

2004s-31

**Estimation de densité
conditionnelle lorsque l'hypothèse
de normalité est insatisfaisante**

Julie Carreau, Yoshua Bengio

Série Scientifique
Scientific Series

Montréal
Mai 2004

© 2004 Julie Carreau, Yoshua Bengio. Tous droits réservés. *All rights reserved.* Reproduction partielle permise avec citation du document source, incluant la notice ©.
Short sections may be quoted without explicit permission, if full credit, including © notice, is given to the source.

CIRANO

Le CIRANO est un organisme sans but lucratif constitué en vertu de la Loi des compagnies du Québec. Le financement de son infrastructure et de ses activités de recherche provient des cotisations de ses organisations-membres, d'une subvention d'infrastructure du ministère de la Recherche, de la Science et de la Technologie, de même que des subventions et mandats obtenus par ses équipes de recherche.

CIRANO is a private non-profit organization incorporated under the Québec Companies Act. Its infrastructure and research activities are funded through fees paid by member organizations, an infrastructure grant from the Ministère de la Recherche, de la Science et de la Technologie, and grants and research mandates obtained by its research teams.

Les organisations-partenaires / The Partner Organizations

PARTENAIRE MAJEUR

- . Ministère du développement économique et régional et de la recherche [MDERR]

PARTENAIRES

- . Alcan inc.
- . Axa Canada
- . Banque du Canada
- . Banque Laurentienne du Canada
- . Banque Nationale du Canada
- . Banque Royale du Canada
- . Bell Canada
- . BMO Groupe Financier
- . Bombardier
- . Bourse de Montréal
- . Caisse de dépôt et placement du Québec
- . Développement des ressources humaines Canada [DRHC]
- . Fédération des caisses Desjardins du Québec
- . GazMétro
- . Hydro-Québec
- . Industrie Canada
- . Ministère des Finances du Québec
- . Pratt & Whitney Canada Inc.
- . Raymond Chabot Grant Thornton
- . Ville de Montréal

- . École Polytechnique de Montréal
- . HEC Montréal
- . Université Concordia
- . Université de Montréal
- . Université du Québec à Montréal
- . Université Laval
- . Université McGill
- . Université de Sherbrooke

ASSOCIE A :

- . Institut de Finance Mathématique de Montréal (IFM²)
- . Laboratoires universitaires Bell Canada
- . Réseau de calcul et de modélisation mathématique [RCM²]
- . Réseau de centres d'excellence MITACS (Les mathématiques des technologies de l'information et des systèmes complexes)

Les cahiers de la série scientifique (CS) visent à rendre accessibles des résultats de recherche effectuée au CIRANO afin de susciter échanges et commentaires. Ces cahiers sont écrits dans le style des publications scientifiques. Les idées et les opinions émises sont sous l'unique responsabilité des auteurs et ne représentent pas nécessairement les positions du CIRANO ou de ses partenaires.

This paper presents research carried out at CIRANO and aims at encouraging discussion and comment. The observations and viewpoints expressed are the sole responsibility of the authors. They do not necessarily represent positions of CIRANO or its partners.

Estimation de densité conditionnelle lorsque l'hypothèse de normalité est insatisfaisante

Julie Carreau^{*}, Yoshua Bengio[†]

Résumé / Abstract

Nous cherchons à modéliser des densités dont la distribution est inconnue mais qui est asymétrique et présente des queues lourdes. Dans ce contexte, l'hypothèse de normalité n'est pas appropriée. Afin de maintenir au minimum le nombre d'hypothèses distributionnelles, nous utilisons une méthode non paramétrique pour modéliser le centre de la distribution. La modélisation est plus difficile dans les queues de la distribution puisque peu d'observations s'y trouvent. Nous nous proposons donc d'utiliser la Pareto généralisée (GPD) pour modéliser les queues de la distribution. La GPD permet d'approximer tous les types de queues de distributions (qu'elles soient finies, exponentielles ou sous-exponentielles). L'estimation des paramètres de la GPD est uniquement basée sur les observations extrêmes. Une observation est définie comme étant extrême si elle dépasse un seuil donné. La principale difficulté de la modélisation avec la GPD réside dans le choix d'un seuil adéquat.

Mots clés : distribution à queue épaisse, Pareto généralisée, estimation de densité conditionnelle.

We aim at modelling fat-tailed densities whose distributions are unknown but are potentially asymmetric. In this context, the standard normality assumption is not appropriate. In order to make as few distributional assumptions as possible, we use a non-parametric algorithm to model the center of the distribution. Density modelling becomes more difficult as we move further in the tail of the distribution since very few observations fall in the upper tail area. Hence we decide to use the generalized Pareto distribution (GPD) to model the tails of the distribution. The GPD can approximate finite, exponential or subexponential tails. The estimation of the parameters of the GPD is based solely on the extreme observations. An observation is defined as being extreme if it is greater than a given threshold. The main difficulty with GPD modelling is to determine an appropriate threshold.

Keywords: fat-tailed distribution, generalized Pareto, conditional density estimation.

^{*} Julie Carreau, Université de Montréal, carreau@iro.umontreal.ca

[†] Yoshua Bengio, Université de Montréal et CIRANO, (514) 343-6804, Yoshua.Bengio@cirano.qc.ca

1 Introduction

Dans le cadre de la régression, les données ont la forme de paires (X, Y) où X est de dimension d et Y est un scalaire. Nous disposons donc d'un ensemble de n observations : $\{(x_1, y_1), \dots, (x_n, y_n)\}$ générées par un processus aléatoire sous-jacent inconnu. Nous cherchons à apprendre une fonction $f(x)$ qui minimise l'erreur quadratique $E[(Y - f(X))^2]$ sur les données d'apprentissage. La fonction f qui minimise l'erreur quadratique est l'espérance conditionnelle $E[Y|X = x]$.

En pratique, nous nous contentons de considérer un ensemble de fonctions paramétrisées $f(x, \theta)$, où θ est le vecteur de paramètres et de chercher θ^* tel que $f(x, \theta^*)$ est proche de $E[Y|X = x]$. Le modèle sous-jacent aux données a la forme suivante :

$$Y = f(X, \theta) + \epsilon,$$

où ϵ est le bruit du modèle dont la distribution est inconnue.

L'hypothèse classique est de supposer que le bruit est distribué selon une loi Normale ($\epsilon \sim \mathcal{N}(0, \sigma^2)$). Cette hypothèse est souvent inappropriée. En effet, la normalité du bruit implique que la distribution modélisée soit symétrique, ait pour support l'axe des réels et que les queues décroissent exponentiellement.

Dans le domaine de l'assurance, les réclamations des assurés sont positives, asymétriques et la présence d'observations extrêmes suggèrent que les queues de la distribution sont sous-exponentielles. La distribution des grandes réclamations (en échelle logarithmique) pour des pertes dues au feu au Danemark est un cas évident de non-normalité, voir la figure 1. La densité empirique est lissée par un noyau gaussien. Ces données sont fournies avec la librairie Evir, voir [1], accompagnant le logiciel gratuit R.

En finance, les rendements de divers actifs est souvent asymétrique du côté négatif ou positif, dépendant des périodes, par exemple à cause de crashes boursiers ou de mouvements de tendance comme la bulle technologique. De nombreux auteurs, entre autres Fama [6] et Mandelbrot [4], ont remarqué que les distributions des titres boursiers présentent des queues épaisses. Les distributions des rendements logarithmiques des parts de BMW, dans le cas positif et négatif, montrent clairement des signes d'asymétrie et de queues légèrement plus épaisses que pour une normale, voir la figure 2. Ces données proviennent également de la librairie Evir.

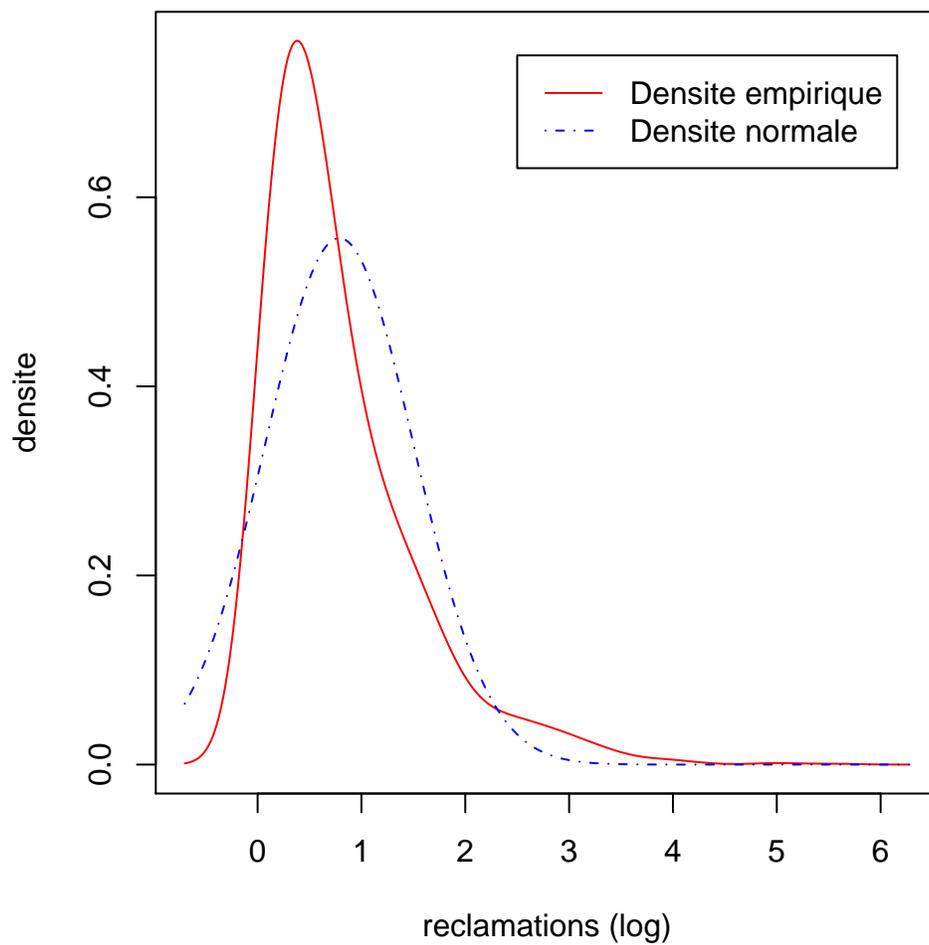


FIG. 1 – Réclamations danoises

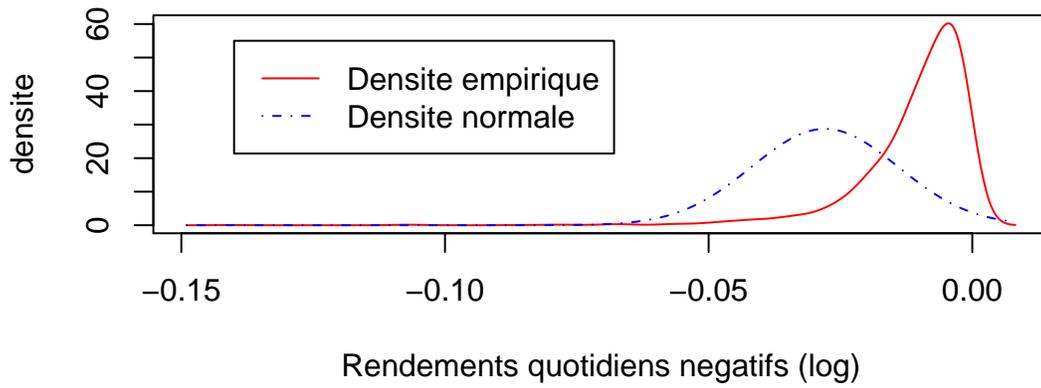
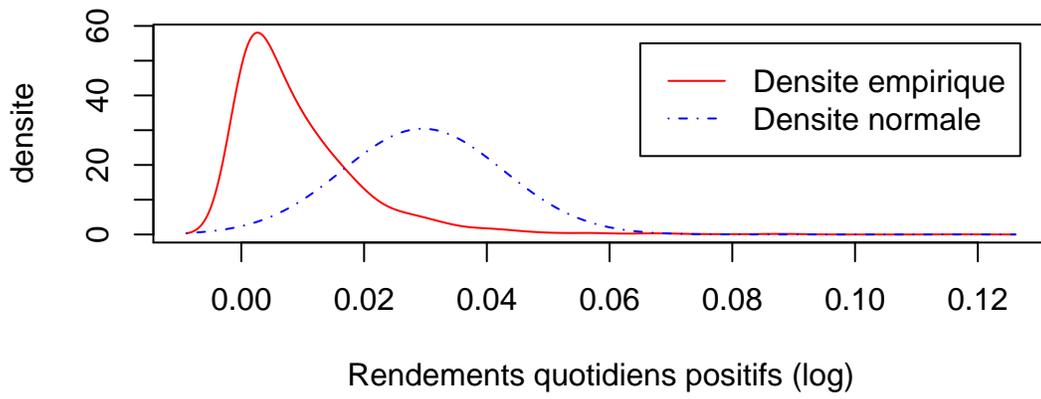


FIG. 2 – Rendements quotidiens des parts de BMW

Toujours en finance, la valeur-à-risque (VaR) est couramment utilisée pour mesurer les risques des instruments financiers (portefeuilles, titres financiers, etc ...). Il s'agit en fait d'un quantile de la distribution des profits et pertes d'un instrument financier p_t à un moment t donné pour un intervalle de temps fixe :

$$P(p_t \leq -\text{VaR}) = \alpha\%.$$

La VaR est donc la perte maximale risquée pour l'instrument financier avec une certitude $\alpha\%$. Les praticiens du domaine financier utilisent souvent l'hypothèse de normalité pour estimer la VaR. Si la distribution de l'instrument financier a en réalité des queues épaisses, l'estimation d'un quantile extrême comme la VaR sera biaisée et la VaR sera sous-estimée. La figure 3 illustre le biais introduit par le choix d'une distribution normale lorsque la vraie distribution est asymétrique pour l'estimation de quantile. Les lignes verticales indiquent où se trouvent les quantiles de 1% pour chaque distribution.

Dans le contexte de ces applications, nous nous intéressons aux problèmes de régression dans la cas où la distribution du bruit est inconnue, présente des queues épaisses et est asymétrique. Il n'est pas possible de simplement éliminer les observations extrêmes puisque celles-ci ne proviennent pas d'erreurs de mesure ; ces observations ont été générées par le processus aléatoire sous-jacent. Les méthodes de régression robuste utilisées en statistique, voir [9] ou [7], supposent une distribution symétrique du bruit.

La distribution de Pareto généralisée permet de modéliser les excès au-delà d'un seuil donné. Cette distribution apparaît comme limite de la fonction de répartition excédentaire :

$$F_u(x) = P(X - u \leq x | X > u),$$

lorsque $u \rightarrow x_f$ où $x_f = \sup_x \{F(x) < 1\}$ et F , la fonction de répartition correspondant à F_u , est dans le domaine d'attraction maximal d'une loi aux valeurs extrêmes.

La Pareto généralisée est donc appropriée pour modéliser les queues des distributions. Elle peut tenir compte des queues finies, exponentielles ou lourdes. Seules les données extrêmes sont utilisées pour estimer les paramètres de la Pareto généralisée ce qui assure un meilleur ajustement du modèle à la queue de la distribution. La difficulté dans l'utilisation de la Pareto généralisée est de déterminer un seuil adéquat lorsque les données sont en nombre fini.

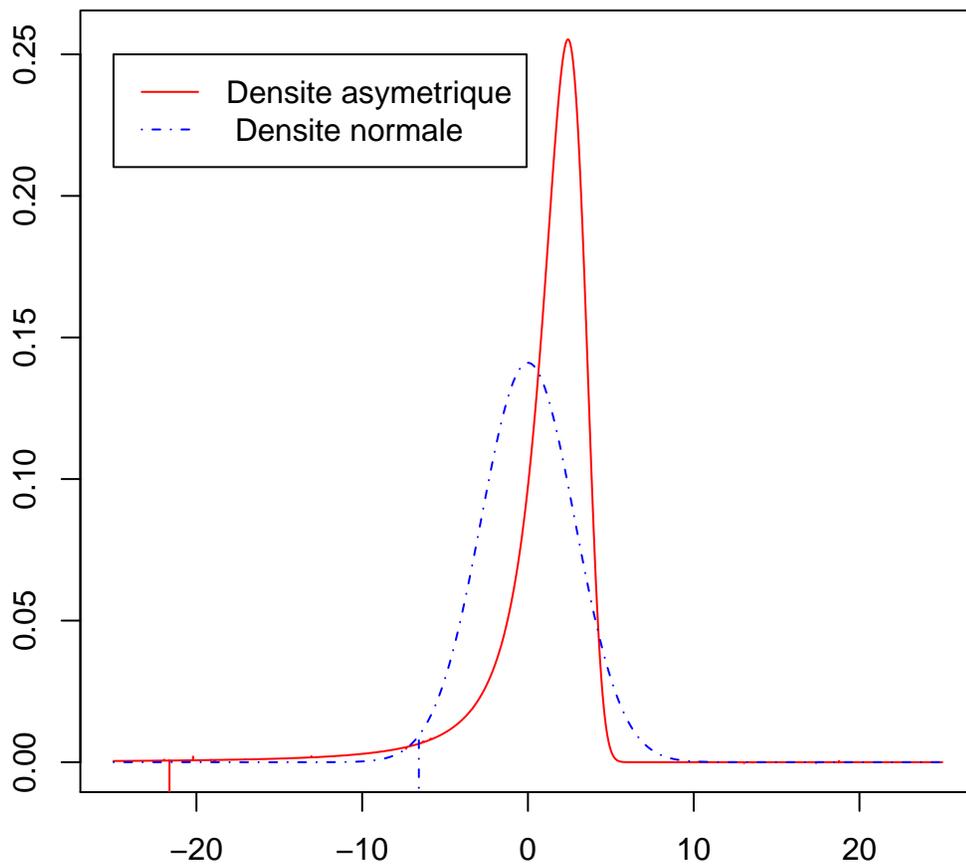


FIG. 3 – L'estimation de quantile de 1% sous une hypothèse distributionnelle erronée peut sous-estimer ce quantile.

La famille des lois α -stables est aussi utilisée pour modéliser la distribution d'observations présentant une queue épaisse. Les lois stables permettent de modéliser la partie centrale de la distribution en limitant l'influence des valeurs extrêmes. Par contre, l'ajustement d'une loi stable ne fournit pas nécessairement une bonne approximation de la queue de la distribution.

Dans la section 2, la théorie reliée aux lois stables, aux distributions de valeurs extrêmes et à la Pareto généralisée est brièvement passée en revue. À la section suivante 3, nous présentons quelques alternatives possibles pour modéliser une densité conditionnelle à ailes lourdes ou pour faire de la régression en présence d'observations extrêmes. Ensuite, dans la section 4, nous proposons une approche différente basée sur la Pareto généralisée, pour modéliser la densité conditionnelle en présence d'observations extrêmes. Quelques expériences préliminaires illustrent cette approche dans la section 5. Nous proposons finalement des extensions possibles de cette méthode dans la section 6.

2 Pareto généralisée

La loi de Pareto généralisée (**GPD**) apparaît dans le contexte de la théorie des valeurs extrêmes. Cette théorie cherche à modéliser les maxima ou minima d'une variable aléatoire. Des analogies intéressantes existent entre la convergence de maxima (minima) de variables aléatoires et la convergence de sommes de variables aléatoires qui mène à la théorie des lois stables.

Dans cette section, un bref aperçu de la théorie des lois stables est présenté dans la sous-section 2.1, suivi par son analogue concernant la théorie des valeurs extrêmes 2.2. Finalement, les propriétés de la distribution de Pareto généralisée sont décrites dans la sous-section 2.3 et les applications pour l'estimation de queue de distribution sont décrites dans la sous-section 2.4.

2.1 Lois stables

Les lois stables sont les uniques lois limites non-dégénérées¹ de sommes de variables aléatoires centrées réduites de façon appropriée. La loi Normale est un cas particulier de lois stables qui comprennent des distributions en forme de

¹Une variable aléatoire est dégénérée si toute la masse de probabilité est concentrée en un seul point.

cloche, symétrique ou non, avec des queues de décroissance exponentielle et sous-exponentielle. La figure 4 illustre les différences entre la loi normale en rouge (correspondant à $\alpha = 2$ et $\beta = 0$) et des lois asymétriques ayant des queues sous-exponentielles (ici ce sont des lois α -stables avec $\alpha < 2$).

Les lois α -stable peuvent être définies par la propriété de **stabilité** de l'addition :

Définition 2.1 Soient X, X_1, X_2 , des variables aléatoires i.i.d. non-dégénérées dont la fonction de répartition est F . F est dite **stable** si elle satisfait l'identité en loi suivante :

$$C_1 X_1 + C_2 X_2 \stackrel{d}{=} b(C_1, C_2) X + a(C_1, C_2)$$

pour $C_1, C_2 \geq 0$ et $b(C_1, C_2)$ et $a(C_1, C_2)$ réels.

Il est possible de montrer que la propriété de stabilité d'une variable aléatoire X dans la définition 2.1 équivaut au fait que X ait la fonction caractéristique suivante :

$$\phi_X(t; \alpha, \beta, c, \mu) = E[e^{itX}] = \exp \left\{ i\mu t - |ct|^\alpha \left[1 - i\beta \frac{t}{|t|} \omega(|t|, \alpha) \right] \right\},$$

où

$$\omega(|t|, \alpha) = \begin{cases} \tan \frac{\pi\alpha}{2} & \text{if } \alpha \neq 1, \\ \frac{-2}{\pi} \log |t| & \text{if } \alpha = 1. \end{cases}$$

La fonction de densité d'une loi stable se retrouve par la transformée de Fourier de la fonction caractéristique car celle-ci est absolument intégrable sur \mathbb{R} .

Les paramètres de la fonction caractéristique d'une loi α -stable décrivent les propriétés suivantes de la distribution :

- $\alpha \in (0, 2]$ est l'exposant caractéristique qui contrôle l'épaisseur de la queue,
- $\beta \in [-1, 1]$ contrôle l'asymétrie ($\beta = 0$ est une distribution symétrique),
- $\mu \in \mathbb{R}$ contrôle l'emplacement et
- $c > 0$ contrôle la dispersion.

Mise à part quelques cas particuliers (à savoir la loi Normale, la loi de Cauchy et la loi de Lévy), il n'y a pas de forme fermée pour la fonction de densité des lois α -stables.

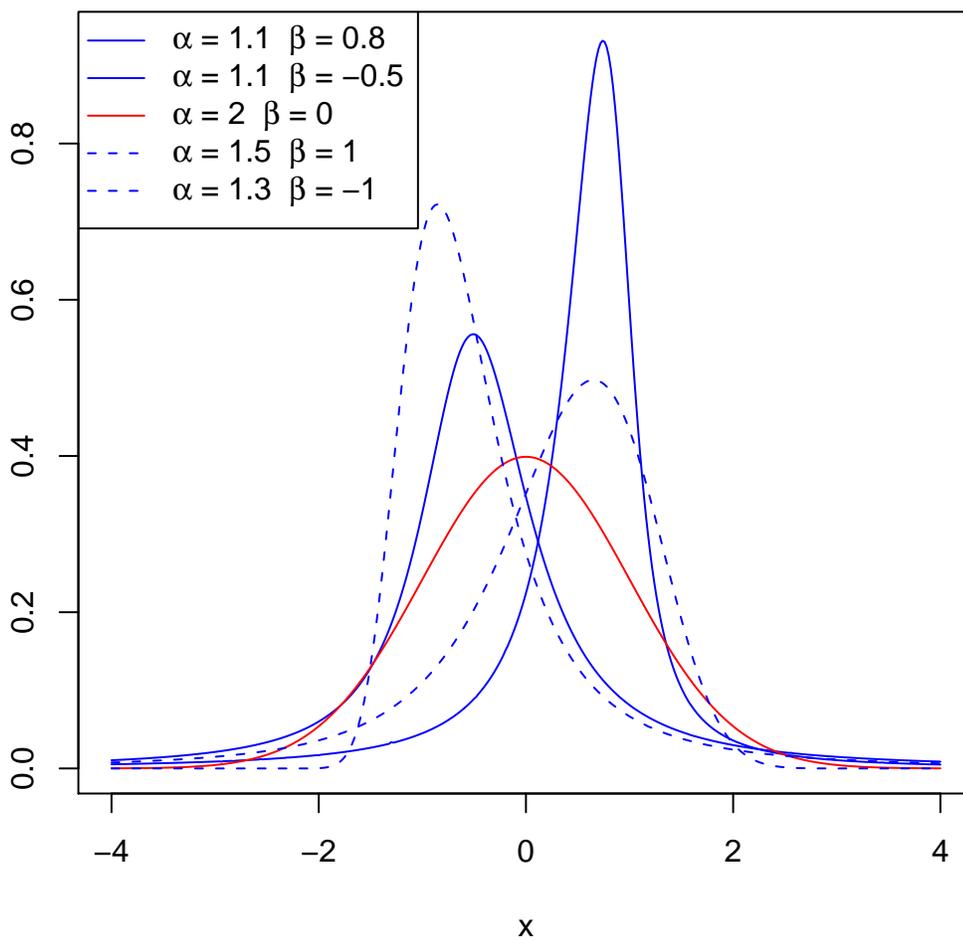


FIG. 4 – Lois α -stables

2.1.1 Domaine d'attraction

Le domaine d'attraction d'une loi α -stable est une forme de généralisation du théorème de la limite centrale pour la loi Normale. Il est défini de la façon suivante :

Définition 2.2 Soient X, X_1, X_2, \dots, X_n , des variables aléatoires i.i.d. non-dégénérées avec fonction de répartition commune F et soit $S_n = \sum_{i=1}^n X_i$. La variable aléatoire X est dans le **domaine d'attraction de \mathcal{S}_α** , ce qui se note $X \in DA(\mathcal{S}_\alpha)$ ou encore $X \in DA(\alpha)$, si $\exists b_n > 0$ et $a_n \in \mathbb{R}$ tels que

$$\frac{S_n - a_n}{b_n} \xrightarrow{d} \mathcal{S}_\alpha.$$

Si $X \in DA(\alpha)$, alors les moments de X ont les caractéristiques suivantes :

$$\begin{aligned} E|X|^\delta &< \infty && \text{pour } \delta \leq \alpha, \\ E|X|^\delta &= \infty && \text{pour } \delta > \alpha \text{ et } \alpha < 2. \end{aligned}$$

2.2 Théorie des valeurs extrêmes

La théorie des valeurs extrêmes classique cherche à comprendre le comportement des maxima ou minima d'une variable aléatoire. La théorie est présentée pour les maxima, celle-ci étant symétrique pour les minima. Les distributions de valeurs extrêmes sont définies par la propriété de stabilité par rapport au maximum :

Définition 2.3 Soient X, X_1, \dots, X_n , des variables aléatoires i.i.d. dont la fonction de répartition est F . Posons $M_n = \max(X_1, \dots, X_n)$. F est dite **max-stable** si elle satisfait l'égalité en distribution suivante, $\forall n > 2$:

$$M_n \stackrel{d}{=} c_n X + d_n,$$

où $c_n > 0$ et $d_n \in \mathbb{R}$

Le théorème fondamental de la théorie des valeurs extrêmes est le théorème de Fisher-Tippett qui stipule les trois lois limites possibles pour les maxima de

variables aléatoires centrés réduits de façon appropriée. Le théorème de Fisher-Tippett est analogue au théorème de la limite centrale et à sa généralisation, voir la section 2.1.1, pour les lois α -stables.

Les trois distributions limites du théorème de Fisher-Tippett sont appelées les distributions de valeurs extrêmes standard. Il y a la distribution de Fréchet $\Phi_\alpha(x)$, la distribution de Weibull $\Psi_\alpha(x)$ et la distribution de Gumbel $\Lambda(x)$. Ces trois distributions sont illustrées à la figure 5.

2.2.1 Domaine d'attraction maximum

Le domaine d'attraction maximum se définit de façon analogue au domaine d'attraction d'une loi stable :

Définition 2.4 *La variable aléatoire X est dans le domaine d'attraction maximum de la fonction de répartition H , ce qui s'écrit $X \in MDA(H)$, si*

$$\exists c_n > 0, d_n \in \mathbb{R} \text{ tels que } c_n^{-1}(M_n - d_n) \xrightarrow{d} H.$$

Il est possible de caractériser le domaine d'attraction maximum de chacune des fonctions de répartition limites du théorème de Fisher-Tippett. Par exemple, pour la loi de Fréchet, nous avons : $X \in MDA(\Phi_\alpha) \Leftrightarrow 1 - F(x) = x^{-\alpha}L(x)$, où $L(x)$ est une fonction à variation lente. De plus, le point extrême supérieur est infini. Les distributions de Pareto et de Cauchy ainsi que les distributions α -stable avec $\alpha < 2$ sont dans le domaine d'attraction maximum de la Fréchet.

2.2.2 Distribution de valeurs extrêmes généralisée

La distribution de valeurs extrêmes généralisée est simplement une reparamétrisation des distributions apparaissant dans le théorème de Fisher-Tippett :

$$H_{\xi;\mu,\beta}(x) = \begin{cases} \exp\left\{-\left(1 + \xi\left(\frac{x-\mu}{\beta}\right)\right)^{-1/\xi}\right\} & \text{si } \xi \neq 0 \\ \exp\left\{-\exp\left\{-\frac{x-\mu}{\beta}\right\}\right\} & \text{si } \xi = 0 \end{cases} \quad (1)$$

Où $1 + \xi\left(\frac{x-\mu}{\beta}\right) > 0$, μ et β sont les paramètres d'emplacement et de dispersion respectivement.

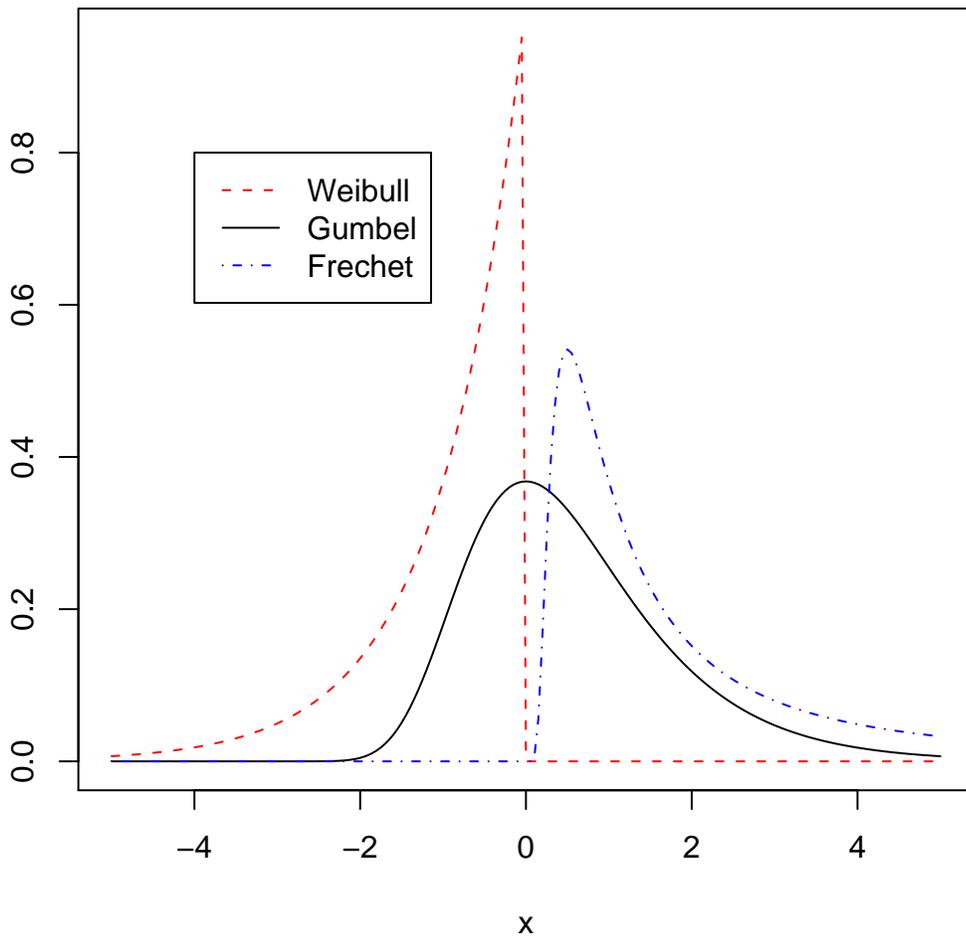


FIG. 5 – Loi des valeurs extrêmes généralisée

2.3 Loi de Pareto généralisée

La fonction de répartition de la Pareto généralisée est donnée, pour $\beta > 0$ et $\xi \in \mathbb{R}$, par :

$$G_{\xi;\beta} = \begin{cases} 1 - (1 + \xi \left(\frac{x}{\beta}\right))^{-1/\xi} & \text{si } \xi \neq 0, \\ 1 - e^{-\frac{x}{\beta}} & \text{si } \xi = 0. \end{cases} \quad (2)$$

où

$$\begin{aligned} x &\geq 0 && \text{si } \xi \geq 0, \\ 0 \leq x &\leq -\frac{\beta}{\xi} && \text{si } \xi < 0. \end{aligned}$$

La fonction de répartition est motivée par le théorème suivant 2.5 , dû à Pickands :

Théorème 2.5 Pour tout $\xi \in \mathbb{R}$:

$$F \in MDA(H_\xi) \Leftrightarrow \lim_{u \uparrow x_F} \sup_{0 < x < x_F - u} |F_u(x) - G_{\xi,\beta(u)}(x)| = 0,$$

pour une fonction positive appropriée β et $F_u(x) = P(X - u \leq x | X > u)$, $x \geq 0$.

La condition $F \in MDA(H_\xi)$ inclut de nombreuses distributions. Le théorème dit que plus on s'intéresse à modéliser l'extrémité de la queue d'une distribution, meilleure est l'approximation par une distribution de Pareto généralisée. La GPD est illustrée à la figure 6 pour différentes valeurs de ξ . C'est ce paramètre ξ qui contrôle l'épaisseur de la queue de la distribution et donc l'existence des moments de la distribution.

2.3.1 Stabilité du seuil

Une variable aléatoire de loi GPD est *stable par rapport au seuil*. C'est-à-dire, soit $X \sim G_{\xi,\beta}$ et soit $u > 0$, alors, pour $x > 0$, $P(X - u \leq x | X > u) = G_{\xi,\beta+\xi u}$.

2.3.2 Estimation des paramètres

Soient les observations $\{x_1, \dots, x_N\}$ et pour un seuil u fixé, soit N_u le nombre d'excès du niveau u . La distribution de Pareto est alors ajustée aux excès $\{y_1, \dots, y_{N_u}\}$ où $y_i = x_{j_i} - u$.

La méthode du maximum de vraisemblance peut être utilisée pour estimer les paramètres de la Pareto généralisée si l'échantillon est assez grand. Par ailleurs,

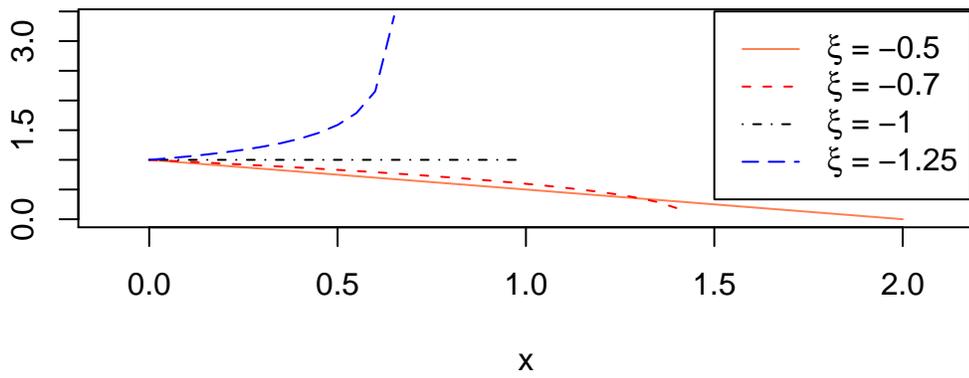
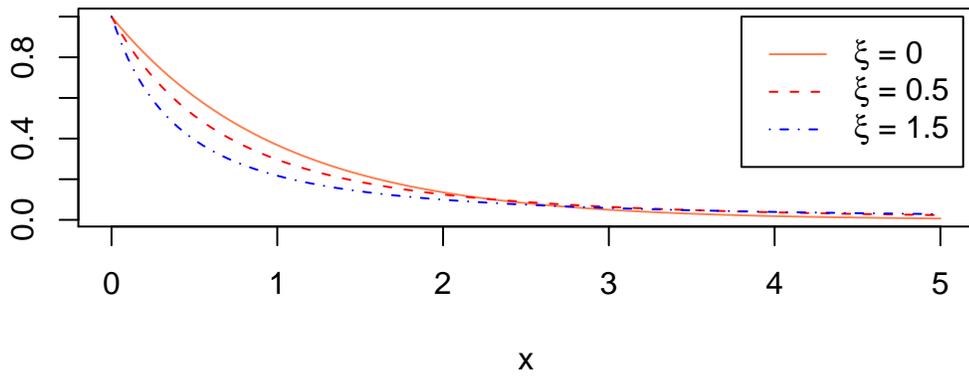


FIG. 6 – Loi de Pareto généralisée

les estimateurs de maximum de vraisemblance sont asymptotiquement normaux et efficaces si $\xi > 1/2$, voir [13] et [14]. La méthode des moments pondérés de façon probabiliste peut aussi être employée, voir [10]. Il faut garder en tête que les estimateurs $\hat{\xi}_u$ et $\hat{\beta}_u$ sont dépendants du seuil u choisi.

2.3.3 Détermination du seuil

Le seuil doit être choisi de façon à faire un compromis : plus le seuil est élevé, plus le biais de l'estimateur est réduit ce qui donne un meilleur modèle, par ailleurs, plus le seuil est bas, plus la variance de l'estimateur est réduite car plus de données participent à l'estimation.

Il n'existe pas de méthodes qui proposent une minimisation automatique du biais et de la variance du modèle. Une méthode utilisée par plusieurs auteurs (voir par exemple, McNeil et Frey [3]) consiste à choisir un seuil aléatoire qui est la $k + 1$ -ième statistique d'ordre. Il y a donc k observations excédentaires et le k peut être déterminé par simulation.

2.4 Estimation de la queue de la distribution

La distribution de Pareto généralisée est employée pour modéliser la queue (supérieure ou inférieure) d'une distribution présentant des valeurs extrêmes. Pour pouvoir faire de l'estimation de quantile ou l'estimation de l'espérance, il faut avoir une formulation qui combine la distribution estimée dans la queue et la distribution centrale.

Une telle formulation est donnée par l'égalité suivante, $\forall x > u : 1 - F(x) = (1 - F(u))(1 - F_u(x - u))$, où $F_u(y) = P\{X - u \leq y | X > u\}$.

L'estimation procède de la façon suivante :

$$1 - F(u) \approx N/n \quad (3)$$

$$1 - F_u(x - u) \approx 1 - G_{\hat{\xi};\hat{\beta}}(x - u) \quad (4)$$

3 Densité conditionnelle à ailes lourdes

Dans certains cas, les observations sont dépendantes d'autres variables qui doivent être prises en compte lors de la modélisation. Il faut donc modéliser la

densité conditionnellement à ces variables. Par ailleurs, les observations peuvent avoir une dépendance temporelle comme c'est le cas dans la plupart des séries financières.

3.1 Loi stable conditionnelle

Lambert et Lindsey [11] ont développé une méthode de calcul de la densité des lois α -stables basée sur la transformée de Fourier. Ceci leur permet d'estimer les paramètres par maximum de vraisemblance. Cette méthode peut être étendue au cadre plus général où les paramètres dépendent des variables explicatives. Les auteurs développent une technique qui s'inspire des modèles linéaires généralisés (voir [12]) et qui permet de faire dépendre de variables explicatives le mode de la distribution au moyen du paramètre d'emplacement μ . Si besoin est (ce qui peut être déterminé par le critère d'information d'Akaike), les autres paramètres peuvent dépendre de variables explicatives à l'aide de fonctions de lien appropriées.

3.2 Pareto généralisée conditionnelle

En utilisant la technique du maximum de vraisemblance, voir la sous-section 2.3.2, pour l'estimation des paramètres de la Pareto généralisée, le modèle peut être étendu facilement au cas où des variables explicatives influençant les observations sont disponibles.

Davison et Smith [2] supposent que $Y_i, i = 1, \dots, N$ sont des variables aléatoires Pareto généralisée indépendentes de paramètres $k_i = k(z_i^T \boldsymbol{\gamma})$ et $\sigma_i = \sigma(\mathbf{x}_i^T \boldsymbol{\beta})$, où \mathbf{z}_i et \mathbf{x}_i sont des vecteurs de variables explicatives connues. Les paramètres $\boldsymbol{\gamma}$ et $\boldsymbol{\beta}$ sont estimés par maximum de vraisemblance.

3.3 Régression robuste pour queues asymétriques

Takeuchi et al. [8] proposent une méthode de régression robuste basée sur la régression de quantiles : pour un niveau p , nous cherchons $\hat{f}_p(X)$ telle que $P(Y < \hat{f}_p(X)|X = x) = p$. Cette méthode s'attaque au problème de régression lorsque le bruit est distribué asymétriquement avec des observations extrêmes.

L'idée principale est que l'espérance conditionnelle $E[Y|X = x]$ que l'estima-

tion cherche à approximer correspond à un quantile de niveau $p_\mu(x)$:

$$p_\mu(x) = F_{Y|X=x}(E[Y|X = x]) = P(Y < E[Y|X = x]|X = x).$$

Puisque p_μ en général n'est pas connu, la régression de quantiles est effectuée pour plusieurs niveaux p_1, \dots, p_n . Ces estimateurs $\hat{f}_{p_i}(X)$, $i = 1, \dots, n$ sont ensuite combinés pour estimer $\hat{f}_{p_\mu}(X)$.

4 Recherche proposée

La méthodologie liée à la distribution de la Pareto généralisée est basée sur le fait que toute fonction de répartition $F(x)$ puisse s'écrire en fonction de $F_u(x)$, où u est dans le domaine de F , lorsque $x > u$. C'est-à-dire :

$$F(x) = \begin{cases} (1 - F(u))F_u(x - u) + F(u) & \text{si } x > u, \\ F(x) & \text{autrement.} \end{cases}$$

Ceci s'étend au cas conditionnel :

$$F_{Y|X=x}(y) = \begin{cases} (1 - F_{Y|X=x}(u(x)))F_{u(x)|X=x}(y - u(x)) + F_{Y|X=x}(u(x)) & \text{si } y > u(x), \\ F_{Y|X=x}(y|x) & \text{autrement.} \end{cases}$$

où $F_{u(x)|X=x}(y) = P(Y - u(X) \leq y | Y > u(X), X = x)$.

La distribution sous le seuil $F_{Y|X=x}(y)$ est modélisée de façon non-paramétrique. La distribution au-delà du seuil est modélisée à l'aide d'une distribution de Pareto généralisée conditionnelle $G_{\xi(x);\beta(x)}(y)$.

4.1 Modélisation de la densité conditionnelle

Une fois qu'une fonction de seuil $u(x)$ adéquate est déterminée, la modélisation se passe en deux étapes : la modélisation de la densité sous le seuil et au-delà du seuil.

4.1.1 Densité conditionnelle centrale

Pour la partie sous le seuil, il s'agit d'utiliser une méthode non-paramétrique. Une méthode traditionnelle non-paramétrique est le mélange de distributions :

$$f(y|X = x) = \sum_{j=1}^M \alpha_j(x) \phi_j(y|X = x),$$

où M est le nombre de composantes, $\phi_j(Y = y|X = x)$ est une fonction de densité paramétrique et les $\alpha_j(x)$ sont les pondérations pour chacune des composantes intervenant dans le mélange. Typiquement, la distribution gaussienne est utilisée comme composante. Les paramètres du modèle, c'est-à-dire $\alpha_j(x)$, $\mu_j(x)$ (emplacement de la Gaussienne j) et $\sigma_j^2(x)$ (variance de la Gaussienne j), sont fournis par un réseau de neurones qui prend en entrée le vecteur x . Si le nombre de composantes dans le mélange et le nombre d'unités cachées sont suffisamment grands, il est possible d'approximer d'aussi proche que voulu n'importe quelle fonction de densité conditionnelle, voir [5].

4.1.2 Densité conditionnelle dans la queue supérieure

La densité conditionnelle au-delà du seuil $u(x)$ est modélisée à l'aide d'une distribution de Pareto généralisée. Les paramètres ξ et β de la Pareto généralisée sont exprimés de façon conditionnelle par les fonctions $\xi(x) = \exp\{\xi^* x\}$ et $\beta(x) = \exp\{\beta^* x\}$, ce qui assure leur positivité. C'est lorsque $\xi > 0$ que la Pareto généralisée possède des queues sous-exponentielles et β étant le paramètre de dispersion est contraint d'être positif, voir la sous-section 2.3.

Soient $\{(x_1, y_1), \dots, (x_n, y_n)\}$ les observations disponibles et soit $u(x)$ le seuil dépendant. Alors les paramètres de la Pareto généralisée conditionnelle sont estimés sur les observations qui représentent un excès au-delà du seuil :

$$\mathcal{E} = \{z_i = (x_i, e_i) | y_i > u(x_i) \wedge e_i = y_i - u(x_i)\}.$$

Soit $l(\theta, z)$, où $\theta = (\xi^*, \beta^*)$ et $z = (x, e)$, la log-vraisemblance négative. Alors :

$$\hat{\theta} = \arg \min_{\theta} l(\theta, z).$$

4.2 Détermination du seuil dépendant

Nous proposons une méthode de sélection automatique du seuil inspirée de la méthode du seuil aléatoire. Dans le cas inconditionnel, il s'agit de fixer un k qui sera le nombre d'excès au-delà du seuil et la $(k + 1)$ -ième statistique d'ordre est alors le seuil aléatoire en question. Dans le cas conditionnel, nous suggérons d'utiliser comme seuil aléatoire un quantile conditionnel de niveau q :

$$f_q(x) = \inf\{y \in \mathbb{R} : F_{Y|X=x}(y) \geq q\}.$$

Lorsque le nombre d'observations est assez grand, il y a environ $1 - q$ % des observations qui dépassent le quantile conditionnel $f_q(x)$.

4.2.1 Estimation du quantile conditionnel

Le quantile conditionnel est estimé par le quantile empirique de Y dans le voisinage de x . Plus précisément :

- Le voisinage de chaque point i de l'ensemble d'entraînement $\mathcal{S}_{\text{train}}$ est calculé :

$$\mathcal{V}_\epsilon^i = \{j \in \{1, \dots, n\} \mid x_j \in [x_i - \epsilon, x_i + \epsilon]\},$$

où ϵ est un hyperparamètre qui contrôle la taille du voisinage.

- Le quantile $f_q(x)$ est estimé à partir de la fonction de répartition empirique calculée sur les données dans le voisinage de i :

$$\{y_j \mid j \in \mathcal{V}_\epsilon^i\}.$$

S'il y a n_i points dans le voisinage du point i , le quantile de niveau q est donné par la k -ième statistique d'ordre de la variable Y dans ce voisinage où k est tel que $\frac{k-1}{n_i} < q \leq \frac{k}{n_i}$.

4.2.2 Sélection de modèle

Le niveau du quantile q est choisi par sélection de modèle : soit \mathcal{Q} , l'ensemble des valeurs considérées pour q , nous devons choisir parmi les modèles $\{\mathcal{M}_q : q \in \mathcal{Q}\}$. Les données $\{(x_1, y_1), \dots, (x_n, y_n)\}$ sont divisées en trois ensembles disjoints : l'ensemble d'entraînement $\mathcal{S}_{\text{train}}$, l'ensemble de validation $\mathcal{S}_{\text{valid}}$ et l'ensemble de test $\mathcal{S}_{\text{test}}$. Les étapes suivantes sont exécutées :

1. Pour chaque $q \in \mathcal{Q}$, le quantile conditionnel $\hat{f}_q(x)$ est estimé sur les données de $\mathcal{S}_{\text{train}}$.

2. Pour chaque $q \in \mathcal{Q}$, le seuil aléatoire $\hat{f}_q(x)$ permet d'obtenir l'estimateur $\hat{\theta} = (\hat{\xi}^*, \hat{\beta}^*)$ des paramètres de la Pareto généralisée conditionnelle par maximum de vraisemblance toujours à l'aide de l'ensemble $\mathcal{S}_{\text{train}}$.
3. Le coût du modèle \mathcal{M}_q est évalué en mesurant la log-vraisemblance négative moyenne sur l'ensemble de validation $\mathcal{S}_{\text{valid}}$.
4. Le modèle retenu \mathcal{M}_{q^*} est celui ayant le plus petit coût estimé sur l'ensemble $\mathcal{S}_{\text{valid}}$.
5. Le coût associé au modèle \mathcal{M}_{q^*} est évalué en calculant la log-vraisemblance négative moyenne sur l'ensemble de test $\mathcal{S}_{\text{test}}$.

4.2.3 Maximum de vraisemblance conditionnel

L'estimation de la Pareto généralisée conditionnelle de l'étape 2 se fait sur les excédents de l'ensemble d'entraînement $\mathcal{S}_{\text{train}}$:

$$\mathcal{E}_{\text{train}} = \{z_i = (x_i, e_i) \in \mathcal{S}_{\text{train}} | y_i > \hat{f}_q(x_i) \wedge e_i = y_i - \hat{f}_q(x_i)\}.$$

4.2.4 Log-vraisemblance négative moyenne

À l'étape 3, le coût du modèle \mathcal{M}_q obtenu est évalué en calculant, sur l'ensemble de validation $\mathcal{S}_{\text{valid}}$, la log-vraisemblance négative moyenne :

$$\frac{1}{n} \sum_{z \in \mathcal{E}_{\text{valid}}} l(\hat{\theta}, z),$$

où $\mathcal{E}_{\text{valid}} = \{z_i = (x_i, e_i) \in \mathcal{S}_{\text{valid}} | y_i > \hat{f}_q(x_i) \wedge e_i = y_i - \hat{f}_q(x_i)\}$. Le fait de prendre la moyenne permet d'éliminer l'influence du nombre d'exemples dans l'évaluation du coût (puisque le nombre d'excédents varie avec le choix du seuil).

Afin de prendre en compte l'impact du choix du seuil sur l'ensemble de l'estimation de densité conditionnelle (centrale et queue supérieure), nous pourrions prendre comme fonction de coût la log-vraisemblance négative sur toutes les données, pas seulement sur les excédents. Cette fonction de coût permettrait de prendre en compte la justesse avec laquelle l'estimation de densité conditionnelle dans la partie centrale capte la vraie distribution conditionnelle sous-jacente.

4.2.5 Coût du modèle

À l'étape 5, pour le niveau de quantile q^* déterminé à l'étape précédente, les excédents sont calculés pour les ensembles $\mathcal{S}_{\text{train}}$ et $\mathcal{S}_{\text{valid}}$:

$$\mathcal{E}_{\text{cost}} = \{z_i = (x_i, e_i) \in \mathcal{S}_{\text{train}} \vee \mathcal{S}_{\text{valid}} | y_i > \hat{f}_{q^*}(x_i) \wedge e_i = y_i - \hat{f}_{q^*}(x_i)\}.$$

Les paramètres de la Pareto généralisée conditionnelle sont ré-estimés sur $\mathcal{E}_{\text{cost}}$ et le coût associé au modèle \mathcal{M}_{q^*} est mesuré sur l'ensemble de test $\mathcal{E}_{\text{test}} = \{z_i = (x_i, e_i) \in \mathcal{S}_{\text{test}} | y_i > \hat{f}_q(x_i) \wedge e_i = y_i - \hat{f}_q(x_i)\}$.

5 Expériences préliminaires

L'algorithme présenté dans ce chapitre a été testé avec des données simulées. Le but de ces expériences préliminaires est de déterminer si la méthode est apte à retrouver les paramètres de la Pareto généralisée conditionnelle lorsque le seuil dépend de x . Les sous-sections 5.1 et 5.2 décrivent la simulation des jeux de données artificielles et les résultats de l'estimation des paramètres de la Pareto généralisée conditionnelle sur ces jeux de données.

Les jeux de données contiennent 20000 observations générés selon le principe suivant : la variable X est distribuée selon une Normale $\mathcal{N}(0, 0.4^2)$, une distribution conditionnelle génère les observations dépendantes pour la partie centrale de la distribution (sous le seuil), la distribution de Pareto généralisée conditionnelle génère les observations dépendantes dans la queue supérieure (au-delà du seuil), une fonction $u(x)$ détermine le seuil dépendant et les excès se produisent selon une variable de Bernoulli de probabilité $p = 0.3$. La variable indépendante X et la variable indépendante Y sont toutes deux scalaires.

5.1 Premier jeu de données

Pour ce jeu de données, la distribution de la partie centrale est une Uniforme dont la borne supérieure est $u(x) = 5x + 10$ et la borne inférieure est $l(x) = 5x$. Le seuil est la borne supérieure de la distribution Uniforme. La figure 7 illustre ce jeu de données.

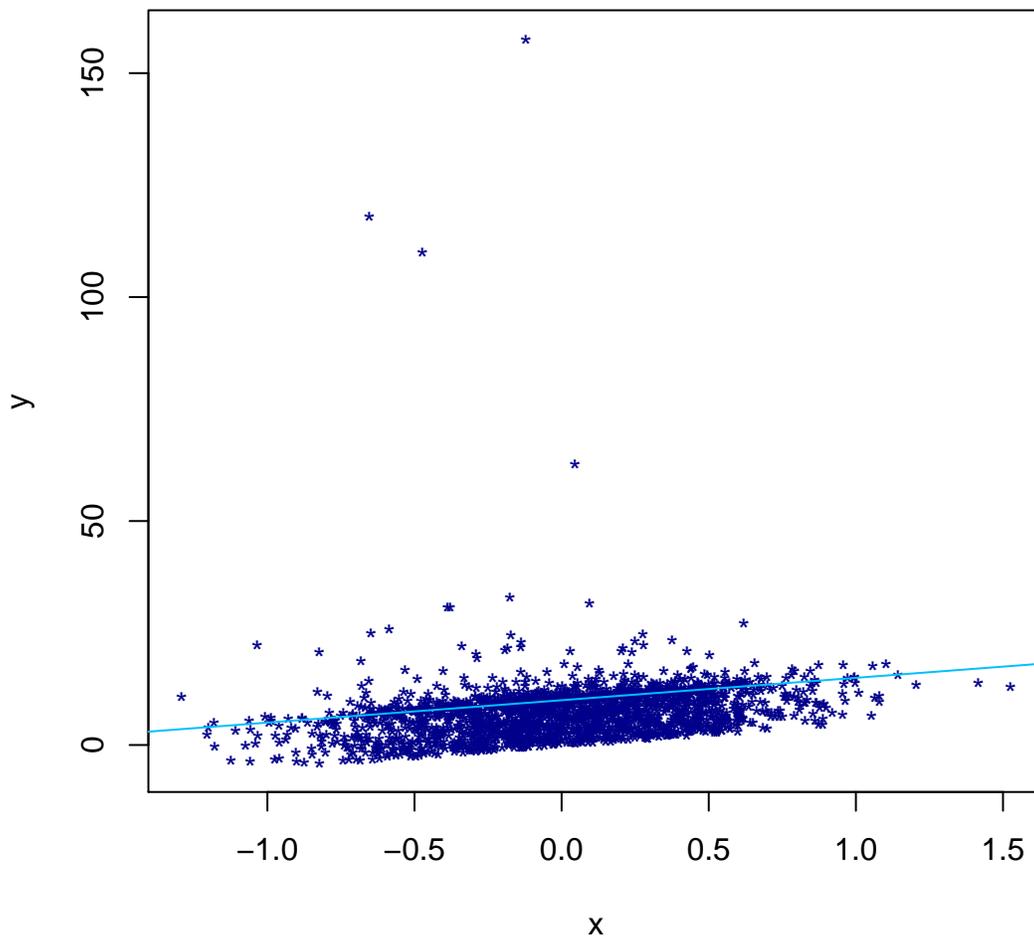


FIG. 7 – Jeu de données #1 : Pareto généralisée conditionnelle au-delà du seuil ; sous le seuil la distribution conditionnelle est $\mathcal{U}[low(x), up(x)]$; la ligne droite est le seuil dépendant $u(x)$.

5.1.1 Estimation

Pour appliquer la méthode d'estimation, nous utilisons un voisinage dont la taille est contrôlée par ϵ qui est fixé à 0.005 (voir la sous-section 4.2.2). L'ensemble Q des niveaux de quantiles explorés est : $\{0.5, 0.55, \dots, 0.90, 0.95\}$.

Les paramètres de la Pareto généralisée conditionnelle ainsi que les estimateurs de la Pareto généralisée conditionnelle sont présentés dans le tableau 1. On y trouve aussi le niveau de quantile choisi q^* et le coût associé au modèle \mathcal{M}_{q^*} .

Indice de queue :	$\xi = \exp(\xi_1 + \xi_2 x)$
Paramètres de queue :	$\xi_1 = -0.5, \xi_2 = -0.4$
Indice de dispersion :	$\beta = \exp(\beta_1 + \beta_2 x)$
Paramètres de dispersion :	$\beta_1 = -0.07, \beta_2 = -0.2$
Quantile retenu :	$q^* = 0.75$
Paramètres de queue estimés :	$\hat{\xi}_1 = -0.6124, \hat{\xi}_2 = -0.3251$
Paramètres de dispersion estimés :	$\hat{\beta}_1 = 0.09995, \hat{\beta}_2 = -0.1912$

TAB. 1 – Jeu de données artificielles #1

5.2 Deuxième jeu de données

Pour ces données, la distribution de la partie centrale est une Weibull de paramètres $\alpha = 1$ et $\beta = 1$, (voir la sous-section 2.2) dont l'espérance est donnée par $u(x) = 5x + 10$. La fonction $u(x)$ donne le seuil dépendant. La figure 8 illustre ce jeu de données.

5.2.1 Estimation

Les paramètres de la Pareto généralisée conditionnelle ainsi que les estimateurs de la Pareto généralisée conditionnelle sont présentés dans le tableau 2. On y trouve aussi le niveau de quantile choisi q^* et le coût associé au modèle \mathcal{M}_{q^*} .

5.2.2 Difficultés d'estimation

L'estimation est très sensible à la présence d'observations extrêmes dans l'ensemble d'entraînement. Davison et Smith [2] mentionnent d'ailleurs que les estimateurs dépendent typiquement des quelques observations extrêmes.

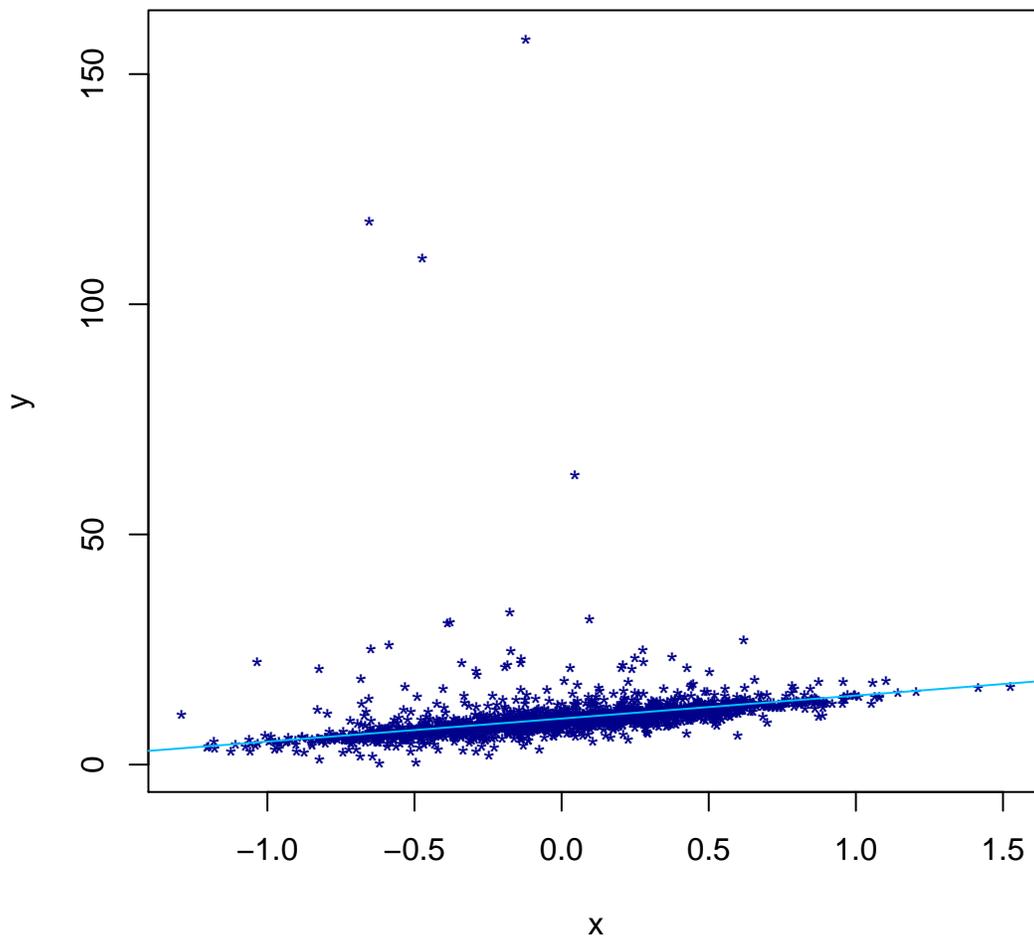


FIG. 8 – Jeu de données #2 : Pareto généralisée conditionnelle au-delà du seuil ; sous le seuil la distribution conditionnelle est $\Psi_{\alpha;\mu(x),\beta}$; la ligne droite est le seuil dépendant $u(x)$.

Indice de queue :	$\xi = \exp(\xi_1 + \xi_2 x)$
Paramètres de queue :	$\xi_1 = -0.5, \xi_2 = -0.4$
Indice de dispersion :	$\beta = \exp(\beta_1 + \beta_2 x)$
Paramètres de dispersion :	$\beta_1 = -0.07, \beta_2 = -0.2$
Quantile retenu :	$q^* = 0.5$
Paramètres de queue estimés :	$\hat{\xi}_1 = -0.4994, \hat{\xi}_2 = -0.4076$
Paramètres de dispersion estimés :	$\hat{\beta}_1 = -0.4006, \hat{\beta}_2 = -0.0021$

TAB. 2 – Jeu de données artificielles #2

Dans nos expériences, ce phénomène est observé lorsque des ensembles d'entraînement différents sont choisis. Les paramètres estimés peuvent alors varier considérablement.

6 Conclusion

L'objectif principal est de développer une méthode d'estimation de densité conditionnelle applicable dans le cas où la distribution est asymétrique avec une queue supérieure de type Pareto. Les distributions de type Pareto dans le cas conditionnel sont définies par la propriété suivante :

$$1 - F_{Y|X=x}(y) = y^{-\alpha(x)} L(y; x),$$

où $y > 0$, $\alpha(x) > 0$ et L est une fonction à variation lente. Cette famille de distributions comprend entre autres les lois α -stable, avec $\alpha < 2$ (voir la sous-section 2.1) et les distributions dans le domaine d'attraction de Fréchet (voir la sous-section 2.2).

6.1 Pareto généralisée conditionnelle

Nous avons proposé dans la section 4 une méthode qui utilise la distribution de Pareto généralisée conditionnelle ajustée au-delà d'un seuil. Comme nous sommes dans un cadre conditionnel, il est logique de considérer que le seuil puisse s'écrire comme une fonction de la variable indépendante. Il n'existe pas de méthodologie communément admise pour déterminer le seuil même dans le cas de l'estimation de densité inconditionnelle. Nous avons développé une stratégie qui permette le

choix du seuil dans le cas d'estimation de densité inconditionnelle et conditionnelle (voir la section 4.2).

Des essais préliminaires de la section 5 indiquent que cette méthode a du potentiel. Plusieurs améliorations pourraient y être apportées :

- L'estimation du quantile conditionnel pourrait être faite par une méthode plus robuste que celle du voisinage ; par exemple, en utilisant un autre type d'estimateur de la densité conditionnelle comme le mélange de gaussiennes mentionné en 4.1.1.
- Jusqu'à maintenant, nous n'avons pas vraiment attaqué le problème de l'estimation de densité dans la partie centrale. Celle-ci pourrait être estimée de façon conjointe à l'estimation de densité dans la partie supérieure afin de permettre la sélection d'un seuil qui soit adéquat pour les deux parties de la distribution (voir la sous-section 4.2.2 pour la description de l'estimation du coût des modèles).

6.1.1 Extensions

Toujours dans l'optique d'ajuster une Pareto généralisée conditionnelle avec un seuil dépendant, une autre stratégie possible serait d'apprendre, par exemple à l'aide d'un réseau de neurones, la fonction qui définit le seuil. Supposons que l'ensemble des fonctions dont nous disposons s'écrive : $u(x, \theta_u)$, où θ_u est un vecteur de paramètres. Pour trouver θ_u^* qui minimise la fonction de coût, il serait préférable de pouvoir calculer la dérivée de la fonction de coût par rapport à θ_u . Il n'est pas clair qu'une formulation existe pour cette dérivée.

En étendant l'idée précédente, nous pourrions considérer les paramètres de la Pareto généralisée conditionnelle comme des fonctions plus générales du type : $\xi(x) = \xi(x, \theta_\xi)$ et $\beta(x) = \beta(x, \theta_\beta)$. Les paramètres θ_ξ et θ_β seraient appris au même titre que les paramètres du seuil θ_u . Autrement dit, un algorithme d'apprentissage comme un réseau de neurones ayant en entrée le vecteur x , fournirait en sortie les valeurs de $\xi(x, \hat{\theta}_\xi)$, $\beta(x, \hat{\theta}_\beta)$ et $u(x, \hat{\theta}_u)$.

Dans un autre ordre d'idées, l'idée du mélange de distributions pourrait être modifiée pour inclure des distributions à queues épaisses. Une ou plusieurs distributions pourraient être des lois α -stables avec $\alpha < 2$. Une autre possibilité serait d'inclure une ou plusieurs des distributions de valeurs extrêmes, ou encore des distributions qui sont dans les domaines d'attraction maximum de ces distribu-

tions. Ces distributions ont l'avantage d'avoir une forme fermée pour la fonction de densité.

Références

- [1] Stephenson A. Evir package for r version 1. <http://www.maths.lancs.ac.uk/stephena/>.
- [2] Davison A.C. and Smith R.L. Models for exceedances over high thresholds. *Journal of the Royal Statistical Society. Series B (Methodological)*, 52(3) :393–442, 1990.
- [3] McNeil A.J. and Frey R. Estimation of tail-related risk measures for heteroscedastic financial time series : an extreme value approach. *Journal of Empirical Finance*, 7 :271–300, 2000.
- [4] Mandelbrot B. The variation of certain speculative prices. *Journal of Business*, 36 :394–419, 1963.
- [5] Bishop C. *Neural Networks for Pattern Recognition*. Oxford, 1995.
- [6] Fama E.F. The behavior of stock market prices. *Journal of Business*, 38 :34–105, 1965.
- [7] Hampel E.M, Ronchetti E.M., Rousseeuw P.J., and Stahel W.A. *Robust Statistics : The Approach based on Influence Functions*. Wiley, 1986.
- [8] Takeuchi I., Bengio Y., and Kanamori T. Robust regression with asymmetric heavy-tail noise distributions. *Neural Computation*, 14(10) :2469–2496, 2002.
- [9] Beirlant J. and Goegebeur Y. Regression with response distributions of pareto-type. *Computational Statistics and Data Analysis*, 42 :595–619, 2003.
- [10] Hosking J.R.M and Wallis J.R. Parameter and quantile estimation for the generalized pareto distribution. *Technometrics*, 29 :339–349, 1987.
- [11] Lambert P. and Lindsey J.K. Analysing financial returns by using regression models based on non-symmetric stable distributions. *Applied Statistics*, 48(3) :409–424, 1999.
- [12] McCullagh P. and Nelder J.A. *Generalized linear models*. Monographs on statistics and applied probability. Chapman and Hall, 1989.

- [13] Smith R.L. Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72 :67–90, 1985.
- [14] Smith R.L. Estimating tails of probability distributions. *The Annals of Statistics*, 15(3) :1174–1207, September 1987.