to Theorem 1 if $(\gamma, \delta, \alpha, \beta) > (\gamma', n\delta', n\alpha', n\beta')$. To accomplish this, the degrees of freedom of the prior distribution are increased from 10 to 1000. Since $n\alpha' = 110$ and now p-cred$^-(\pi_\theta)$ = p-cred$^+(\pi_\theta) = (0, 0, 1001, 0)$, then $(0, 0, 1001, 0) > (0, 0, 110, 0)$ and condition ii) of Theorem 1 is satisfied for any $0 \leq k \leq 10$. It means that the posterior will eventually reject completely any number of outliers as they become more and more extreme. Note that it would be sufficient to increase the degrees of freedom of the prior to 110 in order that p-cred$^-(\pi_\theta)$ = p-cred$^+(\pi_\theta) = (0, 0, 111, 0)$. However, the observations would need to be more extreme to be rejected since the difference of p-credences is small.

In the third case, suppose again that $x_0 = 0$, but now all observations have the same value $x_1 = x_2 = ... = x_{10}$. The posterior expectations of θ are given in

TAB. 2.4. The posterior mean $\mathbb{E}(\theta|\mathrm{x})$ for different values of $x_1 = \ldots = x_{10}$, when $x_0 = 0$.

$x_1 = \ldots = x_{10}$	0	2	10	23	30	100
$\mathbb{E}(\theta\|\mathrm{x})$	0	1.861	9.321	21.630	4.323	1.113

Table 2.4 for $x_1 = 0, 2, 10, 23, 30, 100$, $x_1 = \ldots = x_{10}$. When all observations increase from 0 to a certain threshold (around 23 for this case), the posterior mean also increases. Beyond this threshold, the influence of the outliers decrease to eventually be nul, as they tend to infinity. In this limit case, the posterior mean tends to the prior mean. The interpretation is the same if negative outliers are considered.

2.5. Conclusion

In this paper, the behavior of the posterior density of the location parameter has been investigated when the sample contains outliers. The notion of left and right p-credences has been introduced to characterize respectively the left and right tails of a density. Simple conditions on the tails of the prior and the likelihood, using left and right p-credences, are established to determine the proportion of observations that can be rejected as outliers. We have shown that the posterior distribution converges in law to the posterior that would be obtained from the reduced sample, excluding the outliers, as they tend to $+\infty$ or $-\infty$, at any given rate. An example of combination of predictions of the S&P 500 index return is given.

2.6. Appendix : Proofs

The proofs of Theorems 1 and 2 are given in this Appendix. Since Theorem 1 is a special case of Theorem 2, the proof of the latter is presented first. In Section 2.6.1, the proof of result a) of Theorem 2 is given. The proof of Lemma 3, needed for this proof, is given in Section 2.6.2. The proofs of results b) to e) are given through Sections 2.6.3 to 2.6.6. The proof of Lemma 11, needed in the proof of result e) in Section 2.6.6, is given in Section 2.6.7. Finally, the proof of Theorem 1 is given in Section 2.6.8.

2.6.1. Proof of result a) of Theorem 2

It is assumed that the densities $\pi_\theta, f_1, ..., f_n$ are proper, positive everywhere and bounded above. Then it is easy to show that the marginals $m(\underset{\sim}{x}_k)$ and $m(\underset{\sim}{x}_n)$ are positive and bounded above and that the posterior densities $\pi(\theta|\underset{\sim}{x}_k)$ and $\pi(\theta|\underset{\sim}{x}_n)$ are also proper, positive everywhere and bounded above. Considering that $0 \leq \int_{-h}^{h} \pi(\theta + x_0|\underset{\sim}{x}_k)d\theta = \int_{x_0-h}^{x_0+h} \pi(\theta|\underset{\sim}{x}_k)d\theta \leq 1$ and that $\pi(\theta + x_0|\underset{\sim}{x}_k)$ depends only on the finite distances $x_1 - x_0, ..., x_k - x_0$, it is then possible to show the following lemma.

Lemma 1. $\forall \epsilon > 0$, *there exists a constant* $A_4(\epsilon) > 0$ *such that* $h \geq A_4(\epsilon) \Rightarrow$

$$\int_{x_0-h}^{x_0+h} \pi(\theta|\underset{\sim}{x}_k)d\theta \geq 1 - \epsilon, \int_{-\infty}^{x_0-h} \pi(\theta|\underset{\sim}{x}_k)d\theta \leq \epsilon, \text{ and } \int_{x_0+h}^{\infty} \pi(\theta|\underset{\sim}{x}_k)d\theta \leq \epsilon.$$

Assuming that conditions C1 to C3 are satisfied on the right tail of a proper, positive everywhere and bounded above density f, two other lemmas needed for the proof are given. Note that if conditions C1 to C3 are satisfied on the left tail of f, the lemmas are the same except for the support, written in parentheses.

Lemma 2. $z > A_2$ *and* $\theta > 0$ $(z < -A_2$ *and* $\theta < 0) \Rightarrow f(z + \theta) \leq (M_2)^2 f(z)$.

Proof. It can be shown that if a function is logarithmically convex, then it is also convex. It also can be shown that if the right tail of a proper density is convex, then it is necessarily decreasing. Since f^* is proper and logarithmically convex when $z > A_2$ (see C3), then the right tail of f^* is decreasing, that is $z > A_2 \Rightarrow f^*(z + \theta) < f^*(z), \forall\theta > 0$. Therefore, $z > A_2$ and $\theta > 0 \Rightarrow f(z + \theta) \leq M_2 f^*(z + \theta) \leq M_2 f^*(z) \leq (M_2)^2 f(z)$. Condition C3 is used in the first and last inequalities. The proof for the left tail is similar.

Lemma 3. $h > A_2$, $z > \max[2h, A_1(1, h)]$ *and* $\mathbb{D} = [h, \infty)$ $(z < \min[-2h, -A_1(1, h)]$ *and* $\mathbb{D} = (-\infty, -h]$ *for the left tail)* $\Rightarrow$

$$\int_{\mathbb{D}} \frac{f(z-\theta)f(\theta)}{f(z)} d\theta \leq (M_2)^{10} \text{ and } \frac{f(z-\theta)f(\theta)}{f(z)} \leq (M_2)^{11} \text{ for all } \theta \in \mathbb{D}.$$

Proof. See the Appendix, Section 2.6.2.

Using the fact that the numerators in equations (2.3.1) and (2.3.2) are proportional to $\pi(\theta + x_0|\underset{\sim}{x}_k)$ ($m(\underset{\sim}{x}_k)$ depends only on the constants $x_1 - x_0, ..., x_k - x_0$), equations (2.3.1) and (2.3.2) can respectively be rewritten as follows (assuming $k < m$ and $m < n$) : $\forall \epsilon > 0$, there exists a constant $A_3(\epsilon)$ such that

$$\theta < -A_3(\epsilon) \Rightarrow \frac{\pi(\theta + x_0|\underset{\sim}{x}_k)}{\prod_{i=k+1}^{m} f_i(\theta)} \leq \epsilon \text{ and } \theta > A_3(\epsilon) \Rightarrow \frac{\pi(\theta + x_0|\underset{\sim}{x}_k)}{\prod_{i=m+1}^{n} f_i(\theta)} \leq \epsilon. \quad (2.6.1)$$

Denote $y_i = -(x_i - x_0)$ if $i = k+1, ..., m$ and $y_i = x_i - x_0$ if $i = m+1, ..., n$. It follows that $\underset{\sim}{\phi}_2 = (y_{k+1}, ..., y_m, y_{m+1}, ..., y_n)$ and $\underset{\sim}{\phi}_2 \to \infty$ if $y_i \to \infty$, for $i = k+1, ..., n$, at any given rate. Then

$$\begin{aligned}
\frac{\pi(x_0|\underset{\sim}{x}_k)}{\pi(x_0|\underset{\sim}{x}_n)} &= \frac{m(\underset{\sim}{x}_n)}{m(\underset{\sim}{x}_k)\prod_{i=k+1}^{n} f_i(x_i - x_0)} \\
&= \frac{\int_{-\infty}^{\infty} \pi_\theta(\theta - x_0)\prod_{i=1}^{n} f_i(x_i - \theta)d\theta}{m(\underset{\sim}{x}_k)\prod_{i=k+1}^{n} f_i(x_i - x_0)} \\
&= \frac{\int_{-\infty}^{\infty} \pi(\theta|\underset{\sim}{x}_k)\prod_{i=k+1}^{n} f_i(x_i - \theta)d\theta}{\prod_{i=k+1}^{n} f_i(x_i - x_0)} \\
&= \frac{\int_{-\infty}^{\infty} \pi(\mu + x_0|\underset{\sim}{x}_k)\prod_{i=k+1}^{n} f_i((x_i - x_0) - \mu)d\mu}{\prod_{i=k+1}^{n} f_i(x_i - x_0)} \\
&= \int_{-\infty}^{\infty} \pi(\mu + x_0|\underset{\sim}{x}_k) \prod_{i=k+1}^{m} \frac{f_i(-y_i - \mu)}{f_i(-y_i)} \prod_{i=m+1}^{n} \frac{f_i(y_i - \mu)}{f_i(y_i)} d\mu.
\end{aligned}$$

Then result a) can be rewritten as follows : $\forall \epsilon > 0$, there exists a constant $A_0(\epsilon)$ such that $y_{k+1} > A_0(\epsilon), ..., y_n > A_0(\epsilon) \Rightarrow$

$$1 - \epsilon \leq \int_{-\infty}^{\infty} \pi(\theta + x_0|\underset{\sim}{x}_k) \prod_{i=k+1}^{m} \frac{f_i(-y_i - \theta)}{f_i(-y_i)} \prod_{i=m+1}^{n} \frac{f_i(y_i - \theta)}{f_i(y_i)} d\theta \leq 1 + \epsilon.$$

First choose any $0 < \epsilon < 1$. Note that if the result is true for $0 < \epsilon < 1$, it is necessarily true for any $\epsilon > 0$. Then define

$$\epsilon_0 = \min\left(\left[(1 + \epsilon/3)^{1/(n-k)} - 1\right], \left[1 - (1 - \epsilon/3)^{1/(n-k+1)}\right], (\epsilon/3)M_2^{-11n}\right).$$

Note that $0 < \epsilon_0 < \frac{1}{3}$. Define $h = \max(A_2, A_3(\epsilon_0), A_4(\epsilon_0))$ and then $A_0(\epsilon) = \max(A_1(\epsilon_0, h), 2h)$. Note that $A_0(\epsilon)$ depends only on ϵ. The constant A_1 comes from condition C1, A_2 and M_2 from conditions C2 and C3, A_3 from equation (2.6.1) and A_4 from Lemma 1. The integral is divided in three parts :

$(-\infty, -h], (-h, h]$ and (h, ∞) and consider that $y_{k+1} > A_0(\epsilon), ..., y_n > A_0(\epsilon)$. First consider the integral on $(-h, h]$.

$$\int_{-h}^{h} \pi(\theta + x_0|\underset{\sim}{x}_k) \prod_{i=k+1}^{m} \frac{f_i(-y_i - \theta)}{f_i(-y_i)} \prod_{i=m+1}^{n} \frac{f_i(y_i - \theta)}{f_i(y_i)} d\theta$$

$$\geq \int_{-h}^{h} \pi(\theta + x_0|\underset{\sim}{x}_k) \prod_{i=k+1}^{m} (1 - \epsilon_0) \prod_{i=m+1}^{n} (1 - \epsilon_0) d\theta$$

(C1 is used since $-y_i < -A_1(\epsilon_0, h)$ and $y_i > A_1(\epsilon_0, h)$)

$$= (1 - \epsilon_0)^{n-k} \int_{-h}^{h} \pi(\theta + x_0|\underset{\sim}{x}_k) d\theta$$

$$\geq (1 - \epsilon_0)^{n-k+1}$$

(Lemma 1 is used since $h \geq A_4(\epsilon_0)$)

$$\geq 1 - \epsilon/3.$$

In a similar way, it can be shown that

$$\int_{-h}^{h} \pi(\theta + x_0|\underset{\sim}{x}_k) \prod_{i=k+1}^{m} \frac{f_i(-y_i - \theta)}{f_i(-y_i)} \prod_{i=m+1}^{n} \frac{f_i(y_i - \theta)}{f_i(y_i)} d\theta \leq 1 + \epsilon/3.$$

Consider now (h, ∞) if $m < n$ is assumed.

$$\int_{h}^{\infty} \pi(\theta + x_0|\underset{\sim}{x}_k) \prod_{i=k+1}^{m} \frac{f_i(-y_i - \theta)}{f_i(-y_i)} \prod_{i=m+1}^{n} \frac{f_i(y_i - \theta)}{f_i(y_i)} d\theta$$

$$\leq (M_2)^{2(m-k)} \int_{h}^{\infty} \pi(\theta + x_0|\underset{\sim}{x}_k) \prod_{i=m+1}^{n} \frac{f_i(y_i - \theta)}{f_i(y_i)} d\theta$$

(Lemma 2 is used since $-y_i - \theta < -y_i < -A_2$)

$$\leq \epsilon_0 (M_2)^{2(m-k)} \int_{h}^{\infty} \prod_{i=m+1}^{n} \frac{f_i(y_i - \theta) f_i(\theta)}{f_i(y_i)} d\theta$$

(equation (2.6.1) is used since $\theta \geq A_3(\epsilon_0)$)

$$\leq \epsilon_0 (M_2)^{2(m-k)} (M_2)^{11(n-m-1)} \int_{h}^{\infty} \frac{f_n(y_n - \theta) f_n(\theta)}{f_n(y_n)} d\theta$$

(Lemma 3 is used since $h > A_2$ and $y_i > \max[2h, A_1(1, h)]$)

$$\leq \epsilon_0 (M_2)^{2(m-k)} (M_2)^{11(n-m-1)} (M_2)^{10}$$

(Lemma 3 is used)

$\leq \epsilon_0(M_2)^{11n}$

(since $0 \leq k \leq m \leq n$)

$\leq \epsilon/3.$

Consider (h, ∞) if $m = n$ is assumed.

$$\int_h^\infty \pi(\theta + x_0|\underset{\sim}{x}_k) \prod_{i=k+1}^m \frac{f_i(-y_i - \theta)}{f_i(-y_i)} d\theta$$

$$\leq (M_2)^{2(m-k)} \int_h^\infty \pi(\theta + x_0|\underset{\sim}{x}_k) d\theta$$

(Lemma 2 is used since $-y_i - \theta < -y_i < -A_2$)

$\leq \epsilon_0(M_2)^{2(m-k)}$

(Lemma 1 is used since $h > A_4(\epsilon_0)$)

$\leq \epsilon_0(M_2)^{11n}$

$\leq \epsilon/3.$

In a similar way, it can be shown that

$$\int_{-\infty}^{-h} \pi(\theta + x_0|\underset{\sim}{x}_k) \prod_{i=k+1}^m \frac{f_i(-y_i - \theta)}{f_i(-y_i)} \prod_{i=m+1}^n \frac{f_i(y_i - \theta)}{f_i(y_i)} d\theta \leq \epsilon/3.$$

Considering the three parts of the integral, we showed that

$$\int_{-\infty}^\infty \pi(\theta + x_0|\underset{\sim}{x}_k) \prod_{i=k+1}^m \frac{f_i(-y_i - \theta)}{f_i(-y_i)} \prod_{i=m+1}^n \frac{f_i(y_i - \theta)}{f_i(y_i)} d\theta$$

$$\leq (1 + \epsilon/3) + \epsilon/3 + \epsilon/3$$

$$= 1 + \epsilon,$$

and

$$\int_{-\infty}^\infty \pi(\theta + x_0|\underset{\sim}{x}_k) \prod_{i=k+1}^m \frac{f_i(-y_i - \theta)}{f_i(-y_i)} \prod_{i=m+1}^n \frac{f_i(y_i - \theta)}{f_i(y_i)} d\theta$$

$$\geq (1 - \epsilon/3) + 0 + 0$$

$$> 1 - \epsilon.$$

2.6.2. Proof of Lemma 3

We first introduce four other lemmas needed to prove Lemma 3. Suppose that conditions C1 to C3 are satisfied on the right tail of a proper density f, positive everywhere and bounded above. (If conditions C1 to C3 are satisfied on its left tail, the lemmas are the same, except for the support given in parentheses. The proofs are given only for the right tail, the proofs for the left tail being similar.) Without loss of generality, we assume that the constant M_2 in conditions C2 and C3 is chosen large enough, such that $M_2 > \max\left[\sup_{z\in\mathbb{R}} f(z), g(A_2), 6\right]$.

Lemma 4. *$z > A_2$ ($z < -A_2$ for the left tail) $\Rightarrow f^*(z) > 0$ and $g(z) > 0$.*

Proof. If $f^*(z) = 0$ for a $z > A_2$, then the second part of condition C3 is not satisfied. If $g(z) = 0$ for a $z > A_2$, then condition C2 is not satisfied.

Lemma 5. *$z > A_2$ $(z < -A_2) \Rightarrow f(z) \leq (M_2)^3 g(z)$.*

Proof. Using Lemma 2, if $z > A_2$ then $f(2z) \leq (M_2)^2 f(z)$. Using C2, if $z > A_2$ then $M_2 g(z) \geq \frac{f^2(z)}{f(2z)}$. Therefore $z > A_2 \Rightarrow (M_2)^3 g(z) \geq (M_2)^2 \frac{f^2(z)}{f(2z)} = f(z)\frac{(M_2)^2 f(z)}{f(2z)} \geq f(z)$.

Lemma 6. *$z > A_2 \Rightarrow g(z) < g(A_2)$ $(z < -A_2 \Rightarrow g(z) < g(-A_2))$.*

Proof. Since g is a proper density and it is logarithmically convex (see C3) when $z > A_2$, then the right tail of g is decreasing and then bounded above by $g(A_2)$.

Lemma 7. *For all a, b and z such that $A_2 \leq a \leq b \leq z - A_2$ ($z + A_2 \leq a \leq b \leq -A_2$ for the left tail), $\arg\max_{a\leq\theta\leq b} \frac{f^*(z-\theta)f^*(\theta)}{g(\theta)} \in \{a, b\}$.*

Proof. Since the maximum on a range of a convex function is located at its bounds, it is sufficient to show that $\frac{d^2}{d\theta^2}\log\frac{f^*(z-\theta)f^*(\theta)}{g(\theta)} \geq 0$ for any θ such that $A_2 < \theta < z - A_2$, since the convexity of the logarithm of a function implies the convexity of the function. Then

$$\frac{d^2}{d\theta^2}\log\frac{f^*(z-\theta)f^*(\theta)}{g(\theta)} = \frac{d^2}{d\theta^2}\log f^*(z-\theta) + \frac{d^2}{d\theta^2}\log f^*(\theta) - \frac{d^2}{d\theta^2}\log g(\theta).$$

Using C3, $\frac{d^2}{d\theta^2}\log f^*(\theta) - \frac{d^2}{d\theta^2}\log g(\theta) \geq 0$ for $\theta > A_2$. It can also be shown that $\frac{d^2}{d\theta^2}\log f^*(z-\theta) = \left(\frac{d^2}{dy^2}\log f^*(y)\right)|_{y=z-\theta}$, and using C3, that it is non negative for

$z-\theta > A_2$. Then we showed that $\frac{d^2}{d\theta^2} \log \frac{f^*(z-\theta)f^*(\theta)}{g(\theta)} \geq 0$ if $\theta > A_2$ and $z-\theta > A_2$, that is if $A_2 < \theta < z - A_2$. □

To prove Lemma 3, we divide $[h, \infty)$ in three parts : $[h, z/2], (z/2, z-h]$ and $(z-h, \infty)$. Consider that $h > A_2$ and $z > \max[2h, A_1(1,h)]$. The constants A_1 and A_2 come respectively from conditions C1 and C2.

First consider $h \leq \theta \leq z/2$. Note that $h \leq \theta \leq z/2$, $h > A_2$ and $z > 2h \Rightarrow z > z - A_2 > z - h \geq z - \theta \geq z/2 \geq \theta \geq h > A_2$. Then

$$\frac{f(z-\theta)f(\theta)}{f(z)} \leq (M_2)^3 \frac{f^*(z-\theta)f^*(\theta)}{f^*(z)}$$

(C3 is used since $z-\theta > A_2$, $\theta > A_2$ and $z > A_2$)

$$= (M_2)^3 \left(\frac{f^*(z-\theta)f^*(\theta)}{f^*(z)g(\theta)}\right) g(\theta)$$

$$\leq (M_2)^3 \max\left(\frac{f^*(z-h)f^*(h)}{f^*(z)g(h)}, \frac{f^{*2}(z/2)}{f^*(z)g(z/2)}\right) g(\theta)$$

(Lemma 7 is used since $A_2 < h \leq \theta \leq z/2 < z - A_2$)

$$\leq (M_2)^6 \max\left(\frac{f(z-h)f(h)}{f(z)g(h)}, \frac{f^2(z/2)}{f(z)g(z/2)}\right) g(\theta)$$

(C3 is used since $z-h > A_2, h > A_2, z > A_2$ and $z/2 > A_2$)

$$\leq (M_2)^6 \max\left(\frac{f(z-h)(M_2)^3}{f(z)}, M_2\right) g(\theta)$$

(Lemma 5 is used since $h > A_2$ and C2 is used since $z > A_2$)

$$\leq (M_2)^6 \max\left(2(M_2)^3, M_2\right) g(\theta)$$

(C1 is used since $z > A_1(1,h)$)

$$= 2(M_2)^9 g(\theta)$$

$$\leq 2(M_2)^9 g(A_2)$$

(Lemma 6 is used since $\theta > A_2$)

$$\leq (M_2)^{11}$$

(since $M_2 > 2$ and $M_2 \geq g(A_2)$)

and since $g(\cdot)$ is a proper density,

$$\int_h^{z/2} \frac{f(z-\theta)f(\theta)}{f(z)} d\theta \leq 2(M_2)^9 \int_h^{z/2} g(\theta)d\theta$$
$$\leq 2(M_2)^9.$$

Consider now $z/2 \leq \theta \leq z-h$. It is possible to use the precedent results (when $h \leq \theta \leq z/2$ is considered) if the change of variables $u = z - \theta$ is done, since $h \leq u \leq z/2$. Then

$$\frac{f(z-\theta)f(\theta)}{f(z)} = \frac{f(u)f(z-u)}{f(z)}$$
$$\leq 2(M_2)^9 g(u)$$
$$= 2(M_2)^9 g(z-\theta)$$
$$\leq 2(M_2)^9 g(A_2)$$
(Lemma 6 is used since $z - \theta > A_2$)
$$\leq (M_2)^{11}$$

and since $g(\cdot)$ is a proper density,

$$\int_{z/2}^{z-h} \frac{f(z-\theta)f(\theta)}{f(z)} d\theta \leq 2(M_2)^9 \int_{z/2}^{z-h} g(z-\theta)d\theta$$
$$\leq 2(M_2)^9.$$

Finally consider $\theta \geq z - h$.

$$\frac{f(z-\theta)f(\theta)}{f(z)} \leq \frac{f(z-\theta)(M_2)^2 f(z-h)}{f(z)}$$
(Lemma 2 is used since $\theta \geq z - h > A_2$)
$$\leq 2(M_2)^2 f(z-\theta)$$
(C1 is used since $z > A_1(1,h)$)
$$\leq 2(M_2)^3$$
(since $\sup_{z\in\mathbb{R}} f(z) \leq M_2$ and $M_2 > 1$)
$$\leq (M_2)^{11}$$

and since $f(\cdot)$ is a proper density,

$$\int_{z-h}^{\infty} \frac{f(z-\theta)f(\theta)}{f(z)} d\theta \leq 2(M_2)^2 \int_{z-h}^{\infty} f(z-\theta) d\theta$$

$$\leq 2(M_2)^2$$

$$\leq 2(M_2)^9.$$

If the integrals on the three domains are considered, then

$$\int_{h}^{\infty} \frac{f(z-\theta)f(\theta)}{f(z)} d\theta \leq 6(M_2)^9 \leq (M_2)^{10}.$$

2.6.3. Proof of result b) of Theorem 2

Result b) can be rewritten as follows : $\forall\epsilon > 0, \forall h > 0$ there exists a constant $A_5(\epsilon, h)$ such that $\min[\underset{\sim}{\phi_2}] > A_5(\epsilon, h)$ and $|\theta - x_0| \leq h \Rightarrow 1 - \epsilon \leq \frac{\pi(\theta|\underset{\sim}{x}_n)}{\pi(\theta|\underset{\sim}{x}_k)} \leq 1 + \epsilon$. Note that $\min[\underset{\sim}{\phi_2}]$ stands for $\min[-(x_{k+1} - x_0), ..., -(x_m - x_0), x_{m+1} - x_0, ..., x_n - x_0]$. Result a) of Theorem 2 can also be rewritten as follows : $\forall\epsilon > 0$ there exists a constant $A_0(\epsilon)$ such that $\min[\underset{\sim}{\phi_2}] > A_0(\epsilon) \Rightarrow 1 - \epsilon \leq \frac{\pi(x_0|\underset{\sim}{x}_n)}{\pi(x_0|\underset{\sim}{x}_k)} \leq 1 + \epsilon$.

Choose any $\epsilon > 0$ and any $h > 0$. Then define

$$\epsilon_0 = \min[(1+\epsilon)^{1/(n-k+1)} - 1, 1 - (1-\epsilon)^{1/(n-k+1)}]$$

and $A_5(\epsilon, h) = \max[A_0(\epsilon_0), A_1(\epsilon_0, h)]$. The constants A_0 and A_1 come respectively from the proof of result a) of Theorem 2 and condition C1. Consider that $\min[\underset{\sim}{\phi_2}] > A_5(\epsilon, h)$ and $|\theta - x_0| \leq h$. Then

$$\frac{\pi(\theta|\underset{\sim}{x}_n)}{\pi(\theta|\underset{\sim}{x}_k)} = \frac{m(\underset{\sim}{x}_k)\prod_{i=k+1}^{n} f_i(x_i - \theta)}{m(\underset{\sim}{x}_n)}$$

$$= \frac{\pi(x_0|\underset{\sim}{x}_n)}{\pi(x_0|\underset{\sim}{x}_k)} \left(\prod_{i=k+1}^{n} \frac{f_i([x_i - x_0] - [\theta - x_0])}{f_i(x_i - x_0)} \right)$$

$$\leq (1+\epsilon_0) \prod_{i=k+1}^{n} \frac{f_i([x_i - x_0] - [\theta - x_0])}{f_i(x_i - x_0)}$$

(Result a) is used since $\min[\underset{\sim}{\phi_2}] > A_0(\epsilon_0)$)

$$\leq (1+\epsilon_0) \prod_{i=k+1}^{n} (1+\epsilon_0)$$

(C1 is used since $\min[\underset{\sim}{\phi_2}] > A_1(\epsilon_0, h)$)

$$= (1+\epsilon_0)^{n-k+1}$$
$$\leq 1+\epsilon.$$

In a similar way, it can be shown that $\frac{\pi(\theta|\underset{\sim}{x}_n)}{\pi(\theta|\underset{\sim}{x}_k)} \geq 1-\epsilon$.

2.6.4. Proof of result c) of Theorem 2

Result c) of Theorem 2 says that the posterior density tends to 0, in any finite neighborhood of any outliers x_j, $j \in (k+1, ..., n)$. It can be rewritten as follows : $\forall \epsilon > 0, \forall d > 0$ there exists a constant $A_6(\epsilon, d)$ such that $\min[\underset{\sim}{\phi}_2] > A_6(\epsilon, d)$ and $j \in (k+1, ..., n) \Rightarrow \Pr[|\theta - x_j| \leq d|\underset{\sim}{x}_n] \leq \epsilon$. A lemma analog to Lemma 1 is needed for the proof.

Lemma 8. *$\epsilon > 0, h \geq A_4(\epsilon/2)$ and $\min[\underset{\sim}{\phi}_2] > A_5(\epsilon/2, h) \Rightarrow$*

$$\int_{x_0-h}^{x_0+h} \pi(\theta|\underset{\sim}{x}_n)d\theta \geq 1-\epsilon, \int_{-\infty}^{x_0-h} \pi(\theta|\underset{\sim}{x}_n)d\theta \leq \epsilon \text{ and } \int_{x_0+h}^{\infty} \pi(\theta|\underset{\sim}{x}_n)d\theta \leq \epsilon.$$

Proof. $\int_{x_0-h}^{x_0+h} \pi(\theta|\underset{\sim}{x}_n)d\theta \geq (1-\epsilon/2) \int_{x_0-h}^{x_0+h} \pi(\theta|\underset{\sim}{x}_k)d\theta \geq (1-\epsilon/2)^2 > 1-\epsilon$. Result b) of Theorem 2 is used in the first inequality since $\min[\underset{\sim}{\phi}_2] > A_5(\epsilon/2, h)$ and $|\theta - x_0| \leq h$, and Lemma 1 in used in the second since $h \geq A_4(\epsilon/2)$. Furthermore,

$$\int_{-\infty}^{x_0-h} \pi(\theta|\underset{\sim}{x}_n)d\theta + \int_{x_0+h}^{\infty} \pi(\theta|\underset{\sim}{x}_n)d\theta = \int_{-\infty}^{\infty} \pi(\theta|\underset{\sim}{x}_n)d\theta - \int_{x_0-h}^{x_0+h} \pi(\theta|\underset{\sim}{x}_n)d\theta$$
$$\leq 1-(1-\epsilon)$$
$$= \epsilon.$$

□

Choose any $\epsilon > 0$ and any $d > 0$. Define $h = A_4(\epsilon/2)$ and define $A_6(\epsilon, d) = \max[A_5(\epsilon/2, h), d+h]$, where the constant A_5 comes from the proof of result b) of Theorem 2. Consider that $\min[\underset{\sim}{\phi}_2] > A_6(\epsilon, d)$ and $j \in (k+1, ..., n)$. Since $|x_j - x_0| \in \underset{\sim}{\phi}_2$, it follows that $|x_j - x_0| > d+h$. Then, if $x_j - x_0 > 0$ (that is for $j = m+1, ..., n$),

$$\Pr[|\theta - x_j| \leq d|\underset{\sim}{x}_n] = \int_{x_j-d}^{x_j+d} \pi(\theta|\underset{\sim}{x}_n)d\theta$$
$$\leq \int_{x_j-d}^{\infty} \pi(\theta|\underset{\sim}{x}_n)d\theta$$

$$\leq \int_{x_0+h}^{\infty} \pi(\theta|\underset{\sim}{x}_n)d\theta$$
$$\leq \epsilon.$$

Lemma 8 is used in the last inequality. The proof for $x_j - x_0 < 0$ (that is for $j = k+1, ..., m$) is similar.

2.6.5. Proof of result d) of Theorem 2

The definition of convergence in law of a sequence of random variables $\{Y_s\}_{s=1,2,3,...}$ to a random variable Y, as $s \to \infty$, is given as follows.

Definition 5. $Y_s \xrightarrow{\mathcal{L}} Y$ *if* $\lim_{s\to\infty} \Pr[Y_s \leq d] = \Pr[Y \leq d]$, *for all* d *such that* $\Pr[Y \leq d]$ *is continuous.*

In order to use this definition with $Y_s = (\theta - x_0)|\underset{\sim}{x}_n$ and $Y = (\theta - x_0)|\underset{\sim}{x}_k$, the prior location and the observations are expressed as some functions of the same variable s, denoted by $x_i = h_i(s), i = 0, 1, ..., n$, for any functions $h_i(s)$ on $\mathbb{N}$ which satisfy

i) there exists a constant c_i such that $h_i(s) - h_0(s) = c_i$ for any $s \in \mathbb{N}$, if $i = 1, ..., k$,

ii) $\lim_{s\to\infty}(h_i(s) - h_0(s)) = -\infty$, if $i = k+1, ..., m$,

iii) $\lim_{s\to\infty}(h_i(s) - h_0(s)) = \infty$ if $i = m+1, ..., n$.

The density of Y_s evaluated at the point y is then given by

$$\pi(y + x_0|\underset{\sim}{x}_n) = \frac{\pi_\theta(y)\prod_{i=1}^n f_i(x_i - x_0 - y)}{\int_{-\infty}^{\infty} \pi_\theta(y)\prod_{i=1}^n f_i(x_i - x_0 - y)\,dy}$$
$$= \frac{\pi_\theta(y)\prod_{i=1}^k f_i(c_i - y)\prod_{i=k+1}^n f_i(h_i(s) - h_0(s) - y)}{\int_{-\infty}^{\infty} \pi_\theta(y)\prod_{i=1}^k f_i(c_i - y)\prod_{i=k+1}^n f_i(h_i(s) - h_0(s) - y)\,dy}$$

and the density of Y evaluated at the point y is given by

$$\pi(y + x_0|\underset{\sim}{x}_k) = \frac{\pi_\theta(y)\prod_{i=1}^k f_i(c_i - y)}{\int_{-\infty}^{\infty} \pi_\theta(y)\prod_{i=1}^k f_i(c_i - y)\,dy}.$$

It can be seen that the functions $h_i(s)$ are defined such that $s \to \infty \Leftrightarrow \underset{\sim}{\phi}_2 \to \infty$. Furthermore, it can be seen that the density of $Y = (\theta - x_0)|\underset{\sim}{x}_k$ does not depend on s or $\underset{\sim}{\phi}_2$. Then $Y_s \xrightarrow{\mathcal{L}} Y$ as $s \to \infty$ for any functions $h_i(s)$ which satisfy i), ii) and iii) $\Leftrightarrow (\theta - x_0)|\underset{\sim}{x}_n \xrightarrow{\mathcal{L}} (\theta - x_0)|\underset{\sim}{x}_k$ as $\underset{\sim}{\phi}_2 \to \infty$ at any given rate.

According to Definition 5, the convergence in law is obtained if $\lim_{s\to\infty} \Pr[Y_s \leq d] = \Pr[Y \leq d]$, for all d such that $\Pr[Y \leq d]$ is continuous, or equivalently, if $\lim_{\phi_2\to\infty} \Pr[\theta \leq d+x_0|\underset{\sim}{x}_n] = \Pr[\theta \leq d+x_0|\underset{\sim}{x}_k]$, for all d such that $\Pr[\theta \leq d+x_0|\underset{\sim}{x}_k]$ is continuous. Therefore, the result d) can be rewritten as follows : $\forall\epsilon > 0$ there exists a constant $A_7(\epsilon)$ such that $\min[\underset{\sim}{\phi}_2] > A_7(\epsilon)$ and $d \in \mathbb{R} \Rightarrow$
$|\Pr[\theta - x_0 \leq d|\underset{\sim}{x}_n] - \Pr[\theta - x_0 \leq d|\underset{\sim}{x}_k]| \leq \epsilon$.

Choose any $\epsilon > 0$, define $h = A_4(\epsilon/6)$ and $A_7(\epsilon) = A_5(\epsilon/6, h)$. The constants A_4 and A_5 come respectively from Lemma 1 and the proof of result b) of Theorem 2. The real line is divided in three parts : $(-\infty, -h], (-h, h]$ and (h, ∞), and consider that $\min[\underset{\sim}{\phi}_2] > A_7(\epsilon)$. First consider $d \leq -h$.

$$\begin{aligned}\Pr[\theta - x_0 \leq d|\underset{\sim}{x}_n] &\leq \Pr[\theta \leq x_0 - h|\underset{\sim}{x}_n]\\ &= \int_{-\infty}^{x_0-h} \pi(\theta|\underset{\sim}{x}_n)d\theta\\ &\leq \epsilon/3.\end{aligned}$$

Lemma 8 is used in the last inequality since $h = A_4(\epsilon/6)$ and $\min[\underset{\sim}{\phi}_2] > A_5(\epsilon/6, h)$. In the same way, it can be shown, using Lemma 1, that $\Pr[\theta - x_0 \leq d|\underset{\sim}{x}_k] \leq \epsilon/6$, since $h = A_4(\epsilon/6)$. From this result and from $\Pr[\theta - x_0 \leq d|\underset{\sim}{x}_n] \leq \epsilon/3$, it follows that $|\Pr[\theta - x_0 \leq d|\underset{\sim}{x}_n] - \Pr[\theta - x_0 \leq d|\underset{\sim}{x}_k]| \leq \epsilon/3 < \epsilon$. Now consider $-h < d \leq h$.

$$\begin{aligned}&|\Pr[-h < \theta - x_0 \leq d|\underset{\sim}{x}_n] - \Pr[-h < \theta - x_0 \leq d|\underset{\sim}{x}_k]|\\ &\qquad\leq \int_{x_0-h}^{x_0+d} |\pi(\theta|\underset{\sim}{x}_n) - \pi(\theta|\underset{\sim}{x}_k)|\, d\theta\\ &\qquad= \int_{x_0-h}^{x_0+d} \pi(\theta|\underset{\sim}{x}_k) \left|\frac{\pi(\theta|\underset{\sim}{x}_n)}{\pi(\theta|\underset{\sim}{x}_k)} - 1\right| d\theta\\ &\qquad\leq \epsilon/6 \int_{x_0-h}^{x_0+d} \pi(\theta|\underset{\sim}{x}_k)d\theta\\ &\qquad\leq \epsilon/6.\end{aligned}$$

Result b) of Theorem 2 is used in the second inequality since $\min[\underset{\sim}{\phi}_2] > A_5(\epsilon/6, h)$ and $|\theta - x_0| \leq h$. Therefore,

$$|\Pr[\theta - x_0 \leq d|\underset{\sim}{x}_n] - \Pr[\theta - x_0 \leq d|\underset{\sim}{x}_k]|$$

$$\leq |\Pr[\theta - x_0 \leq -h|\underset{\sim}{x}_n] - \Pr[\theta - x_0 \leq -h|\underset{\sim}{x}_k]|$$
$$+ |\Pr[-h < \theta - x_0 \leq d|\underset{\sim}{x}_n] - \Pr[-h < \theta - x_0 \leq d|\underset{\sim}{x}_k]|$$
$$\leq \epsilon/3 + \epsilon/6$$
$$= \epsilon/2$$
$$< \epsilon.$$

Finally consider $d > h$.

$$\begin{aligned}\Pr[h < \theta - x_0 \leq d|\underset{\sim}{x}_n] \leq \Pr[\theta > x_0 + h|\underset{\sim}{x}_n] \\ = \int_{x_0+h}^{\infty} \pi(\theta|\underset{\sim}{x}_n)d\theta \\ \leq \epsilon/3.\end{aligned}$$

Lemma 8 is used in the last inequality since $h = A_4(\epsilon/6)$ and $\min[\underset{\sim}{\phi}_2] > A_5(\epsilon/6, h)$. In the same way, it can be shown, using Lemma 1, that $\Pr[h < \theta - x_0 \leq d|\underset{\sim}{x}_k] \leq \epsilon/6$, since $h = A_4(\epsilon/6)$. From this result and from $\Pr[h < \theta - x_0 \leq d|\underset{\sim}{x}_n] \leq \epsilon/3$, it follows that

$$|\Pr[h < \theta - x_0 \leq d|\underset{\sim}{x}_n] - \Pr[h < \theta - x_0 \leq d|\underset{\sim}{x}_k]| \leq \epsilon/3.$$

Finally, from this result and from $|\Pr[\theta - x_0 \leq h|\underset{\sim}{x}_n] - \Pr[\theta - x_0 \leq h|\underset{\sim}{x}_k]| \leq \epsilon/2$, it follows that $|\Pr[\theta - x_0 \leq d|\underset{\sim}{x}_n] - \Pr[\theta - x_0 \leq d|\underset{\sim}{x}_k]| \leq \epsilon/2 + \epsilon/3 < \epsilon$.

2.6.6. Proof of result e) of Theorem 2

First we introduce three lemmas needed for the proof.

Lemma 9.

$$\mathbb{E}^{\pi(\theta|\underset{\sim}{x}_k)}[|w(\theta - x_0)|] \leq M_1,$$

where $M_1 = \frac{M_2^k}{m(\underset{\sim}{x}_k)}\, \mathbb{E}^{\pi_0(\theta)}[|w(\theta)|] < \infty$.

Proof.

$$\begin{aligned}\mathbb{E}^{\pi(\theta|\underset{\sim}{x}_k)}[|w(\theta - x_0)|] &= \int_{-\infty}^{\infty} |w(\theta - x_0)|\, \pi_0(\theta - x_0) \frac{\prod_{i=1}^{k} f_i(x_i - \theta)}{m(\underset{\sim}{x}_k)} d\theta \\ &\leq \frac{M_2^k}{m(\underset{\sim}{x}_k)} \int_{-\infty}^{\infty} |w(\theta - x_0)|\, \pi_0(\theta - x_0) d\theta \\ &= M_1.\end{aligned}$$

The fact that $f_1, ..., f_n$ are bounded above by M_2 (see Section 2.6.2) is used in the inequality. Furthermore, since $\mathbb{E}^{\pi_\theta(\theta)}[|w(\theta)|] < \infty$ is assumed in Theorem 2 and $0 < m(\underset{\sim}{x}_k) < \infty$, it follows that $M_1 < \infty$. □

Considering that

$$0 \leq \int_{-h}^{h} |w(\theta)|\, \pi(\theta + x_0|\underset{\sim}{x}_k)d\theta = \int_{x_0-h}^{x_0+h} |w(\theta - x_0)|\, \pi(\theta|\underset{\sim}{x}_k)d\theta \leq M_1$$

and that $|w(\theta)|\, \pi(\theta + x_0|\underset{\sim}{x}_k)$ depends only on the finite distances $x_1 - x_0, ...,$ $x_k - x_0$, it is then possible to show the following lemma.

Lemma 10. $\forall \epsilon > 0$, *there exists a constant* $A_9(\epsilon) > 0$ *such that* $h \geq A_9(\epsilon) \Rightarrow$

$$\int_{-\infty}^{x_0-h} |w(\theta - x_0)|\, \pi(\theta|\underset{\sim}{x}_k)d\theta \leq \epsilon \text{ and } \int_{x_0+h}^{\infty} |w(\theta - x_0)|\, \pi(\theta|\underset{\sim}{x}_k)d\theta \leq \epsilon.$$

Note that the condition $\mathbb{E}^{\pi_\theta(\theta)}[|w(\theta)|] < \infty$ is conservative but is appropriate whatever the number of outliers is. It could be possible to relax it using the condition $\mathbb{E}^{\pi(\theta|\underset{\sim}{x}_k)}[|w(\theta - x_0)|] < \infty$, considering also the non-outlier observations.

A last lemma is needed and its proof is given in Section 2.6.7 of this Appendix.

Lemma 11. $\forall \epsilon > 0$, *there exists a constant* $A_8(\epsilon)$ *such that* $\min[\underset{\sim}{\phi}_2] > A_8(\epsilon) \Rightarrow$

$$\left|\mathbb{E}^{\pi(\theta|\underset{\sim}{x}_n)}[|w(\theta - x_0)|] - \mathbb{E}^{\pi(\theta|\underset{\sim}{x}_k)}[|w(\theta - x_0)|]\right| < \epsilon.$$

Lemma 11 is similar to the result e) of Theorem 2, except it considers the absolute value of $w(\theta - x_0)$.

Consider now the result e) of Theorem 2, which can be rewritten as follows. $\forall \epsilon > 0$, there exists a constant $A_0(\epsilon)$ such that $\min[\underset{\sim}{\phi}_2] > A_0(\epsilon) \Rightarrow$

$$\left|\mathbb{E}^{\pi(\theta|\underset{\sim}{x}_n)}[w(\theta - x_0)] - \mathbb{E}^{\pi(\theta|\underset{\sim}{x}_k)}[w(\theta - x_0)]\right| < \epsilon.$$

Choose any $\epsilon > 0$. Define $\epsilon_0 = \epsilon/7$, $h = A_9(\epsilon_0)$ and
$A_0(\epsilon) = \max[A_5(\epsilon_0/M_1, h), A_8(\epsilon_0)]$, where the constant $A_5(\epsilon_0/M_1, h)$ comes from the proof of result b) of Theorem 2, which was rewritten as follows : $\forall \epsilon > 0, \forall h > 0$, there exists a constant $A_5(\epsilon, h)$ such that $\min[\underset{\sim}{\phi}_2] > A_5(\epsilon, h)$ and $|\theta - x_0| \leq h \Rightarrow \left|\frac{\pi(\theta|\underset{\sim}{x}_n)}{\pi(\theta|\underset{\sim}{x}_k)} - 1\right| \leq \epsilon$.

Then

$$\left|\mathbb{E}^{\pi(\theta|\underset{\sim}{x}_n)}[w(\theta - x_0)] - \mathbb{E}^{\pi(\theta|\underset{\sim}{x}_k)}[w(\theta - x_0)]\right|$$

$$= \left| \int_{-\infty}^{\infty} w(\theta - x_0)\pi(\theta|\underset{\sim}{x}_n)d\theta - \int_{-\infty}^{\infty} w(\theta - x_0)\pi(\theta|\underset{\sim}{x}_k)d\theta \right|$$

$$\leq \int_{-\infty}^{x_0-h} |w(\theta - x_0)|\, \pi(\theta|\underset{\sim}{x}_n)d\theta + \int_{x_0+h}^{\infty} |w(\theta - x_0)|\, \pi(\theta|\underset{\sim}{x}_n)d\theta$$

$$+ \left| \int_{x_0-h}^{x_0+h} w(\theta - x_0)\pi(\theta|\underset{\sim}{x}_n)d\theta - \int_{x_0-h}^{x_0+h} w(\theta - x_0)\pi(\theta|\underset{\sim}{x}_k)d\theta \right|$$

$$+ \int_{-\infty}^{x_0-h} |w(\theta - x_0)|\, \pi(\theta|\underset{\sim}{x}_k)d\theta + \int_{x_0+h}^{\infty} |w(\theta - x_0)|\, \pi(\theta|\underset{\sim}{x}_k)d\theta$$

$$\leq \mathbb{E}^{\pi(\theta|\underset{\sim}{x}_n)}[|w(\theta - x_0)|] - \int_{x_0-h}^{x_0+h} |w(\theta - x_0)|\, \pi(\theta|\underset{\sim}{x}_n)d\theta$$

$$+ \int_{x_0-h}^{x_0+h} |w(\theta - x_0)|\, |\pi(\theta|\underset{\sim}{x}_n) - \pi(\theta|\underset{\sim}{x}_k)|\, d\theta + 2\epsilon_0.$$

Lemma 10 is used in the last inequality since $h = A_9(\epsilon_0)$. Now

$$\int_{x_0-h}^{x_0+h} |w(\theta - x_0)|\, |\pi(\theta|\underset{\sim}{x}_n) - \pi(\theta|\underset{\sim}{x}_k)|\, d\theta$$

$$= \int_{x_0-h}^{x_0+h} |w(\theta - x_0)|\, \pi(\theta|\underset{\sim}{x}_k) \left| \frac{\pi(\theta|\underset{\sim}{x}_n)}{\pi(\theta|\underset{\sim}{x}_k)} - 1 \right| d\theta$$

$$\leq \frac{\epsilon_0}{M_1} \int_{x_0-h}^{x_0+h} |w(\theta - x_0)|\, \pi(\theta|\underset{\sim}{x}_k)d\theta$$

(result b) of Theorem 2 is used since $A_0(\epsilon) \geq A_5(\epsilon_0/M_1, h)$)

$$\leq \frac{\epsilon_0}{M_1} \mathbb{E}^{\pi(\theta|\underset{\sim}{x}_k)}[|w(\theta - x_0)|]$$

$$\leq \epsilon_0$$

(Lemma 9 is used).

Finally

$$\mathbb{E}^{\pi(\theta|\underset{\sim}{x}_n)}[|w(\theta - x_0)|] - \int_{x_0-h}^{x_0+h} |w(\theta - x_0)|\, \pi(\theta|\underset{\sim}{x}_n)d\theta$$

$$\leq \left| \mathbb{E}^{\pi(\theta|\underset{\sim}{x}_n)}[|w(\theta - x_0)|] - \mathbb{E}^{\pi(\theta|\underset{\sim}{x}_k)}[|w(\theta - x_0)|] \right|$$

$$+ \left| \mathbb{E}^{\pi(\theta|\underset{\sim}{x}_k)}[|w(\theta - x_0)|] - \int_{x_0-h}^{x_0+h} |w(\theta - x_0)|\, \pi(\theta|\underset{\sim}{x}_k)d\theta \right|$$

$$+ \left| \int_{x_0-h}^{x_0+h} |w(\theta - x_0)|\, \pi(\theta|\underset{\sim}{x}_k)d\theta - \int_{x_0-h}^{x_0+h} |w(\theta - x_0)|\, \pi(\theta|\underset{\sim}{x}_n)d\theta \right|$$

$$\leq \epsilon_0$$

$$+ \int_{-\infty}^{x_0-h} |w(\theta - x_0)|\, \pi(\theta|\underset{\sim}{x}_k) d\theta + \int_{x_0+h}^{\infty} |w(\theta - x_0)|\, \pi(\theta|\underset{\sim}{x}_k) d\theta$$

$$+ \int_{x_0-h}^{x_0+h} |w(\theta - x_0)|\, |\pi(\theta|\underset{\sim}{x}_n) - \pi(\theta|\underset{\sim}{x}_k)|\, d\theta$$

(Lemma 11 is used since $A_0(\epsilon) \geq A_8(\epsilon_0)$)

$\leq \epsilon_0 + \epsilon_0 + \epsilon_0 + \epsilon_0$

(Lemma 10 and the preceding result are used)

$= 4\epsilon_0.$

Then we showed that $\left|\mathbb{E}^{\pi(\theta|\underset{\sim}{x}_n)}[w(\theta - x_0)] - \mathbb{E}^{\pi(\theta|\underset{\sim}{x}_k)}[w(\theta - x_0)]\right| \leq 7\epsilon_0 = \epsilon.$

2.6.7. Proof of Lemma 11

We want to show that

$$\lim_{\underset{\sim}{\phi}_2 \to \infty} \mathbb{E}^{\pi(\theta|\underset{\sim}{x}_n)}[|w(\theta - x_0)|] = \mathbb{E}^{\pi(\theta|\underset{\sim}{x}_k)}[|w(\theta - x_0)|],$$

or equivalently

$$\lim_{\underset{\sim}{\phi}_2 \to \infty} \mathbb{E}^{\pi(\theta|\underset{\sim}{x}_n)}[|w(\theta - x_0)| + 1] = \mathbb{E}^{\pi(\theta|\underset{\sim}{x}_k)}[|w(\theta - x_0)| + 1].$$

Define the density π^* as $\pi^*(\theta) = \frac{(|w(\theta)|+1)\pi_\theta(\theta)}{\int_{-\infty}^{\infty}(|w(\theta)|+1)\pi_\theta(\theta)d\theta}$, or considering the location parameter x_0, as $\pi^*(\theta - x_0) = \frac{(|w(\theta-x_0)|+1)\pi_\theta(\theta-x_0)}{\int_{-\infty}^{\infty}(|w(\theta-x_0)|+1)\pi_\theta(\theta-x_0)d\theta}$. Since the prior density π_θ is proper and $\mathbb{E}^{\pi_\theta(\theta)}[|w(\theta)|] < \infty$, the denominator is finite and π^* is a proper density. The density π^* is also positive everywhere and bounded above, since the numerator is positive and the prior density π_θ and the function $|w(\theta)|\,\pi_\theta(\theta)$ are bounded above, as assumed in Theorem 2.

It is then possible to use the result a) of Theorem 2 using the density π^* as the prior instead of π_θ. If the conditions given by equations (2.3.1) and (2.3.2) are used with π^* instead of π_θ, they are equivalent to the conditions given by equations (2.3.1) to (2.3.4) using π_θ. Result a) using π_θ as the prior is equivalent to

$$\lim_{\underset{\sim}{\phi}_2 \to \infty} \frac{m(\underset{\sim}{x}_k) \prod_{i=k+1}^{n} f_i(x_i - x_0)}{m(\underset{\sim}{x}_n)} = 1,$$

and result a) using π^* as the prior is equivalent to

$$\lim_{\phi_2 \to \infty} \frac{\left(\prod_{i=k+1}^{n} f_i(x_i - x_0)\right) \int_{-\infty}^{\infty} \pi^*(\theta - x_0) \prod_{i=1}^{k} f_i(x_i - \theta) d\theta}{\int_{-\infty}^{\infty} \pi^*(\theta - x_0) \prod_{i=1}^{n} f_i(x_i - \theta) d\theta} = 1$$

$$\Leftrightarrow \lim_{\phi_2 \to \infty}$$

$$\frac{\left(\prod_{i=k+1}^{n} f_i(x_i - x_0)\right) \int_{-\infty}^{\infty} (|w(\theta - x_0)| + 1) \pi_\theta(\theta - x_0) \prod_{i=1}^{k} f_i(x_i - \theta) d\theta}{\int_{-\infty}^{\infty} (|w(\theta - x_0)| + 1) \pi_\theta(\theta - x_0) \prod_{i=1}^{n} f_i(x_i - \theta) d\theta} = 1.$$

The result can now be shown.

$$\begin{aligned}
&\mathbb{E}^{\pi(\theta|\underset{\sim}{x}_n)}[|w(\theta - x_0)| + 1] \\
&= \frac{\int_{-\infty}^{\infty} (|w(\theta - x_0)| + 1) \pi_\theta(\theta - x_0) \prod_{i=1}^{n} f_i(x_i - \theta) d\theta}{m(\underset{\sim}{x}_n)} \\
&= \frac{\int_{-\infty}^{\infty} (|w(\theta - x_0)| + 1) \pi_\theta(\theta - x_0) \prod_{i=1}^{n} f_i(x_i - \theta) d\theta}{\left(\prod_{i=k+1}^{n} f_i(x_i - x_0)\right) \int_{-\infty}^{\infty} (|w(\theta - x_0)| + 1) \pi_\theta(\theta - x_0) \prod_{i=1}^{k} f_i(x_i - \theta) d\theta} \\
&\times \frac{m(\underset{\sim}{x}_k) \prod_{i=k+1}^{n} f_i(x_i - x_0)}{m(\underset{\sim}{x}_n)} \\
&\times \frac{\int_{-\infty}^{\infty} (|w(\theta - x_0)| + 1) \pi_\theta(\theta - x_0) \prod_{i=1}^{k} f_i(x_i - \theta) d\theta}{m(\underset{\sim}{x}_k)}.
\end{aligned}$$

If the limit as $\underset{\sim}{\phi}_2 \to \infty$ is taken, the first two terms in the last expression are 1 according to result a) using respectively π^* and π_θ as the prior. The last term is $\mathbb{E}^{\pi(\theta|\underset{\sim}{x}_k)}[|w(\theta - x_0)| + 1]$, which prove the result.

2.6.8. Proof of Theorem 1

Since Theorem 1 is an application of Theorem 2, it is sufficient to show that if conditions of the former are satisfied, then conditions of the latter are also satisfied. The context is the same for both theorems, and in particular, the densities $\pi_\theta, f_1, ..., f_n$ are assumed to be proper, positive everywhere and bounded above in both cases. The conditions needed for results a) to d) in Theorem 1 are p-cred$^-(f_i)$ = p-cred$^+(f_i) = (\gamma', \delta', \alpha', \beta')$, $i = 1, ..., n$, p-cred$^-(\pi_\theta)$ = p-cred$^+(\pi_\theta) = (\gamma, \delta, \alpha, \beta)$ and

i) $\gamma' < 1$, $k \geq n/2$ or

ii) $\gamma' < 1$, $k < n/2$, $(\gamma, \delta, \alpha, \beta) > (\gamma', \delta'(n - 2k), \alpha'(n - 2k), \beta'(n - 2k))$.

The asymptotic behavior for the left and right tails is the same for each density since their left and right p-credences are the same. Furthermore, the behavior is the same for all densities f_i, $i = 1, ..., n$. Then the conditions needed for results a) to d) in Theorem 2 can be simplified as follows. Conditions C1 to C3 must be satisfied on the right tail of f_n and equations (2.3.1) and (2.3.2) must be satisfied. To simplify the notation, f_n is denoted as f.

First we show that condition C1 is satisfied on the right tail of f, which can be written as follows. For any constant $h > 0$ and for all θ such that $|\theta| \leq h$, $\lim_{z\to\infty} \frac{f(z+\theta)}{f(z)} = 1$. Furthermore, since p-cred$^+(f) = (\gamma', \delta', \alpha', \beta')$, there exists a constant $K_1 > 0$ such that $\lim_{z\to\infty} \frac{f(z)}{\mathrm{e}^{-\delta' z^{\gamma'}} z^{-\alpha'} \log^{-\beta'} z} = K_1$. If $h > 0$ and $|\theta| \leq h$, then

$$
\begin{aligned}
\lim_{z\to\infty} \frac{f(z+\theta)}{f(z)} &= \lim_{z\to\infty} \frac{f(z+\theta)}{\mathrm{e}^{-\delta'(z+\theta)^{\gamma'}} (z+\theta)^{-\alpha'} \log^{-\beta'}(z+\theta)} \\
&\qquad \times \frac{\mathrm{e}^{-\delta' z^{\gamma'}} z^{-\alpha'} \log^{-\beta'} z}{f(z)} \\
&\qquad \times \frac{\mathrm{e}^{-\delta'(z+\theta)^{\gamma'}} (z+\theta)^{-\alpha'} \log^{-\beta'}(z+\theta)}{\mathrm{e}^{-\delta' z^{\gamma'}} z^{-\alpha'} \log^{-\beta'} z} \\
&= \lim_{z\to\infty} \frac{K_1}{K_1} \frac{\mathrm{e}^{-\delta'(z+\theta)^{\gamma'}} (z+\theta)^{-\alpha'} \log^{-\beta'}(z+\theta)}{\mathrm{e}^{-\delta' z^{\gamma'}} z^{-\alpha'} \log^{-\beta'} z} \\
&= \lim_{z\to\infty} \mathrm{e}^{-\delta'((z+\theta)^{\gamma'} - z^{\gamma'})} \left(\frac{z}{z+\theta}\right)^{\alpha'} \left(\frac{\log z}{\log(z+\theta)}\right)^{\beta'} \\
&= \lim_{z\to\infty} \mathrm{e}^{-\delta'((z+\theta)^{\gamma'} - z^{\gamma'})}.
\end{aligned}
$$

It is easy to check in the last equality that the last two terms tend to 1 as z tends to infinity. Furthermore, using the Taylor series development of $(z+\theta)^{\gamma'} - z^{\gamma'}$, it can be shown that the last expression tends to 1 as z tends to infinity if and only if $\gamma' < 1$, which is a condition of Theorem 1.

Now we show that conditions C2 and C3 are satisfied on the right tail of f, which can be written as follows. There exist constants A_2 and $M_2 > 1$ and a proper density g such that for all $z > A_2$,

C2 : $\frac{f^2(z/2)}{f(z)g(z/2)} \leq M_2$,

C3 : $\frac{d^2}{dz^2} \log f^*(z) \geq \frac{d^2}{dz^2} \log g(z) \geq 0$, where f^* is f or any other proper densities which satisfy $\frac{1}{M_2} \leq \frac{f(z)}{f^*(z)} \leq M_2$ for all $z > A_2$.

Define $f^*(z) = p(z|\gamma', \delta', \alpha', \beta', z_0)$, with any $1 < z_0 < A_2$, that is a GEP density with the same left and right p-credences as $f(z)$. The tails behavior of f and f^* are the same and both are proper densities. Define

$$g(z) = \begin{cases} (1+|z|)^{-3}; & \text{if } \gamma' > 0, \delta' > 0, \\ f^*(z); & \text{if } \gamma' = 0, \delta' = 0. \end{cases}$$

The density g is proper since $(1+|z|)^{-3}$ and f^* are also proper densities.

Consider the first case, when $0 < \gamma' < 1, \delta' > 0$ and $g(z) = (1+|z|)^{-3}$. If the normalizing constant of f^* is denoted by K_2, then $z > A_2 \Rightarrow$

$$\begin{aligned} \frac{f^2(z/2)}{f(z)g(z/2)} &\leq M_2^3 \frac{f^{*2}(z/2)}{f^*(z)g(z/2)} \\ &= M_2^3 K_2 \frac{\left(\mathrm{e}^{-\delta'(z/2)^{\gamma'}} (z/2)^{-\alpha'} \log^{-\beta'}(z/2)\right)^2}{\left(\mathrm{e}^{-\delta' z^{\gamma'}} z^{-\alpha'} \log^{-\beta'} z\right) (1+z/2)^{-3}} \\ &= M_2^3 K_2 \,\mathrm{e}^{-\delta'(2^{1-\gamma'}-1)z^{\gamma'}} (z/4)^{-\alpha'} \left(\frac{\log^2(z/2)}{\log z}\right)^{-\beta'} (1+z/2)^3, \end{aligned}$$

and $\lim_{z\to\infty} M_2^3 K_2 \,\mathrm{e}^{-\delta'(2^{1-\gamma'}-1)z^{\gamma'}} (z/4)^{-\alpha'} \left(\frac{\log^2(z/2)}{\log z}\right)^{-\beta'} (1+z/2)^3 = 0$ since the dominant term is the exponential one and it tends to 0 as $z \to \infty$ since $\gamma' > 0, \delta' > 0$ and $2^{1-\gamma'} - 1 > 0 \Leftrightarrow \gamma' < 1$. It is sufficient to show that condition C2 is satisfied since the last expression is decreasing for $z > A_2$ if A_2 is chosen large enough, which means that $\frac{f^2(z/2)}{f(z)g(z/2)}$ is bounded by a constant.

Furthermore, if $z > z_0$ (which is the case if $z > A_2$ since $z_0 < A_2$), it can be shown that

$$\begin{aligned} \frac{d^2}{dz^2} \log f^*(z) &= \frac{d^2}{dz^2} \log \left[\mathrm{e}^{-\delta' z^{\gamma'}} z^{-\alpha'} \log^{-\beta'} z\right] \\ &= \frac{d^2}{dz^2} \left[-\delta' z^{\gamma'} - \alpha' \log z - \beta' \log(\log z)\right] \\ &= \gamma'(1-\gamma')\delta' z^{\gamma'-2} + \frac{\alpha'}{z^2} + \frac{\beta'(\log z + 1)}{z^2 \log^2 z} \\ &= \frac{1}{z^2}\left[\gamma'(1-\gamma')\delta' z^{\gamma'} + \alpha' + \frac{\beta'}{\log z} + \frac{\beta'}{\log^2 z}\right]. \end{aligned}$$

Furthermore, $\frac{d^2}{dz^2} \log g(z) = \frac{3}{(1+|z|)^2} > 0$ for any values of z. Finally, if $z > z_0$,

$$\frac{d^2}{dz^2} \log f^*(z) - \frac{d^2}{dz^2} \log g(z)$$
$$= \frac{1}{z^2}\left[\gamma'(1-\gamma')\delta' z^{\gamma'} + \alpha' + \frac{\beta'}{\log z} + \frac{\beta'}{\log^2 z} - \frac{3z^2}{(1+z)^2}\right].$$

The term in brackets goes to $+\infty$ as $z \longrightarrow \infty$ if $\gamma'(1-\gamma')\delta' > 0$, that is if $0 < \gamma' < 1$ and $\delta' > 0$, which show that $\frac{d^2}{dz^2} \log f^*(z) - \frac{d^2}{dz^2} \log g(z) \geq 0$ if z is large enough. Then conditions C2 and C3 are satisfied if $\gamma' > 0, \delta' > 0$.

Consider now the second case, when $\gamma' = 0, \delta' = 0$ and $g(z) = f^*(z)$. Then $z > A_2 \Rightarrow$

$$\begin{aligned}
\frac{f^2(z/2)}{f(z)g(z/2)} &\leq M_2^3 \frac{f^{*2}(z/2)}{f^*(z)f^*(z/2)} \\
&= M_2^3 \frac{f^*(z/2)}{f^*(z)} \\
&= M_2^3 \frac{(z/2)^{-\alpha'} \log^{-\beta'}(z/2)}{z^{-\alpha'} \log^{-\beta'} z} \\
&= M_2^3 2^{\alpha'} \left(\frac{\log(z/2)}{\log z}\right)^{-\beta'},
\end{aligned}$$

and $\lim_{z\to\infty} M_2^3 2^{\alpha'} \left(\frac{\log(z/2)}{\log z}\right)^{-\beta'} = M_2^3 2^{\alpha'}$ It is sufficient to show that condition C2 is satisfied since the last expression is bounded above for $z > A_2$ if A_2 is chosen large enough, which means that $\frac{f^2(z/2)}{f(z)g(z/2)}$ is bounded by a constant.

Furthermore, if $z > z_0$, it can be shown that

$$\frac{d^2}{dz^2} \log f^*(z) = \frac{1}{z^2}\left[\alpha' + \frac{\beta'}{\log z} + \frac{\beta'}{\log^2 z}\right].$$

The term in brackets converge to α' as $z \longrightarrow \infty$. Since f^* is a proper density and $\gamma' = \delta' = 0$, it follows that $\alpha' \geq 1$, which show that $\frac{d^2}{dz^2} \log f^*(z) = \frac{d^2}{dz^2} \log g(z) \geq 0$ if z is large enough. Then conditions C2 and C3 are also satisfied if $\gamma' = 0, \delta' = 0$.

Consider now equations (2.3.1) and (2.3.2) in Theorem 2. Assuming that conditions in Theorem 1 are satisfied, equation (2.3.2) can be rewritten as follows, if $m < n$.

$$\lim_{\theta\to\infty} \frac{\pi_\theta(\theta)\prod_{i=1}^k f_i(x_i - x_0 - \theta)}{\prod_{i=m+1}^n f_i(\theta)} = 0$$

$$\Leftrightarrow \lim_{\theta\to\infty} \frac{\pi_\theta(\theta)\prod_{i=1}^{k} f_i(-\theta)}{\prod_{i=m+1}^{n} f_i(\theta)} \frac{\prod_{i=1}^{k} f_i(x_i - x_0 - \theta)}{\prod_{i=1}^{k} f_i(-\theta)} = 0$$

$$\Leftrightarrow \lim_{\theta\to\infty} \frac{\pi_\theta(\theta)\prod_{i=1}^{k} f_i(-\theta)}{\prod_{i=m+1}^{n} f_i(\theta)} = 0$$

(C1 is used since p-cred$^{-}(f_i) = (\gamma', \delta', \alpha', \beta'), i = 1, ..., n,$ and $\gamma' < 1$)

$$\Leftrightarrow \lim_{\theta\to\infty} \frac{p(\theta|\gamma,\delta,\alpha,\beta,z_0)p^k(\theta|\gamma',\delta',\alpha',\beta',z_0')}{p^{n-m}(\theta|\gamma',\delta',\alpha',\beta',z_0')} = 0$$

$$\Leftrightarrow \lim_{\theta\to\infty} \frac{p(\theta|\gamma,\delta,\alpha,\beta,z_0)}{p^{n-m-k}(\theta|\gamma',\delta',\alpha',\beta',z_0')} = 0, \qquad (2.6.2)$$

where $p(\theta|\gamma,\delta,\alpha,\beta,z_0)$ and $p(\theta|\gamma',\delta',\alpha',\beta',z_0')$ are GEP densities with respectively the same left and right p-credences as π_θ and f_i, for any $z_0 > 1$ and $z_0' > 1$. The symmetry of the GEP density about 0 is used in the last equations. In the same way, equation (2.3.1) can be rewritten as follows, if $k < m$.

$$\lim_{\theta\to-\infty} \frac{\pi_\theta(\theta)\prod_{i=1}^{k} f_i(x_i - x_0 - \theta)}{\prod_{i=k+1}^{m} f_i(\theta)} = 0$$

$$\Leftrightarrow \lim_{\theta\to\infty} \frac{p(\theta|\gamma,\delta,\alpha,\beta,z_0)p^k(\theta|\gamma',\delta',\alpha',\beta',z_0')}{p^{m-k}(\theta|\gamma',\delta',\alpha',\beta',z_0')} = 0$$

$$\Leftrightarrow \lim_{\theta\to\infty} \frac{p(\theta|\gamma,\delta,\alpha,\beta,z_0)}{p^{m-2k}(\theta|\gamma',\delta',\alpha',\beta',z_0')} = 0. \qquad (2.6.3)$$

As long as the number of observations around x_0 is larger or equal to the number of outliers on the left of x_0 (that is $k \geq m - k \Leftrightarrow m - 2k \leq 0$) and larger or equal to the number of outliers on the right of x_0 (that is $k \geq n - m \Leftrightarrow n - m - k \leq 0$), equations (2.3.1) and (2.3.2) are satisfied, whatever the left and right p-credences are, since the tails of a proper GEP density go to 0. This is equivalent to $k \geq \max[m - k, n - m]$. For instance, if $k = n/3$, $m - k = n/3$ and $n - m = n/3$, then $k \geq \max[m - k, n - m]$ is satisfied. In this case, the posterior can reject up to one third of the observations considered as left outliers plus another one third considered as right outliers, for a total of $\frac{2}{3}n$ outliers.

However Theorem 1 considers that the direction of outliers is unpredictable, so the results must hold even for the cases where all the $n - k$ outliers are on the right of x_0 ($m = k$) or on the left of x_0 ($m = n$). Equation (2.6.2) when $m = k$

and equation (2.6.3) when $m = n$ become one unique equation given by

$$\lim_{\theta\to\infty} \frac{p(\theta|\gamma,\delta,\alpha,\beta,z_0)}{p^{n-2k}(\theta|\gamma',\delta',\alpha',\beta',z_0')} = 0. \tag{2.6.4}$$

It can be shown that if equation (2.6.4) is satisfied, then equations (2.3.1) and (2.3.2) are also satisfied for any values of k and m such that $0 \le k \le m \le n$.

Consider now two cases : $k \ge n/2$ and $k < n/2$. Since $k \ge n/2 \Leftrightarrow n - 2k \le 0$, equation (2.6.4) is satisfied if $k \ge n/2$ for any p-credence, which corresponds to condition i) in Theorem 1.

If $k < n/2$, or equivalently $n - 2k > 0$, then according to Proposition 2, equation (2.6.4) is satisfied if

$$\begin{aligned}&\text{p-cred}^+(p(\theta|\gamma,\delta,\alpha,\beta,z_0)) > \text{p-cred}^+(p^{n-2k}(\theta|\gamma',\delta',\alpha',\beta',z_0'))\\ \Leftrightarrow\ &(\gamma,\delta,\alpha,\beta) > (\gamma',(n-2k)\delta',(n-2k)\alpha',(n-2k)\beta').\end{aligned}$$

The last equivalence is true since it can be shown that p-cred$^+(\prod_{i=1}^n f_i(\theta)) = (\gamma, \sum_{i=1}^n \delta_i, \sum_{i=1}^n \alpha_i, \sum_{i=1}^n \beta_i)$ if p-cred$^+(f_i) = (\gamma, \delta_i, \alpha_i, \beta_i)$, for $i = 1, ..., n$. It means that equation (2.6.4) is satisfied if $k < n/2$ and $(\gamma,\delta,\alpha,\beta) > (\gamma',(n-2k)\delta',(n-2k)\alpha',(n-2k)\beta')$, which corresponds to condition ii) in Theorem 1.

For result e) of Theorem 1, the function $w(\theta)$ of Theorem 2 is set to $w(\theta) = \theta^p$, where p is any positive integer. The condition $\mathbb{E}^{\pi_\theta(\theta)}[|w(\theta)|] < \infty$ in Theorem 2 is then satisfied since $\mathbb{E}^{\pi_\theta(\theta)}[|\theta|^p] < \infty$. Furthermore, the function $|w(\theta)|\,\pi_\theta(\theta)$ is bounded above since $|\theta|^p\,\pi_\theta(\theta)$ is bounded above.

Finally, it can be shown, in the same way as above, that if the equation given by

$$\lim_{\theta\to\infty} \frac{\theta^p p(\theta|\gamma,\delta,\alpha,\beta,z_0)}{p^{n-2k}(\theta|\gamma',\delta',\alpha',\beta',z_0')} = 0 \tag{2.6.5}$$

is satisfied, then conditions given by equations (2.3.3) and (2.3.4) in Theorem 2 are also satisfied for any values of k and m such that $0 \le k \le m \le n$. Consider again two cases : $k \ge n/2$ and $k < n/2$. If $k \ge n/2$ or equivalently $n-2k \le 0$, then equation (2.6.5) is satisfied for any p-credence since $\lim_{\theta\to\infty} \theta^p p(\theta|\gamma,\delta,\alpha,\beta,z_0) = 0$ using the condition $\mathbb{E}^{\pi_\theta(\theta)}[|\theta|^p] < \infty$.

If $k < n/2$, or equivalently $n - 2k > 0$, then according to Proposition 2, equation (2.6.5) is satisfied if

$$\text{p-cred}^+(\theta^p p(\theta|\gamma, \delta, \alpha, \beta, z_0)) > \text{p-cred}^+(p^{n-2k}(\theta|\gamma', \delta', \alpha', \beta', z_0'))$$
$$\Leftrightarrow (\gamma, \delta, \alpha - p, \beta) > (\gamma', (n-2k)\delta', (n-2k)\alpha', (n-2k)\beta').$$

It means that equation (2.6.5) is also satisfied if $k < n/2$ and $(\gamma, \delta, \alpha, \beta) > (\gamma', (n-2k)\delta', (n-2k)\alpha' + p, (n-2k)\beta')$, which corresponds to the condition needed for result e) in Theorem 1.

2.7. Acknowledgments

The authors are thankful to the NSERC (Natural Sciences and Engineering Research Council of Canada), the FQRNT (Le Fonds québécois de la recherche sur la nature et les technologies) and the SOA (Society of Actuaries) for their financial support.

2.8. References

ANGERS, J.-F. (1992) Use of Student-t prior for the estimation of normal means : A computational approach, *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, *IMS Lecture Notes - Monograph Series*, Vol. **29**, pp. 273-283.

ANGERS, J.-F. (2000) P-Credence and outliers, *Metron*, **58**, 81-108.

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

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.P. (1973) Posterior expectations for large observations, *Biometrika*, **60**, 664-667.

DE FINETTI, B. (1961) The Bayesian approach to the rejection of outliers, *Proceedings of the Fourth Berkeley Symposium on Probability and Statistics*, (**Vol. 1**), Berkeley : University of California Press, pp. 199-210.

DESGAGNÉ, A. and ANGERS, J.-F. (2003) *Computational aspect of the generalized exponential power density*, Technical report CRM-2918, Université de Montréal (http ://www.crm.umontreal.ca/pub/Rapports/2900-2999/2918.pdf).

DESGAGNÉ, A. et ANGERS, J.-F. (2005) Importance Sampling with the Generalized Exponential Power Density, *Statistics and Computing*, **15**, 189-195.

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.M. (1974) On coherence, inadmissibility and inference about many parameters in the theory of least squares, *Studies, Bayesian Econometrics and Statistics*, eds. S. E. Fienberg and A. Zellner, 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, *Bayesian Statistic III*, eds. Bernardo, J.M., DeGroot M.H., Lindley, D.V., 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.

Chapitre 3

OUTLIERS FOR SCALE PARAMETER INFERENCE AND POSITIVE OBSERVATIONS

Cet article sera soumis pour publication dans la revue *Insurance and Mathematics.* Le premier auteur est Alain Desgagné et le coauteur est le directeur de recherche Jean-François Angers. La contribution de Alain Desgagné à cet article consiste en la conception, recherche, développement, programmation et rédaction de toutes les parties de l'article, sous la supervision du directeur de recherche.

Abstract

The use of heavy-tailed distributions is a valuable tool in developing robust Bayesian procedures, limiting the influence of outliers on posterior inference. In this paper, the behavior of the posterior density of the scale parameter is investigated when the sample contains outliers and positive observations. The notion of left and right log-credences is introduced to characterize respectively the left and right tails of a density defined on $\mathbb{R}^+$. Simple conditions on the tails of the prior and the likelihood, using left and right log-credences, are established to determine the proportion of observations that can be rejected as outliers. It is shown that the posterior distribution converges in law to the posterior that would be obtained from the reduced sample, excluding the outliers, as they tend to 0 or infinity, at any given rate. An example of combination of the predictions of volatility of the S&P 500 index return is presented.

Key words : Bayesian inference, Outlier, Heavy-tailed modeling, Generalized exponential power family, Scale parameter.

3.1. Introduction

The use of heavy-tailed distributions is a valuable tool in developing robust Bayesian procedures, limiting the influence of outliers on posterior inference. Outlier rejection in Bayesian analysis has been studied for inference on the location parameter by several authors, see for instance De Finetti (1961), Dawid (1973), Hill (1974), O'Hagan (1979), O'Hagan (1988), Meinhold and Singpurwalla (1989), O'Hagan (1990), Angers and Berger (1991), Carlin and Polson (1991), Angers (1992), Fan and Berger (1992), Geweke (1994), Angers (1996), Angers (2000), Desgagné and Angers (2005b).

In Desgagné and Angers (2005b), outlier rejection for inference on the location parameter is studied and results of convergence of the posterior are given for any number of observations and outliers, where the outliers tend to plus or minus infinity at any given rate. Conditions on the prior and the likelihood are given in order to determine the number of outliers that the posterior can eventually reject.

In this paper, outlier rejection for inference on the scale parameter with positive observations is studied in the same context given in Desgagné and Angers (2005b). The use of heavy-tailed distributions is still a condition of rejection of outliers. However, the tails must be even heavier in inference on the scale parameter.

In Section 3.2, the relation between outliers rejection and log-credence is considered. In Section 3.2.1, the notion of left and right log-credences is introduced in order to characterize and order the tails of a density defined on $\mathbb{R}^+$. This measure is similar to p-credence (see Desgagné and Angers, 2005b), but the accent is put on heavier tailed distributions, such as polynomial, logarithmic and logarithmic of the logarithmic tails behavior.

In Section 3.2.2, the behavior of the posterior density of the scale parameter is investigated when the sample contains outliers. Simple conditions on the tails of the prior and the likelihood, using left and right log-credences, are established to determine the proportion of observations that can be rejected as outliers. We show that the posterior distribution converges in law to the posterior that would be obtained from the reduced sample, excluding the outliers, as they tend to 0 or ∞ at any given rate.

In Section 3.3, the same results of convergence are given when we specify which ones of the observations are considered as outliers and if they are large or small outliers. The conditions on the tails of the prior and the likelihood are generalized to encompass densities excluded by left and right log-credences. The influence of the left and right tails of each observation's density and the prior is established clearly in these conditions.

In Section 3.4, an example of combination of predictions of the volatility of the S&P 500 index return is given. While the results given in this paper have a strong theorical component, it is shown in this example that they also have useful and easy applications in real context.

3.2. Outliers rejection using log-credence

In this section, conditions on prior and likelihood are established using log-credence to obtain robust Bayesian inference on the scale parameter. The influence of the outliers on the posterior density is expected to decrease when the outliers become extreme.

In Section 3.2.1, log-credence is defined to characterize the tails of a density defined on $\mathbb{R}^+$. In Section 3.2.2, conditions are presented to obtain convergence of the posterior density based on all observations to the posterior excluding the outliers, as the outliers tend to 0 or ∞.

3.2.1. A measure of the tails : left and right log-credences

Conditions of robustness concern mainly the tails of the prior and observations' densities. For densities defined on $\mathbb{R}$, Desgagné and Angers (2005b) defined the left and right p-credences to characterize their left and right tails. These measures compare the tails of a density to those of the GEP (generalized exponential power) density with parameters $(\gamma, \delta, \alpha, \beta)$. If $f(z)$ is the density of a random variable Z, then right p-credence is denoted by p-cred$^+(f)$ or p-cred$^+(Z)$ and is defined as follows.

Definition 6. *A density f has right p-credence $(\gamma, \delta, \alpha, \beta)$ if there exists a constant $K > 0$ such that*

$$\lim_{z \to \infty} \frac{f(z)}{e^{-\delta|z|^{\gamma}} |z|^{-\alpha} \log^{-\beta} |z|} = K.$$

The definition of left p-credence is identical, except $\lim_{z\to\infty}$ is replaced by $\lim_{z\to-\infty}$. It is denoted by p-cred$^-(f)$ or p-cred$^-(Z)$.

We introduce a similar measure for the tails of a density defined on $\mathbb{R}^+$, called left and right log-credences. If $f(z)$ is the density of a random variable Z defined on $\mathbb{R}^+$, then right log-credence is denoted by log-cred$^+(f)$ or log-cred$^+(Z)$ and is defined as follows.

Definition 7. *A density f has right log-credence $(\gamma, \delta, \alpha, \beta)$ if there exists a constant $K > 0$ such that*

$$\lim_{z \to \infty} \frac{f(z)}{\frac{1}{z} e^{-\delta|\log z|^{\gamma}} |\log z|^{-\alpha} \log^{-\beta} |\log z|} = K.$$

The definition of left log-credence is identical, except $\lim_{z\to\infty}$ is replaced by $\lim_{z\to 0}$. Once right (or left) log-credence of two densities have been determined, a dominance relation can be established to compare and order their tails, as described in Proposition 3.

Proposition 3. *Let f and g be two densities defined on $\mathbb{R}^+$ such that* log-cred$^+(f) = (\gamma, \delta, \alpha, \beta)$ *and* log-cred$^+(g) = (\gamma', \delta', \alpha', \beta')$.

i) *If $\gamma' = \gamma, \delta' = \delta, \alpha' = \alpha, \beta' = \beta$, then the right log-credences of f and g are equal, which is denoted by $(\gamma', \delta', \alpha', \beta') = (\gamma, \delta, \alpha, \beta)$. Their right tails are equivalent, which means that $\lim_{z\to\infty} \frac{f(z)}{g(z)} = k$ for a positive constant k.*

ii) *The right log-credence of g is smaller than that of f, which is denoted by $(\gamma', \delta', \alpha', \beta') < (\gamma, \delta, \alpha, \beta)$, if*

a) $\gamma' < \gamma$,

b) $\gamma' = \gamma, \delta' < \delta$,

c) $\gamma' = \gamma, \delta' = \delta, \alpha' < \alpha$,

d) or $\gamma' = \gamma, \delta' = \delta, \alpha' = \alpha, \beta' < \beta$.

The right tail of g strictly dominates the right tail of f, which means that $\lim_{z\to\infty} \frac{f(z)}{g(z)} = 0$.

The left tails of two densities are compared and ordered in a similar way using left log-credence, replacing $\lim_{z\to\infty}$ by $\lim_{z\to 0}$. The left tail of the density with the smallest left p-credence dominates the left tail of the other density. Note that log-credence and p-credence are ordered exactly in the same way and their proofs are similar, see Angers (2000).

The right (or left) log-credence of a density is $(\gamma, \delta, \alpha, \beta)$ if its right (or left) tail has the same behavior as a log GEP density with parameters $(\gamma, \delta, \alpha, \beta)$. We say that a random variable Z has a log GEP density if $X = \log Z$ has a GEP density. Therefore, if $p(x|\gamma, \delta, \alpha, \beta, z_0)$ is the density of a GEP random variable X, then $q(z|\gamma, \delta, \alpha, \beta, z_0) = \frac{1}{z} p(\log z|\gamma, \delta, \alpha, \beta, z_0)$ is the density of a log GEP random variable $Z = \mathrm{e}^X$, that is

$$q(z|\gamma, \delta, \alpha, \beta, z_0) \propto \frac{1}{z}\,\mathrm{e}^{-\delta \max(|\log z|, z_0)^\gamma} \max(|\log z|\,, z_0)^{-\alpha} \log^{-\beta} \max(|\log z|\,, z_0).$$

The interpretation of the behavior of the left tail of a density defined on $\mathbb{R}^+$ is a little different since the domain is bounded by 0. At 0, three cases are possible, that is $\lim_{z\to 0} f(z)$ is either 0, a positive constant K or ∞. For any density f with log-cred⁻$(f) = (\gamma, \delta, \alpha, \beta)$, it can be shown, using Proposition 3, that

i) $\lim_{z\to 0} f(z) = 0 \Leftrightarrow (\gamma, \delta, \alpha, \beta) > (1, 1, 0, 0)$,

ii) $\exists\, 0 < K < \infty$ such that $\lim_{z\to 0} f(z) = K \Leftrightarrow (\gamma, \delta, \alpha, \beta) = (1, 1, 0, 0)$,

iii) $\lim_{z\to 0} f(z) = \infty \Leftrightarrow (\gamma, \delta, \alpha, \beta) < (1, 1, 0, 0)$.

As it is shown in the next section, $\gamma < 1$ is required to obtain robust inference, which means that only densities with left tail going to infinity will provide robust inference, at least for the rejection of small outliers. Note that even if the left tail is going to infinity, the density can be proper. The gamma density with a shape parameter smaller than 1 is a well-known example.

The log-credence and p-credence are both defined for polynomial and logarithmic behavior. The polynomial term for the right log-credence is $z^{-(\delta+1)}$ when $\gamma = 1$ and $z > 1$ and for the left log-credence it is $z^{\delta-1}$ when $\gamma = 1$ and $z < 1$. Exponential behavior proportional to $\mathrm{e}^{-\delta|z|^\gamma}$ is considered in p-credence while exponential behavior proportional to $\mathrm{e}^{-\delta|\log z|^\gamma}$ is considered in log-credence. As it is shown in the next section, it is not useful to consider behavior of the type $e^{-\delta|z|^\gamma}$ in log-credence for outliers rejection in scale parameter inference, since this type of tails is not sufficiently heavy. A term of logarithm of logarithm in log-credence makes it possible to consider very heavy tails distributions.

As example, the right log-credence of a gamma density, given by $f(z) \propto z^{\delta-1}\,e^{-\lambda z}, \lambda > 0, \delta > 0, z > 0$, is not defined since its right tail is not sufficiently heavy. However its left log-credence is defined and is given by $(1, \delta, 0, 0)$. As mentionned above, its left tail will go to infinity, a positive constant or 0 depending on whether $\delta < 1$, $\delta = 1$ or $\delta > 1$. Note that the right log-credence of an inverse gamma distribution is the left log-credence of the gamma distribution, and its left log-credence is not defined. The log normal density, given by $f(z) \propto \frac{1}{z}\,e^{-0.5(\log z)^2}, z > 0$, has left and right log-credences given by $(2, 0.5, 0, 0)$ and its left tail goes to 0. The shifted Pareto density, given by $f(z) \propto (z_0 + z)^{-a}, z_0 > 0, a > 1, z > 0$, has a right log-credence given by $(1, a-1, 0, 0)$ and its left log-credence is given by $(1, 1, 0, 0)$, since $\lim_{z\to 0} f(z)$ is a positive constant. The shifted log-Pareto density, given by $f(z) \propto \frac{1}{z_0+z}\log^{-b}(z_0+z), b > 1, z_0 > 1, z > 0$, has a right log-credence given by $(0, 0, b, 0)$ and its left log-credence is given by $(1, 1, 0, 0)$. Note that distributions with log-credence given by $(1, \delta, 0, 0)$, $\delta > 0$, are called heavy-tailed distributions and those with log-credence given by $(0, 0, \alpha, 0)$, $\alpha > 1$, are called super heavy-tailed distributions (see Reiss and Thomas, 1997). Finally, the most heavy tails are given by distributions with log-credence of $(0, 0, 1, \beta)$, $\beta > 1$.

3.2.2. Outliers rejection using left and right log-credences

In this section, the behavior of the posterior density of the scale parameter is investigated when the sample contains outliers, that is when there is a conflict between some extreme observations and the information provided by the prior and the other observations. Note that extreme values in this context may mean large, but also small positive values relative to the others, that is $\max(z, 1/z)$ is large. Using left and right log-credences, conditions on the tails of the prior density and the likelihood are established in order to obtain robust posterior inference. The influence of the outliers decreases as they become more extreme and eventually the outliers are rejected. Consider the following Bayesian context.

i) Let $X_1, ..., X_n$ be n random variables conditionally independent given the scale parameter σ with the conditional densities of $X_i|\sigma$ given by $\frac{1}{\sigma} f_i(\frac{x_i}{\sigma})$, where $X_i \in \mathbb{R}^+, \sigma \in \mathbb{R}^+$, $i = 1, ..., n$.

ii) The prior density of σ is $\frac{1}{x_0}\pi_\sigma(\frac{\sigma}{x_0})$, where $x_0 \in \mathbb{R}^+$ is a known scale parameter.

The densities $\pi_\sigma, f_1, ..., f_n$ are assumed to be proper and positive everywhere. Furthermore, the functions $z\pi_\sigma(z)$ and $zf_i(z), i = 1, ..., n$ are assumed bounded above for all $z > 0$. We assume that the sample consists of a group of k observations and of $n-k$ outliers, $0 \leq k \leq n$. Assume without loss of generality that the outliers are the observations denoted by $x_{k+1}, ..., x_n$. The ratios $\frac{x_1}{x_0}, ..., \frac{x_k}{x_0}$ are assumed fixed. The ratios of each outlier $x_{k+1}, ..., x_n$ and any values of the group composed of $x_0, x_1, ..., x_k$ tend to 0 or ∞, i.e. $\max(\frac{x_{k+1}}{x_0}, \frac{x_0}{x_{k+1}}) \to \infty, ..., \max(\frac{x_n}{x_0}, \frac{x_0}{x_n}) \to \infty$. Generally, we consider that $x_0, x_1, ..., x_k$ are fixed and that the outliers tend to 0 or ∞. However, we could consider that $x_0, x_1, ..., x_k$ are not fixed, as long as they are moving as a group such that the ratios $\frac{x_i}{x_0}$ are fixed, $i = 1, ..., k$ and as long as their ratios with the outliers tend to 0 or ∞.

Let the posterior density of σ be denoted by $\pi(\sigma|\underset{\sim}{x}_n)$ if all n observations are considered, and denoted by $\pi(\sigma|\underset{\sim}{x}_k)$ if only non-outlier observations denoted by $x_1, ..., x_k$ are considered. Let also the marginal density of $\underset{\sim}{x}_n = (x_1, ..., x_n)$ be denoted by $m(\underset{\sim}{x}_n)$ and the marginal density of $\underset{\sim}{x}_k = (x_1, ..., x_k)$ be denoted by $m(\underset{\sim}{x}_k)$. Let the vector of the ratios of the outliers and x_0 be denoted by $\underset{\sim}{\phi}_1 = \left(\max(\frac{x_{k+1}}{x_0}, \frac{x_0}{x_{k+1}}), ..., \max(\frac{x_n}{x_0}, \frac{x_0}{x_n})\right)$. The notation $\underset{\sim}{\phi}_1 \to \infty$ means that each term of the vector tends to ∞ at any given rate. Finally, let the $(100 \times p)^{th}$ percentile of a density f be denoted by $Q^f(p)$, where $Q^f(p)$ is such that $\int_{-\infty}^{Q^f(p)} f(z)dz = p$.

Theorem 3. *Suppose that* log-cred$^-(f_i)$ = log-cred$^+(f_i) = (\gamma', \delta', \alpha', \beta')$, $i = 1, ..., n$ *and* log-cred$^-(\pi_\sigma)$ = log-cred$^+(\pi_\sigma) = (\gamma, \delta, \alpha, \beta)$. *For any integer k such that $0 \leq k \leq n$ and for any $x_0, x_1, ..., x_k$ such that $\frac{x_1}{x_0}, ..., \frac{x_k}{x_0}$ are fixed, if*

i) $\gamma' < 1$, $k \geq n/2$ *or*

ii) $\gamma' < 1$, $k < n/2$, $(\gamma, \delta, \alpha, \beta) > (\gamma', \delta'(n-2k), \alpha'(n-2k), \beta'(n-2k))$

then

a) $\lim_{\phi_1 \to \infty} \frac{\pi(x_0|\underline{x}_n)}{\pi(x_0|\underline{x}_k)} = 1$,

b) *for any* $h > 1$ *and for all* σ *such that* $\frac{1}{h} \leq \frac{\sigma}{x_0} \leq h$, $\lim_{\phi_1 \to \infty} \frac{\pi(\sigma|\underline{x}_n)}{\pi(\sigma|\underline{x}_k)} = 1$,

c) *for any* $h > 1$ *and* $j \in (k+1, ..., n)$, $\lim_{\phi_1 \to \infty} \Pr[\frac{1}{h} \leq \frac{\sigma}{x_j} \leq h|\underline{x}_n] = 0$,

d) $(\sigma/x_0)|\underline{x}_n \xrightarrow{\mathcal{L}} (\sigma/x_0)|\underline{x}_k$ *as* $\phi_1 \to \infty$, *where the density of the random variables* $(\sigma/x_0)|\underline{x}_n$ *and* $(\sigma/x_0)|\underline{x}_k$ *evaluated at the point* y *are given by* $x_0\pi(yx_0|\underline{x}_n)$ *and* $x_0\pi(yx_0|\underline{x}_k)$,

e) *for any* $0 < p < 1$, $\lim_{\phi_1 \to \infty} Q^{x_0\pi(yx_0|\underline{x}_n)}(p) = Q^{x_0\pi(yx_0|\underline{x}_k)}(p)$.

Proof. See the Appendix, Section 3.6.8.

Note that for each density $\pi_\sigma, f_1, ..., f_n$, the left and right tails have the same behavior. This condition is needed to obtain the same robustness, whether the outliers are on the left or on the right of x_0. Furthermore, log-credences of $f_1, ..., f_n$ are assumed to be identical. This condition ensures the same robustness against any extreme observations among $x_1, x_2, ..., x_n$.

The condition $\gamma' < 1$ ensures that left and right log-credences of $f_1, ..., f_n$ are sufficiently small, or equivalently that the density's tails of each potential outlier are sufficiently heavy. This also ensures that the posterior can reject up to $[n/2]$ outliers, where $[a]$ stands for the integer part of a. The conditions tell us that the posterior can reject more than $[n/2]$ outliers if the left and right log-credences of the prior are sufficiently large relatively to those of the likelihood. If one is not interested into putting too much confidence on the prior, then choosing $(\gamma, \delta, \alpha, \beta) \leq (\gamma', \delta', \alpha', \beta')$ ensures that the posterior can reject up to $[n/2]$ outliers, but not necessarily $[n/2] + 1$, if for instance all the outliers are on the right of x_0. At the other extreme, if one wants to put a large confidence in the prior, then choosing $(\gamma, \delta, \alpha, \beta) > (\gamma', n\delta', n\alpha', n\beta')$ ensures rejection of up to n outliers, even for the extreme cases where all the outliers are on the right of x_0, or all outliers are on the left of x_0. An intermediate choice of $(\gamma, \delta, \alpha, \beta)$ such that $(\gamma', \delta', \alpha', \beta') < (\gamma, \delta, \alpha, \beta) \leq (\gamma', n\delta', n\alpha', n\beta')$ gives intermediate results. Note that for specific directions of the outliers, it could be possible that the number of outliers rejected by the posterior is larger than the number suggested by Theorem 3. This case is addressed in Section 3.3 with Theorem 4.

Asymptotic behavior of the posterior density is established through results a) to d), as $\max(\frac{x_i}{x_0}, \frac{x_0}{x_i}) \to \infty$ at any given rate, for $i = k+1, ..., n$. Note that if x_0 is fixed, the limit can be rewritten as $\max(x_i, x_i^{-1}) \to \infty$, $i = k+1, ..., n$.

Result a) says that the influence of the outliers on the posterior density evaluated at x_0 is asymptotically null. Note that result a) can be rewritten as $\lim_{\underset{\sim}{\phi}_1 \to \infty} \frac{x_0^k m(\underset{\sim}{x}_k) \prod_{i=k+1}^{n} f_i(x_i/x_0)}{x_0^n m(\underset{\sim}{x}_n)} = 1$ since

$$\pi(x_0|\underset{\sim}{x}_n) = \frac{1}{m(\underset{\sim}{x}_n)} \frac{1}{x_0} \pi_\sigma(1) \prod_{i=1}^{n} \frac{1}{x_0} f_i(x_i/x_0)$$

and

$$\pi(x_0|\underset{\sim}{x}_k) = \frac{1}{m(\underset{\sim}{x}_k)} \frac{1}{x_0} \pi_\sigma(1) \prod_{i=1}^{k} \frac{1}{x_0} f_i(x_i/x_0).$$

It says that the asymptotic behaviors of $x_0^k m(\underset{\sim}{x}_k) \prod_{i=k+1}^{n} f_i(x_i/x_0)$ and $x_0^n m(\underset{\sim}{x}_n)$ are equivalent. (Note that only $x_0^n m(\underset{\sim}{x}_n)$ and $\prod_{i=k+1}^{n} f_i(x_i/x_0)$ depend on $\underset{\sim}{\phi}_1$, while $x_0^k m(\underset{\sim}{x}_k)$ depends only on the constants $x_1/x_0, ..., x_k/x_0$.) Result b) says that the influence of the outliers on the posterior density in any neighborhood bounded by any finite multiples of x_0 (or $x_1, ..., x_k$) is asymptotically null. Result c) says that the posterior density tends to 0 in a neighborhood bounded by any finite multiples of the outliers. Finally, the convergence in distribution of the posterior density to the posterior excluding outliers is established in result d), as $\underset{\sim}{\phi}_1 \to \infty$. An equivalent result is given by $\lim_{\underset{\sim}{\phi}_1 \to \infty} \Pr[\sigma/x_0 \leq d|\underset{\sim}{x}_n] = \Pr[\sigma/x_0 \leq d|\underset{\sim}{x}_k]$, for any $d > 0$ (see the Appendix, Section 3.6.5).

If the ratio of a given observation x_j $(j > k)$ and the center of $x_0, x_1, ..., x_k$ increases but remains smaller than a certain threshold, the influence of this observation on the posterior density usually increases. However, if this ratio increases beyond the threshold, the influence of the observation begins to decrease to eventually be null. The interpretation is similar if the ratio decreases to 0.

Finally, result e) says that the influence of the outliers on the $(100p)^{th}$ posterior percentile is asymptotically null. It can be shown that if conditions of Theorem 3 are satisfied, all the posterior moments of σ exist if $\gamma > 1$ but none of them exist if $\gamma < 1$, since the tails of the posterior are too heavy. However, setting $\gamma > 1$

makes the posterior converging to the prior and rejecting all the observations in case of conflict between the prior and all observations. If one does not want to put such large confidence in the prior, robust measures based on posterior percentiles can rather be used, such that the median or the interquartile distance.

3.3. Outliers rejection with general conditions

While conditions of Theorem 3 can be satisfied for a large class of densities using left and right log-credences, it is still possible to relax and to generalize these conditions such that they can be satisfied for any distributions. In this section, conditions in Theorem 3 are given without using left and right log-credences. Conditions are also relaxed such that the left and right tails of each density may have their own asymptotic behavior. It is now possible to consider any distributions, such that the gamma distribution for instance, where the right log-credence is not defined and the behavior of its left and right tails is different. Conditions are also given such that it is possible to specify which observation is an outlier and what is its direction. The new theorem with relaxed conditions is given in Theorem 4 in Section 3.3.2.

Even if conditions in Theorem 3 are a special case of the conditions in Theorem 4, they are still very general and useful in practice since it is easier to determine if the conditions are satisfied with the help of left and right log-credences.

However, Theorem 4 is interesting from a theorical point of view. For instance, it is possible to see the influence of each density's tail in the rejection of outliers. Since Theorem 3 is a special case of Theorem 4, it could be possible to extend the definition of left and right log-credences to include a larger class of densities, as long as the conditions of Theorem 4 are satisfied.

Some conditions of Theorem 4 are introduced in Section 3.3.1. These conditions concern mainly the thickness and regularity of the tails of a density defined on $\mathbb{R}^+$.

3.3.1. Conditions of thickness and regularity for the tails of a density defined on $\mathbb{R}^+$

The tails of the likelihood must satisfy certain conditions of thickness and regularity when robust inference is expected. In Theorem 3, these conditions are given using log-credence. They are more general in Theorem 4. (Note that the conditions are the same for the left and right tails, except for the support of the density which is given in parentheses for the left tail.)

Three conditions of thickness and regularity for the tails of a density $f(z)$ defined on $z > 0$ are given by conditions C1 to C3 as follows. The density f is assumed to be proper and positive everywhere for all $z > 0$. Furthermore, the function $zf(z)$ is assumed bounded above for all $z > 0$.

C1 : $\forall \epsilon > 0, \forall h > 1$, there exists a constant $A_1(\epsilon, h)$ such that $z > A_1(\epsilon, h)$ ($z < A_1^{-1}(\epsilon, h)$ for the left tail) and $\frac{1}{h} \leq \sigma \leq h \Rightarrow 1 - \epsilon \leq \frac{\sigma f(\sigma z)}{f(z)} \leq 1 + \epsilon$.

For conditions C2 and C3, there exist constants A_2 and $M_2 > 1$ and a proper density g such that for all $z > A_2$ ($z < A_2^{-1}$ for the left tail),

C2 : $\frac{f^2(\sqrt{z})}{\sqrt{z} f(z) g(\sqrt{z})} \leq M_2$,

C3 : $z^2 \frac{d^2}{dz^2} \log f^*(z) + z \frac{d}{dz} \log f^*(z) \geq z^2 \frac{d^2}{dz^2} \log g(z) + z \frac{d}{dz} \log g(z) \geq 0$,

where f^* is f or any other proper densities which satisfy $\frac{1}{M_2} \leq \frac{f(z)}{f^*(z)} \leq M_2$ for all $z > A_2$ ($z < A_2^{-1}$ for the left tail).

Condition C1 ensures that the tails are sufficiently heavy. For example, if $f(z)$ is the density of a log normal distribution, $\lim_{z\to\infty} \frac{2f(2z)}{f(z)} = 0$ and condition C1 is not satisfied. If $f(z)$ is the density of a shifted log-Pareto density on $\mathbb{R}^+$, given by $f(z) \propto \frac{1}{z_0+z} \log^{-b}(z_0 + z)$, $\lim_{z\to\infty} \frac{\sigma f(\sigma z)}{f(z)} = 1$, for any fixed $\sigma > 0$ and condition C1 is satisfied on the right tail. However condition C1 is not satisfied on its left tail. For conditions C2 and C3 on the right tail, the density

$$g(z) = \begin{cases} \frac{\epsilon}{2} z^{-1}(1 + |\log z|)^{-(1+\epsilon)}; & \text{if } \lim_{z\to\infty} f(z) z \, |\log z|^{1+\epsilon} = 0, \\ f^*(z); & \text{otherwise,} \end{cases}$$

is usually appropriate, for any choice of $\epsilon > 0$. The same density $g(z)$ is also usually appropriate when the left tail is considered, except $\lim_{z\to\infty}$ is replaced by $\lim_{z\to 0}$ in the first row.

The density f^* may be chosen as f or any other proper densities with a right (or left) tail of the same order. Condition C2 can be rewritten as $\frac{(\sqrt{z}f(\sqrt{z}))^2}{zf(z)\sqrt{z}g(\sqrt{z})} \leq M_2$. This condition ensures that the ratio of the function $zf(z)$ evaluated at $\sqrt{z}$ and at z is bounded, as z increases in the right tail or decreases in the left tail. This ensures that the tail of f is sufficiently heavy. Condition C3 ensures a certain smoothness in the tails of f^* and g. It can be shown that this condition ensures that the tails are convex.

3.3.2. Outlier rejection

Consider the Bayesian context given in Section 3.2.2.

i) Let $X_1, ..., X_n$ be n random variables conditionally independent given σ with the conditional densities of $X_i|\sigma$ given by $\frac{1}{\sigma}f_i(\frac{x_i}{\sigma})$, where $X_i \in \mathbb{R}^+, \sigma \in \mathbb{R}^+$, $i = 1, ..., n$.

ii) The prior density of σ is $\frac{1}{x_0}\pi_\sigma(\frac{\sigma}{x_0})$, where $x_0 \in \mathbb{R}^+$ is a known scale parameter.

The densities $\pi_\sigma, f_1, ..., f_n$ are assumed to be proper and positive everywhere. Furthermore, the functions $z\pi_\sigma(z)$ and $zf_i(z), i = 1, ..., n$ are assumed bounded above for all $z > 0$. We assume that the sample consists of a group of k observations such that the ratios $\frac{x_i}{x_0}$ are fixed, $m - k$ outliers on the left of x_0 and $n - m$ outliers on the right of x_0 $(0 \leq k \leq m \leq n)$. The ratios of the outliers over x_0 tend to 0 if they are on the left of x_0 and tend to ∞ if they on the right of x_0. Assume without loss of generality that the left outliers are the observations denoted by $x_{k+1}, ..., x_m$ and the right outliers are the observations denoted by $x_{m+1}, ..., x_n$. Finally, let the vector of the ratios of the outliers and x_0 be denoted by $\underset{\sim}{\phi_2} = \left((\frac{x_{k+1}}{x_0})^{-1}, ..., (\frac{x_m}{x_0})^{-1}, \frac{x_{m+1}}{x_0}, ..., \frac{x_n}{x_0}\right)$. The notation $\underset{\sim}{\phi_2} \to \infty$ means that each term of the vector tends to ∞ at any given rate.

Theorem 4. *For any integer k and m such that $0 \leq k \leq m \leq n$ and for any $x_0, x_1, ..., x_k$ such that $\frac{x_1}{x_0}, ..., \frac{x_k}{x_0}$ are fixed, if conditions C1 to C3 are satisfied on the left tails of $f_{k+1}, ..., f_m$ and on the right tails of $f_{m+1}, ..., f_n$, and if*

$$\lim_{\sigma \to 0} \frac{\sigma \pi(\sigma) \prod_{i=1}^{k} \frac{1}{\sigma} f_i(\frac{1}{\sigma} \frac{x_i}{x_0})}{\prod_{i=k+1}^{m} \sigma f_i(\sigma)} = 0 \text{ when } k < m, \tag{3.3.1}$$

and

$$\lim_{\sigma \to \infty} \frac{\sigma \pi(\sigma) \prod_{i=1}^{k} \frac{1}{\sigma} f_i(\frac{1}{\sigma} \frac{x_i}{x_0})}{\prod_{i=m+1}^{n} \sigma f_i(\sigma)} = 0 \text{ when } m < n, \tag{3.3.2}$$

then

a) $\lim_{\phi_2 \to \infty} \frac{\pi(x_0 | \underline{x}_n)}{\pi(x_0 | \underline{x}_k)} = 1$,

b) for any $h > 1$ and for all σ such that $\frac{1}{h} \leq \frac{\sigma}{x_0} \leq h$, $\lim_{\phi_2 \to \infty} \frac{\pi(\sigma | \underline{x}_n)}{\pi(\sigma | \underline{x}_k)} = 1$,

c) for any $h > 1$ and $j \in (k+1, ..., n)$, $\lim_{\phi_2 \to \infty} \Pr[\frac{1}{h} \leq \frac{\sigma}{x_j} \leq h | \underline{x}_n] = 0$,

d) $(\sigma/x_0)|\underline{x}_n \xrightarrow{\mathcal{L}} (\sigma/x_0)|\underline{x}_k$ *as $\phi_1 \to \infty$, where the density of the random variables $(\sigma/x_0)|\underline{x}_n$ and $(\sigma/x_0)|\underline{x}_k$ evaluated at the point y are given by $x_0 \pi(y x_0 | \underline{x}_n)$ and $x_0 \pi(y x_0 | \underline{x}_k)$.*

In addition, for any function $w(\cdot)$ on $\mathbb{R}^+$ such that $\mathbb{E}^{\pi_\sigma(\sigma)}[|w(\sigma)|] < \infty$ and $\sigma |w(\sigma)| \pi_\sigma(\sigma)$ is bounded above, if

$$\lim_{\sigma \to 0} \frac{\sigma w(\sigma) \pi(\sigma) \prod_{i=1}^{k} \frac{1}{\sigma} f_i(\frac{1}{\sigma} \frac{x_i}{x_0})}{\prod_{i=k+1}^{m} \sigma f_i(\sigma)} = 0 \text{ when } k < m, \tag{3.3.3}$$

and

$$\lim_{\sigma \to \infty} \frac{\sigma w(\sigma) \pi(\sigma) \prod_{i=1}^{k} \frac{1}{\sigma} f_i(\frac{1}{\sigma} \frac{x_i}{x_0})}{\prod_{i=m+1}^{n} \sigma f_i(\sigma)} = 0 \text{ when } m < n, \tag{3.3.4}$$

then

e) $\lim_{\phi_2 \to \infty} \mathbb{E}^{\pi(\sigma | \underline{x}_n)}[w(\frac{\sigma}{x_0})] = \mathbb{E}^{\pi(\sigma | \underline{x}_k)}[w(\frac{\sigma}{x_0})]$.

Proof. See the Appendix, Sections 3.6.1 to 3.6.7.

Note that if $k = 0$, $\prod_{i=1}^{k} \frac{1}{\sigma} f_i(\frac{1}{\sigma} \frac{x_i}{x_0})$ is set to 1. Conditions C1 to C3 ensure that the density's tail exhibits some smoothness and it is sufficiently heavy. Since the numerators in the conditions given by equations (3.3.1) and (3.3.2) are proportional to $\sigma x_0 \pi(\sigma x_0 | \underset{\sim}{x}_k)$, these two conditions can be interpreted as follows : the left tail of the function $\prod_{i=k+1}^{m} \sigma f_i(\sigma)$ is heavier than the left tail of $\sigma x_0 \pi(\sigma x_0 | \underset{\sim}{x}_k)$ and the right tail of $\prod_{i=m+1}^{n} \sigma f_i(\sigma)$ is heavier than the right tail of $\sigma x_0 \pi(\sigma x_0 | \underset{\sim}{x}_k)$. Note that the density evaluated at the point σ, given by $x_0 \pi(\sigma x_0 | \underset{\sim}{x}_k)$, depends only on the fixed ratios $x_1/x_0, ..., x_k/x_0$. There is no conditions on the right tail of the left outliers' densities or on the left tail of the right outliers' densities.

The results a) to d) are identical to those given in Theorem 3, except that $\lim_{\phi_1 \to \infty}$ is replaced by $\lim_{\underset{\sim}{\phi}_2 \to \infty}$. The results are valid for a specific direction of the outliers. Note that if x_0 is fixed, the limit can be rewritten as $x_{k+1} \to 0, ..., x_m \to 0, x_{m+1} \to \infty, ..., x_n \to \infty$.

Finally result e) establishes the convergence of the posterior expectation of any function such that the prior expectation of the absolute value of the function exists, if equations (3.3.3) and (3.3.4) are satisfied. The interpretation of these two equations is identical to that of equations (3.3.1) and (3.3.2), where the prior $\pi_\sigma(\sigma)$ has been replaced by a density proportional to $|w(\sigma)| \, \pi_\sigma(\sigma)$, which is proper since $\mathbb{E}^{\pi_\sigma(\sigma)}[|w(\sigma)|] < \infty$. Note that the function $\sigma \, |w(\sigma)| \, \pi_\sigma(\sigma)$ must be bounded above.

3.4. Example

3.4.1. Context

Combining forecasting is an effective method used by many practitioners, see for example Clement (1989), and Min and Zellner (1993). Suppose that a portfolio manager needs a prediction on the volatility of the monthly returns of the S&P 500 index for the next year, where the volatility is measured by the standard deviation of the next twelve monthly returns. He asks 5 experts for their prediction on the volatility as well as a 95% confidence interval on this prediction. The manager wants to combine this information with his prior beliefs using the Bayesian model described in Section 3.2.2. According to this setting, the manager chooses $\frac{1}{x_0}\pi_\sigma(\frac{\sigma}{x_0}) = \frac{1}{s_0 \sigma} T_3\left(\log[(\sigma/x_0)^{1/s_0}]\right)$ and $\frac{1}{\sigma} f_i(\frac{x_i}{\sigma}) = \frac{1}{s_i x_i} T_3\left(\log[(x_i/\sigma)^{1/s_i}]\right)$ for $i = 1, ..., 5$, where $T_3(\cdot)$ is a Student density with 3 degrees of freedom, $x_i > 0$,

$s_i > 0$ for $i = 0, ..., 5$, and $\sigma > 0$. As it is shown below, the parameter s_i is a measure of the confidence of the prediction x_i. In the next section, it is shown how these densities are constructed.

3.4.2. Exponential transformation of symmetric densities defined on $\mathbb{R}$

It is possible to create a new class of densities defined on $\mathbb{R}^+$ using the exponential transformation of symmetric densities defined on $\mathbb{R}$. The most well-known distribution of this type is the log normal distribution. In this example, the Student is used instead of the normal distribution. We start with a density g_1 defined on $\mathbb{R}$ and symmetric about 0, as the Student density for instance. If the density of Z_1 is $g_1(z_1)$, then the density of $Z_2 = (e^{Z_1})^s$ is $g_2(z_2) = \frac{1}{sz_2}g_1(\log z_2^{1/s})$, where $z_1 \in \mathbb{R}$, $z_2 > 0$ and $s > 0$. Since the density g_1 is such that $g_1(-z_1) = g_1(z_1)$ for all $z_1 \in \mathbb{R}$, it follows that the density g_2 is such that $z_2 g_2(z_2) = \frac{1}{z_2}g_2(\frac{1}{z_2})$ for all $z_2 > 0$. It is then easy to show that the median of Z_2 is 1 for any values of the parameter s. Finally, the density of $Z_3 = mZ_2$ is $g_3(z_3) = \frac{1}{sz_3}g_1(\log[(z_3/m)^{1/s}])$, where $z_3 > 0$, $s > 0$ and $m > 0$. The median of Z_3 is m for any values of the parameter s. In this example $g_1(\cdot) = T_3(\cdot)$, $g_2(z_2)$ corresponds to $\pi_\sigma(\sigma)$ and $g_3(z_3)$ to $\frac{1}{x_0}\pi_\sigma(\frac{\sigma}{x_0})$ if $z_2 = z_3 = \sigma$, $s = s_0$ and $m = x_0$ and $g_2(z_2)$ corresponds to $f_i(x_i)$ and $g_3(z_3)$ to $\frac{1}{\sigma}f_i(\frac{x_i}{\sigma})$ if $z_2 = z_3 = x_i$, $s = s_i$ and $m = \sigma$.

If $Q_Z(p)$ is the $(100 \times p)^{th}$ percentile of a random variable Z, such that $\Pr[Z \leq Q_Z(p)] = p$, then for any $0 \leq \alpha \leq 1$, $\Pr[Q_{Z_3}(\alpha/2) \leq Z_3 \leq Q_{Z_3}(1 - \alpha/2)] = 1 - \alpha$. Then it can be shown that the median of Z_3 is the geometric mean of the $(100 \times \alpha/2)^{th}$ and $(100 \times (1 - \alpha/2))^{th}$ percentiles of Z_3, that is $m = \sqrt{Q_{Z_3}(\alpha/2)Q_{Z_3}(1 - \alpha/2)}$ for any $0 \leq \alpha \leq 1$. It follows that $\Pr[Q_{Z_3}(\alpha/2) \leq Z_3 \leq m] = \Pr[m \leq Z_3 \leq Q_{Z_3}(1 - \alpha/2)]$. Furthermore,

$$\begin{aligned}
1 - \alpha &= \Pr[Q_{Z_3}(\alpha/2) \leq Z_3 \leq Q_{Z_3}(1 - \alpha/2)] \\
&= \Pr[Q_{Z_3}(\alpha/2) \leq m(e^{Z_1})^s \leq Q_{Z_3}(1 - \alpha/2)] \\
&= \Pr\left[\log[(m^{-1}Q_{Z_3}(\alpha/2))^{1/s}] \leq Z_1 \leq \log[(m^{-1}Q_{Z_3}(1 - \alpha/2))^{1/s}]\right] \\
&= \Pr\left[-\frac{1}{2s}\log\frac{Q_{Z_3}(1 - \alpha/2)}{Q_{Z_3}(\alpha/2)} \leq Z_1 \leq \frac{1}{2s}\log\frac{Q_{Z_3}(1 - \alpha/2)}{Q_{Z_3}(\alpha/2)}\right]
\end{aligned}$$

$$= \Pr[-Q_{Z_1}(1-\alpha/2) \leq Z_1 \leq Q_{Z_1}(1-\alpha/2)].$$

The fourth equality comes from $m = \sqrt{Q_{Z_3}(\alpha/2)Q_{Z_3}(1-\alpha/2)}$. Then for any given m, a and b such that $m = \sqrt{ab}$ and for any symmetric densities g_1, it is possible to choose the parameter s in order to have $1-\alpha$ of the mass of the density $g_3(z_3)$ lying between a and b if the equality $\frac{1}{2s}\log(b/a) = Q_{Z_1}(1-\alpha/2)$ is satisfied, or equivalently if $s = \frac{\log\sqrt{b/a}}{Q_{Z_1}(1-\alpha/2)}$.

3.4.3. Data

The collected data are given in Table 3.1. The predictions of the volatility for the prior and the experts are respectively corresponding to $x_0, x_1, ..., x_5$ and are given in the first row. The left and right bounds of the 95% confidence interval for each source of information are denoted by a_i and b_i $(i = 0, 1, ..., 5)$ and are given in the next two rows. Note that the bounds are around the predictions such that their geometric mean is equal to the prediction, that is $x_i = \sqrt{a_i b_i}$. In the last row, the parameters s_i are given for each source of information, such that $s_i = \frac{\log\sqrt{b_i/a_i}}{Q_{Z_1}(0.975)}$, where $Q_{Z_1}(0.975) = 3.182$ represents the 97.5^{th} percentile of a Student distribution with 3 degrees of freedom. Note that all numbers are expressed in percentage.

TAB. 3.1. Prior and experts' predictions and 95% confidence intervals

	Prior	**Experts**				
i=	0	1	2	3	4	5
Prediction (x_i)	4.50	1.60	4.00	4.60	5.20	9.75
Left bound (a_i)	1.80	0.80	2.50	2.88	3.25	7.50
Right bound (b_i)	11.25	3.20	6.40	7.36	8.32	12.68
s_i	0.288	0.218	0.148	0.148	0.148	0.082

It follows that $\Pr[a_i \leq X_i \leq b_i | \sigma = x_i] = 95\%$, $i = 1, ..., 5$ and $\Pr[a_0 \leq \sigma \leq b_0 | x_0] = 95\%$. The prior information is based on historical data. The standard deviation of the monthly returns of the S&P 500 has been computed for each of the last 20 years (1984 to 2004) and their geometric mean has been calculated to 4.5%. Around 95% of them lies between 1.8% and 11.25%.

Since σ represents the median of the observations $X_i|\sigma$ (the predictions of the experts), the manager is interested to estimate σ as the final prediction. Inference on σ is performed using the posterior density of σ, given by

$$\pi(\sigma|x_1, x_2, x_3, x_4, x_5) = \frac{\frac{1}{x_0}\pi(\frac{\sigma}{x_0})\prod_{i=1}^{5}\frac{1}{\sigma}f_i(\frac{x_i}{\sigma})}{\int_0^\infty \frac{1}{x_0}\pi(\frac{\sigma}{x_0})\prod_{i=1}^{5}\frac{1}{\sigma}f_i(\frac{x_i}{\sigma})d\sigma}.$$

3.4.4. Results

Data shows that experts 2, 3 and 4 provided similar predictions and confidence. The parameter s_i can be interpreted as a measure of the confidence of expert i on his prediction since it is a function of the ratio of the upper and lower bounds of the prediction. A small s_i indicates a relatively large confidence on the prediction. The prior prediction is also similar, but the confidence interval is larger than that of experts 2, 3 and 4. However, the information provided by experts 1 and 5 seems different and possibly in conflict with the other sources of information. The prediction of experts 1 and 5 are respectively the smallest and the largest with 1.6 and 9.75 and the overlap of their confidence interval with the other intervals is small.

All the information is combined through the posterior distribution of σ and the final prediction is given by the posterior estimation of σ. Since the right tail of the posterior density of σ is too heavy, the posterior moments do not exist and the posterior median of σ is used. The calculation is done using Monte Carlo simulations with importance sampling, see Desgagné and Angers (2005a).

Theorem 3 specifies how the posterior will behave in presence of outliers. It can be shown that log-cred$^-(f_i)$ = log-cred$^+(f_i) = (0,0,4,0)$, $i = 1,\ldots,5$ and log-cred$^-(\pi)$ = log-cred$^+(\pi) = (0,0,4,0)$. With these log-credences, only condition i) of Theorem 3 can be satisfied, i.e. $\gamma' < 1$ and $k \leq n/2$, which means that results a) to e) of Theorem 3 hold as long as the number of outliers is less or equal to 2. If experts 1 and 5 provided conflicting information, their influence on the model will be limited and would decrease if the conflict was increasing.

The results are given in the first row of Table 3.2. In the second row, the results are also computed for the same model where the Student is replaced by the standard normal density. For the log normal density, it can be shown that log-cred$^-(f_i)$= log-cred$^+(f_i) = (2, (2s_i^2)^{-1}, 0, 0)$, $i = 1, ..., 5$ and that log-cred$^-(\pi)$= log-cred$^+(\pi) = (2, (2s_0^2)^{-1}, 0, 0)$. With these log-credences, conditions of Theorem 3 are not satisfied and the influence of experts 1 and 5 is expected to be more important.

TAB. 3.2. Posterior predictions of the volatility

Model	Experts 2 to 4 in the model	Expert 1 is added	Expert 5 is added	Experts 1 and 5 are added
Log Student	4.6	4.4	4.8	4.6
Log normal	4.6	4.0	6.6	6.0

In the first column, the results are given when only the prior and experts 2, 3 and 4 are considered. The posterior median of σ is evaluated to 4.6% for both models, which makes sense since there is no conflicting information. In the next columns, results are given when experts 1 and 5 are added separately and together. Experts 1 and 5, separately or together, have a small influence on the posterior median of σ in the first model, most of the information provided by these two experts is considered conflicting by the model and therefore is rejected. However their influence is much more important in the second model. For instance, if expert 5 is added, the prediction of the model is 6.6%.

3.5. Conclusion

In this paper, the behavior of the posterior density of the scale parameter has been investigated when the sample contains outliers. The notion of left and right log-credences has been introduced to characterize respectively the left and right tails of a density defined on $\mathbb{R}^+$. Simple conditions on the tails of the prior and the likelihood, using left and right log-credences, are established to determine the proportion of observations that can be rejected as outliers. We have shown that the posterior distribution converges in law to the posterior that would be obtained from the reduced sample, excluding the outliers, as they tend to 0 or ∞, at any given rate. An example of combination of predictions of volatility of the S&P 500 index return is given.

3.6. Appendix : Proofs

The proofs of Theorems 3 and 4 are given in this appendix. Since Theorem 3 is a special case of Theorem 4, the proof of the latter is presented first. In Section 3.6.1, the proof of result a) of Theorem 4 is given. The proof of Lemma 15, needed in the preceding proof, is given in Section 3.6.2. The proofs of results b) to e) are given through Sections 3.6.3 to 3.6.6. The proof of Lemma 23, needed in the proof of result e), is given in Section 3.6.7. Finally, the proof of Theorem 3 is given in Section 3.6.8.

3.6.1. Proof of result a) of Theorem 4

It is assumed that the densities $\pi_\sigma, f_1, ..., f_n$ are proper and positive everywhere. Furthermore, the functions $z\pi_\sigma(z)$ and $zf_i(z), i = 1, ..., n$ are bounded above for all $z > 0$. Then it is easy to show that $m(\underset{\sim}{x}_k)\prod_{i=1}^{k} x_i$ and $m(\underset{\sim}{x}_n)\prod_{i=1}^{n} x_i$ are positive and bounded above for any positive $x_1, ..., x_n$, that the posteriors $\pi(\sigma|\underset{\sim}{x}_k)$ and $\pi(\sigma|\underset{\sim}{x}_n)$ are proper and positive densities, and that $\sigma\pi(\sigma|\underset{\sim}{x}_k)$ and $\sigma\pi(\sigma|\underset{\sim}{x}_n)$ are bounded above for all $\sigma > 0$, for any positive $x_1, ..., x_n$. Considering that $0 \leq \int_{1/h}^{h} x_0\pi(\sigma x_0|\underset{\sim}{x}_k)d\sigma = \int_{x_0/h}^{x_0 h} \pi(\sigma|\underset{\sim}{x}_k)d\sigma \leq 1$ and that $x_0\pi(\sigma x_0|\underset{\sim}{x}_k)$ depends only on the finite ratios $\frac{x_1}{x_0}, ..., \frac{x_k}{x_0}$, it is then possible to show the following lemma.

Lemma 12. *$\forall\epsilon > 0$, there exists a constant $A_4(\epsilon) > 1$ such that $h \geq A_4(\epsilon) \Rightarrow$*

$$\int_{x_0/h}^{x_0 h} \pi(\sigma|\underset{\sim}{x}_k)d\sigma \geq 1 - \epsilon, \int_{0}^{x_0/h} \pi(\sigma|\underset{\sim}{x}_k)d\sigma \leq \epsilon, \text{ and } \int_{x_0 h}^{\infty} \pi(\sigma|\underset{\sim}{x}_k)d\sigma \leq \epsilon.$$

Assuming conditions C1 to C3 are satisfied on the right tail of a proper and positive everywhere density f such that $zf(z)$ is bounded above for all $z > 0$, two other lemmas needed for the proof are given. Note that if conditions C1 to C3 are satisfied on the left tail of f, the lemmas are the same, except for the support, written in parentheses.

Lemma 13. $z > A_2 \Rightarrow \frac{d}{dz}\log z f^*(z) \leq 0$ *and* $\frac{d}{dz}\log z g(z) \leq 0$ $(z < A_2^{-1} \Rightarrow \frac{d}{dz}\log z f^*(z) \geq 0$ *and* $\frac{d}{dz}\log z g(z) \geq 0)$.

Proof : Let $h(x) = \mathrm{e}^x f^*(\mathrm{e}^x)$ be a density defined on $\mathbb{R}$. It can be shown that

i) $\int_{-\infty}^{\infty} h(x)dx = \int_0^{\infty} f^*(z)dz = 1$,

ii) $\frac{d}{dx}\log h(x)|_{x=\log z} = z\frac{d}{dz}\log z f^*(z)$,

iii) $\frac{d^2}{dx^2}\log h(x)|_{x=\log z} = z^2\frac{d^2}{dz^2}\log f^*(z) + z\frac{d}{dz}\log f^*(z)$.

First condition C3 and iii) imply that $\frac{d^2}{dx^2}\log h(x) \geq 0$ if $x > \log A_2$. Since $h(x)$ is a log-convex function on this range, it can be shown that $h(x)$ is also convex for $x > \log A_2$. It also can be shown that if a right tail of a proper density is convex, then it is necessarily decreasing. Since h is proper using i), it follows that $h(x)$ is decreasing for $x > \log A_2$, which is equivalent to $\frac{d}{dx}\log h(x) \leq 0$ for $x > \log A_2$. Using ii), it is also equivalent to $\frac{d}{dz}\log z f^*(z) \leq 0$ for $z > A_2$. The proofs for the left tail and for g are similar.

Lemma 14. $z > A_2$ *and* $\sigma > 1$ $(z < A_2^{-1}$ *and* $\sigma < 1) \Rightarrow \sigma f(\sigma z) \leq (M_2)^2 f(z)$.

Proof. Using Lemma 13, the right tail of $zf^*(z)$ is decreasing, that is $z > A_2 \Rightarrow \sigma z f^*(\sigma z) < z f^*(z), \forall \sigma > 1$. Therefore, $z > A_2 \Rightarrow \sigma f(\sigma z) \leq M_2 \sigma f^*(\sigma z) \leq M_2 f^*(z) \leq (M_2)^2 f(z)$. Condition C3 is used in the first and last inequalities. The proof for the left tail is similar.

Lemma 15. $h > A_2$, $z > \max[h^2, A_1(1,h)]$ *and* $\mathbb{D} = [h, \infty)$ $(z < \min[h^{-2}, A_1^{-1}(1,h)]$ *and* $\mathbb{D} = (0, 1/h]$ *for the left tail)* $\Rightarrow$

$$\int_{\mathbb{D}} \frac{f(\frac{z}{\sigma})f(\sigma)}{\sigma f(z)} d\sigma \leq (M_2)^{10} \text{ and } \frac{f(\frac{z}{\sigma})f(\sigma)}{f(z)} \leq (M_2)^{11} \text{ for all } \sigma \in \mathbb{D}.$$

Proof. See the Appendix, Section 3.6.2.

Using the fact that the numerators in equations (3.3.1) and (3.3.2) are proportional to $\sigma x_0 \pi(\sigma x_0 | \underset{\sim}{x}_k)$ $(x_0^k m(\underset{\sim}{x}_k)$ depends only on the constants $\frac{x_1}{x_0}, \ldots, \frac{x_k}{x_0})$, equations (3.3.1) and (3.3.2) can respectively be rewritten as follows (assuming $k < m$ and $m < n$) : $\forall \epsilon > 0$, there exists a constant $A_3(\epsilon)$ such that

$$\sigma < A_3^{-1}(\epsilon) \Rightarrow \frac{\sigma x_0 \pi(\sigma x_0 | \underset{\sim}{x}_k)}{\prod_{i=k+1}^{m} \sigma f_i(\sigma)} \leq \epsilon \text{ and } \sigma > A_3(\epsilon) \Rightarrow \frac{\sigma x_0 \pi(\sigma x_0 | \underset{\sim}{x}_k)}{\prod_{i=m+1}^{n} \sigma f_i(\sigma)} \leq \epsilon. \quad (3.6.1)$$

Denote $y_i = (\frac{x_i}{x_0})^{-1}$ if $i = k+1, ..., m$ and $y_i = \frac{x_i}{x_0}$ if $i = m+1, ..., n$. It follows that $\underset{\sim}{\phi}_2 = (y_{k+1}, ..., y_m, y_{m+1}, ..., y_n)$ and $\underset{\sim}{\phi}_2 \to \infty$ if $y_i \to \infty$, for $i = k+1, ..., n$, at any given rate. Then

$$\begin{aligned}
\frac{\pi(x_0|\underset{\sim}{x}_k)}{\pi(x_0|\underset{\sim}{x}_n)} &= \frac{m(\underset{\sim}{x}_n)}{m(\underset{\sim}{x}_k)\prod_{i=k+1}^n \frac{1}{x_0} f_i(\frac{x_i}{x_0})} \\
&= \frac{\int_0^\infty \frac{1}{x_0}\pi_\sigma(\frac{\sigma}{x_0})\prod_{i=1}^n \frac{1}{\sigma} f_i(\frac{x_i}{\sigma})d\sigma}{m(\underset{\sim}{x}_k)\prod_{i=k+1}^n \frac{1}{x_0} f_i(\frac{x_i}{x_0})} \\
&= \frac{\int_0^\infty \pi(\sigma|\underset{\sim}{x}_k)\prod_{i=k+1}^n \frac{1}{\sigma} f_i(\frac{x_i}{\sigma})d\sigma}{\prod_{i=k+1}^n \frac{1}{x_0} f_i(\frac{x_i}{x_0})} \\
&= \frac{\int_0^\infty x_0\pi(\sigma' x_0|\underset{\sim}{x}_k)\prod_{i=k+1}^n \frac{1}{\sigma' x_0} f_i(\frac{x_i}{x_0\sigma'})d\sigma'}{\prod_{i=k+1}^n \frac{1}{x_0} f_i(\frac{x_i}{x_0})} \\
&= \int_0^\infty x_0\pi(\sigma' x_0|\underset{\sim}{x}_k)\prod_{i=k+1}^m \frac{\frac{1}{\sigma'} f_i(\frac{1}{y_i\sigma'})}{f_i(\frac{1}{y_i})}\prod_{i=m+1}^n \frac{\frac{1}{\sigma'} f_i(\frac{y_i}{\sigma'})}{f_i(y_i)}d\sigma'.
\end{aligned}$$

Then result a) can be rewritten as follows : $\forall\epsilon > 0$, there exists a constant $A_0(\epsilon)$ such that $y_{k+1} > A_0(\epsilon), ..., y_n > A_0(\epsilon) \Rightarrow$

$$1-\epsilon \le \int_0^\infty x_0\pi(\sigma x_0|\underset{\sim}{x}_k)\prod_{i=k+1}^m \frac{\frac{1}{\sigma} f_i(\frac{1}{y_i\sigma})}{f_i(\frac{1}{y_i})}\prod_{i=m+1}^n \frac{\frac{1}{\sigma} f_i(\frac{y_i}{\sigma})}{f_i(y_i)}d\sigma \le 1+\epsilon.$$

First choose any $0 < \epsilon < 1$. Note that if the result is true for $0 < \epsilon < 1$, it is necessarily true for any $\epsilon > 0$. Then define

$$\epsilon_0 = \min\left(\left[(1+\epsilon/3)^{1/(n-k)} - 1\right], \left[1 - (1-\epsilon/3)^{1/(n-k+1)}\right], (\epsilon/3)M_2^{-11n}\right).$$

Note that $0 < \epsilon_0 < \frac{1}{3}$. Define $h = \max(A_2, A_3(\epsilon_0), A_4(\epsilon_0))$ and $A_0(\epsilon) = \max(A_1(\epsilon_0, h), h^2)$. Note that $A_0(\epsilon)$ depends only on ϵ. The constant A_1 comes from condition C1, A_2 and M_2 from conditions C2 and C3, A_3 from equation (3.6.1) and A_4 from Lemma 12. The integral is divided in three parts : $(0, \frac{1}{h}]$, $(\frac{1}{h}, h]$ and (h, ∞) and consider that $y_{k+1} > A_0(\epsilon), ..., y_n > A_0(\epsilon)$. First consider the integral on $(\frac{1}{h}, h]$.

$$\begin{aligned}
&\int_{1/h}^h x_0\pi(\sigma x_0|\underset{\sim}{x}_k)\prod_{i=k+1}^m \frac{\frac{1}{\sigma} f_i(\frac{1}{y_i\sigma})}{f_i(\frac{1}{y_i})}\prod_{i=m+1}^n \frac{\frac{1}{\sigma} f_i(\frac{y_i}{\sigma})}{f_i(y_i)}d\sigma \\
&\qquad \ge \int_{1/h}^h x_0\pi(\sigma x_0|\underset{\sim}{x}_k)\prod_{i=k+1}^m (1-\epsilon_0)\prod_{i=m+1}^n (1-\epsilon_0)d\sigma
\end{aligned}$$

$$\text{(C1 is used since } \frac{1}{y_i} < A_1^{-1}(\epsilon_0, h) \text{ and } y_i > A_1(\epsilon_0, h))$$

$$= (1-\epsilon_0)^{n-k} \int_{1/h}^{h} x_0 \pi(\sigma x_0 | \underset{\sim}{x}_k) d\sigma$$

$$= (1-\epsilon_0)^{n-k} \int_{x_0/h}^{x_0 h} \pi(\sigma | \underset{\sim}{x}_k) d\sigma$$

$$\geq (1-\epsilon_0)^{n-k+1}$$

$$\text{(Lemma 12 is used since } h \geq A_4(\epsilon_0))$$

$$\geq 1 - \epsilon/3.$$

In a similar way, it can be shown that

$$\int_{1/h}^{h} x_0 \pi(\sigma x_0 | \underset{\sim}{x}_k) \prod_{i=k+1}^{m} \frac{\frac{1}{\sigma} f_i(\frac{1}{y_i \sigma})}{f_i(\frac{1}{y_i})} \prod_{i=m+1}^{n} \frac{\frac{1}{\sigma} f_i(\frac{y_i}{\sigma})}{f_i(y_i)} d\sigma \leq 1 + \epsilon/3.$$

Consider now (h, ∞) if $m < n$ is assumed.

$$\int_{h}^{\infty} x_0 \pi(\sigma x_0 | \underset{\sim}{x}_k) \prod_{i=k+1}^{m} \frac{\frac{1}{\sigma} f_i(\frac{1}{y_i \sigma})}{f_i(\frac{1}{y_i})} \prod_{i=m+1}^{n} \frac{\frac{1}{\sigma} f_i(\frac{y_i}{\sigma})}{f_i(y_i)} d\sigma$$

$$\leq (M_2)^{2(m-k)} \int_{h}^{\infty} x_0 \pi(\sigma x_0 | \underset{\sim}{x}_k) \prod_{i=m+1}^{n} \frac{\frac{1}{\sigma} f_i(\frac{y_i}{\sigma})}{f_i(y_i)} d\sigma$$

$$\text{(Lemma 14 is used since } \frac{1}{y_i \sigma} < \frac{1}{y_i} < A_2^{-1})$$

$$\leq \epsilon_0 (M_2)^{2(m-k)} \int_{h}^{\infty} \frac{1}{\sigma} \prod_{i=m+1}^{n} \frac{f_i(\frac{y_i}{\sigma}) f_i(\sigma)}{f_i(y_i)} d\sigma$$

$$\text{(equation (3.6.1) is used since } \sigma \geq A_3(\epsilon_0))$$

$$\leq \epsilon_0 (M_2)^{2(m-k)} (M_2)^{11(n-m-1)} \int_{h}^{\infty} \frac{f_n(\frac{y_n}{\sigma}) f_n(\sigma)}{\sigma f_n(y_n)} d\sigma$$

$$\text{(Lemma 15 is used since } h > A_2 \text{ and } y_i > \max[h^2, A_1(1, h)])$$

$$\leq \epsilon_0 (M_2)^{2(m-k)} (M_2)^{11(n-m-1)} (M_2)^{10}$$

$$\text{(Lemma 15 is used)}$$

$$\leq \epsilon_0 (M_2)^{11n}$$

$$\text{(since } 0 \leq k \leq m \leq n)$$

$$\leq \epsilon/3.$$

Consider (h, ∞) if $m = n$ is assumed.

$$\int_h^\infty x_0 \pi(\sigma x_0 | \underset{\sim}{x}_k) \prod_{i=k+1}^{m} \frac{\frac{1}{\sigma} f_i(\frac{1}{y_i \sigma})}{f_i(\frac{1}{y_i})} d\sigma$$

$$\leq (M_2)^{2(m-k)} \int_h^\infty x_0 \pi(\sigma x_0 | \underset{\sim}{x}_k) d\sigma$$

$$\left(\text{Lemma 14 is used since } \frac{1}{y_i \sigma} < \frac{1}{y_i} < A_2^{-1}\right)$$

$$= (M_2)^{2(m-k)} \int_{x_0 h}^\infty \pi(\sigma | \underset{\sim}{x}_k) d\sigma$$

$$\leq \epsilon_0 (M_2)^{2(m-k)}$$

$$(\text{Lemma 12 is used since } h > A_4(\epsilon_0))$$

$$\leq \epsilon_0 (M_2)^{11n}$$

$$\leq \epsilon/3.$$

In a similar way, it can be shown that

$$\int_0^{1/h} x_0 \pi(\sigma x_0 | \underset{\sim}{x}_k) \prod_{i=k+1}^{m} \frac{\frac{1}{\sigma} f_i(\frac{1}{y_i \sigma})}{f_i(\frac{1}{y_i})} \prod_{i=m+1}^{n} \frac{\frac{1}{\sigma} f_i(\frac{y_i}{\sigma})}{f_i(y_i)} d\sigma \leq \epsilon/3.$$

Considering the three parts of the integral, we showed that

$$\int_0^\infty x_0 \pi(\sigma x_0 | \underset{\sim}{x}_k) \prod_{i=k+1}^{m} \frac{\frac{1}{\sigma} f_i(\frac{1}{y_i \sigma})}{f_i(\frac{1}{y_i})} \prod_{i=m+1}^{n} \frac{\frac{1}{\sigma} f_i(\frac{y_i}{\sigma})}{f_i(y_i)} d\sigma$$

$$\leq (1 + \epsilon/3) + \epsilon/3 + \epsilon/3$$

$$= 1 + \epsilon,$$

and

$$\int_0^\infty x_0 \pi(\sigma x_0 | \underset{\sim}{x}_k) \prod_{i=k+1}^{m} \frac{\frac{1}{\sigma} f_i(\frac{1}{y_i \sigma})}{f_i(\frac{1}{y_i})} \prod_{i=m+1}^{n} \frac{\frac{1}{\sigma} f_i(\frac{y_i}{\sigma})}{f_i(y_i)} d\sigma$$

$$\geq (1 - \epsilon/3) + 0 + 0$$

$$> 1 - \epsilon.$$

3.6.2. Proof of Lemma 15

We first introduce four other lemmas needed to prove Lemma 15. Suppose that conditions C1 to C3 are satisfied on the right tail of a proper and positive everywhere density f such that $zf(z)$ is bounded above for all $z > 0$. (If conditions C1 to C3 are satisfied on its left tail, the lemmas are the same, except for the support given in parentheses. The proofs are given only for the right tail, the proofs for the left tail being similar.) Without loss of generality, we assume that the constant M_2 in conditions C2 and C3 is chosen large enough, such that $M_2 > \max[\sup_{z>0} zf(z), A_2 g(A_2), 6]$.

Lemma 16. $z > A_2$ *(*$z < A_2^{-1}$ *for the left tail* *)* $\Rightarrow f^*(z) > 0$ *and* $g(z) > 0$.

Proof. If $f^*(z) = 0$ for a $z > A_2$, then the second part of condition C3 is not satisfied. If $g(z) = 0$ for a $z > A_2$, then condition C2 is not satisfied.

Lemma 17. $z > A_2\, (z < A_2^{-1}) \Rightarrow f(z) \leq (M_2)^3 g(z)$.

Proof. Using Lemma 14, if $z > A_2$ then $zf(z^2) \leq (M_2)^2 f(z)$. Using C2, if $z > A_2$ then $M_2 g(z) \geq \frac{f^2(z)}{zf(z^2)}$. Therefore $z > A_2 \Rightarrow (M_2)^3 g(z) \geq (M_2)^2 \frac{f^2(z)}{zf(z^2)} = f(z)\frac{(M_2)^2 f(z)}{zf(z^2)} \geq f(z)$.

Lemma 18. $z > A_2 \Rightarrow zg(z) < A_2 g(A_2) \quad \left(z < A_2^{-1} \Rightarrow zg(z) < A_2^{-1} g(A_2^{-1})\right)$.

Proof. Using Lemma 13, the right tail of $zg(z)$ is decreasing if $z > A_2$ (increasing for the left tail if $z < A_2^{-1}$).

Lemma 19. *For all a, b and z such that* $A_2 \leq a \leq b \leq zA_2^{-1}$ *(*$zA_2 \leq a \leq b \leq A_2^{-1}$ *for the left tail),* $\arg\max_{a \leq \sigma \leq b} \frac{f^*(z/\sigma) f^*(\sigma)}{\sigma g(\sigma)} \in \{a, b\}$.

Proof. First let $h(\theta) = e^\theta f^*(e^\theta)$ and $q(\theta) = e^\theta g(e^\theta)$ be two proper density defined on $\mathbb{R}$ and define $x = \log z$ and $\theta = \log \sigma$. It can be shown that $\arg\max_{a \leq \sigma \leq b} \frac{f^*(z/\sigma) f^*(\sigma)}{\sigma g(\sigma)} \in \{a, b\}$ for all a, b and z such that $A_2 \leq a \leq b \leq zA_2^{-1} \Leftrightarrow \arg\max_{\log a \leq \theta \leq \log b} \frac{h(x-\theta) h(\theta)}{q(\theta)} \in \{\log a, \log b\}$ for all a, b and x such that $\log A_2 \leq \log a \leq \log b \leq x - \log A_2$.

Since the maximum on a range of a convex function is located at its bounds, it is sufficient to show that $\frac{d^2}{d\theta^2} \log \frac{h(x-\theta)h(\theta)}{q(\theta)} \geq 0$ for any θ such that $\log A_2 < \theta < x - \log A_2$, since the log-convexity of a function implies its convexity. Then

$$\frac{d^2}{d\theta^2} \log \frac{h(x-\theta)h(\theta)}{q(\theta)} = \frac{d^2}{d\theta^2} \log h(x-\theta) + \frac{d^2}{d\theta^2} \log h(\theta) - \frac{d^2}{d\theta^2} \log q(\theta).$$

Using Lemma 13 iii), condition C3 can be rewritten as $\frac{d^2}{dx^2} \log h(x) \geq \frac{d^2}{dx^2} \log q(x) \geq 0$ for $x > \log A_2$. It follows that $\frac{d^2}{d\theta^2} \log h(\theta) - \frac{d^2}{d\theta^2} \log q(\theta) \geq 0$ for

$\theta > \log A_2$. It can also be shown that $\frac{d^2}{d\theta^2}\log h(x-\theta) = (\frac{d^2}{dy^2}\log h(y))|_{y=x-\theta}$, and using C3, that it is non negative for $x-\theta > \log A_2$. Then we showed that $\frac{d^2}{d\theta^2}\log\frac{h(x-\theta)h(\theta)}{q(\theta)} \geq 0$ if $\theta > \log A_2$ and $x-\theta > \log A_2$, that is if $\log A_2 < \theta < x - \log A_2$. □

To prove Lemma 15, we divide $[h,\infty)$ in three parts : $[h.\sqrt{z}]$, $(\sqrt{z}, z/h]$ and $(z/h,\infty)$. Consider that $h > A_2$ and $z > \max[h^2, A_1(1,h)]$. The constants A_1 and A_2 come respectively from conditions C1 and C2.

First consider $h \leq \sigma \leq \sqrt{z}$. Note that $h \leq \sigma \leq \sqrt{z}$, $h > A_2$ and $z > h^2 \Rightarrow z > z/A_2 > z/h \geq z/\sigma \geq \sqrt{z} \geq \sigma \geq h > A_2$. Then

$$\frac{f(\frac{z}{\sigma})f(\sigma)}{f(z)} \leq (M_2)^3\frac{f^*(\frac{z}{\sigma})f^*(\sigma)}{f^*(z)}$$

(C3 is used since $z/\sigma > A_2$, $\sigma > A_2$ and $z > A_2$)

$$= (M_2)^3\left(\frac{f^*(\frac{z}{\sigma})f^*(\sigma)}{f^*(z)\sigma g(\sigma)}\right)\sigma g(\sigma)$$

$$\leq (M_2)^3 \max\left(\frac{f^*(\frac{z}{h})f^*(h)}{f^*(z)hg(h)}, \frac{f^{*2}(\sqrt{z})}{f^*(z)\sqrt{z}g(\sqrt{z})}\right)\sigma g(\sigma)$$

(Lemma 19 is used since $A_2 < h \leq \sigma \leq \sqrt{z} < z/A_2$)

$$\leq (M_2)^6 \max\left(\frac{f(\frac{z}{h})f(h)}{f(z)hg(h)}, \frac{f^2(\sqrt{z})}{f(z)\sqrt{z}g(\sqrt{z})}\right)\sigma g(\sigma)$$

(C3 is used since $z/h > A_2, h > A_2, z > A_2$ and $\sqrt{z} > A_2$)

$$\leq (M_2)^6 \max\left(\frac{\frac{1}{h}f(\frac{z}{h})(M_2)^3}{f(z)}, M_2\right)\sigma g(\sigma)$$

(Lemma 17 is used since $h > A_2$ and C2 is used since $z > A_2$)

$$\leq (M_2)^6 \max\left(2(M_2)^3, M_2\right)\sigma g(\sigma)$$

(C1 is used since $z > A_1(1,h)$)

$$= 2(M_2)^9\sigma g(\sigma)$$

$$\leq 2(M_2)^9 A_2 g(A_2)$$

(Lemma 18 is used since $\sigma > A_2$)

$$\leq (M_2)^{11}$$

(since $M_2 > 2$ and $M_2 \geq A_2 g(A_2)$)

and since $g(\cdot)$ is a proper density,

$$\int_h^{\sqrt{z}} \frac{f(\frac{z}{\sigma})f(\sigma)}{\sigma f(z)} d\sigma \leq 2(M_2)^9 \int_h^{\sqrt{z}} g(\sigma)d\sigma$$
$$\leq 2(M_2)^9.$$

Consider now $\sqrt{z} \leq \sigma \leq \frac{z}{h}$. It is possible to use the precedent results (when $h \leq \sigma \leq \sqrt{z}$ is considered) if the change of variables $u = \frac{z}{\sigma}$ is done, since $h \leq u \leq \sqrt{z}$. Then

$$\begin{aligned}\frac{f(\frac{z}{\sigma})f(\sigma)}{f(z)} &= \frac{f(u)f(\frac{z}{u})}{f(z)} \\ &\leq 2(M_2)^9 u g(u) \\ &= 2(M_2)^9 \frac{z}{\sigma} g(\frac{z}{\sigma}) \\ &\leq 2(M_2)^9 A_2 g(A_2) \\ &\quad \text{(Lemma 18 is used since } z/\sigma > A_2) \\ &\leq (M_2)^{11}\end{aligned}$$

and since $g(\cdot)$ is a proper density,

$$\begin{aligned}\int_{\sqrt{z}}^{\frac{z}{h}} \frac{f(\frac{z}{\sigma})f(\sigma)}{\sigma f(z)} d\sigma &\leq 2(M_2)^9 \int_{\sqrt{z}}^{\frac{z}{h}} \frac{z}{\sigma^2} g(\frac{z}{\sigma}) d\sigma \\ &= 2(M_2)^9 \int_h^{\sqrt{z}} g(u)du \\ &\leq 2(M_2)^9.\end{aligned}$$

Finally consider $\sigma \geq \frac{z}{h}$.

$$\begin{aligned}\frac{f(\frac{z}{\sigma})f(\sigma)}{f(z)} &\leq \frac{f(\frac{z}{\sigma})\frac{z}{\sigma h}(M_2)^2 f(\frac{z}{h})}{f(z)} \\ &\quad \text{(Lemma 14 is used since } \sigma \geq z/h > A_2) \\ &= \frac{\frac{1}{h}f(\frac{z}{h})}{f(z)}(M_2)^2 f(\frac{z}{\sigma})\frac{z}{\sigma} \\ &\leq 2(M_2)^2 f(\frac{z}{\sigma})\frac{z}{\sigma} \\ &\quad \text{(C1 is used since } z > A_1(1,h)) \\ &\leq 2(M_2)^3\end{aligned}$$

$$(\text{since } \sup_{z>0} zf(z) \leq M_2 \text{ and } M_2 > 1)$$

$$\leq (M_2)^{11}$$

and since $f(\cdot)$ is a proper density,

$$\begin{aligned}\int_{\frac{z}{h}}^{\infty} \frac{f(\frac{z}{\sigma})f(\sigma)}{\sigma f(z)} d\sigma &\leq 2(M_2)^2 \int_{\frac{z}{h}}^{\infty} f(\frac{z}{\sigma})\frac{z}{\sigma^2} d\sigma \\ &= 2(M_2)^2 \int_0^h f(u)du \\ &\leq 2(M_2)^2 \\ &\leq 2(M_2)^9.\end{aligned}$$

If the integrals on the three domains are considered, then

$$\int_h^{\infty} \frac{f(\frac{z}{\sigma})f(\sigma)}{\sigma f(z)} d\sigma \leq 6(M_2)^9 \leq (M_2)^{10}.$$

3.6.3. Proof of result b) of Theorem 4

Result b) can be rewritten as follows : $\forall \epsilon > 0, \forall h > 1$ there exists a constant $A_5(\epsilon, h)$ such that $\min[\underset{\sim}{\phi}_2] > A_5(\epsilon, h)$ and $\frac{1}{h} \leq \frac{\sigma}{x_0} \leq h \Rightarrow 1 - \epsilon \leq \frac{\pi(\sigma|\underset{\sim}{x}_n)}{\pi(\sigma|\underset{\sim}{x}_k)} \leq 1 + \epsilon$. Note that $\min[\underset{\sim}{\phi}_2]$ stands for $\min\left[(\frac{x_{k+1}}{x_0})^{-1}, ..., (\frac{x_m}{x_0})^{-1}, \frac{x_{m+1}}{x_0}, ..., \frac{x_n}{x_0}\right]$. Result a) of Theorem 4 can also be rewritten as follows : $\forall \epsilon > 0$ there exists a constant $A_0(\epsilon)$ such that $\min[\underset{\sim}{\phi}_2] > A_0(\epsilon) \Rightarrow 1 - \epsilon \leq \frac{\pi(x_0|\underset{\sim}{x}_n)}{\pi(x_0|\underset{\sim}{x}_k)} \leq 1 + \epsilon$.

Choose any $\epsilon > 0$ and any $h > 1$. Then define

$$\epsilon_0 = \min[(1+\epsilon)^{1/(n-k+1)} - 1, 1 - (1-\epsilon)^{1/(n-k+1)}]$$

and $A_5(\epsilon, h) = \max[A_0(\epsilon_0), A_1(\epsilon_0, h)]$. The constants A_0 and A_1 come respectively from the proof of result a) of Theorem 4 and condition C1. Consider that $\min[\underset{\sim}{\phi}_2] > A_5(\epsilon, h)$ and $\frac{1}{h} \leq \frac{\sigma}{x_0} \leq h$. Then

$$\begin{aligned}\frac{\pi(\sigma|\underset{\sim}{x}_n)}{\pi(\sigma|\underset{\sim}{x}_k)} &= \frac{m(\underset{\sim}{x}_k)\prod_{i=k+1}^n \frac{1}{\sigma} f_i(\frac{x_i}{\sigma})}{m(\underset{\sim}{x}_n)} \\ &= \frac{\pi(x_0|\underset{\sim}{x}_n)}{\pi(x_0|\underset{\sim}{x}_k)} \left(\prod_{i=k+1}^{n} \frac{\frac{1}{\sigma} f_i(\frac{x_i}{x_0}\frac{x_0}{\sigma})}{\frac{1}{x_0} f_i(\frac{x_i}{x_0})} \right)\end{aligned}$$

$$\leq (1+\epsilon_0) \prod_{i=k+1}^{n} \frac{\frac{x_0}{\sigma} f_i(\frac{x_i}{x_0}\frac{x_0}{\sigma})}{f_i(\frac{x_i}{x_0})}$$

(Result a) is used since $\min[\underset{\sim}{\phi_2}] > A_0(\epsilon_0)$)

$$\leq (1+\epsilon_0) \prod_{i=k+1}^{n} (1+\epsilon_0)$$

(C1 is used since $\min[\underset{\sim}{\phi_2}] > A_1(\epsilon_0, h)$)

$$= (1+\epsilon_0)^{n-k+1}$$

$$\leq 1+\epsilon.$$

In a similar way, it can be shown that $\frac{\pi(\sigma|\underset{\sim}{x}_n)}{\pi(\sigma|\underset{\sim}{x}_k)} \geq 1-\epsilon$.

3.6.4. Proof of result c) of Theorem 4

Result c) of Theorem 4 says that the posterior density tends to 0 in a neighborhood bounded by any finite multiples of the outliers x_j, $j \in (k+1, ..., n)$. It can be rewritten as follows : $\forall\epsilon > 0, \forall d > 1$ there exists a constant $A_6(\epsilon, d)$ such that $\min[\underset{\sim}{\phi_2}] > A_6(\epsilon, d)$ and $j \in (k+1, ..., n) \Rightarrow \Pr[\frac{1}{d} \leq \frac{\sigma}{x_j} \leq d|\underset{\sim}{x}_n] \leq \epsilon$. A lemma analog to Lemma 12 is needed for the proof.

Lemma 20. $\epsilon > 0, h \geq A_4(\epsilon/2)$ *and* $\min[\underset{\sim}{\phi_2}] > A_5(\epsilon/2, h) \Rightarrow$

$$\int_{x_0/h}^{x_0 h} \pi(\sigma|\underset{\sim}{x}_n)d\sigma \geq 1-\epsilon, \int_0^{x_0/h} \pi(\sigma|\underset{\sim}{x}_n)d\sigma \leq \epsilon \text{ and } \int_{x_0 h}^{\infty} \pi(\sigma|\underset{\sim}{x}_n)d\sigma \leq \epsilon.$$

Proof. $\int_{x_0/h}^{x_0 h} \pi(\sigma|\underset{\sim}{x}_n)d\sigma \geq (1-\epsilon/2)\int_{x_0/h}^{x_0 h} \pi(\sigma|\underset{\sim}{x}_k)d\sigma \geq (1-\epsilon/2)^2 > 1-\epsilon$. Result b) of Theorem 4 is used in the first inequality since $\min[\underset{\sim}{\phi_2}] > A_5(\epsilon/2, h)$ and $\frac{1}{h} \leq \frac{\sigma}{x_0} \leq h$, and Lemma 12 in used in the second since $h \geq A_4(\epsilon/2)$. Furthermore, $\int_0^{x_0/h} \pi(\sigma|\underset{\sim}{x}_n)d\sigma + \int_{x_0 h}^{\infty} \pi(\sigma|\underset{\sim}{x}_n)d\sigma = \int_0^{\infty} \pi(\sigma|\underset{\sim}{x}_n)d\sigma - \int_{x_0/h}^{x_0 h} \pi(\sigma|\underset{\sim}{x}_n)d\sigma \leq 1-(1-\epsilon) = \epsilon$. □

Choose any $\epsilon > 0$ and any $d > 1$. Define $h = A_4(\epsilon/2)$ and define $A_6(\epsilon, d) = \max[A_5(\epsilon/2, h), dh]$, where the constant A_5 comes from the proof of result b) of Theorem 4. Consider that $\min[\underset{\sim}{\phi_2}] > A_6(\epsilon, d)$ and $j \in (k+1, ..., n)$. Since $\max(\frac{x_j}{x_0}, \frac{x_0}{x_j}) \in \underset{\sim}{\phi_2}$, it follows that $\max(\frac{x_j}{x_0}, \frac{x_0}{x_j}) > dh$. Then, if $x_j/x_0 > 1$ (that is for $j = m+1, ..., n$),

$$\Pr[\frac{1}{d} \leq \frac{\sigma}{x_j} \leq d | \underset{\sim}{x}_n] = \int_{x_j/d}^{x_j d} \pi(\sigma|\underset{\sim}{x}_n) d\sigma$$
$$\leq \int_{x_j/d}^{\infty} \pi(\sigma|\underset{\sim}{x}_n) d\sigma$$
$$\leq \int_{x_0 h}^{\infty} \pi(\sigma|\underset{\sim}{x}_n) d\sigma$$
$$\leq \epsilon.$$

Lemma 20 is used in the last inequality. The proof for $x_j/x_0 < 1$ (that is for $j = k+1, ..., m$) is similar.

3.6.5. Proof of result d) of Theorem 4

The definition of convergence in law of a sequence of random variables $\{Y_s\}_{s=1,2,3,...}$ to a random variable Y, as $s \to \infty$, is given as follows.

Definition 8. $Y_s \xrightarrow{\mathcal{L}} Y$ *if* $\lim_{s\to\infty} \Pr[Y_s \leq d] = \Pr[Y \leq d]$, *for all* d *such that* $\Pr[Y \leq d]$ *is continuous.*

In order to use this definition with $Y_s = (\sigma/x_0)|\underset{\sim}{x}_n$ and $Y = (\sigma/x_0)|\underset{\sim}{x}_k$, the observations and the scale parameter of the prior are expressed as some functions of the same variable s, denoted by $x_i = h_i(s), i = 0, 1, ..., n$, for any functions $h_i(s)$ on $\mathbb{N}$ which satisfy

- **i)** there exists a constant c_i such that $h_i(s)/h_0(s) = c_i$ for any $s \in \mathbb{N}$, if $i = 1, ..., k$,
- **ii)** $\lim_{s\to\infty}(h_i(s)/h_0(s)) = 0$, if $i = k+1, ..., m$,
- **iii)** $\lim_{s\to\infty}(h_i(s)/h_0(s)) = \infty$ if $i = m+1, ..., n$.

The density of Y_s evaluated at the point y is then given by

$$x_0 \pi(y x_0 | \underset{\sim}{x}_n) = \frac{\pi_\sigma(y) \prod_{i=1}^{n} \frac{1}{y} f_i(\frac{1}{y}\frac{x_i}{x_0})}{\int_0^\infty \pi_\sigma(y) \prod_{i=1}^{n} \frac{1}{y} f_i(\frac{1}{y}\frac{x_i}{x_0})\, dy}$$
$$= \frac{\pi_\sigma(y) \prod_{i=1}^{k} \frac{1}{y} f_i(\frac{c_i}{y}) \prod_{i=k+1}^{n} \frac{1}{y} f_i(\frac{1}{y}\frac{h_i(s)}{h_0(s)})}{\int_0^\infty \pi_\sigma(y) \prod_{i=1}^{k} \frac{1}{y} f_i(\frac{c_i}{y}) \prod_{i=k+1}^{n} \frac{1}{y} f_i(\frac{1}{y}\frac{h_i(s)}{h_0(s)})\, dy}$$

and the density of Y evaluated at the point y is given by

$$x_0\pi(yx_0|\underset{\sim}{x}_k) = \frac{\pi_\sigma(y)\prod_{i=1}^k \frac{1}{y} f_i(\frac{c_i}{y})}{\int_0^\infty \pi_\sigma(y)\prod_{i=1}^k \frac{1}{y} f_i(\frac{c_i}{y}), dy}.$$

It can be seen that the functions $h_i(s)$ are defined such that $s \to \infty \Leftrightarrow \underset{\sim}{\phi}_2 \to \infty$. Furthermore, it can be seen that the density of $Y = (\sigma/x_0)|\underset{\sim}{x}_k$ does not depend on s or $\underset{\sim}{\phi}_2$. Then $Y_s \xrightarrow{\mathcal{L}} Y$ as $s \to \infty$ for any functions $h_i(s)$ which satisfy i), ii) and iii) $\Leftrightarrow (\sigma/x_0)|\underset{\sim}{x}_n \xrightarrow{\mathcal{L}} (\sigma/x_0)|\underset{\sim}{x}_k$ as $\underset{\sim}{\phi}_2 \to \infty$ at any given rate.

According to Definition 8, the convergence in law is obtained if $\lim_{s\to\infty} \Pr[Y_s \leq d] = \Pr[Y \leq d]$, for all d such that $\Pr[Y \leq d]$ is continuous, or equivalently, if $\lim_{\underset{\sim}{\phi}_2\to\infty} \Pr[\sigma \leq x_0 d|\underset{\sim}{x}_n] = \Pr[\sigma \leq x_0 d|\underset{\sim}{x}_k]$, for all d such that $\Pr[\sigma \leq x_0 d|\underset{\sim}{x}_k]$ is continuous. Therefore, the result d) can be rewritten as follows : $\forall \epsilon > 0$ there exists a constant $A_7(\epsilon)$ such that $\min[\underset{\sim}{\phi}_2] > A_7(\epsilon)$ and $d > 0 \Rightarrow$
$|\Pr[\sigma/x_0 \leq d|\underset{\sim}{x}_n] - \Pr[\sigma/x_0 \leq d|\underset{\sim}{x}_k]| \leq \epsilon$.

Choose any $\epsilon > 0$, define $h = A_4(\epsilon/6)$ and $A_7(\epsilon) = A_5(\epsilon/6, h)$. The constants A_4 and A_5 come respectively from Lemma 12 and the proof of result b) of Theorem 4. The real line is divided in three parts : $(0, 1/h], (1/h, h]$ and (h, ∞) and we consider that $\min[\underset{\sim}{\phi}_2] > A_7(\epsilon)$. First consider $d \leq 1/h$.

$$\begin{aligned}\Pr[\sigma/x_0 \leq d|\underset{\sim}{x}_n] &\leq \Pr[\sigma \leq x_0/h|\underset{\sim}{x}_n] \\ &= \int_0^{x_0/h} \pi(\sigma|\underset{\sim}{x}_n)d\sigma \\ &\leq \epsilon/3.\end{aligned}$$

Lemma 20 is used in the last inequality since $h = A_4(\epsilon/6)$ and $\min[\underset{\sim}{\phi}_2] > A_5(\epsilon/6, h)$. In the same way, it can be shown, using Lemma 12, that $\Pr[\sigma/x_0 \leq d|\underset{\sim}{x}_k] \leq \epsilon/6$, since $h = A_4(\epsilon/6)$. From this result and from $\Pr[\sigma/x_0 \leq d|\underset{\sim}{x}_n] \leq \epsilon/3$, it follows that $|\Pr[\sigma/x_0 \leq d|\underset{\sim}{x}_n] - \Pr[\sigma/x_0 \leq d|\underset{\sim}{x}_k]| \leq \epsilon/3 < \epsilon$. Now consider $1/h < d \leq h$.

$$\begin{aligned}&|\Pr[1/h < \sigma/x_0 \leq d|\underset{\sim}{x}_n] - \Pr[1/h < \sigma/x_0 \leq d|\underset{\sim}{x}_k]| \\ &\qquad \leq \int_{x_0/h}^{x_0 d} |\pi(\sigma|\underset{\sim}{x}_n) - \pi(\sigma|\underset{\sim}{x}_k)|\, d\sigma\end{aligned}$$

$$= \int_{x_0/h}^{x_0 d} \pi(\sigma|\underset{\sim}{x}_k) \left| \frac{\pi(\sigma|\underset{\sim}{x}_n)}{\pi(\sigma|\underset{\sim}{x}_k)} - 1 \right| d\sigma$$

$$\leq \epsilon/6 \int_{x_0/h}^{x_0 d} \pi(\sigma|\underset{\sim}{x}_k) d\sigma$$

$$\leq \epsilon/6.$$

Result b) of Theorem 4 can be used in the second inequality since $\min[\underset{\sim}{\phi}_2] > A_5(\epsilon/6, h)$ and $1/h \leq \sigma/x_0 \leq h$. Therefore,

$$|\Pr[\sigma/x_0 \leq d|\underset{\sim}{x}_n] - \Pr[\sigma/x_0 \leq d|\underset{\sim}{x}_k]|$$

$$\leq |\Pr[\sigma/x_0 \leq 1/h|\underset{\sim}{x}_n] - \Pr[\sigma/x_0 \leq 1/h|\underset{\sim}{x}_k]|$$

$$+ |\Pr[1/h < \sigma/x_0 \leq d|\underset{\sim}{x}_n] - \Pr[1/h < \sigma/x_0 \leq d|\underset{\sim}{x}_k]|$$

$$\leq \epsilon/3 + \epsilon/6$$

$$= \epsilon/2$$

$$< \epsilon.$$

Finally consider $d > h$.

$$\Pr[h < \sigma/x_0 \leq d|\underset{\sim}{x}_n] \leq \Pr[\sigma > x_0 h|\underset{\sim}{x}_n]$$

$$= \int_{x_0 h}^{\infty} \pi(\sigma|\underset{\sim}{x}_n) d\sigma$$

$$\leq \epsilon/3.$$

Lemma 20 is used in the last inequality since $h = A_4(\epsilon/6)$ and $\min[\underset{\sim}{\phi}_2] > A_5(\epsilon/6, h)$. In the same way, it can be shown, using Lemma 12, that $\Pr[h < \sigma/x_0 \leq d|\underset{\sim}{x}_k] \leq \epsilon/6$, since $h = A_4(\epsilon/6)$. From this result and from $\Pr[h < \sigma/x_0 \leq d|\underset{\sim}{x}_n] \leq \epsilon/3$, it follows that

$$|\Pr[h < \sigma/x_0 \leq d|\underset{\sim}{x}_n] - \Pr[h < \sigma/x_0 \leq d|\underset{\sim}{x}_k]| \leq \epsilon/3.$$

Finally, from this result and from $|\Pr[\sigma/x_0 \leq h|\underset{\sim}{x}_n] - \Pr[\sigma/x_0 \leq h|\underset{\sim}{x}_k]| \leq \epsilon/2$, it follows that $|\Pr[\sigma/x_0 \leq d|\underset{\sim}{x}_n] - \Pr[\sigma/x_0 \leq d|\underset{\sim}{x}_k]| \leq \epsilon/2 + \epsilon/3 < \epsilon$.

3.6.6. Proof of result e) of Theorem 4

First we introduce three lemmas needed for the proofs.

Lemma 21.

$$\mathbb{E}^{\pi(\sigma|\underset{\sim}{x}_k)}[|w(\sigma/x_0)|] \leq M_1,$$

where $M_1 = \mathbb{E}^{\pi_\sigma(\sigma)}[|w(\sigma)|] \frac{M_2^k}{m(\underset{\sim}{x}_k)\prod_{i=1}^k x_i} < \infty.$

Proof.

$$\begin{aligned}
\mathbb{E}^{\pi(\sigma|\underset{\sim}{x}_k)}[|w(\sigma/x_0)|] &= \int_0^\infty |w(\sigma/x_0)| \frac{1}{x_0}\pi_\sigma(\sigma/x_0)\frac{\prod_{i=1}^k \frac{1}{\sigma} f_i(x_i/\sigma)}{m(\underset{\sim}{x}_k)} d\sigma \\
&= \int_0^\infty |w(\sigma/x_0)| \frac{1}{x_0}\pi_\sigma(\sigma/x_0)\frac{\prod_{i=1}^k \frac{x_i}{\sigma} f_i(x_i/\sigma)}{m(\underset{\sim}{x}_k)\prod_{i=1}^k x_i} d\sigma \\
&\leq \frac{M_2^k}{m(\underset{\sim}{x}_k)\prod_{i=1}^k x_i}\int_0^\infty |w(\sigma/x_0)| \frac{1}{x_0}\pi_\sigma(\sigma/x_0) d\sigma \\
&= \frac{M_2^k}{m(\underset{\sim}{x}_k)\prod_{i=1}^k x_i}\int_0^\infty |w(\sigma)| \pi_\sigma(\sigma) d\sigma \\
&= M_1.
\end{aligned}$$

The fact that $zf_1(z), ..., zf_n(z)$ are bounded above by M_2 is used in the inequality. Furthermore, since $\mathbb{E}^{\pi_\sigma(\sigma)}[|w(\sigma)|] < \infty$ is assumed in Theorem 4 and $0 < m(\underset{\sim}{x}_k)\prod_{i=1}^k x_i < \infty$, it follows that $M_1 < \infty$. □

Considering that

$$0 \leq \int_{1/h}^{h} |w(\sigma)|\, x_0 \pi(\sigma x_0|\underset{\sim}{x}_k) d\sigma = \int_{x_0/h}^{x_0 h} |w(\sigma/x_0)|\, \pi(\sigma|\underset{\sim}{x}_k) d\sigma \leq M_1$$

and that $|w(\sigma)|\, x_0\pi(\sigma x_0|\underset{\sim}{x}_k)$ depends only on the finite ratios $\frac{x_1}{x_0}, ..., \frac{x_k}{x_0}$, it is then possible to show the following lemma.

Lemma 22. $\forall \epsilon > 0$, *there exists a constant* $A_9(\epsilon) > 0$ *such that* $h \geq A_9(\epsilon) \Rightarrow$

$$\int_0^{x_0/h} |w(\sigma/x_0)|\, \pi(\sigma|\underset{\sim}{x}_k) d\sigma \leq \epsilon \text{ and } \int_{x_0 h}^{\infty} |w(\sigma/x_0)|\, \pi(\sigma|\underset{\sim}{x}_k) d\sigma \leq \epsilon.$$

Note that condition $\mathbb{E}^{\pi_\sigma(\sigma)}[|w(\sigma)|] < \infty$ is conservative but is appropriate whatever the number of outliers is. It could be possible to relax it using the condition $\mathbb{E}^{\pi(\sigma|\underset{\sim}{x}_k)}[|w(\sigma/x_0)|] < \infty$, considering also the non-outliers observations altogether with the prior. A last lemma is needed.

Lemma 23. $\forall \epsilon > 0$, *there exists a constant* $A_8(\epsilon)$ *such that* $\min[\underset{\sim}{\phi}_2] > A_8(\epsilon) \Rightarrow$

$$\left|\mathbb{E}^{\pi(\sigma|\underset{\sim}{x}_n)}[|w(\sigma/x_0)|] - \mathbb{E}^{\pi(\sigma|\underset{\sim}{x}_k)}[|w(\sigma/x_0)|]\right| < \epsilon.$$

Lemma 23 is similar to the result e) of Theorem 4, except it considers the absolute value of $w(\sigma/x_0)$. Its proof is given in Section 3.6.7.

Consider now the result e) of Theorem 4, which can be rewritten as follows. $\forall\epsilon > 0$, there exists a constant $A_0(\epsilon)$ such that $\min[\underset{\sim}{\phi_2}] > A_0(\epsilon) \Rightarrow$

$$\left|\mathbb{E}^{\pi(\sigma|\underset{\sim}{x}_n)}[w(\sigma/x_0)] - \mathbb{E}^{\pi(\sigma|\underset{\sim}{x}_k)}[w(\sigma/x_0)]\right| < \epsilon.$$

Choose any $\epsilon > 0$. Define $\epsilon_0 = \epsilon/7$, $h = A_9(\epsilon_0)$ and $A_0(\epsilon) = \max[A_5(\epsilon_0/M_1, h), A_8(\epsilon_0)]$, where the constant $A_5(\epsilon_0/M_1, h)$ comes from the proof of result b) of Theorem 4, which was rewritten as follows : $\forall\epsilon > 0, \forall h > 1$, there exists a constant $A_5(\epsilon, h)$ such that $\min[\underset{\sim}{\phi_2}] > A_5(\epsilon, h)$ and $1/h \leq \sigma/x_0 \leq h \Rightarrow \left|\frac{\pi(\sigma|\underset{\sim}{x}_n)}{\pi(\sigma|\underset{\sim}{x}_k)} - 1\right| \leq \epsilon$.

Then

$$\begin{aligned}
&\left|\mathbb{E}^{\pi(\sigma|\underset{\sim}{x}_n)}[w(\sigma/x_0)] - \mathbb{E}^{\pi(\sigma|\underset{\sim}{x}_k)}[w(\sigma/x_0)]\right| \\
&= \left|\int_0^\infty w(\sigma/x_0)\pi(\sigma|\underset{\sim}{x}_n)d\sigma - \int_0^\infty w(\sigma/x_0)\pi(\sigma|\underset{\sim}{x}_k)d\sigma\right| \\
&\leq \int_0^{x_0/h} |w(\sigma/x_0)|\,\pi(\sigma|\underset{\sim}{x}_n)d\sigma + \int_{x_0h}^\infty |w(\sigma/x_0)|\,\pi(\sigma|\underset{\sim}{x}_n)d\sigma \\
&+ \left|\int_{x_0/h}^{x_0h} w(\sigma/x_0)\pi(\sigma|\underset{\sim}{x}_n)d\sigma - \int_{x_0/h}^{x_0h} w(\sigma/x_0)\pi(\sigma|\underset{\sim}{x}_k)d\sigma\right| \\
&+ \int_0^{x_0/h} |w(\sigma/x_0)|\,\pi(\sigma|\underset{\sim}{x}_k)d\sigma + \int_{x_0h}^\infty |w(\sigma/x_0)|\,\pi(\sigma|\underset{\sim}{x}_k)d\sigma \\
&\leq \mathbb{E}^{\pi(\sigma|\underset{\sim}{x}_n)}[|w(\sigma/x_0)|] - \int_{x_0/h}^{x_0h} |w(\sigma/x_0)|\,\pi(\sigma|\underset{\sim}{x}_n)d\sigma \\
&+ \int_{x_0/h}^{x_0h} |w(\sigma/x_0)|\,|\pi(\sigma|\underset{\sim}{x}_n) - \pi(\sigma|\underset{\sim}{x}_k)|\,d\sigma + 2\epsilon_0.
\end{aligned}$$

Lemma 22 is used in the last inequality since $h = A_9(\epsilon_0)$. Now

$$\begin{aligned}
&\int_{x_0/h}^{x_0h} |w(\sigma/x_0)|\,|\pi(\sigma|\underset{\sim}{x}_n) - \pi(\sigma|\underset{\sim}{x}_k)|\,d\sigma \\
&= \int_{x_0/h}^{x_0h} |w(\sigma/x_0)|\,\pi(\sigma|\underset{\sim}{x}_k)\left|\frac{\pi(\sigma|\underset{\sim}{x}_n)}{\pi(\sigma|\underset{\sim}{x}_k)} - 1\right| d\sigma \\
&\leq \frac{\epsilon_0}{M_1}\int_{x_0/h}^{x_0h} |w(\sigma/x_0)|\,\pi(\sigma|\underset{\sim}{x}_k)d\sigma
\end{aligned}$$

(result b) of Theorem 4 is used since $A_0(\epsilon) \geq A_5(\epsilon_0/M_1, h)$)

$$\leq \frac{\epsilon_0}{M_1} \mathbb{E}^{\pi(\sigma|\underset{\sim}{x}_k)}[|w(\sigma/x_0)|]$$

$$\leq \epsilon_0.$$

(Lemma 21 is used)

Finally

$$\mathbb{E}^{\pi(\sigma|\underset{\sim}{x}_n)}[|w(\sigma/x_0)|] - \int_{x_0/h}^{x_0 h} |w(\sigma/x_0)|\, \pi(\sigma|\underset{\sim}{x}_n) d\sigma$$

$$\leq \left| \mathbb{E}^{\pi(\sigma|\underset{\sim}{x}_n)}[|w(\sigma/x_0)|] - \mathbb{E}^{\pi(\sigma|\underset{\sim}{x}_k)}[|w(\sigma/x_0)|] \right|$$

$$+ \left| \mathbb{E}^{\pi(\sigma|\underset{\sim}{x}_k)}[|w(\sigma/x_0)|] - \int_{x_0/h}^{x_0 h} |w(\sigma/x_0)|\, \pi(\sigma|\underset{\sim}{x}_k) d\sigma \right|$$

$$+ \left| \int_{x_0/h}^{x_0 h} |w(\sigma/x_0)|\, \pi(\sigma|\underset{\sim}{x}_k) d\sigma - \int_{x_0/h}^{x_0 h} |w(\sigma/x_0)|\, \pi(\sigma|\underset{\sim}{x}_n) d\sigma \right|$$

$$\leq \epsilon_0$$

$$+ \int_0^{x_0/h} |w(\sigma/x_0)|\, \pi(\sigma|\underset{\sim}{x}_k) d\sigma + \int_{x_0 h}^{\infty} |w(\sigma/x_0)|\, \pi(\sigma|\underset{\sim}{x}_k) d\sigma$$

$$+ \int_{x_0/h}^{x_0 h} |w(\sigma/x_0)|\, |\pi(\sigma|\underset{\sim}{x}_n) - \pi(\sigma|\underset{\sim}{x}_k)|\, d\sigma$$

(Lemma 23 is used since $A_0(\epsilon) \geq A_8(\epsilon_0)$)

$$\leq \epsilon_0 + \epsilon_0 + \epsilon_0 + \epsilon_0$$

(Lemma 22 and the preceding result are used)

$$= 4\epsilon_0.$$

Then we showed that $\left| \mathbb{E}^{\pi(\sigma|\underset{\sim}{x}_n)}[w(\sigma/x_0)] - \mathbb{E}^{\pi(\sigma|\underset{\sim}{x}_k)}[w(\sigma/x_0)] \right| \leq 7\epsilon_0 = \epsilon.$

3.6.7. Proof of Lemma 23

We want to show that

$$\lim_{\underset{\sim}{\phi}_2 \to \infty} \mathbb{E}^{\pi(\sigma|\underset{\sim}{x}_n)}[|w(\sigma/x_0)|] = \mathbb{E}^{\pi(\sigma|\underset{\sim}{x}_k)}[|w(\sigma/x_0)|],$$

or equivalently

$$\lim_{\underset{\sim}{\phi}_2 \to \infty} \mathbb{E}^{\pi(\sigma|\underset{\sim}{x}_n)}[|w(\sigma/x_0)| + 1] = \mathbb{E}^{\pi(\sigma|\underset{\sim}{x}_k)}[|w(\sigma/x_0)| + 1].$$

Define the density π^* as $\pi^*(\sigma) = \frac{(|w(\sigma)|+1)\pi_\sigma(\sigma)}{\int_0^\infty (|w(\sigma)|+1)\pi_\sigma(\sigma)d\sigma}$ or

$$\frac{1}{x_0}\pi^*(\sigma/x_0) = \frac{(|w(\sigma/x_0)|+1)\frac{1}{x_0}\pi_\sigma(\sigma/x_0)}{\int_0^\infty (|w(\sigma/x_0)|+1)\frac{1}{x_0}\pi_\sigma(\sigma/x_0)d\sigma}.$$

Since the prior π_σ is proper and $\mathbb{E}^{\pi_\sigma(\sigma)}[|w(\sigma)|] < \infty$, the denominator is finite and π^* is a proper density. The density π^* is positive everywhere since the numerator is positive and $\sigma\pi^*(\sigma)$ is bounded above since the functions $\sigma\pi_\sigma(\sigma)$ and $\sigma\,|w(\sigma)|\,\pi_\sigma(\sigma)$ are also bounded above, as assumed in Theorem 4.

It is then possible to use the result a) of Theorem 4 using the density π^* as prior instead of π_σ. If the conditions given by equations (3.3.1) and (3.3.2) are used with π^* instead of π_σ, they are equivalent to the conditions given by equations (3.3.1) to (3.3.4) using π_σ. Result a) using π_σ as prior is equivalent to

$$\lim_{\underset{\sim}{\phi}_2\to\infty} \frac{m(\underset{\sim}{x}_k)\prod_{i=k+1}^n \frac{1}{x_0} f_i(x_i/x_0)}{m(\underset{\sim}{x}_n)} = 1,$$

and result a) using π^* as prior is equivalent to

$$\lim_{\underset{\sim}{\phi}_2\to\infty} \frac{\left(\prod_{i=k+1}^n \frac{1}{x_0} f_i(x_i/x_0)\right)\int_0^\infty \frac{1}{x_0}\pi^*(\sigma/x_0)\prod_{i=1}^k \frac{1}{\sigma} f_i(x_i/\sigma)d\sigma}{\int_0^\infty \frac{1}{x_0}\pi^*(\sigma/x_0)\prod_{i=1}^n \frac{1}{\sigma} f_i(x_i/\sigma)d\sigma} = 1$$

$$\Leftrightarrow \lim_{\underset{\sim}{\phi}_2\to\infty}$$

$$\frac{\left(\prod_{i=k+1}^n \frac{1}{x_0} f_i(x_i/x_0)\right)\int_0^\infty (|w(\sigma/x_0)|+1)\frac{1}{x_0}\pi_\sigma(\sigma/x_0)\prod_{i=1}^k \frac{1}{\sigma} f_i(x_i/\sigma)d\sigma}{\int_0^\infty (|w(\sigma/x_0)|+1)\frac{1}{x_0}\pi_\sigma(\sigma/x_0)\prod_{i=1}^n \frac{1}{\sigma} f_i(x_i/\sigma)d\sigma} = 1.$$

The result can now be shown.

$$\begin{aligned}
&\mathbb{E}^{\pi(\sigma|\underset{\sim}{x}_n)}[|w(\sigma/x_0)|+1] \\
&= \frac{\int_0^\infty (|w(\sigma/x_0)|+1)\frac{1}{x_0}\pi_\sigma(\sigma/x_0)\prod_{i=1}^n \frac{1}{\sigma} f_i(x_i/\sigma)d\sigma}{m(\underset{\sim}{x}_n)} \\
&= \frac{\int_0^\infty (|w(\sigma/x_0)|+1)\frac{1}{x_0}\pi_\sigma(\sigma/x_0)\prod_{i=1}^n \frac{1}{\sigma} f_i(x_i/\sigma)d\sigma}{\left(\prod_{i=k+1}^n \frac{1}{x_0} f_i(x_i/x_0)\right)\int_0^\infty (|w(\sigma/x_0)|+1)\frac{1}{x_0}\pi_\sigma(\sigma/x_0)\prod_{i=1}^k \frac{1}{\sigma} f_i(x_i/\sigma)d\sigma} \\
&\times \frac{m(\underset{\sim}{x}_k)\prod_{i=k+1}^n \frac{1}{x_0} f_i(x_i/x_0)}{m(\underset{\sim}{x}_n)} \\
&\times \frac{\int_0^\infty (|w(\sigma/x_0)|+1)\frac{1}{x_0}\pi_\sigma(\sigma/x_0)\prod_{i=1}^k \frac{1}{\sigma} f_i(x_i/\sigma)d\sigma}{m(\underset{\sim}{x}_k)}.
\end{aligned}$$

If the limit as $\underset{\sim}{\phi}_2 \to \infty$ is taken, the first two terms in the last expression are 1 according to results a) using respectively π^* and π_σ as prior. The last term is $\mathbb{E}^{\pi(\sigma|\underset{\sim}{x}_k)}[|w(\sigma/x_0)| + 1]$, which prove the result.

3.6.8. Proof of Theorem 3

Since Theorem 3 is an application of Theorem 4, it is sufficient to show that if conditions of the former are satisfied, then conditions of the latter are also satisfied. The context is the same for both theorems, and in particular, the densities $\pi_\sigma, f_1, ..., f_n$ are assumed to be proper, positive everywhere and $z\pi_\sigma(z)$, $zf_i(z), i = 1, ..., n$ are bounded above. The conditions needed for results a) to d) in Theorem 3 are log-cred⁻(f_i) = log-cred$^+(f_i) = (\gamma', \delta', \alpha', \beta')$, $i = 1, ..., n$, log-cred⁻(π_σ) = log-cred$^+(\pi_\sigma) = (\gamma, \delta, \alpha, \beta)$ and

i) $\gamma' < 1$, $k \geq n/2$ or

ii) $\gamma' < 1$, $k < n/2$, $(\gamma, \delta, \alpha, \beta) > (\gamma', \delta'(n - 2k), \alpha'(n - 2k), \beta'(n - 2k))$.

Since the left and right log-credences are the same for each density and the log-credence is the same for all densities f_i of the observations i, $i = 1, ..., n$, then the conditions needed for results a) to d) in Theorem 4 can be simplified as follows. Conditions C1 to C3 are satisfied on the right tail of f_n and equations (3.3.1) and (3.3.2) are satisfied. To simplify the notation, f_n is simply denoted by f.

First we show that condition C1 is satisfied on the right tail of f, which can be written as follows. For any constant $h > 1$ and for all σ such that $\frac{1}{h} \leq \sigma \leq h$, $\lim_{z\to\infty} \frac{z\sigma f(z\sigma)}{zf(z)} = 1$. Furthermore, since log-cred$^+(f) = (\gamma', \delta', \alpha', \beta')$, there exists a constant $K_1 > 0$ such that $\lim_{z\to\infty} \frac{zf(z)}{e^{-\delta'(\log z)^{\gamma'}}(\log z)^{-\alpha'}\log^{-\beta'}(\log z)} = K_1$. If $h > 1$ and $\frac{1}{h} \leq \sigma \leq h$, then

$$
\begin{aligned}
\lim_{z\to\infty} \frac{z\sigma f(z\sigma)}{zf(z)} &= \lim_{z\to\infty} \frac{z\sigma f(z\sigma)}{e^{-\delta'(\log(z\sigma))^{\gamma'}}(\log(z\sigma))^{-\alpha'}\log^{-\beta'}(\log(z\sigma))} \\
&\quad \times \frac{e^{-\delta'(\log z)^{\gamma'}}(\log z)^{-\alpha'}\log^{-\beta'}(\log z)}{zf(z)} \\
&\quad \times \frac{e^{-\delta'(\log(z\sigma))^{\gamma'}}(\log(z\sigma))^{-\alpha'}\log^{-\beta'}(\log(z\sigma))}{e^{-\delta'(\log z)^{\gamma'}}(\log z)^{-\alpha'}\log^{-\beta'}(\log z)} \\
&= \lim_{z\to\infty} \frac{K_1}{K_1} \frac{e^{-\delta'(\log(z\sigma))^{\gamma'}}(\log(z\sigma))^{-\alpha'}\log^{-\beta'}(\log(z\sigma))}{e^{-\delta'(\log z)^{\gamma'}}(\log z)^{-\alpha'}\log^{-\beta'}(\log z)}
\end{aligned}
$$

$$= \lim_{z\to\infty} \mathrm{e}^{-\delta'[(\log(z\sigma))^{\gamma'}-(\log z)^{\gamma'}]}\left(\frac{\log z}{\log(z\sigma)}\right)^{\alpha'}\left(\frac{\log(\log z)}{\log(\log(z\sigma))}\right)^{\beta'}$$

$$= \lim_{x\to\infty} \mathrm{e}^{-\delta'[(x+\theta)^{\gamma'}-x^{\gamma'}]}\left(\frac{x}{x+\theta}\right)^{\alpha'}\left(\frac{\log x}{\log(x+\theta)}\right)^{\beta'}$$

where $x = \log z$ and $\theta = \log \sigma, -\log h \leq \theta \leq \log h$

$$= \lim_{x\to\infty} \mathrm{e}^{-\delta'[(x+\theta)^{\gamma'}-x^{\gamma'}]}.$$

It is easy to check in the last equality that the last two terms tend to 1 as x tends to infinity. Furthermore, using the Taylor series development of $(x+\theta)^{\gamma'} - x^{\gamma'}$, it can be shown that the last expression tends to 1 as x tends to infinity if and only if $\gamma' < 1$, which is a condition of Theorem 3.

Now we show that conditions C2 and C3 are satisfied on the right tail of f, which can be written as follows. There exist constants A_2 and $M_2 > 1$ and a proper density g such that for all $z > A_2$,

C2 : $\frac{f^2(\sqrt{z})}{\sqrt{z}f(z)g(\sqrt{z})} \leq M_2$,

C3 : $z^2\frac{d^2}{dz^2}\log f^*(z) + z\frac{d}{dz}\log f^*(z) \geq z^2\frac{d^2}{dz^2}\log g(z) + z\frac{d}{dz}\log g(z) \geq 0$,

where f^* is f or any other proper densities which satisfy $\frac{1}{M_2} \leq \frac{f(z)}{f^*(z)} \leq M_2$ for all $z > A_2$. Note that, as mentioned in Lemma 13, condition C3 is equivalent to

C3 : $\frac{d^2}{dx^2}\log \mathrm{e}^x f^*(\mathrm{e}^x)|_{x=\log z} \geq \frac{d^2}{dx^2}\log \mathrm{e}^x g(\mathrm{e}^x)|_{x=\log z} \geq 0$.

Define $f^*(z) = \frac{1}{z}p(\log z|\gamma', \delta', \alpha', \beta', z_0) = q(z|\gamma', \delta', \alpha', \beta', z_0)$, $z > 0$, with any $z_0 > 1$, where $p(\cdot|\gamma', \delta', \alpha', \beta', z_0)$ is a GEP density defined on $\mathbb{R}$ with left and right p-credences given by $(\gamma', \delta', \alpha', \beta')$ and $q(\cdot|\gamma', \delta', \alpha', \beta', z_0)$ is a log-GEP density defined on $\mathbb{R}^+$ with left and right log-credences also given by $(\gamma', \delta', \alpha', \beta')$. The symmetry of f^* about 1 is given by $zf^*(z) = \frac{1}{z}f^*(\frac{1}{z})$. Therefore, $f^*(z)$ is a log-GEP density with the same left and right log-credences as $f(z)$. The tails behavior of f and f^* are the same and both are proper densities. Define, for $z > 0$,

$$g(z) = \begin{cases} \frac{1}{z}(1+|\log z|)^{-3}; & \text{if } \gamma' > 0, \delta' > 0, \\ f^*(z); & \text{if } \gamma' = 0, \delta' = 0. \end{cases}$$

The density g is proper since $\frac{1}{z}(1+|\log z|)^{-3}$ and f^* are also proper densities.

Consider the first case, when $0 < \gamma' < 1, \delta' > 0$ and $g(z) = \frac{1}{z}(1 + |\log z|)^{-3}$. Then

$$\lim_{z\to\infty} \frac{f^2(\sqrt{z})}{\sqrt{z}f(z)g(\sqrt{z})}$$

$$\lim_{z\to\infty} \frac{(\sqrt{z}f(\sqrt{z}))^2}{zf(z)\sqrt{z}g(\sqrt{z})}$$

$$= \lim_{z\to\infty} \frac{K_1^2}{K_1} \frac{\left(\mathrm{e}^{-\delta'(\log\sqrt{z})^{\gamma'}}(\log\sqrt{z})^{-\alpha'}\log^{-\beta'}(\log\sqrt{z})\right)^2}{\left(\mathrm{e}^{-\delta'(\log z)^{\gamma'}}(\log z)^{-\alpha'}\log^{-\beta'}(\log z)\right)\,(1 + (\log\sqrt{z}))^{-3}}$$

$$= \lim_{z\to\infty} K_1\,\mathrm{e}^{-\delta'(2^{1-\gamma'}-1)(\log z)^{\gamma'}}\left(\frac{\log z}{4}\right)^{-\alpha'}\left(\frac{\log^2(\log\sqrt{z})}{\log(\log z)}\right)^{-\beta'}$$

$$\times\,(1 + (\log\sqrt{z}))^3$$

$$= 0.$$

The dominant term is the exponential one and it tends to 0 as $z \to \infty$ since $\gamma' > 0, \delta' > 0$ and $2^{1-\gamma'} - 1 > 0 \Leftrightarrow \gamma' < 1$. It is sufficient to show that condition C2 is satisfied since $zf(z)$ and $zg(z)$ are both positive and bounded above functions, with monotonous tails.

Furthermore, if $z > z_0$, it can be shown that

$$z^2\frac{d^2}{dz^2}\log f^*(z) + z\frac{d}{dz}\log f^*(z)$$

$$= \left(\frac{d^2}{dx^2}\log[\mathrm{e}^x f^*(\mathrm{e}^x)]\right)|_{x=\log z}$$

$$= \left(\frac{d^2}{dx^2}\log\left[\mathrm{e}^{-\delta' x^{\gamma'}}x^{-\alpha'}\log^{-\beta'}x\right]\right)|_{x=\log z}$$

$$= \left(\frac{d^2}{dx^2}\left[-\delta' x^{\gamma'} - \alpha'\log x - \beta'\log(\log x)\right]\right)|_{x=\log z}$$

$$= \left(\gamma'(1-\gamma')\delta' x^{\gamma'-2} + \frac{\alpha'}{x^2} + \frac{\beta'(\log x + 1)}{x^2\log^2 x}\right)|_{x=\log z}$$

$$= \left(\frac{1}{x^2}\left[\gamma'(1-\gamma')\delta' x^{\gamma'} + \alpha' + \frac{\beta'}{\log x} + \frac{\beta'}{\log^2 x}\right]\right)|_{x=\log z}$$

$$= \frac{1}{(\log z)^2}\left[\gamma'(1-\gamma')\delta'(\log z)^{\gamma'} + \alpha' + \frac{\beta'}{\log(\log z)} + \frac{\beta'}{\log^2(\log z)}\right].$$

Furthermore,

$$z^2\frac{d^2}{dz^2}\log g(z) + z\frac{d}{dz}\log g(z) = \left(\frac{d^2}{dx^2}\log[\mathrm{e}^x g(\mathrm{e}^x)]\right)|_{x=\log z}$$

$$= \left(\frac{3}{(1+|x|)^2} \right) |_{x=\log z}$$
$$= \frac{3}{(1+|\log z|)^2}$$
$$> 0$$

for any value of $z > 0$. Finally, if $z > z_0$,

$$z^2 \frac{d^2}{dz^2} \log f^*(z) + z \frac{d}{dz} \log f^*(z) - z^2 \frac{d^2}{dz^2} \log g(z) - z \frac{d}{dz} \log g(z)$$
$$= \frac{1}{(\log z)^2} \left[\gamma'(1-\gamma')\delta'(\log z)^{\gamma'} + \alpha' + \frac{\beta'}{\log(\log z)} + \frac{\beta'}{\log^2(\log z)} - \frac{3(\log z)^2}{(1+\log z)^2} \right].$$

The term in brackets goes to $+\infty$ as $z \to \infty$ if $\gamma'(1-\gamma')\delta' > 0$, that is if $0 < \gamma' < 1$ and $\delta' > 0$, which show that $z^2 \frac{d^2}{dz^2} \log f^*(z) + z \frac{d}{dz} \log f^*(z) \geq z^2 \frac{d^2}{dz^2} \log g(z) + z \frac{d}{dz} \log g(z) \geq 0$ if z is large enough. Then conditions C2 and C3 are satisfied if $\gamma' > 0, \delta' > 0$.

Consider now the second case, when $\gamma' = 0, \delta' = 0$ and $g(z) = f^*(z)$. If the normalizing constant of f^* is denoted by K_2, then

$$\lim_{z \to \infty} \frac{f^2(\sqrt{z})}{\sqrt{z} f(z) g(\sqrt{z})} = \lim_{z \to \infty} \frac{(\sqrt{z} f(\sqrt{z}))^2}{z f(z) \sqrt{z} f^*(\sqrt{z})}$$
$$= \lim_{z \to \infty} \frac{\sqrt{z} f(\sqrt{z})}{z f(z)} \frac{K_1 p(\log \sqrt{z} | 0, 0, \alpha', \beta', z_0')}{K_2 p(\log \sqrt{z} | 0, 0, \alpha', \beta', z_0')}$$
$$= \lim_{z \to \infty} \frac{K_1 \sqrt{z} f(\sqrt{z})}{K_2 z f(z)}$$
$$= \lim_{z \to \infty} \frac{K_1}{K_2} \frac{(\log \sqrt{z})^{-\alpha'} \log^{-\beta'}(\log \sqrt{z})}{(\log z)^{-\alpha'} \log^{-\beta'}(\log z)}$$
$$= \lim_{z \to \infty} \frac{K_1}{K_2} 2^{\alpha'} \left(\frac{\log(\log \sqrt{z})}{\log(\log z)} \right)^{-\beta'}$$
$$= \lim_{z \to \infty} \frac{K_1}{K_2} 2^{\alpha'} \left(\frac{\log(\log z) - \log 2}{\log(\log z)} \right)^{-\beta'}$$
$$= \frac{K_1}{K_2} 2^{\alpha'}.$$

Furthermore, if $z > z_0$, it can be shown that

$$z^2 \frac{d^2}{dz^2} \log f^*(z) + z \frac{d}{dz} \log f^*(z) = \frac{1}{(\log z)^2} \left[\alpha' + \frac{\beta'}{\log(\log z)} + \frac{\beta'}{\log^2(\log z)} \right].$$

The term in brackets converge to α' as $z \longrightarrow \infty$. Since f^* is a proper density and $\gamma' = \delta' = 0$, it follows that $\alpha' \geq 1$, which show that $z^2 \frac{d^2}{dz^2} \log f^*(z) + z\frac{d}{dz} \log f^*(z) \geq z^2 \frac{d^2}{dz^2} \log g(z) + z\frac{d}{dz} \log g(z) \geq 0$ if z is large enough. Then conditions C2 and C3 are also satisfied if $\gamma' = 0, \delta' = 0$.

Consider now equations (3.3.1) and (3.3.2) in Theorem 4. Using conditions in Theorem 3, equation (3.3.2) can be rewritten as follows, if $m < n$.

$$\lim_{\sigma \to \infty} \frac{\sigma \pi_\sigma(\sigma) \prod_{i=1}^{k} \frac{1}{\sigma} f_i(\frac{x_i}{x_0 \sigma})}{\prod_{i=m+1}^{n} \sigma f_i(\sigma)} = 0$$

$$\Leftrightarrow \lim_{\sigma \to \infty} \frac{\sigma \pi_\sigma(\sigma) \prod_{i=1}^{k} \frac{1}{\sigma} f_i(\frac{1}{\sigma})}{\prod_{i=m+1}^{n} \sigma f_i(\sigma)} \frac{\prod_{i=1}^{k} \frac{1}{\sigma} f_i(\frac{x_i}{x_0 \sigma})}{\prod_{i=1}^{k} \frac{1}{\sigma} f_i(\frac{1}{\sigma})} = 0$$

$$\Leftrightarrow \lim_{\sigma \to \infty} \frac{\sigma \pi_\sigma(\sigma) \prod_{i=1}^{k} \frac{1}{\sigma} f_i(\frac{1}{\sigma})}{\prod_{i=m+1}^{n} \sigma f_i(\sigma)} = 0$$

(C1 is used since log-cred$^{-}(f_i) = (\gamma', \delta', \alpha', \beta'), i = 1, ..., n,$ and $\gamma' < 1$)

$$\Leftrightarrow \lim_{\sigma \to \infty} \frac{\sigma q(\sigma | \gamma, \delta, \alpha, \beta, z_0)(\frac{1}{\sigma} q(\frac{1}{\sigma} | \gamma', \delta', \alpha', \beta', z_0'))^k}{(\sigma q(\sigma | \gamma', \delta', \alpha', \beta', z_0'))^{n-m}} = 0$$

$$\Leftrightarrow \lim_{\sigma \to \infty} \frac{q(\sigma | \gamma, \delta, \alpha, \beta, z_0)}{\frac{1}{\sigma}(\sigma q(\sigma | \gamma', \delta', \alpha', \beta', z_0'))^{n-m-k}} = 0, \tag{3.6.2}$$

where $q(\sigma | \gamma, \delta, \alpha, \beta, z_0)$ and $q(\sigma | \gamma', \delta', \alpha', \beta', z_0')$ are log-GEP densities with respectively the same left and right log-credences as π_σ and f_i, for any $z_0 > 1$ and $z_0' > 1$. The symmetry of the log-GEP density about 1 is used in the last equation, that is $\sigma q(\sigma | \gamma', \delta', \alpha', \beta', z_0') = \frac{1}{\sigma} q(\frac{1}{\sigma} | \gamma', \delta', \alpha', \beta', z_0')$. In the same way, equation (3.3.1) can be rewritten as follows, if $k < m$.

$$\lim_{\sigma \to 0} \frac{\sigma \pi_\sigma(\sigma) \prod_{i=1}^{k} \frac{1}{\sigma} f_i(\frac{x_i}{x_0 \sigma})}{\prod_{i=m+1}^{n} \sigma f_i(\sigma)} = 0$$

$$\Leftrightarrow \lim_{\sigma \to 0} \frac{\sigma q(\sigma | \gamma, \delta, \alpha, \beta, z_0)(\frac{1}{\sigma} q(\frac{1}{\sigma} | \gamma', \delta', \alpha', \beta', z_0'))^k}{(\sigma q(\sigma | \gamma', \delta', \alpha', \beta', z_0'))^{m-k}} = 0$$

$$\Leftrightarrow \lim_{\sigma \to 0} \frac{q(\sigma | \gamma, \delta, \alpha, \beta, z_0)}{\frac{1}{\sigma}(\sigma q(\sigma | \gamma', \delta', \alpha', \beta', z_0'))^{m-2k}} = 0. \tag{3.6.3}$$

As long as there are more observations around x_0 than outliers on the left of x_0 (that is $m - 2k \leq 0$) and there are more observations than outliers on the right of x_0 (that is $n - m - k \leq 0$), equations (3.3.1) and (3.3.2) are satisfied, whatever the left and right log-credences are, since the tails of $\sigma h(\sigma)$ go to 0, for any proper log-GEP densities h. This is equivalent to $k \geq \max[m-k, n-m]$. For instance, if $k=a, m=2a, n=3a$, where a is any positive integer, then $k=n/3$, $m-k=n/3$ and $n-m=n/3$, which means that $k \geq \max[m-k, n-m]$ is satisfied. In this case, the posterior can reject up to one third of the observations as left outliers plus another one third as right outliers, for a total of $\frac{2}{3}n$ outliers.

However Theorem 3 considers that the direction of outliers is unpredictable, so the results must hold even for the extreme cases where all the $n-k$ outliers are on the right of x_0 $(m=k)$ or on the left of x_0 $(m=n)$. Equation (3.6.2) when $m=k$ and equation (3.6.3) when $m=n$ become one unique equation given by

$$\lim_{\sigma \to 0} \frac{q(\sigma|\gamma, \delta, \alpha, \beta, z_0)}{\frac{1}{\sigma}(\sigma q(\sigma|\gamma', \delta', \alpha', \beta', z_0'))^{n-2k}} = 0. \tag{3.6.4}$$

It can be shown that if equation (3.6.4) is satisfied, then equations (3.3.1) and (3.3.2) are also satisfied for any values of k and m such that $0 \leq k \leq m \leq n$. Consider now two cases : $k \geq n/2$ and $k < n/2$. Since $k \geq n/2 \Leftrightarrow n-2k \leq 0$, equation (3.6.4) is satisfied if $k \geq n/2$ for any log-credences, which corresponds to condition i) in Theorem 3.

If $k < n/2$, or equivalently $n-2k > 0$, then according to Proposition 3, equation (3.6.4) is satisfied if

$$\begin{aligned} &\text{log-cred}^+(q(\sigma|\gamma, \delta, \alpha, \beta, z_0)) > \text{log-cred}^+(\frac{1}{\sigma}(\sigma q(\sigma|\gamma', \delta', \alpha', \beta', z_0'))^{n-2k}) \\ &\Leftrightarrow (\gamma, \delta, \alpha, \beta) > (\gamma', (n-2k)\delta', (n-2k)\alpha', (n-2k)\beta'). \end{aligned}$$

The last equivalence is true since it can be shown that

$$\text{log-cred}^+\left(\frac{1}{\sigma}\prod_{i=1}^{n} \sigma f_i(\sigma)\right) = (\gamma, \sum_{i=1}^{n} \delta_i, \sum_{i=1}^{n} \alpha_i, \sum_{i=1}^{n} \beta_i)$$

if $\text{log-cred}^+(f_i) = (\gamma, \delta_i, \alpha_i, \beta_i)$. It means that equation (3.6.4) is satisfied if $k < n/2$ and $(\gamma, \delta, \alpha, \beta) > (\gamma', (n-2k)\delta', (n-2k)\alpha', (n-2k)\beta')$, which corresponds to condition ii) in Theorem 3.

Finally, result e) of Theorem 3 is a corollary of result d), which says that

$$\lim_{\phi_1 \to \infty} \Pr[\sigma/x_0 \leq d|\underline{x}_n] = \Pr[\sigma/x_0 \leq d|\underline{x}_k],$$

for all d where $\Pr[\sigma/x_0 \leq d|\underline{x}_k]$ is continuous. Let $(\sigma/x_0)|\underline{x}_n$ be denoted by $Y|\underline{x}_n$ and $(\sigma/x_0)|\underline{x}_k$ be denoted by $Y|\underline{x}_k$, then the density of $Y|\underline{x}_n$ is $x_0\pi(yx_0|\underline{x}_n)$ and the density of $Y|\underset{\sim}{x}_k$ is $x_0\pi(yx_0|\underset{\sim}{x}_k)$. Then result d) is equivalent to

$$\lim_{\underset{\sim}{\phi}_1\to\infty} \Pr[Y \le d|\underset{\sim}{x}_n] = \Pr[Y \le d|\underset{\sim}{x}_k],$$

for all d where $\Pr[Y \le d|\underset{\sim}{x}_k]$ is continuous.

If $d = Q^{x_0\pi(yx_0|\underset{\sim}{x}_k)}(p)$, then

$$\lim_{\underset{\sim}{\phi}_1\to\infty} \Pr[Y \le Q^{x_0\pi(yx_0|\underset{\sim}{x}_k)}(p)|\underset{\sim}{x}_n] = \Pr[Y \le Q^{x_0\pi(yx_0|\underset{\sim}{x}_k)}(p)|\underset{\sim}{x}_k]$$

$$\Leftrightarrow \lim_{\underset{\sim}{\phi}_1\to\infty} \Pr[Y \le Q^{x_0\pi(yx_0|\underset{\sim}{x}_k)}(p)|\underset{\sim}{x}_n] = p.$$

Furthermore, by definition we have

$$\Pr[Y \le Q^{x_0\pi(yx_0|\underset{\sim}{x}_n)}(p)|\underset{\sim}{x}_n] = p$$

$$\Leftrightarrow \lim_{\underset{\sim}{\phi}_1\to\infty} \Pr[Y \le Q^{x_0\pi(yx_0|\underset{\sim}{x}_n)}(p)|\underset{\sim}{x}_n] = p,$$

which means that $Q^{x_0\pi(yx_0|\underset{\sim}{x}_n)}(p) \to Q^{x_0\pi(yx_0|\underset{\sim}{x}_k)}(p)$ as $\underset{\sim}{\phi}_1 \to \infty$.

3.7. Acknowledgments

The authors are thankful to the NSERC (Natural Sciences and Engineering Research Council of Canada), the FQRNT (Le Fonds québécois de la recherche sur la nature et les technologies) and the SOA (Society of Actuaries) for their financial support.

3.8. References

ANGERS, J.-F. (1992) Use of Student-t prior for the estimation of normal means : A computational approach, *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, *IMS Lecture Notes - Monograph Series*, Vol. **29**, pp. 273-283.

ANGERS, J.-F. (2000) P-Credence and outliers, *Metron*, **58**, 81-108.

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

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

CLEMENT, R.T. (1989) Combining forecasts : A review and an annotated bibliography, *International Journal of Forecasting*, **5**, 559-583.

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

DE FINETTI, B. (1961) The Bayesian approach to the rejection of outliers, *Proceedings of the Fourth Berkeley Symposium on Probability and Statistics*, **(Vol. 1)**, Berkeley : University of California Press, pp. 199-210.

DESGAGNÉ, A. et ANGERS, J.-F. (2005) Importance Sampling with the Generalized Exponential Power Density, *Statistics and Computing*, **15**, 189-195.

DESGAGNÉ, A. and ANGERS, J.-F. (2005b), Outliers and choice of the prior for location parameter inference, to appear in *Ph.D. Thesis of Alain Desgagné*, submitted to *Metron*.

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.M. (1974) On coherence, inadmissibility and inference about many parameters in the theory of least squares, *Studies, Bayesian Econometrics and Statistics*, eds. S. E. Fienberg and A. Zellner, 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.

MIN, C.-K. and ZELLNER, A. (1993) Bayesian and non-Bayesian methods for combining models and forecasts with applications to forecasting international growth rates, *Journal of Econometrics*, **56**, 89-118.

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, *Bayesian Statistic III*, eds. Bernardo, J.M., DeGroot M.H., Lindley, D.V., 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.

REISS, R.-D. and THOMAS, M. (1997) *Statistical analysis of extreme values*, Birkhauser Verlag, Basel.

CONCLUSION

Dans cette thèse, nous avons développé des méthodes statistiques bayésiennes robustes, qui demeurent efficaces même si des valeurs aberrantes viennent contaminer les données, à l'aide des densités à ailes relevées. À cette fin, nous avons utilisé la densité GEP, d'abord comme fonction d'importance, puis comme densité de référence. Cette densité définie sur les réels a des ailes ayant un comportement de type exponentiel, polynomial et logarithmique, ce qui permet d'obtenir des densités à ailes relevées et même très relevées (super heavy tails). La plupart des densités connues ont des ailes ayant le même comportement que celles de la densité GEP. Pour la modélisation des observations sur les réels positifs, nous avons utilisé la densité log GEP comme densité de référence.

Dans le premier article, la densité GEP a été proposée comme fonction d'importance dans les simulations Monte Carlo dans le contexte de l'estimation des moments *a posteriori* d'un paramètre de position. Il peut être difficile de choisir une fonction d'importance appropriée et souvent cela doit être fait sur la base du cas par cas. Pour toute densité *a priori* et toute vraisemblance dont la p-crédence est définie, les paramètres de la densité GEP sont obtenus par des équations données dans cet article, en fonction de n'importe quel échantillon. Le choix de la densité GEP nous permet d'obtenir des résultats fiables, même s'il y a des sources d'information conflictuelles. De plus, puisque la p-crédence de la densité GEP est légèrement inférieure à celle de la densité *a posteriori*, les simulations Monte Carlo demeurent efficaces. Finalement, nous montrons comment simuler des observations provenant de la densité GEP à l'aide de la méthode de transformation inverse.

Dans le deuxième article, le comportement de la densité *a posteriori* d'un paramètre de position a été étudié lorsqu'un échantillon comprend des valeurs aberrantes. La notion de p-crédence à gauche et à droite a été introduite afin de caractériser respectivement les ailes de gauche et de droite d'une densité. Des conditions simples sur les ailes de la densité *a priori* et de la vraisemblance, en utilisant la p-crédence à gauche et à droite, ont été établies afin de déterminer la proportion d'observations pouvant être rejetées lorsque celles-ci sont extrêmes. Nous avons démontré que la densité *a posteriori* converge en loi vers la densité *a posteriori* obtenue à partir d'un échantillon excluant les valeurs aberrantes, à mesure que celles-ci tendent vers plus ou moins l'infini, à n'importe quel taux. Un exemple de combinaison de prévisions du rendement de l'indice financier S&P 500 a été présenté. Nous avons également généralisé les conditions de convergence afin d'inclure des densités dont la p-crédence n'est pas nécessairement définie.

Dans le troisième article, le comportement de la densité *a posteriori* d'un paramètre d'échelle a été étudié lorsqu'un échantillon d'observations positives comprend des valeurs aberrantes. La notion de log-crédence a été introduite afin de caractériser respectivement les ailes de gauche et de droite d'une densité définie sur $\mathbb{R}^+$. Des conditions simples sur les ailes de la densité *a priori* et de la vraisemblance, semblables à celles du précédent article sur le paramètre de position mais utilisant la log-crédence à gauche et à droite, ont été établies afin de déterminer la proportion d'observations pouvant être rejetées lorsque celles-ci sont extrêmes. Nous avons démontré que la densité *a posteriori* converge en loi vers la densité *a posteriori* obtenue à partir d'un échantillon excluant les valeurs aberrantes, à mesure que celles-ci tendent vers 0 ou l'infini, à n'importe quel taux. Un exemple de combinaison de prévisions de la volatilité des rendements de l'indice financier S&P 500 a été présenté. Les conditions de convergence ont également été généralisées afin d'inclure des densités dont la log-crédence n'est pas nécessairement définie.