



Estimation de mesures de risque pour des distributions elliptiques conditionnées

Antoine Usseglio-Carleve

► To cite this version:

Antoine Usseglio-Carleve. Estimation de mesures de risque pour des distributions elliptiques conditionnées. Analyse numérique [cs.NA]. Université de Lyon, 2018. Français. NNT : 2018LYSE1094 . tel-01831860v2

HAL Id: tel-01831860

<https://hal.science/tel-01831860v2>

Submitted on 25 Sep 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



N° d'ordre NNT : 2018LYSE1094

THÈSE DE DOCTORAT DE L'UNIVERSITÉ DE LYON

opérée au sein de
l'Université Claude Bernard Lyon 1

École Doctorale ED512
Infomaths

Spécialité de doctorat :
Discipline : Mathématiques

Soutenue publiquement le 26/06/2018, par :
Antoine Usseglio-Carleve

Estimation de mesures de risque pour des distributions elliptiques conditionnées

Devant le jury composé de :

Guillou Armelle, Professeure, Université de Strasbourg
Hashorva Enkelejd, Professeur ordinaire, Université de Lausanne
Bel Liliane, Professeure, AgroParisTech
Ghoudi Kilani, Professeur, United Arab Emirates University
Girard Stéphane, Directeur de Recherche, INRIA Grenoble Rhone-Alpes
Soulier Philippe, Professeur, Université Paris 10

Rapporteuse
Rapporteur
Examinatrice
Examineur
Examineur
Président du jury

Maume-Deschamps Véronique, Professeure, Université Lyon 1
Rulhière Didier, Maître de conférence, Université Lyon 1

Directrice de thèse
Co-directeur de thèse

RÉSUMÉ

Cette thèse s'intéresse à l'estimation de certaines mesures de risque d'une variable aléatoire $Y \in \mathbb{R}$ en présence d'une covariable $\mathbf{X} \in \mathbb{R}^d$. Pour cela, on va considérer que le vecteur $(\mathbf{X}, Y)$ suit une loi elliptique. Ce modèle est adapté par exemple à la prédiction d'un champ elliptique étant données des observations de ce dernier.

Dans un premier temps, on va s'intéresser aux quantiles de Y sachant $\mathbf{X} = \mathbf{x}$. On va alors tester d'abord un modèle de régression quantile assez répandu dans la littérature, pour lequel on obtient des résultats théoriques que l'on discutera. Face aux limites d'un tel modèle, en particulier pour des niveaux de quantile dits extrêmes, on proposera une nouvelle approche plus adaptée. Des résultats asymptotiques sont donnés, appuyés par une étude numérique puis par un exemple sur des données réelles.

Dans un second chapitre, on s'intéressera à une autre mesure de risque appelée expectile. La structure du chapitre est sensiblement la même que celle du précédent, à savoir le test d'un modèle de régression inadapté aux expectiles extrêmes, pour lesquels on propose une approche méthodologique puis statistique. De plus, en mettant en évidence le lien entre les quantiles et expectiles extrêmes, on s'aperçoit que d'autres mesures de risque extrêmes sont étroitement liées aux quantiles extrêmes. On se concentrera sur deux familles appelées L_p -quantiles et mesures d'Haezendonck-Goovaerts, pour lesquelles on propose des estimateurs extrêmes. Une étude numérique est également fournie.

Enfin, le dernier chapitre propose quelques pistes pour traiter le cas où la taille de la covariable $\mathbf{X}$ est grande. En constatant que nos estimateurs définis précédemment étaient moins performants dans ce cas, on s'inspire alors de quelques méthodes d'estimation en grande dimension pour proposer d'autres estimateurs. Une étude numérique permet d'avoir un aperçu de leurs performances.

Mots-clés : *Distributions elliptiques ; Quantiles extrêmes ; Expectiles ; Régression quantile ; Régression expectile ; Théorie des valeurs extrêmes.*

ABSTRACT

This PhD thesis focuses on the estimation of some risk measures for a random variable $Y \in \mathbb{R}$ with a covariate vector $\mathbf{X} \in \mathbb{R}^d$. For that purpose, we will consider that the random vector $(\mathbf{X}, Y)$ is elliptically distributed.

In a first time, we deal with the quantiles of Y given $\mathbf{X} = \mathbf{x}$. We thus firstly investigate a quantile regression model, widespread in the litterature, for which we get theoretical results. Indeed, such a model has some limitations, especially when the quantile level is said to be extreme. Therefore, we propose another more suited approach. Asymptotic results are given, illustrated by a simulation study and a real data example.

In a second chapter, we focus on another risk measure called expectile. The structure of the chapter is essentially the same as the previous one. Indeed, we first use a regression model that is not adapted to extreme expectiles, for which a methodological and statistical approach is proposed. Furthermore, highlighting the link between extreme quantiles and expectiles, we realize that other extreme risk measures are closely related to extreme quantiles. We will focus on two families called L_p -quantiles and Haezendonck-Goovaerts risk measures, for which we propose extreme estimators. A simulation study is also provided.

Finally, the last chapter is devoted to the case where the size of the covariate vector $\mathbf{X}$ is large. By noticing that our previous estimators perform poorly in this case, we rely on some high dimensional estimation methods to propose other estimators. A simulation study gives a visual overview of their performances.

Keywords : *Elliptical distributions ; Extreme quantiles ; Expectiles ; Quantile regression ; Expectile regression ; Extreme value theory.*

TABLE DES MATIÈRES

I Introduction	1
II Outils mathématiques	5
II.1. Mesures de risque	5
II.2. Variation régulière	7
II.3. Distributions elliptiques	9
II.3.a. Premières définitions	9
II.3.b. Mélanges d'échelle gaussiens consistants	10
II.3.c. Exemples	12
II.3.c.1. Loi Normale	13
II.3.c.2. Loi Normale contaminée	13
II.3.c.3. Loi de Student	13
II.3.c.4. Loi de Laplace	13
II.3.c.5. Lois de Slash et de Slash généralisée	14
II.3.d. Bilan	15
III Quantiles conditionnels	17
III.1. Problématique	17
III.1.a. Cas général	18
III.1.b. Quelques quantiles conditionnels théoriques	18
III.1.b.1. Loi Normale	18
III.1.b.2. Loi de Student	19
III.2. Le modèle de régression	20
III.2.a. Présentation du modèle	21
III.2.b. Résultats dans le cadre elliptique consistant	21
III.2.c. Commentaires	25
III.3. Quantiles extrêmes	27
III.3.a. Approche méthodologique	27
III.3.b. Estimation des paramètres extrémaux	33
III.3.c. Estimation des quantiles extrêmes	41
III.3.c.1. Conditions	41
III.3.c.2. Quantiles intermédiaires	42
III.3.c.3. Hauts quantiles	46
III.4. Etude numérique	50
III.4.a. Quantiles intermédiaires	51
III.4.b. Hauts quantiles	53
III.4.c. Un exemple de données réelles	54

IV Expectiles et autres mesures conditionnelles	57
IV.1. Introduction	57
IV.1.a. La notion d'expectile	57
IV.1.b. Algorithmes	61
IV.1.b.1. Algorithme MM	62
IV.1.b.2. Algorithme de point fixe	63
IV.1.c. Quelques formes théoriques d'expectiles conditionnels	64
IV.1.c.1. Loi Normale	65
IV.1.c.2. Loi Student	65
IV.1.c.3. Loi Normale contaminée	65
IV.2. Le modèle de régression	66
IV.2.a. Présentation du modèle	66
IV.2.b. Résultats dans le cadre elliptique consistant	66
IV.2.c. Commentaires	69
IV.3. Expectiles extrêmes	70
IV.3.a. Approche méthodologique	70
IV.3.b. Estimation des paramètres extrémaux	78
IV.3.c. Estimation des expectiles extrêmes	81
IV.3.c.1. Expectiles intermédiaires	81
IV.3.c.2. Hauts Expectiles	83
IV.3.d. Liens avec les quantiles extrêmes	84
IV.4. Autres mesures de risque extrêmes basées sur le quantile	87
IV.4.a. L_p -quantiles	87
IV.4.b. Mesures d'Haezendonck-Goovaerts	91
IV.5. Etude numérique	94
IV.5.a. Expectiles intermédiaires	94
IV.5.b. Hauts expectiles	96
IV.5.c. TVaR intermédiaires	97
IV.5.d. TVaR hautes	99
V Le cas de la grande dimension	101
V.1. Approximation gaussienne	101
V.2. Estimation des paramètres en (très) grande dimension	104
V.2.a. Les méthodes de sparsité	105
V.2.b. Les méthodes de shrinkage	108
V.2.b.1. Le cas de $\mathbf{S}_n$	109
V.2.b.2. Le cas de $\mathbf{V}_n$	113
V.3. Etude numérique	115
V.3.a. En petite dimension	115
V.3.b. En "moyenne" dimension	116
V.3.c. En grande dimension	117

TABLE DES MATIÈRES

vii

VI Conclusion et perspectives

119

Bibliographie

121

A Fonctions et lois usuelles

129

B Conditions du noyau K

131

INTRODUCTION

Ces dernières années, la notion de mesures de risque a pris une importance considérable dans de nombreux domaines tels que l'environnement, ainsi que la finance et l'assurance. Par exemple, la notion de temps de retour est fondamentale dans l'étude des risques naturels tels que les tremblements de terre ou inondations. De même, depuis la crise financière de 2008, les autorités de contrôle prudentielles imposent aux banques ainsi qu'aux compagnies d'assurances une certaine solidité financière. Le bilan annuel d'une entreprise étant aléatoire, il est nécessaire d'appréhender le risque de cet aléa. Les mesures de risque, qui ne sont ni plus ni moins que des fonctions d'un espace de probabilité dans $\mathbb{R}$, remplissent ce rôle, et suscitent un intérêt croissant dans la littérature. Les mesures dites "extrêmes" s'intéressent spécifiquement aux événements rares, et leur estimation pose alors un certain nombre de problèmes, si les données disponibles ne sont pas suffisantes. Ainsi, si l'on doit par exemple estimer la fréquence d'une catastrophe nucléaire telle que celles de Tchernobyl ou Fukushima, les données disponibles sont très limitées. Pour en revenir à l'exemple des compagnies d'assurance, les directives Solvabilité 2 obligent ces dernières à être en capacité de rembourser tous leurs souscripteurs dans 99.5% des cas. Si on définit le quantile de niveau $\alpha \in]0, 1[$ d'une variable aléatoire réelle Y comme l'infimum de $\{y : \mathbb{P}(Y > y) \geq \alpha\}$, il faut donc de fait estimer un quantile d'ordre 0.995. L'estimation de quantiles extrêmes est un sujet très répandu dans la littérature récente. Si on se penche par exemple sur les papiers de [El Methni et al., 2012] ou [de Valk, 2016], il existe plusieurs méthodes d'estimation de ce type.

Cependant, le quantile a été récemment critiqué comme mesure de risque. [Artzner et al., 1999] lui reproche notamment de ne pas posséder la propriété de sous-additivité. En d'autres termes, le quantile de la somme de deux portefeuilles peut potentiellement dépasser la somme des quantiles des deux portefeuilles, ce qui est incohérent avec le principe d'agrégation des risques, ou phénomène de diversification. Une telle limite a conduit de nombreux auteurs à proposer de nouvelles mesures de risque alternatives. Une première possibilité, proposée par [Newey and Powell, 1987], est de remplacer le quantile par une mesure de risque nommée expectile. Cette mesure est de plus en plus utilisée, notamment par [Taylor, 2008] qui se sert de l'expectile pour estimer une autre mesure de risque nommée Expected Shortfall, ou encore par [Cai and Weng, 2016] qui propose un programme de réassurance optimal basé sur un critère d'expectile. Le papier [Sobotka et al., 2013] propose même une étude sur la malnutrition en Inde en utilisant les expectiles. L'estimation d'expectiles extrêmes est également traitée dans [Daouia et al., 2017a]. Conformément à [Chen, 1996], les quantile et expectile peuvent être regroupés dans une famille de mesures de risque nommée L_p -quantiles. En considérant une variable aléatoire Y , les L_p -quantiles

de niveau $\alpha \in]0, 1[$ de Y sont les solutions du problème de minimisation

$$q_{p,\alpha}(Y) = \arg \min_{y \in \mathbb{R}} \mathbb{E} [(1 - \alpha)|Y - y|_-^p + \alpha|Y - y|_+^p].$$

Les cas $p = 1$ et 2 correspondent respectivement au quantile et à l'expectile. Certains intérêts des L_p -quantiles ont notamment été mis en évidence dans [Daouia et al., 2017b], et un exemple d'application peut être trouvé dans [Chambers and Tzavidis, 2006]. En parallèle, il est proposé dans [Artzner et al., 1999] de s'intéresser au comportement de la distribution au delà du quantile. La mesure nommée Tail Value at Risk, ou TVaR est alors proposée de la manière suivante : si Y est une variable aléatoire réelle, et $q_\alpha(Y)$ est la quantile de niveau $\alpha \in]0, 1[$ de Y , alors la Tail Value at Risk de niveau α de Y , notée $\text{TVaR}_\alpha(Y)$, est la quantité

$$\text{TVaR}_\alpha(Y) = \frac{1}{1 - \alpha} \int_\alpha^1 q_u(Y) d_u.$$

Dans [Bargès et al., 2009], un problème d'allocation de capital est résolu avec la TVaR. Plus récemment, [Di Bernardino and Prieur, 2017] ont proposé l'estimation de Conditional Tail Expectation (une mesure coïncidant avec la TVaR dans le cas des distributions continues) sur des jeux de données environnementaux. A l'instar du quantile, nommé Value at Risk en finance, ou VaR, certains auteurs ont proposé de généraliser la TVaR à une classe de mesures de risque. [Haezendonck and Goovaerts, 1982] ont alors proposé les mesures dites d'Haezendonck-Goovaerts.

Au regard du paragraphe précédent, il apparait que la littérature sur l'estimation de mesures de risque d'une variable aléatoire Y est assez fournie. Notre but dans cette thèse sera de nous intéresser à l'estimation de mesures de risque extrêmes de Y en présence d'une co-variable $\mathbf{X}$. Dans ce contexte, quelques papiers récents existent et sont notamment basés sur des estimateurs à noyau : [Daouia et al., 2013] s'est focalisé sur les quantiles, quand [El Methni et al., 2018] s'est focalisé sur des mesures de risque nommées Conditional Tail Moments, une autre généralisation de la TVaR. Dans ce travail, nous considérerons le cas où le vecteur $(\mathbf{X}, Y)$ suit une loi elliptique. Les distributions elliptiques, introduites dans [Kelker, 1970], ont la représentation stochastique suivante : si $\mathbf{Z}$ est un vecteur aléatoire elliptique de $\mathbb{R}^d$, alors

$$\mathbf{Z} \stackrel{\mathcal{L}}{=} \boldsymbol{\mu} + R\boldsymbol{\Lambda}\mathbf{U}^{(d)}$$

où $\boldsymbol{\mu} \in \mathbb{R}^d$, $\boldsymbol{\Lambda} \in \mathbb{R}^{d \times d}$, $\mathbf{U}^{(d)}$ est le vecteur uniformément distribué sur la sphère unité de dimension d , et R est une variable aléatoire positive indépendante de $\mathbf{U}^{(d)}$. Ces dernières possèdent de nombreuses applications en finance, comme en théorie de portefeuille dans [Owen and Rabinovitch, 1983], en allocation d'actifs dans [Xiao and Valdez, 2015], ou encore en environnement dans [Thibaud and Opitz, 2015]. Ces distributions ont pour vocation de généraliser la loi gaussienne, en considérant des lois symétriques aux propriétés différentes, notamment sur la queue de distribution. D'un point de vue pratique, de

telles distributions possèdent des propriétés qui rendent l'étude plus simple à plusieurs égards. En effet, la symétrie ainsi que la stabilité par la somme (et par conséquent la stabilité des marginales) présentent un intérêt pratique certain. De plus, la densité d'un vecteur elliptique se ramène finalement à une fonction de $\mathbb{R}$ dans $\mathbb{R}$, ce qui rend l'estimation de cette dernière bien plus aisée que celle d'une fonction multivariée. Concernant les distributions conditionnelles des lois elliptiques, il existe quelques papiers traitant le sujet comme [Hashorva, 2007], qui fournit des relations extrêmes entre les fonctions de répartition conditionnelle et inconditionnelle des lois elliptiques, ou encore [Hashorva, 2006b] qui donne quelques propriétés sur la queue de distribution conditionnelle. Notre but sera alors de fournir des méthodes d'estimation efficaces des quantiles (et autres mesures de risque) conditionnels, en particulier pour des niveaux extrêmes.

Enfin, il est une autre question qui attirera notre attention, et elle concerne le cas où la dimension du vecteur de covariables $\mathbf{X}$ est en grande dimension. Ce problème a donné lieu à de nombreuses recherches, notamment dans les modèles de régression. Par exemple, les méthodes Sliced Inverse Regression (SIR), introduites dans [Li, 1991], traitent cette question et ont des applications comme [Li and Zhang, 2010] en génomique. Plus récemment, [Chiancone et al., 2017] ont mentionné le fait que de telles méthodes étaient basées sur une hypothèse gaussienne implicite, et proposé de remplacer la loi gaussienne par une loi Student, de manière à mieux traiter les distributions à queue lourde. Dans notre cadre, nous proposerons des approximations de quantiles conditionnels lorsque la covariable est en grande dimension.

La structure de cette thèse est la suivante. Dans le Chapitre II nous donnons un aperçu des outils mathématiques utilisés dans cette thèse. Nous introduirons les notions de mesure de risque et de variation régulière, puis nous rappellerons les définitions et propriétés des distributions elliptiques. Ce dernier paragraphe est principalement une revue de la littérature existante sur les elliptiques, et fournit quelques exemples illustrés de distributions.

Le Chapitre III se focalise sur les quantiles conditionnels. Dans un premier temps, la problématique est définie, et quelques exemples de formules de quantiles conditionnels théoriques sont donnés. Par la suite, nous proposons une première approche pour approximer des quantiles conditionnels via le modèle de régression

$$q_{\alpha}^{RQ}(Y|\mathbf{X} = \mathbf{x}) = \boldsymbol{\beta}^{\top} \mathbf{x} + \beta_0.$$

A travers une étude théorique de ce modèle dans le cadre des distributions elliptiques, il ressort que ce modèle est clairement inefficace, en particulier lorsque le niveau de quantile est extrême. On propose alors dans une troisième section une autre approche adaptée aux quantiles extrêmes. Le quantile conditionnel $q_{\alpha}(Y|\mathbf{X} = \mathbf{x})$ est alors approximé par un quantile inconditionnel dépendant de deux paramètres $\eta \in \mathbb{R}$ et $\ell(\mathbf{x}) \in]0, +\infty[$.

$$q_{\alpha\uparrow}^E(Y|\mathbf{X} = \mathbf{x}) = q_{\alpha, \eta, \ell(\mathbf{x})}(Y).$$

On propose alors d'estimer ces deux paramètres η et $\ell(\mathbf{x})$, pour lesquels on donne des résultats de normalité asymptotique, puis on en déduit les résultats asymptotiques de notre estimateur de $q_{\alpha\uparrow}^E(Y|\mathbf{X} = \mathbf{x})$. L'efficacité de notre approche est alors testée dans une étude numérique. Ce chapitre se base notamment sur les articles [Maume-Deschamps et al., 2017b] et [Usseglio-Carleve, 2017].

Le Chapitre IV a une structure similaire au précédent, en remplaçant le quantile par l'expectile. En effet, le formalisme de l'expectile, sa littérature ainsi que quelques méthodes de calcul sont fournies au début du chapitre. A l'instar du quantile, le modèle de régression

$$e_{\alpha}^{RE}(Y|\mathbf{X} = \mathbf{x}) = \boldsymbol{\beta}^{\top} \mathbf{x} + \beta_0$$

est proposé pour les expectiles conditionnels, et ne connaît pas un meilleur succès, d'où l'approche dévouée aux expectiles extrêmes qui s'ensuit. En se basant sur l'approximation faite pour les quantiles, on propose la suivante

$$e_{\alpha\uparrow}^E(Y|\mathbf{X} = \mathbf{x}) = e_{\alpha, \eta, \ell^e(\mathbf{x})}(Y).$$

Un estimateur pour $\ell^e(\mathbf{x})$ est proposé, et ses résultats asymptotiques sont donnés. On en déduit un estimateur de $e_{\alpha\uparrow}^E$ asymptotiquement normal sous certaines conditions. On met également en évidence le lien déjà établi entre les quantiles et expectiles extrêmes, qui nous permet de définir d'autres estimateurs pour nos expectiles conditionnels. En guise de perspective, deux familles de mesures de risque sont données en exemple, et des estimateurs conditionnels sont introduits pour ces mesures, en se basant sur leur proportionnalité asymptotique avec les quantiles. Leurs propriétés asymptotiques sont également discutées. Enfin, une étude numérique permet d'illustrer tous ces estimateurs. Une partie de ce chapitre se trouve dans [Maume-Deschamps et al., 2018] et [Usseglio-Carleve, 2017]. Le Chapitre V s'intéresse au cas où la taille de la covariable est grande. On montre dans un premier temps que dans un tel cas de figure, la loi conditionnelle converge vers une loi limite assez simple, qui nécessite uniquement l'estimation des paramètres de moyenne et de covariance. En particulier, il faut estimer la matrice de covariance inverse, appelée également matrice de précision, en grande dimension. Ce sujet étant extrêmement ouvert dans la littérature, nous proposons une revue des méthodes d'estimation des matrices de covariance et précision, telles que les approches sparse ou de shrinkage, puis appliquons ces méthodes dans notre problématique. Un exemple numérique est d'ailleurs proposé pour conclure.

OUTILS MATHÉMATIQUES

Ce premier chapitre propose un aperçu des outils mathématiques que nous allons utiliser tout au long de la thèse. Il est scindé en trois sections. Une première détaille la notion de mesure de risque d'une variable aléatoire, ainsi que les principales propriétés associées. La seconde traite des variations régulières, et nous sera particulièrement utile lorsqu'on parlera de mesure de risque extrême. Enfin, la dernière section sera consacrée aux distributions elliptiques.

II.1. Mesures de risque

La notion de mesure de risque est assez simple à appréhender. En effet, si on considère $\mathcal{X}$ un ensemble de variables aléatoires réelles, la définition est immédiate.

Définition 1 (Mesure de risque). *Une mesure de risque ρ est une fonction de $\mathcal{X}$ dans $\mathbb{R}$.*

Les cas de l'espérance et de la variance sont deux exemples bien connus de mesures de risque. On peut également considérer les quantiles de niveau $\alpha \in]0, 1[$.

Définition 2 (Quantiles). *Soit $Y \in \mathcal{X}$. Le quantile de niveau $\alpha \in]0, 1[$ de Y est défini comme :*

$$q_\alpha(Y) = \inf\{y : \mathbb{P}(Y \leq y) \geq \alpha\}.$$

[Artzner et al., 1999] propose également plusieurs exemples. La définition suivante donne quelques propriétés des mesures de risque.

Définition 3 (Propriétés des mesures de risque). *Une mesure de risque $\rho : \mathcal{X} \rightarrow \mathbb{R}$ est dite :*

- *homogène si $\rho(\lambda Y) = \lambda \rho(Y)$, $\forall \lambda \geq 0, Y \in \mathcal{X}$,*
- *sous-additive si $\rho(Y_1 + Y_2) \leq \rho(Y_1) + \rho(Y_2)$, $\forall Y_1, Y_2 \in \mathcal{X}$,*
- *invariante par translation si $\rho(Y + m) = \rho(Y) + m$, $\forall m \in \mathbb{R}, Y \in \mathcal{X}$,*
- *monotone si $\rho(Y_1) \leq \rho(Y_2)$, $\forall Y_1 \leq Y_2$ p.s.,*
- *cohérente si elle vérifie les 4 propriétés précédentes.*

La propriété de cohérence est relativement appréciée d'un point de vue pratique. En effet, la sous-additivité met par exemple en évidence le phénomène de diversification des risques. Il paraît par exemple logique que le risque d'un portefeuille d'actifs financiers soit moins élevé que la somme des risques individuels des actifs. Cependant, bien que vérifiant

les propriétés d'homogénéité, invariance par translation et monotonie, les quantiles ne sont pas des mesures de risque cohérentes, car elles ne sont pas sous-additives. Des auteurs tels que [Artzner et al., 1999] ont alors proposé des mesures alternatives cohérentes, l'une d'entre elles étant la Tail Value at Risk.

Définition 4 (Tail Value at Risk). *Soit $Y \in \mathcal{X}$. La Tail Value at Risk de niveau $\alpha \in]0, 1[$ de Y est définie comme :*

$$TVaR_\alpha(Y) = \frac{1}{1-\alpha} \int_{\alpha}^1 q_u(Y) d_u.$$

La $TVaR_\alpha$ est donc la moyenne des quantiles au delà de α . Si cette dernière est bien cohérente, elle ne vérifie en revanche pas la propriété d'élicitabilité. Cette notion, introduite dans [Ziegel, 2016], se définit de la manière suivante.

Définition 5 (Elicitabilité). *Une mesure de risque $\rho : \mathcal{X} \rightarrow \mathbb{R}$ est dite élicitable si il existe une fonction de coût $s : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}_+$ telle que*

$$\mathbb{E}[s(\rho(X), X)] \leq \mathbb{E}[s(x, X)],$$

pour tout $X \in \mathcal{X}$ et $x \in \mathbb{R}$.

Il découle de la Définition 5 qu'une mesure de risque X élicitable relativement à une fonction s peut être exprimée simplement comme une solution du problème

$$\rho(X) \in \arg \min_{x \in \mathbb{R}} \mathbb{E}[s(x, X)].$$

Cette notion est donc importante pour définir des procédures de backtesting.

Par exemple, on peut énoncer un résultat de [Gneiting, 2011] qui découle de ceux de [Thomson, 1979] et [Saerens, 2000], et qui montre que les quantiles de niveau $\alpha \in]0, 1[$ sont élicitables, avec une fonction de scoring de la forme

$$s(x, y) = (\mathbf{1}_{\{x \geq y\}} - \alpha) (g(x) - g(y)),$$

où g est une fonction croissante. En revanche, la $TVaR$ n'est pas élicitable.

Dans cette thèse, nous nous intéressons à des quantiles dans le Chapitre III puis à des $TVaR$ dans le Chapitre IV. On s'intéressera aussi longuement aux expectiles, l'unique mesure de risque à la fois élicitable et cohérente (voir [Bellini et al., 2014]). Nous verrons également que toutes ces mesures peuvent être chacune généralisées, soit par les L_p -quantiles (qui sont élicitables), soit par les mesures d'Haezendonck-Goovaerts (qui sont cohérentes) que nous définirons dans le Chapitre IV.

II.2. Variation régulière

Dans cette section, on introduit la notion de variation régulière pour les variables aléatoires. Cette famille, inspirée des fonctions à variation régulière de Karamata (voir [Bingham et al., 1989]), est définie de la manière suivante.

Définition 6 (Variation régulière). *Soit Y une variable aléatoire réelle de fonction de survie $\bar{F}_Y(t) = \mathbb{P}(Y > t)$. Y est dite à variation régulière d'indice $\gamma > 0$, et on notera $Y \in \mathcal{RV}(\gamma)$, si*

$$\lim_{t \rightarrow +\infty} \frac{\bar{F}_Y(t\omega)}{\bar{F}_Y(t)} = \omega^{-\frac{1}{\gamma}}, \forall \omega > 0. \quad (\text{II.1})$$

Le cas limite $\gamma \rightarrow 0^+$ rassemble des distributions dites à queue légère. Cette classe de distributions contient notamment la loi normale, la loi exponentielle (et plus généralement les lois gamma) ou encore la loi de Laplace. De même, les distributions vérifiant l'Equation (6) avec $\gamma > 0$ pourront être qualifiées de lois à queue lourde.

On peut également réécrire la Définition 6 à l'aide de la densité de Y . En effet, Y est à variation régulière d'indice $\gamma > 0$ si

$$\lim_{t \rightarrow +\infty} \frac{f_Y(t\omega)}{f_Y(t)} = \omega^{-\frac{1}{\gamma}-1}, \forall \omega > 0.$$

Enfin, on peut également exprimer la notion de variation régulière avec la fonction quantile de Y de la manière suivante :

$$\lim_{t \rightarrow +\infty} \frac{F_Y^{-1}\left(1 - \frac{1}{t\omega}\right)}{F_Y^{-1}\left(1 - \frac{1}{t}\right)} = \omega^\gamma, \forall \omega > 0.$$

Le théorème de représentation de Karamata (voir [Bingham et al., 1989]) précise qu'une variable aléatoire Y satisfaisant (6) vérifie $\bar{F}_Y(t) = t^{-1/\gamma}L(t)$, où $L : \mathbb{R} \rightarrow \mathbb{R}_+$ est une fonction à variation lente, i.e. qui vérifie $L(t\omega)/L(t) \rightarrow 1$ lorsque $t \rightarrow +\infty$, pour tout $\omega > 0$. Sa fonction de survie se comporte donc presque comme un monôme au voisinage de $+\infty$. Pour être plus précis dans le comportement asymptotique de la fonction de survie, on introduit ci-dessous la notion de variation régulière de second ordre.

Définition 7 (Variation régulière de second ordre). *Soit Y une variable aléatoire réelle de fonction de survie $\bar{F}_Y(t) = \mathbb{P}(Y > t)$. Y est dite à variation régulière de second ordre d'indices $\gamma > 0$ et $\rho \leq 0$, et on notera $Y \in 2\mathcal{RV}(\gamma, \rho)$, si il existe une fonction A de signe constant telle que $A(t) \rightarrow 0$ quand $t \rightarrow +\infty$, et*

$$\lim_{t \rightarrow +\infty} \frac{\frac{\bar{F}_Y(t\omega)}{\bar{F}_Y(t)} - \omega^{-1/\gamma}}{A\left(\frac{1}{\bar{F}_Y(t)}\right)} = \omega^{-1/\gamma} \frac{\omega^{\rho/\gamma} - 1}{\gamma\rho}, \forall \omega > 0. \quad (\text{II.2})$$

Notons que dans le cas $\rho = 0$, le terme $(\omega^{\rho/\gamma} - 1)/(\gamma\rho)$ est remplacé par $\ln(\omega)/\gamma^2$. Cette notion peut s'avérer particulièrement intéressante, notamment dans le cas $\rho < 0$. En effet, comme expliqué dans [de Haan and Ferreira, 2006], si $Y \in 2\mathcal{RV}(\gamma, \rho)$ avec $\rho < 0$, sa fonction de survie sera équivalente à un monôme au voisinage de $+\infty$. En d'autres termes, $\bar{F}_Y(t) \sim c_1 t^{-1/\gamma}$, $c_1 \in \mathbb{R}$ quand $t \rightarrow +\infty$ ou de manière équivalente $F_Y^{-1}(1 - \frac{1}{t}) \sim c_2 t^\gamma$, $c_2 \in \mathbb{R}$, si on réécrit la condition (7) avec la fonction quantile de la manière suivante :

$$\lim_{t \rightarrow +\infty} \frac{\frac{F_Y^{-1}(1 - \frac{1}{t\omega})}{F_Y^{-1}(1 - \frac{1}{t})} - \omega^\gamma}{A(t)} = \omega^\gamma \frac{\omega^\rho - 1}{\rho}, \forall \omega > 0.$$

Cette définition est également valable pour le cas $\rho = 0$, en remplaçant $(\omega^\rho - 1)/\rho$ par $\ln(\omega)$.

Cependant, il semble important de préciser qu'une fonction de survie asymptotiquement équivalente à un monôme n'implique pas nécessairement la condition de second ordre (7). La loi de Pareto en est la preuve (voir Table II.1).

La Table II.1 donne quelques exemples bien connus de distributions à queue lourde, avec leur indice de premier ordre γ , ainsi que celui de second ordre ρ lorsqu'il existe. Les lois Student, Slash et Slash généralisée sont définies dans la Section II.3. Quant aux distributions Pareto, Burr et Fisher, elles sont fournies dans l'Annexe A.

Loi	γ	ρ
Student, $\nu > 0$	$\frac{1}{\nu}$	$-\frac{2}{\nu}$
Pareto, $a > 0$	$\frac{1}{a}$	$\times$
Burr, $a > 0, b > 0$	$\frac{1}{ab}$	$-\frac{1}{b}$
Fisher, $\nu_1, \nu_2 > 0$	$\frac{2}{\nu_2}$	$-\frac{2}{\nu_2}$
Slash, $a > 0$	$\frac{1}{a}$	$\times$
Slash généralisée, $a > 0, m \in \mathbb{N}^*$	$\frac{1}{a}$	$-\frac{1}{a}$

TABLE II.1 – Exemples de distributions à queue lourde.

Il est à noter que de telles distributions sont à considérer avec plus de précautions. En effet, si $Y \in \mathcal{RV}(\gamma)$, alors le moment d'ordre $k > 0$ de Y existe si et seulement si $\gamma < \frac{1}{k}$. Cette notion de variation régulière nous sera utile lorsque nous considérerons des mesures de risque dites extrêmes, telles que des quantiles de niveau α proche de 1.

II.3. Distributions elliptiques

Dans cette thèse, nous nous concentrons sur les distributions elliptiques. Cette section propose donc un panorama des définitions et propriétés de cette famille de distributions.

II.3.a. Premières définitions

Les distributions elliptiques, introduites pour la première fois dans [Kelker, 1970], ont pour but de généraliser la distribution gaussienne, i.e. de construire des distributions symétriques aux propriétés différentes, comme par exemple des distributions possédant une queue de distribution lourde. Ces distributions sont par ailleurs fréquemment utilisées en finance, et notamment en théorie de portefeuille : nous pouvons citer comme exemples [Owen and Rabinovitch, 1983] ou [Xiao and Valdez, 2015].

Elles sont particulièrement intéressantes, car possèdent des propriétés remarquables que nous détaillerons le long de ce chapitre. La première d'entre elle est une fonction caractéristique assez simple. En effet, la définition suivante, que l'on peut trouver dans [Cambanis et al., 1981], introduit la notion de distribution elliptique via sa fonction caractéristique.

Définition 8 (Distribution elliptique). *Un vecteur $\mathbf{X}$ de dimension $d \in \mathbb{N}^*$ est dit elliptique si et seulement si il existe un unique vecteur $\boldsymbol{\mu} \in \mathbb{R}^d$, une matrice semi-définie positive $\boldsymbol{\Sigma} \in \mathbb{R}^{d \times d}$, et une fonction $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}$ telle que la fonction caractéristique de $\mathbf{X} - \boldsymbol{\mu}$ est*

$$\mathbb{E} [\exp\{it^\top (\mathbf{X} - \boldsymbol{\mu})\}] = \phi(t^\top \boldsymbol{\Sigma} t), t \in \mathbb{R}^d.$$

Un tel vecteur elliptique sera noté dans un premier temps $\mathbf{X} \sim \mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$. Il convient de préciser que $\boldsymbol{\Sigma}$ n'est pas nécessairement la matrice de covariance de $\mathbf{X}$. En effet, la matrice de covariance peut d'abord ne pas exister (comme pour la distribution de Cauchy par exemple), et lorsqu'elle existe, les deux matrices sont alors proportionnelles (voir Equation (II.3.a.) plus loin). Pour la suite, nous nous intéresserons à des distributions dites non dégénérées, c'est à dire que nous supposerons que la matrice $\boldsymbol{\Sigma}$ est inversible. Le théorème suivant est fondamental en théorie des distributions elliptiques. Proposé dans [Cambanis et al., 1981], il définit les distributions elliptiques par une représentation stochastique très simple.

Théorème 1 (Représentation stochastique). *Le vecteur $\mathbf{X}$ est elliptique, $\mathbf{X} \sim \mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$, si et seulement si*

$$\mathbf{X} \stackrel{\mathcal{L}}{=} \boldsymbol{\mu} + R\boldsymbol{\Lambda}\mathbf{U}^{(d)} \tag{II.3}$$

où $\boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top = \boldsymbol{\Sigma}$, $\mathbf{U}^{(d)}$ est le vecteur uniformément distribué sur $\mathcal{S}^{d-1}$ la sphère unité de dimension d , et R est une variable aléatoire positive indépendante de $\mathbf{U}^{(d)}$.

Nous pouvons alors noter, de manière équivalente, $\mathbf{X} \sim \mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$ ou $\mathbf{X} \sim \mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, R)$ selon si on définit l'elliptique par sa fonction caractéristique ou par la variable aléatoire R , que nous appellerons rayon. Il est évident que la représentation [1] n'est pas unique. En effet, pour toute matrice proportionnelle à $\boldsymbol{\Sigma}$, par une homothétie de R , on obtient la même distribution. Ajoutons d'ailleurs que la matrice de covariance de $\mathbf{X}$, mentionnée précédemment, peut être exprimée en fonction de R et $\boldsymbol{\Sigma}$ de la manière suivante :

$$\mathbb{V}[\mathbf{X}] = \frac{\mathbb{E}[R^2]}{d} \boldsymbol{\Sigma}. \quad (\text{II.4})$$

Notons également que $\mathcal{S}^{d-1}$ dépend d'une norme. Bien que [Gupta and Song, 1997] ait considéré l'ensemble des normes L_p , la grande majorité des auteurs s'intéresse principalement au cas $p = 2$. Pour la suite, nous choisirons aussi la norme euclidienne. Le théorème suivant, que l'on peut trouver dans [Frahm, 2004], fournit la densité d'un vecteur elliptique.

Théorème 2 (Densité elliptique). *Soit $\mathbf{X} \sim \mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, R)$ un vecteur elliptique de dimension $d \in \mathbb{N}$. Sa densité est donnée par*

$$f_{\mathbf{X}}(\mathbf{x}) = |\boldsymbol{\Sigma}|^{-\frac{1}{2}} c_d g_d \left((\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right), \mathbf{x} \in \mathbb{R}^d, \quad (\text{II.5})$$

où $c_d g_d(t) = (2\pi^{d/2})^{-1} \Gamma(d/2) t^{-(d-1)/2} f_R(\sqrt{t})$, et $f_R(t)$ est la densité de R .

On voit que la densité de $\mathbf{X}$ est directement liée à la densité de son rayon R . Le coefficient c_d est alors appelé coefficient de normalisation, tandis que g_d prend généralement le nom de générateur. Enfin, il apparaît que la densité dépend fortement de la quantité $(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})$, que nous appellerons distance de Mahalanobis de $\mathbf{x}$.

II.3.b. Mélanges d'échelle gaussiens consistants

Par la suite, nous allons considérer une classe particulière de distributions elliptiques, caractérisées dans [Kano, 1994]. Ces distributions sont appelées consistantes ou cohérentes en dimension, et permettent en particulier de définir les champs elliptiques. Par souci de simplicité, nous les appellerons mélanges d'échelle gaussiens consistants (MEGC).

Définition 9 (Mélange d'échelle gaussien consistant). *Soit $(\mathbf{X}_d)_{d \in \mathbb{N}^*}$ une famille de lois elliptiques de dimension $d \in \mathbb{N}^*$ et de fonction caractéristique $(\phi_d)_{d \in \mathbb{N}^*}$. On dit que $(\mathbf{X}_d)_{d \in \mathbb{N}^*}$ est un MEGC si la fonction ϕ_d ne dépend pas de d .*

La Définition [9] est donnée en utilisant la fonction caractéristique de $\mathbf{X}$. On peut également définir la notion de mélange d'échelle gaussien par le rayon R , via la proposition suivante de [Kano, 1994].

Proposition 3 (Représentation des MEGC). *Soit $(\mathbf{X}_d)_{d \in \mathbb{N}}$ une famille de lois elliptiques de dimension $d \in \mathbb{N}^*$ et de rayon R_d . $(\mathbf{X}_d)_{d \in \mathbb{N}}$ est alors un MEGC si et seulement si R_d admet la représentation stochastique suivante :*

$$R_d \stackrel{\mathcal{L}}{=} \chi_d \xi, \quad (\text{II.6})$$

où χ_d est la racine carrée d'une loi du χ^2 à d degrés de liberté, ξ est une variable aléatoire qui ne dépend pas de d , et χ_d , ξ et $\mathbf{U}^{(d)}$ de [\[1\]](#) sont mutuellement indépendantes.

Des distributions appartenant à cette famille sont alors déterminées par $\boldsymbol{\mu} \in \mathbb{R}^d$, la matrice $\boldsymbol{\Sigma} \in \mathbb{R}^{d \times d}$, ainsi qu'une variable aléatoire $\xi \in \mathbb{R}$ indépendante de la dimension d , que l'on nommera variable d'échelle. On notera $\mathbf{X} \sim \mathcal{EC}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \xi)$. Pour résumer, un tel vecteur admet la représentation $\mathbf{X} \stackrel{\mathcal{L}}{=} \boldsymbol{\mu} + \chi_d \xi \boldsymbol{\Lambda} \mathbf{U}^{(d)} \stackrel{\mathcal{L}}{=} \boldsymbol{\mu} + \xi \mathcal{N}(0, \boldsymbol{\Sigma})$. Il s'agit finalement d'une loi normale à variance aléatoire $\xi^2 \boldsymbol{\Sigma}$, où ξ est indépendant de la dimension d . La proposition suivante donne une propriété importante des MEGC, à savoir la stabilité des sous-vecteurs, ainsi que la stabilité par la somme. Une preuve peut être trouvée dans [\[Hult and Lindskog, 2002\]](#).

Proposition 4 (Loi de la somme). *Soit $\mathbf{Z} \sim \mathcal{EC}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \xi)$. Alors pour tout $\mathbf{c} \in \mathbb{R}^d$, on a $\mathbf{c}^\top \mathbf{Z} \sim \mathcal{EC}_1(\mathbf{c}^\top \boldsymbol{\mu}, \mathbf{c}^\top \boldsymbol{\Sigma} \mathbf{c}, \xi)$. De plus, si $\mathbf{Z} = (\mathbf{X}, \mathbf{Y})$ avec $\mathbf{X} \in \mathbb{R}^{d_1}$, $\mathbf{Y} \in \mathbb{R}^{d_2}$ tels que $d_1 + d_2 = d$, on écrit*

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_X & \boldsymbol{\Sigma}_{XY} \\ \boldsymbol{\Sigma}_{YX} & \boldsymbol{\Sigma}_Y \end{pmatrix}, \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_X \\ \boldsymbol{\mu}_Y \end{pmatrix}.$$

Alors $\mathbf{X} \sim \mathcal{EC}_{d_1}(\boldsymbol{\mu}_X, \boldsymbol{\Sigma}_X, \xi)$ et $\mathbf{Y} \sim \mathcal{EC}_{d_2}(\boldsymbol{\mu}_Y, \boldsymbol{\Sigma}_Y, \xi)$.

Ainsi, tout sous-vecteur d'un vecteur MEGC est défini par la même variable d'échelle ξ . Notons que le fait que ξ soit indépendant de la dimension est la clé de cette proposition. Dans un cadre spatial, cette propriété est indispensable pour pouvoir construire des champs aléatoires elliptiques, que l'on peut trouver dans [\[Opitz, 2016\]](#) par exemple. La définition ci dessous caractérise les champs elliptiques.

Définition 10 (Champ elliptique). *Soit $(Z_t)_{t \in T}$ un champ aléatoire indexé sur $T \subset \mathbb{R}^d$. On dira que $(Z_t)_{t \in T}$ est un champ ξ -elliptique si*

$$(Z_t)_{t \in T} = \mu_t + \xi \mathcal{N}(0, \sigma(.,.))$$

où $\mathcal{N}(0, \sigma(.,.))$ est un champ gaussien centré, et de fonction de covariance $\sigma : T \times T \rightarrow \mathbb{R}$.

Dans [\[Maume-Deschamps et al., 2017b\]](#) et [\[Maume-Deschamps et al., 2018\]](#), on se place justement dans un cadre spatial pour la prédiction de mesures de risque conditionnelles, qui se réduit finalement à un problème multivarié. En terme d'application, on peut imaginer la prédiction de température, où les covariables seraient des stations de mesure, comme dans [\[El Methni et al., 2014\]](#), par exemple. Etant donné l'objectif de ce travail, nous consacrons la dernière proposition aux propriétés des vecteurs elliptiques conditionnés par certaines de leurs composantes.

Proposition 5 (Loi conditionnelle). Soit $\mathbf{Z} = (\mathbf{X}, \mathbf{Y}) \sim \mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, R)$ avec $\mathbf{X} \in \mathbb{R}^{d_1}$, $\mathbf{Y} \in \mathbb{R}^{d_2}$ tels que $d_1 + d_2 = d$. On écrit

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{\mathbf{X}} & \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{Y}} \\ \boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{X}} & \boldsymbol{\Sigma}_{\mathbf{Y}} \end{pmatrix}, \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_{\mathbf{X}} \\ \boldsymbol{\mu}_{\mathbf{Y}} \end{pmatrix}.$$

La distribution conditionnelle de $\mathbf{Y} | (\mathbf{X} = \mathbf{x})$ est elliptique de paramètres

$$\boldsymbol{\mu}_{\mathbf{Y}|\mathbf{x}} = \boldsymbol{\mu}_{\mathbf{Y}} + \boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{X}} \boldsymbol{\Sigma}_{\mathbf{X}}^{-1} (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{X}}), \boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{x}} = \boldsymbol{\Sigma}_{\mathbf{Y}} - \boldsymbol{\Sigma}_{\mathbf{Y}\mathbf{X}} \boldsymbol{\Sigma}_{\mathbf{X}}^{-1} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{Y}} \quad (\text{II.7})$$

et de rayon R^* défini comme suit :

$$R^* \stackrel{\mathcal{L}}{=} R \sqrt{1 - \beta} \left\{ R \sqrt{\beta} \mathbf{U}^{(d_1)} = \mathbf{C}_{\mathbf{X}}^{-1} (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{X}}) \right\} \quad (\text{II.8})$$

où $\mathbf{C}_{\mathbf{X}}$ est la racine de Cholesky de $\boldsymbol{\Sigma}_{\mathbf{X}}$, et β est une loi Beta de paramètres $d_1/2$ et $d_2/2$. Enfin, si $\mathbf{Z}$ est un MEGC, i.e. si $R \stackrel{\mathcal{L}}{=} \chi_d \xi$, la densité conditionnelle de $\mathbf{Y} | (\mathbf{X} = \mathbf{x})$ est donnée par

$$f_{\mathbf{Y}|\mathbf{x}}(\mathbf{y}|\mathbf{x}) = |\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{x}}|^{-1/2} c_{\mathbf{Y}|\mathbf{x}} g_d \left(M(\mathbf{x}) + (\mathbf{y} - \boldsymbol{\mu}_{\mathbf{Y}|\mathbf{x}})^\top \boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{x}}^{-1} (\mathbf{y} - \boldsymbol{\mu}_{\mathbf{Y}|\mathbf{x}}) \right) \quad (\text{II.9})$$

où $c_{\mathbf{Y}|\mathbf{x}} = c_d / (c_{d_1} g_{d_1}(M(\mathbf{x})))$ et $M(\mathbf{x}) = (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{X}})^\top \boldsymbol{\Sigma}_{\mathbf{X}}^{-1} (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{X}})$.

La première partie de la Proposition 5 i.e. les Equations (5) et (5), est valable pour toutes les distributions elliptiques. En revanche, la formule de densité conditionnelle (5) ne s'applique que pour les MEGC.

Preuve : Les détails de l'obtention des Equations (5) et (5) peuvent être trouvés respectivement dans [Cambanis et al., 1981] et [Frahm, 2004]. La densité conditionnelle (5) est quant à elle immédiate. En effet, on a $f_{\mathbf{Y}|\mathbf{x}}(\mathbf{y}|\mathbf{x}) = f_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, \mathbf{y}) / f_{\mathbf{X}}(\mathbf{x})$. En utilisant la Proposition 4 et la stabilité des sous-vecteurs, alors

$$f_{\mathbf{Y}|\mathbf{x}}(\mathbf{y}|\mathbf{x}) = \frac{|\boldsymbol{\Sigma}|^{-1/2} c_d g_d \left(((\mathbf{x}, \mathbf{y}) - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} ((\mathbf{x}, \mathbf{y}) - \boldsymbol{\mu}) \right)}{|\boldsymbol{\Sigma}_{\mathbf{X}}|^{-1/2} c_{d_1} g_{d_1}(M(\mathbf{x}))}$$

En remarquant que $((\mathbf{x}, \mathbf{y}) - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} ((\mathbf{x}, \mathbf{y}) - \boldsymbol{\mu}) = M(\mathbf{x}) + (\mathbf{y} - \boldsymbol{\mu}_{\mathbf{Y}|\mathbf{x}})^\top \boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{x}}^{-1} (\mathbf{y} - \boldsymbol{\mu}_{\mathbf{Y}|\mathbf{x}})$ et $|\boldsymbol{\Sigma}| = |\boldsymbol{\Sigma}_{\mathbf{X}}| |\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{x}}|$, on obtient le résultat. $\square$

II.3.c. Exemples

Dans cette section, nous donnons quelques exemples de Mélanges d'Echelle Gaussiens Consistants, leurs densités ainsi que les variables d'échelle ξ qui les caractérisent.

II.3.c.1. Loi Normale

La loi Normale est probablement le cas le plus connu, notamment pour sa simplicité. En effet, dans ce cas, la variable d'échelle ξ vaut 1 presque sûrement. Dès lors, en dimension $d \in \mathbb{N}$, sa densité est donnée par

$$\frac{1}{(2\pi)^{d/2}} \exp \left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \right), \mathbf{x} \in \mathbb{R}^d.$$

II.3.c.2. Loi Normale contaminée

La première idée pour la variable d'échelle ξ pourrait être une loi discrète sur m atomes $\theta_1, \dots, \theta_m$. Pour tout $i \in \{1, \dots, m\}$, ξ prendrait alors la valeur θ_i avec probabilité p_i tel que $\sum_i p_i = 1$. De telles distributions sont appelées des lois normales contaminées, et peuvent être trouvées dans [Tukey, 1960]. Leur densité est alors donnée par

$$\frac{1}{(2\pi)^{d/2}} \sum_{k=1}^m p_k \theta_k^{-d} \exp \left(-\frac{1}{2\theta_k^2}(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \right), \mathbf{x} \in \mathbb{R}^d.$$

Les lois normales contaminées permettent modifier la structure "en cloche" de la loi normale, mais ne changent en aucun cas les propriétés de la queue de distribution, qui reste légère.

II.3.c.3. Loi de Student

La loi de Student à $\nu > 0$ degrés de liberté, détaillée dans [Nadarajah and Kotz, 2004], admet quant à elle une variable d'échelle continue, donnée par la représentation $\xi \stackrel{\mathcal{L}}{=} \sqrt{\nu/\chi_\nu^2}$. Sa densité est par conséquent

$$\frac{\Gamma\left(\frac{d+\nu}{2}\right)}{(\nu\pi)^{d/2}\Gamma\left(\frac{d+\nu}{2}\right)} \left(1 + \frac{1}{\nu}(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right)^{-\frac{d+\nu}{2}}, \mathbf{x} \in \mathbb{R}^d.$$

Une différence majeure de la loi Student, par rapport à la loi normale, est sa queue de distribution, bien plus lourde. Cependant, il n'est pas difficile de voir que la loi Student tend vers la loi normale lorsque le degré de liberté ν tend vers l'infini.

Le cas $\nu = 1$ correspond à la loi de Cauchy, connue pour être une loi stable.

II.3.c.4. Loi de Laplace

Comme résumé dans [Eltoft et al., 2006], prenons $\xi \stackrel{\mathcal{L}}{=} \sqrt{\mathcal{E}(1/\lambda)}$, $\lambda > 0$ où $\mathcal{E}(1/\lambda)$ est une loi exponentielle de paramètre $1/\lambda$, et nous obtenons une loi de Laplace de paramètre

λ , de densité

$$\frac{2}{\lambda(2\pi)^{d/2}} \frac{K_{d/2-1} \left(\sqrt{\frac{2}{\lambda}} (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right)}{\left[\sqrt{\frac{\lambda}{2}} (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right]^{d/2-1}}, \quad \mathbf{x} \in \mathbb{R}^d - \{\boldsymbol{\mu}\},$$

où $K_m(x) = \frac{\pi}{2} \frac{I_{-m}(x) - I_m(x)}{\sin(m\pi)}$ est la fonction de Bessel de second type d'ordre m modifiée, avec $I_m(x) = \sum_{k=0}^{+\infty} \frac{1}{k! \Gamma(k+m+1)} \left(\frac{x}{2}\right)^{2k+m}$ (voir [Abramowitz et al., 1966]). La variable d'échelle étant à queue légère, la loi de Laplace a les mêmes propriétés de queue de la loi normale.

II.3.c.5. Lois de Slash et de Slash généralisée

Introduite dans [Wang and Genton, 2006], la loi de Slash admet comme échelle ξ une loi de Pareto $\mathcal{P}(1, a)$, $a > 0$ de densité

$$\frac{a}{x^{a+1}}, x \geq 1.$$

Une loi de Slash de paramètres $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ et a a donc comme densité

$$\frac{2^{a/2-1} a \Gamma\left(\frac{d+a}{2}\right) \chi_{d+a}^2 \left((\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right)}{\pi^{d/2} \left[(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right]^{(d+a)/2}}, \quad \mathbf{x} \in \mathbb{R}^d - \{\boldsymbol{\mu}\}, \quad (\text{II.10})$$

où $\chi_{d+a}^2(\cdot)$ est la fonction de répartition d'une loi du χ^2 à $d+a$ degrés de liberté.

Ce type de distribution possède des propriétés de queue similaires aux lois de Student données précédemment, à savoir une queue de distribution lourde. De plus, il tend vers la loi normale lorsque a tend vers l'infini. C'est donc également une bonne alternative à la distribution gaussienne. Cependant, on a pu constater dans la Table [I.1] qu'elle ne vérifiait pas la propriété de variation régulière de second ordre. On peut alors modifier la loi Slash, en remplaçant la loi de Pareto $\mathcal{P}(1, a)$, $a > 0$ par l'inverse d'une loi Beta de paramètres $a > 0$ et $m \in \mathbb{N}^*$. En utilisant la densité de la loi Beta, donnée en Annexe [A] on obtient la densité suivante pour une pareille loi :

$$\frac{1}{\pi^{d/2} B(a, m)} \sum_{k=0}^{m-1} \binom{k}{m-1} (-1)^k \frac{\Gamma\left(\frac{a+k+d}{2}\right) \chi_{a+k+d}^2 \left((\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right)}{2^{1-(a+k)/2} \left[(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right]^{(a+k+d)/2}}, \quad \mathbf{x} \in \mathbb{R}^d - \{\boldsymbol{\mu}\},$$

où B est la fonction Bêta définie dans l'Annexe [A] et $\chi_{a+k+d}^2(\cdot)$ est la fonction de répartition d'une loi du χ^2 à $a+k+d$ degrés de liberté. On voit bien, à l'aide de la densité ([I.3.c.5.]), que le cas $m = 1$ mène à la loi de Slash. Cette distribution sera nommée par la suite loi de Slash généralisée.

Le paragraphe suivant est un bilan de cette section sur les distributions elliptiques.

II.3.d. Bilan

La Table II.2 propose un bilan des lois elliptiques que nous venons d'expliciter, en donnant la constante c_d , le générateur g_d (comme défini dans l'Equation (2)), ainsi que la variable d'échelle ξ , issu de la représentation (3). Notons que cette classe de distributions elliptiques consistantes est assez large, la variable d'échelle ξ n'ayant comme restriction que d'être positive et indépendante de la loi du χ_d^2 dans la représentation (3).

Loi	c_d	$g_d(t)$	ξ
Normale	$\frac{1}{(2\pi)^{\frac{d}{2}}}$	$\exp(-\frac{t}{2})$	1
Normale contaminée, $m \in \mathbb{N}$, $\boldsymbol{\theta}$, $\mathbf{p}$	$\frac{1}{(2\pi)^{\frac{d}{2}}}$	$\sum_{k=1}^m \pi_k \theta_k^{-d} \exp\left(-\frac{1}{2\theta_k^2} t\right)$	$\sum_{k=1}^m p_k \delta_{\theta_k}$
Student, $\nu > 0$	$\frac{\Gamma(\frac{d+\nu}{2})}{\Gamma(\frac{\nu}{2})} \frac{1}{(\nu\pi)^{\frac{d}{2}}}$	$(1 + \frac{t}{\nu})^{-\frac{d+\nu}{2}}$	$\sqrt{\frac{\nu}{\chi_\nu^2}}$
Laplace, $\lambda > 0$	$\frac{2}{\lambda(2\pi)^{d/2}}$	$\frac{K_{d/2-1}(\sqrt{\frac{2}{\lambda}} t)}{(\sqrt{\frac{\lambda}{2}} t)^{d/2-1}}$	$\sqrt{\mathcal{E}(\frac{1}{\lambda})}$
Slash, $a > 0$	$\frac{2^{\frac{a}{2}-1} a \Gamma(\frac{d+a}{2})}{\pi^{\frac{d}{2}}}$	$\frac{\chi_{d+a}^2(t)}{t^{\frac{d+a}{2}}}$	$\mathcal{P}(1, a)$
Slash généralisée, $a > 0, m \in \mathbb{N}^*$	$\frac{1}{\pi^{d/2} B(a, m)}$	$\sum_{k=0}^{m-1} \binom{k}{m-1} (-1)^k \frac{\Gamma(\frac{a+k+d}{2})}{2^{1-(a+k)/2}} \frac{\chi_{a+k+d}^2(t)}{t^{(a+k+d)/2}}$	$\frac{1}{\beta(a, m)}$

TABLE II.2 – Exemples de Mélanges d'Echelle Gaussiens Consistants

Un formulaire des fonctions K , Γ , B , ainsi que des lois $\mathcal{E}$, χ_ν^2 , $\mathcal{P}$ et β est fourni en Annexe A.

Dans le but de donner un aperçu visuel de ces distributions, la Figure II.1 représente leur densité dans le cas sphérique bivarié, i.e. lorsque $d = 2$, $\boldsymbol{\Sigma} = \mathbf{I}_2$ et $\boldsymbol{\mu} = \mathbf{0}_{\mathbb{R}^2}$.

Finalement, nous avons vu dans ce chapitre que les distributions elliptiques étaient un outil assez simple à utiliser, et plutôt répandu dans la littérature, qui permettait de conserver les propriétés de stabilité par combinaison linéaire ainsi que celles de projection de la loi gaussienne, avec la possibilité supplémentaire d'ajuster la queue de distribution. Cependant, la représentation stochastique (5) montre que l'étude des distributions conditionnelles n'est pas immédiate, et a suscité beaucoup d'intérêt. En effet, il existe de nombreux papiers traitant du comportement des vecteurs elliptiques lorsqu'une des composantes dépasse un certain seuil. A titre d'exemples, [Fougères and Soulier, 2010] ou [Hashorva and Jaworski, 2012] proposent de telles études.

Dans cette thèse, nous nous intéresserons au cas où le vecteur est conditionné par l'égalité de certaines de ses composantes. On considérera donc un vecteur $\mathbf{Z} = (\mathbf{X}, Y) \sim$

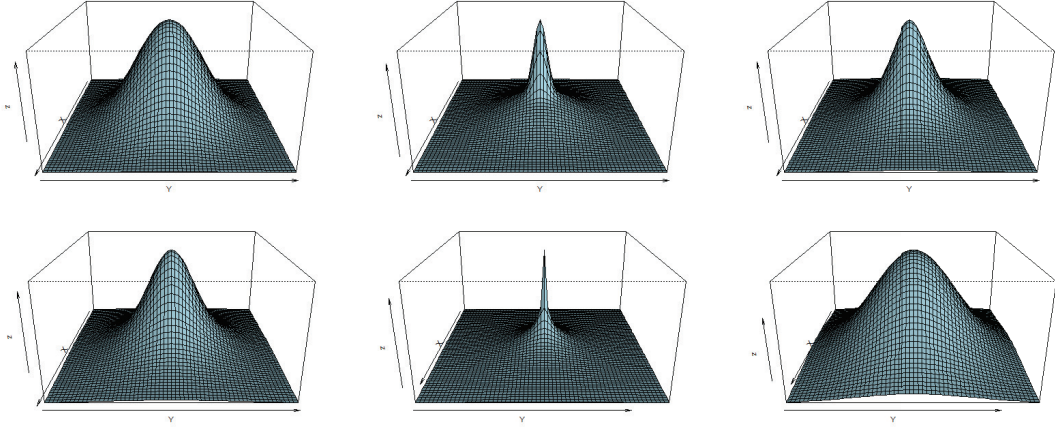


FIGURE II.1 – Densités bivariées (de haut en bas, de gauche à droite) : loi Normale, loi Normale contaminée avec $m = 2$, $p_1 = 0.9$, $p_2 = 0.1$, $\theta_1 = 1$, $\theta_2 = 0.2$, loi de Cauchy, loi de Student avec $\nu = 2$, loi de Laplace avec $\lambda = 2$ et loi de Slash avec $a = 2$.

$\mathcal{EC}_{d+1}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \xi)$ avec $\mathbf{X} \in \mathbb{R}^d$, $Y \in \mathbb{R}$ tels que

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_X & \boldsymbol{\Sigma}_{XY} \\ \boldsymbol{\Sigma}_{YX} & \Sigma_Y \end{pmatrix}, \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_X \\ \mu_Y \end{pmatrix},$$

et s'intéressera à la loi conditionnelle $Y|\mathbf{X} = \mathbf{x}$, $\mathbf{x} \in \mathbb{R}^d$. Dans cette optique, [Hashorva, 2007] a proposé des relations asymptotiques entre les répartitions des rayons conditionnel et inconditionnel. Si R est à variation régulière d'indice $\gamma > 0$, alors

$$\overline{F}_{R^*}(t) \underset{t \rightarrow +\infty}{\sim} \frac{c(M(\mathbf{x}))}{1 + d\gamma} t^{-d} \overline{F}_R(t). \quad (\text{II.11})$$

Notre but sera d'étudier certaines mesures de risque sur cette loi. La première d'entre elles, à laquelle le chapitre suivant est consacré, est le quantile. Il s'agit pour cela d'inverser la relation (II.3.d.), ce qui n'est pas immédiat. Nous verrons par la suite qu'il est possible de déduire d'autres mesures à partir du quantile.

QUANTILES CONDITIONNELS

Le quantile, parfois appelé Value at Risk en finance et assurance, est probablement la plus connue des mesures de risque. La raison de cette popularité réside notamment dans sa simplicité. En effet, si on considère une variable aléatoire réelle X de fonction de répartition F , le quantile de niveau $\alpha \in]0, 1[$, noté $q_\alpha(X)$, est défini comme suit.

$$q_\alpha(X) = \inf\{x : F(x) \geq \alpha\}.$$

Dans le cas où la fonction F est strictement croissante, on utilise également la notation $q_\alpha(X) = F^{-1}(\alpha)$. En d'autres termes, pour un niveau α , le quantile est la valeur pour laquelle $100\alpha\%$ des valeurs d'un échantillon X sont inférieures. C'est pourquoi cette mesure connaît de nombreuses applications, notamment en finance (voir par exemple [Jorion, 1996]) où il paraît intéressant si ce n'est nécessaire d'avoir une région de confiance concernant la perte d'un portefeuille. Plus récemment, les directives Solvabilité II obligent les compagnies d'assurance à dominer un quantile de niveau 0.995 sur leurs dépenses de remboursement. Dans le domaine de l'environnement, la mesure du quantile est également utile pour l'appréhension des catastrophes naturelles notamment. Par exemple, [Jagger and Elsner, 2009] l'emploie pour modéliser l'intensité des cyclones.

Dans cette section, comme amorcé dans l'introduction, on s'intéressera aux quantiles conditionnels d'une variable aléatoire en présence d'une covariable, dans le cadre elliptique. Pour cela, on définira dans un premier temps le cadre, puis on donnera quelques cas simples où l'on est capable de calculer des quantiles conditionnels. On proposera par la suite le modèle de régression comme première approche, pour lequel nous porterons un regard critique, qui nous mènera à l'étude plus particulière des quantiles dits extrêmes. Une application numérique est enfin proposée, avec notamment une étude sur des données réelles. Une grande partie de ce chapitre est publié dans [Maume-Deschamps et al., 2017b], une autre partie peut être trouvée dans [Usseglio-Carleve, 2017].

III.1. Problématique

On définit ici de manière formelle les quantiles conditionnels dans le cas d'un vecteur elliptique consistant. Dans un premier temps, on donne une forme générale pour ces quantiles, puis on donne quelques exemples pour lesquels on obtient des formules fermées.

III.1.a. Cas général

On rappelle $\mathbf{Z} = (\mathbf{X}, Y) \sim \mathcal{EC}_{d+1}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \xi)$ avec $\mathbf{X} \in \mathbb{R}^d$, $Y \in \mathbb{R}$ tels que

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{\mathbf{X}} & \boldsymbol{\Sigma}_{\mathbf{X}Y} \\ \boldsymbol{\Sigma}_{Y\mathbf{X}} & \Sigma_Y \end{pmatrix}, \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_{\mathbf{X}} \\ \mu_Y \end{pmatrix},$$

et on notera Φ_Y la fonction de répartition de la loi Y centrée réduite, i.e. de $(Y - \mu_Y)/\sqrt{\Sigma_Y}$. De la même manière, $\Phi_{Y|\mathbf{x}}$ sera la fonction de répartition de la loi (elliptique, voir Proposition 5) conditionnelle centrée réduite $(Y|(\mathbf{X} = \mathbf{x}) - \mu_{Y|\mathbf{x}})/\sqrt{\Sigma_{Y|\mathbf{X}}}$. Rappelons que les quantités $\mu_{Y|\mathbf{x}}$ et $\Sigma_{Y|\mathbf{X}}$ sont introduites dans l'Equation (5). En utilisant ces notations, il n'est pas difficile avec les propriétés d'homogénéité puis d'invariance par translation du quantile, de fournir une expression générale pour le quantile de niveau $\alpha \in]0, 1[$ de $Y|(\mathbf{X} = \mathbf{x})$.

$$q_\alpha(Y|\mathbf{X} = \mathbf{x}) = \mu_{Y|\mathbf{x}} + \sqrt{\Sigma_{Y|\mathbf{X}}} \Phi_{Y|\mathbf{x}}^{-1}(\alpha). \quad (\text{III.1})$$

Si les quantités $\mu_{Y|\mathbf{x}}$ et $\Sigma_{Y|\mathbf{X}}$ paraissent relativement simples à exploiter au vu de l'Equation (5), le quantile conditionnel $\Phi_{Y|\mathbf{x}}^{-1}$ semble beaucoup moins évident. En effet, d'un point de vue statistique, les deux premiers paramètres peuvent être estimés par la méthode des moments voire par des M-estimateurs de [Maronna, 1976] ou [Tyler, 1987]. En revanche, le quantile conditionnel ne peut être estimé directement.

Le but dans ce chapitre sera donc de proposer des méthodes d'estimation de $\Phi_{Y|\mathbf{x}}^{-1}$. Pour cela, on va dans un premier temps proposer un modèle de régression, où le quantile de $Y|(\mathbf{X} = \mathbf{x})$ sera approximé par une combinaison affine de la covariable $\mathbf{X}$. Nous donnerons les paramètres de régression optimaux au sens d'une fonction de perte asymétrique, puis porterons un regard critique sur cette méthode appliquée à ce cadre elliptique. De plus, son inefficacité dans le cas des niveaux de quantile extrêmes (i.e. lorsque $\alpha \rightarrow 0$ ou $\alpha \rightarrow 1$) nous mènera à proposer une autre approche pour de tels quantiles.

III.1.b. Quelques quantiles conditionnels théoriques

De manière à pouvoir appréhender la performance des approximations et estimations que l'on va proposer par la suite, on donne dans ce paragraphe quelques formules théoriques de quantiles conditionnels. Les deux exemples traités sont la loi Normale et la loi de Student.

III.1.b.1. Loi Normale

Le cas gaussien est évidemment le cas le plus simple. En effet, la distribution conditionnelle de $Y|\mathbf{X} = \mathbf{x}$ reste gaussienne, de paramètres $\mu_{Y|\mathbf{x}}$ et $\Sigma_{Y|\mathbf{X}}$. Le quantile de niveau

α s'obtient alors immédiatement de la manière suivante :

$$q_\alpha(Y|\mathbf{X} = \mathbf{x}) = \mu_{Y|\mathbf{x}} + \sqrt{\Sigma_{Y|\mathbf{X}}} \Phi^{-1}(\alpha)$$

où Φ est la fonction de répartition de la loi normale centrée réduite. Les problématiques de krigeage (voir par exemple [Ligas and Kulczycki, 2010]) utilisent beaucoup cette notion. En effet, l'idée est de considérer un champ gaussien observé en un nombre $d \in \mathbb{N}$ de points, et de prédire la valeur du champ Y à un autre point. Les quantités $\mu_{Y|\mathbf{x}}$ et $\Sigma_{Y|\mathbf{X}}$ sont alors appelées moyenne et variance de krigeage.

III.1.b.2. Loi de Student

Dans le cas des lois de Student, la distribution conditionnelle n'est pas de même nature, contrairement au cas gaussien. Cependant, il est possible d'obtenir une formule pour les quantiles conditionnels. Pour cela, on donne le lemme suivant, issu de [Nadarajah and Kotz, 2004], qui fournit la densité de la loi conditionnelle.

Lemme 6 (Densité Student conditionnelle). *Soit $\mathbf{Z} \in \mathbb{R}^d$ un vecteur Student à ν degrés de liberté de paramètres $\boldsymbol{\mu} \in \mathbb{R}^d$ et $\boldsymbol{\Sigma} \in \mathbb{R}^{d \times d}$. Considérons $\mathbf{Z} = (\mathbf{X}, \mathbf{Y})^\top$ avec $\mathbf{X} \in \mathbb{R}^{d_1}$, $\mathbf{Y} \in \mathbb{R}^{d_2}$, et $d_1 + d_2 = d$. Alors la densité du vecteur aléatoire $\mathbf{Y} | (\mathbf{X} = \mathbf{x})$ est donnée par*

$$f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = \frac{\Gamma\left(\frac{\nu+d}{2}\right)}{\{(\nu+d_1)\pi\}^{\frac{d_2}{2}} \Gamma\left(\frac{\nu+d_1}{2}\right) |\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}|^{\frac{1}{2}}} \left\{ 1 + \frac{1}{\nu} \frac{M(\mathbf{y}|\mathbf{x})}{1 + \frac{1}{\nu} M(\mathbf{x})} \right\}^{-\frac{\nu+d}{2}} \times \left(\frac{\frac{\nu+d_1}{\nu}}{1 + \frac{1}{\nu} M(\mathbf{x})} \right)^{\frac{d_2}{2}}, \quad (\text{III.2})$$

où $M(\mathbf{y}|\mathbf{x})$ et $M(\mathbf{x})$ sont les distances de Mahalanobis suivantes :

$$M(\mathbf{y}|\mathbf{x}) = (\mathbf{y} - \boldsymbol{\mu}_{\mathbf{Y}|\mathbf{x}})^\top \boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1} (\mathbf{y} - \boldsymbol{\mu}_{\mathbf{Y}|\mathbf{x}}), \quad M(\mathbf{x}) = (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{X}})^\top \boldsymbol{\Sigma}_{\mathbf{X}}^{-1} (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{X}}).$$

Avec la densité précédente, il est alors possible d'obtenir une formule explicite pour le quantile conditionnel dans le cas Student.

Proposition 7 (Quantile Student conditionnel). *Soit $\mathbf{Z} \in \mathbb{R}^{d+1}$ un vecteur Student à ν degrés de liberté et paramètres $\boldsymbol{\mu} \in \mathbb{R}^{d+1}$ et $\boldsymbol{\Sigma} \in \mathbb{R}^{(d+1) \times (d+1)}$. On considère $\mathbf{Z} = (\mathbf{X}, Y)^\top$ avec $\mathbf{X} \in \mathbb{R}^d$ et $Y \in \mathbb{R}$. Le quantile conditionnel de niveau α de $Y | (\mathbf{X} = \mathbf{x})$ est alors donné par*

$$q_\alpha(Y|\mathbf{X} = \mathbf{x}) = \mu_{Y|\mathbf{x}} + \sqrt{\Sigma_{Y|\mathbf{X}}} \sqrt{\frac{\nu}{\nu+d}} \sqrt{1 + \frac{1}{\nu} M(\mathbf{x})} \Phi_{\nu+d}^{-1}(\alpha). \quad (\text{III.3})$$

où Φ_m est la fonction de répartition de la loi de Student à m degrés de liberté centrée réduite.

Preuve : On considère la densité [\[6\]](#), avec $d_1 = d$ et $d_2 = 1$. On a $M(y|\mathbf{x}) = (y - \mu_{Y|\mathbf{x}})^2 / \Sigma_{Y|\mathbf{x}} \in \mathbb{R}$. Soit q_α le quantile conditionnel de $Y|(\mathbf{X} = \mathbf{x})$. Il vérifie

$$\int_{-\infty}^{q_\alpha} f_{Y|\mathbf{X}}(y|\mathbf{x}) dy = \alpha.$$

D'où

$$\frac{\Gamma\left(\frac{\nu+d+1}{2}\right) \{(\nu+d)\pi\}^{-1/2}}{\Gamma\left(\frac{\nu+d}{2}\right) |\Sigma_{Y|\mathbf{X}}|^{1/2}} \left(\frac{\frac{\nu+d}{\nu}}{1 + M(\mathbf{x})/\nu} \right)^{1/2} \int_{-\infty}^{q_\alpha} \left\{ 1 + \frac{1}{\nu+d} \frac{\frac{\nu+d}{\nu} M(y|\mathbf{x})}{1 + M(\mathbf{x})/\nu} \right\}^{-\frac{\nu+d+1}{2}} dy = \alpha$$

ou, avec

$$\begin{cases} t &= \sqrt{\frac{\nu+d}{\nu}}(y - \mu_{Y|\mathbf{x}}) \left(\sqrt{\Sigma_{Y|\mathbf{X}}} \sqrt{1 + M(\mathbf{x})/\nu} \right)^{-1} \\ t_\alpha &= \sqrt{\frac{\nu+d}{\nu}}(q_\alpha - \mu_{Y|\mathbf{x}}) \left(\sqrt{\Sigma_{Y|\mathbf{X}}} \sqrt{1 + M(\mathbf{x})/\nu} \right)^{-1} \end{cases},$$

$$\frac{\Gamma\left(\frac{\nu+d+1}{2}\right)}{\{(\nu+d)\pi\}^{1/2} \Gamma\left(\frac{\nu+d}{2}\right)} \times \int_{-\infty}^{t_\alpha} \left(1 + \frac{t^2}{\nu+d} \right)^{-\frac{\nu+d+1}{2}} dt = \alpha.$$

Soit $\Phi_{\nu+d}$ la fonction de répartition Student à $d + \nu$ degrés de liberté. Il est évident que

$$\Phi_{\nu+d} \left\{ \frac{\sqrt{\frac{\nu+d}{\nu}}(q_\alpha - \mu_{Y|\mathbf{x}})}{\sqrt{\Sigma_{Y|\mathbf{X}}} \sqrt{1 + M(\mathbf{x})/\nu}} \right\} = \alpha.$$

Par conséquent,

$$q_\alpha = \mu_{Y|\mathbf{x}} + \sqrt{\frac{\nu}{\nu+d}} \sqrt{\Sigma_{Y|\mathbf{X}}} \sqrt{1 + \frac{1}{\nu} M(\mathbf{x})} \Phi_{\nu+d}^{-1}(\alpha).$$

La preuve est alors complète. $\square$

On se servira notamment des quantiles de Student conditionnels dans l'étude numérique de la Section [III.4](#).

III.2. Le modèle de régression

Si on vient de voir qu'il est possible d'obtenir des formules explicites de quantile conditionnel pour quelques cas particuliers, on ne dispose pas de pareil résultat dans le cas général des distributions elliptiques. Une première idée serait alors de considérer un modèle de régression quantile.

III.2.a. Présentation du modèle

La régression quantile, introduite par [Koenker and Bassett, 1978], est une méthode assez répandue pour estimer des quantiles conditionnels. Le principe consiste à approximer le quantile de niveau α de $Y|(\mathbf{X} = \mathbf{x})$ par une combinaison affine de $\mathbf{x}$.

$$q_\alpha^{RQ}(Y|\mathbf{X} = \mathbf{x}) = \boldsymbol{\beta}^{*\top} \mathbf{x} + \beta_0^*. \quad (\text{III.4})$$

La quantité $q_\alpha^{RQ}(Y|\mathbf{X})$ est appelée prédicteur de régression quantile. Les coefficients de régression $\boldsymbol{\beta}^*$ et β_0^* sont des solutions du problème de minimisation

$$(\boldsymbol{\beta}^*, \beta_0^*) = \arg \min_{\boldsymbol{\beta} \in \mathbb{R}^d, \beta_0 \in \mathbb{R}} \mathbb{E} [\mathcal{S}_\alpha(Y - \boldsymbol{\beta}^\top \mathbf{X} - \beta_0)] \quad (\text{III.5})$$

avec la fonction $\mathcal{S}_\alpha$ que l'on appellera fonction de score (voir [Grant and Gneiting, 2012]). Si on note $t^+ = t \mathbb{1}_{\{t > 0\}}$, on a l'expression suivante :

$$\mathcal{S}_\alpha(t) = (\alpha - 1)t + t^+. \quad (\text{III.6})$$

Tel qu'est formulé le problème dans l'Equation (III.2.a.), il semble que ce modèle impose que les moments d'ordre 1 existent, ce qui exclurait des distributions telles que la loi de Cauchy. Dans le cas de MEGC, cette condition est équivalente à $\mathbb{E}[\xi] < +\infty$.

La plupart du temps, ces paramètres optimaux sont obtenus directement par des méthodes numériques comme l'algorithme MM ([Hunter and Lange, 2000]) ou le gradient stochastique ([Zheng, 2011]), sans que l'on ait d'idée sur la forme de ces paramètres. Dès lors, il est compliqué de donner la distribution du prédicteur de régression quantile.

Dans le cas des Mélanges d'Echelle Gaussiens Consistants, nous allons expliciter ces paramètres optimaux, et par conséquent être capables de donner la distribution du prédicteur $q_\alpha^{RQ}(Y|\mathbf{X})$.

III.2.b. Résultats dans le cadre elliptique consistant

Dans cette section, nous donnons la forme explicite des paramètres $\boldsymbol{\beta}^*$ et β_0^* présentés plus haut pour des MEGC. Pour cela, on a besoin, au vu de l'Equation (III.2.a.), de l'espérance de moments tronqués. On donne alors le lemme suivant, utile à la résolution du problème.

Lemme 8. Soit $(X_1, X_2) \sim \mathcal{EC}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \xi)$ un MEGC bivarié, avec des paramètres :

$$\boldsymbol{\mu} = (\mu_1, \mu_2)^\top, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

Alors

$$\mathbb{E} [X_1 \mathbb{1}_{\{X_2 > 0\}}] = \mu_1 \Phi_{X_1} \left(\frac{\mu_2}{\sigma_2} \right) + \rho\sigma_1 \int_{\mu_2/\sigma_2}^{\infty} c_1 z g_1(z^2) dz. \quad (\text{III.7})$$

Preuve : Soit f_{X_1, X_2} la densité du vecteur aléatoire (X_1, X_2) . Alors

$$\mathbb{E} [X_1 \mathbb{1}_{\{X_2 > 0\}}] = \int_{-\infty}^{\infty} \int_0^{\infty} x_1 f_{X_1, X_2}(x_1, x_2) dx_1 dx_2.$$

On a

$$\int_{-\infty}^{\infty} x f_{X_1, X_2}(x_1, x_2) dx_1 = f_{X_2}(x_2) \int_{-\infty}^{\infty} x_1 \frac{f_{X_1, X_2}(x_1, x_2)}{f_{X_2}(x_2)} dx_1 = f_{X_2}(x_2) \int_{-\infty}^{\infty} x_1 f_{X_1|X_2}(x_1|x_2) dx_1.$$

On déduit de l'Equation (5) que

$$\int_{-\infty}^{\infty} x_1 f_{X_1, X_2}(x_1, x_2) dx = f_{X_2}(x_2) \mathbb{E} [X_1 | X_2 = x_2] = f_{X_2}(x_2) \left\{ \mu_1 + \rho \frac{\sigma_1}{\sigma_2} (x_2 - \mu_2) \right\}.$$

On intégrant par rapport à x_2 , on obtient

$$\mathbb{E} [X_1 \mathbb{1}_{\{X_2 > 0\}}] = \int_0^{\infty} f_{X_2}(x_2) \left\{ \mu_1 + \rho \frac{\sigma_1}{\sigma_2} (x_2 - \mu_2) \right\} dx_2.$$

On rappelle que $f_{X_2}(x_2) = (c_1/\sigma_2) g_1\{(x_2 - \mu_2)^2/\sigma_2^2\}$. Ce qui conduit à

$$\mathbb{E} [X_1 \mathbb{1}_{\{X_2 > 0\}}] = \mu_1 \int_0^{\infty} f_{X_2}(x_2) dx_2 + \rho \frac{\sigma_1}{\sigma_2} c_1 \int_0^{\infty} \frac{x_2 - \mu_2}{\sigma_2} g_1 \left\{ \frac{(x_2 - \mu_2)^2}{\sigma_2^2} \right\} dx_2.$$

En opérant le changement de variable $z = (x_2 - \mu_2)/\sigma_2$, on obtient

$$\mathbb{E} [X_1 \mathbb{1}_{\{X_2 > 0\}}] = \mu_1 \Phi_{X_1} \left(\frac{\mu_2}{\sigma_2} \right) + \rho \sigma_1 \int_{-\mu_2/\sigma_2}^{\infty} c_1 z g_1(z^2) dz.$$

Par symétrie des lois elliptiques, on en déduit (8). $\square$

En utilisant le Lemme 8 et quelques calculs supplémentaires, il apparaît que nous sommes capables, dans le cas des MEGC, de fournir des formules explicites pour les paramètres optimaux β^* et β_0^* du problème (III.2.a.). Ce résultat est notamment dû à la propriété de stabilité par la somme de la Proposition 4. De plus, il est facile de donner la distribution du prédictor $q_\alpha^{RQ}(Y|\mathbf{X})$. C'est le propos du théorème suivant, que nous discutons par la suite.

Théorème 9 (Prédicteur de Régression Quantile). *Soit $\mathbf{Z} = (\mathbf{X}, Y)^\top \sim \mathcal{EC}_{d+1}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \xi)$ avec $\mathbb{E}[\xi] < +\infty$. Le vecteur de Régression Quantile $(\boldsymbol{\beta}^*, \beta_0^*)$ de $Y|(\mathbf{X} = \mathbf{x})$, qui satisfait l'Equation (III.2.a.), est donné par*

$$\boldsymbol{\beta}^* = \boldsymbol{\Sigma}_X^{-1} \boldsymbol{\Sigma}_{XY}, \quad \beta_0^* = \mu_Y - \boldsymbol{\Sigma}_{YX} \boldsymbol{\Sigma}_X^{-1} \boldsymbol{\mu}_X + \sqrt{\boldsymbol{\Sigma}_{Y|X}} \Phi_Y^{-1}(\alpha).$$

Le prédicteur de régression quantile de niveau $\alpha \in [0, 1]$ est donné par

$$q_\alpha^{RQ}(Y|\mathbf{X} = \mathbf{x}) = \mu_{Y|\mathbf{x}} + \sqrt{\boldsymbol{\Sigma}_{Y|X}} \Phi_Y^{-1}(\alpha).$$

De plus, la distribution de $q_\alpha^{RQ}(Y|\mathbf{X})$ est

$$q_\alpha^{RQ}(Y|\mathbf{X}) \sim \mathcal{EC}_1(\mu_Y + \sqrt{\boldsymbol{\Sigma}_{Y|X}} \Phi_Y^{-1}(\alpha), \boldsymbol{\Sigma}_{YX} \boldsymbol{\Sigma}_X^{-1} \boldsymbol{\Sigma}_{XY}, \xi).$$

Preuve : Dans un premier temps, introduisons quelques notations : soit ρ_j le coefficient de corrélation entre X_j et la variable aléatoire $Y - \boldsymbol{\beta}^{*\top} \mathbf{X} - \beta_0^*$. En effet, $X_j \sim \mathcal{EC}_1(\mu_j, \sigma_j^2, \xi)$. De plus, $Y - \boldsymbol{\beta}^{*\top} \mathbf{X} - \beta_0^* \sim \mathcal{EC}_1(\mu_Y - \boldsymbol{\beta}^{*\top} \boldsymbol{\mu}_X - \beta_0^*, (-\boldsymbol{\beta}^*, 1)^\top \boldsymbol{\Sigma} (-\boldsymbol{\beta}^{*\top}, 1), \xi)$. On a alors, pour tout $j = 1, \dots, d$, $(X_j, Y - \boldsymbol{\beta}^{*\top} \mathbf{X} - \beta_0^*) \sim \mathcal{EC}_2(\mathbf{m}_j, \boldsymbol{\Omega}_j, \xi)$ où :

$$\boldsymbol{\Omega}_j = \begin{pmatrix} \sigma_j^2 & \rho_j \sigma_j \sqrt{(-\boldsymbol{\beta}^*, 1)^\top \boldsymbol{\Sigma} (-\boldsymbol{\beta}^{*\top}, 1)} \\ \rho_j \sigma_j \sqrt{(-\boldsymbol{\beta}^*, 1)^\top \boldsymbol{\Sigma} (-\boldsymbol{\beta}^{*\top}, 1)} & (-\boldsymbol{\beta}^*, 1)^\top \boldsymbol{\Sigma} (-\boldsymbol{\beta}^{*\top}, 1) \end{pmatrix}.$$

Résolvons maintenant le problème de minimisation (III.2.a.). En utilisant la fonction $\mathcal{S}_\alpha$ donnée dans l'Equation (III.2.a.), et sa dérivée $\mathcal{S}'_\alpha(x) = (\alpha - 1) + \mathbb{1}_{\{x > 0\}}$ le problème est équivalent à

$$\begin{cases} (1 - \alpha) \mathbb{E}[\mathbf{X}] - \mathbb{E}[\mathbf{X} \mathbb{1}_{\{Y - \boldsymbol{\beta}^{*\top} \mathbf{X} - \beta_0^* > 0\}}] = 0, \\ (1 - \alpha) - \mathbb{E}[\mathbb{1}_{\{Y - \boldsymbol{\beta}^{*\top} \mathbf{X} - \beta_0^* > 0\}}] = 0. \end{cases} \quad (\text{III.8})$$

Etant donné que $Y - \boldsymbol{\beta}^{*\top} \mathbf{X} - \beta_0^* \sim \mathcal{EC}_1(\bar{\mu}, \bar{\sigma}^2, \xi)$ avec $\bar{\mu} = \mu_Y - \boldsymbol{\beta}^{*\top} \boldsymbol{\mu}_X - \beta_0^*$ et $\bar{\sigma}^2 = (-\boldsymbol{\beta}^*, 1)^\top \boldsymbol{\Sigma} (-\boldsymbol{\beta}^*, 1)$, alors $\mathbb{E}[\mathbb{1}_{\{Y - \boldsymbol{\beta}^{*\top} \mathbf{X} - \beta_0^* > 0\}}] = \mathbb{P}(Y - \boldsymbol{\beta}^{*\top} \mathbf{X} - \beta_0^* > 0)$, i.e., d'où

$$\mathbb{E}[\mathbb{1}_{\{Y - \boldsymbol{\beta}^{*\top} \mathbf{X} - \beta_0^* > 0\}}] = 1 - \Phi_Y(-\bar{\mu}/\bar{\sigma}) = \Phi_Y(\bar{\mu}/\bar{\sigma}).$$

Ensuite, en utilisant l'Equation (8) du Lemme 8 on a, pour tout $j \in \{1, \dots, d\}$,

$$\mathbb{E}[X_j \mathbb{1}_{\{Y - \boldsymbol{\beta}^{*\top} \mathbf{X} - \beta_0^* > 0\}}] = \mu_j \Phi_Y(\bar{\mu}/\bar{\sigma}) + \rho_j \sigma_j \int_{\bar{\mu}/\bar{\sigma}}^{\infty} c_1 z g_1(z^2) dz.$$

Finalement, le système (III.2.b.) peut se réécrire comme suit :

$$\begin{cases} \forall_{j \in \{1, \dots, d\}} & \mu_j \{(1 - \alpha) - \Phi_Y(\bar{\mu}/\bar{\sigma})\} - \rho_j \sigma_j \int_{\bar{\mu}/\bar{\sigma}}^{\infty} c_1 z g_1(z^2) dz = 0, \\ & (1 - \alpha) - \Phi_Y(\bar{\mu}/\bar{\sigma}) = 0. \end{cases}$$

Par conséquent,

$$\forall_{j \in \{1, \dots, d\}} \quad \rho_j \sigma_j \int_{\bar{\mu}/\bar{\sigma}}^{\infty} c_1 z g_1(z^2) dz = 0.$$

On a alors $\rho_1 = \dots = \rho_d = 0$. Il n'est alors pas difficile d'en déduire $\beta^* = \Sigma_X^{-1} \Sigma_{XY}$. En effet, $(\mathbf{X}, Y - \beta^{*\top} \mathbf{X} - \beta_0^*) \sim \mathcal{EC}_{d+1}(\mathbf{m}, \Omega, \xi)$ avec

$$\Omega = \begin{pmatrix} \Sigma_X & \mathbf{0}_{\mathbb{R}^d} \\ \mathbf{0}_{\mathbb{R}^d}^\top & (-\beta^*, 1)^\top \Sigma (-\beta^{*\top}, 1) \end{pmatrix}.$$

Son déterminant est donc $(-\beta^*, 1)^\top \Sigma (-\beta^{*\top}, 1) |\Sigma_X|$. En utilisant les moments conditionnels de $Y - \beta^{*\top} \mathbf{X} - \beta_0^* | \mathbf{X}$ donnés dans la Proposition 5 on obtient la relation suivante :

$$\Sigma_{Y|\mathbf{X}} = \Sigma_Y - \Sigma_{Y\mathbf{X}} \Sigma_X^{-1} \Sigma_{X\mathbf{Y}} = (-\beta^*, 1)^\top \Sigma (-\beta^{*\top}, 1).$$

En développant $(-\beta^*, 1)^\top \Sigma (-\beta^{*\top}, 1) = \Sigma_Y + \beta^{*\top} \Sigma_X \beta^* - 2\beta^{*\top} \Sigma_{XY}$, l'équation précédente devient

$$2\beta^{*\top} \Sigma_{XY} - \beta^{*\top} \Sigma_X \beta^* - \Sigma_{Y\mathbf{X}} \Sigma_X^{-1} \Sigma_{X\mathbf{Y}} = 0.$$

La fonction convexe $2\beta^{*\top} \Sigma_{XY} - \beta^{*\top} \Sigma_X \beta^* - \Sigma_{Y\mathbf{X}} \Sigma_X^{-1} \Sigma_{X\mathbf{Y}}$ atteint clairement son minimum en $\beta^* = \Sigma_X^{-1} \Sigma_{XY}$.

Concernant β_0^* , il satisfait

$$\beta_0^* = \arg \min_{\beta_0 \in \mathbb{R}} \mathbb{E} [\mathcal{S}_\alpha(Y - \beta^{*\top} \mathbf{X} - \beta_0)].$$

Finalement, β_0^* est le quantile de niveau α de la variable aléatoire $Y - \beta^{*\top} \mathbf{X}$. Il est évident que $Y - \beta^{*\top} \mathbf{X} \sim \mathcal{EC}_1(\mu_Y - \Sigma_{Y\mathbf{X}} \Sigma_X^{-1} \mu_X, \Sigma_{Y|\mathbf{X}} = \Sigma_Y - \Sigma_{Y\mathbf{X}} \Sigma_X^{-1} \Sigma_{X\mathbf{Y}}, \xi)$, $Y - \beta^{*\top} \mathbf{X}$ étant une transformation affine de $(\mathbf{X}, Y)$. De plus, avec l'expression de β^* , $\mathbb{E}[Y - \beta^{*\top} \mathbf{X}] = \mu_Y - \Sigma_{Y\mathbf{X}} \Sigma_X^{-1} \mu_X$. Concernant le second moment, il est égal à $(-\beta^*, 1)^\top \Sigma (-\beta^{*\top}, 1)$. Donc

$$(-\beta^*, 1)^\top \Sigma (-\beta^{*\top}, 1) = \sigma_Y^2 - 2 \sum_{i=1}^n \beta_i^* \rho_{ix} \sigma_i \sigma_Y + \sum_{j=1}^n \beta_j^* \sum_{i=1}^n \beta_i^* \rho_{ij} \sigma_i \sigma_j = \sigma_Y^2 - 2\beta^{*\top} \Sigma_{XY} + \beta^{*\top} \Sigma_X \beta^*.$$

En remplaçant $\beta^* = \Sigma_X^{-1} \Sigma_{XY}$, on montre la relation suivante :

$$(-\beta^*, 1)^\top \Sigma (-\beta^{*\top}, 1) = \sigma_Y^2 - 2\Sigma_{Y\mathbf{X}} \Sigma_X^{-1} \Sigma_{X\mathbf{Y}} + \Sigma_{Y\mathbf{X}} \Sigma_X^{-1} \Sigma_X \Sigma_X^{-1} \Sigma_{X\mathbf{Y}} = \sigma_Y^2 - \Sigma_{Y\mathbf{X}} \Sigma_X^{-1} \Sigma_{X\mathbf{Y}} = \Sigma_{Y|\mathbf{X}}.$$

En utilisant la forme générale d'un quantile elliptique, on obtient

$$\beta_0^* = \mu_Y - \Sigma_{YX} \Sigma_X^{-1} \mu_X + \sqrt{\Sigma_{Y|X}} \Phi_Y^{-1}(\alpha).$$

Nous disposons maintenant de nos paramètres optimaux β^* et β_0^* . Nous pouvons donc exprimer notre prédicteur de régression quantile de Y sachant $\mathbf{X} = \mathbf{x}$, en reprenant l'Equation (III.2.a.) :

$$q_\alpha^{RQ}(Y|\mathbf{X} = \mathbf{x}) = \beta^* \mathbf{x} + \beta_0^* = \mu_Y + \Sigma_{YX} \Sigma_X^{-1} (\mathbf{x} - \mu_X) + \sqrt{\Sigma_{Y|X}} \Phi_Y^{-1}(\alpha).$$

On reconnaît, sur le côté droit, l'expression de $\mu_{Y|\mathbf{X}}$ donnée dans l'Equation (5). Concernant sa distribution,

$$q_\alpha^{RQ}(Y|\mathbf{X}) = \mu_Y + \Sigma_{YX} \Sigma_X^{-1} (\mathbf{X} - \mu_X) + \sqrt{\Sigma_{Y|X}} \Phi_Y^{-1}(\alpha),$$

avec $\mathbf{X} \sim \mathcal{EC}_d(\mu_X, \Sigma_X, \xi)$. Comme combinaison affine de $\mathbf{X}$, en utilisant la Proposition 4 on a la relation en loi $q_\alpha^{RQ}(Y|\mathbf{X}) \sim \mathcal{EC}_1\{\mu_Y + \sqrt{\Sigma_{Y|X}} \Phi_Y^{-1}(\alpha), (\Sigma_{YX} \Sigma_X^{-1}) \Sigma_X (\Sigma_X^{-1} \Sigma_{XY}), \xi\}$. D'où le résultat. $\square$

III.2.c. Commentaires

Les résultats obtenus dans la section précédente donnent lieu à plusieurs remarques. Dans un premier temps, on pourrait s'interroger sur l'efficacité du modèle dans le cas linéaire, i.e. si on prend $\beta_0^* = 0$ et que le prédicteur $q_\alpha^{RQ}(Y|\mathbf{X})$ n'est finalement qu'une combinaison linéaire des covariables $\mathbf{X}$. Dans ce cas, on voit que le prédicteur $q_\alpha^{RQ}(Y|\mathbf{X}) = \beta^* \mathbf{X}$ ne dépend pas du niveau de quantile α . Ce modèle est donc évidemment à exclure, excepté dans le cas $\alpha = 1/2$, i.e. celui de la médiane. En effet, la formule de moyenne conditionnelle donnée dans l'Equation (5) montre que le prédicteur de régression quantile coïncide avec la moyenne conditionnelle dans le cas $\alpha = 1/2$, la moyenne et la médiane étant confondues dans le cas des distributions elliptiques. On peut également remarquer que si $\Sigma_{XY} = \mathbf{0}_{\mathbb{R}^d}$, ce qui signifie que $\mathbf{X}$ et Y sont décorrélés, le prédicteur $q_\alpha^{RQ}(Y|\mathbf{X})$ est nul, ce qui est un problème. Pour toutes ces raisons, le modèle linéaire est à exclure, et le terme affine β_0^* semble indispensable.

Concentrons nous maintenant sur le modèle affine. En comparant le prédicteur $q_\alpha^{RQ}(Y|\mathbf{X} = \mathbf{x})$ avec l'Equation (III.1.a.) du quantile conditionnel théorique, on s'aperçoit que les formules sont quasiment les mêmes, à un détail près. Le $\Phi_{Y|\mathbf{x}}^{-1}$ de $q_\alpha(Y|\mathbf{X} = \mathbf{x})$ est remplacé par Φ_Y^{-1} dans $q_\alpha^{RQ}(Y|\mathbf{X} = \mathbf{x})$, c'est à dire que le quantile conditionnel est remplacé par l'inconditionnel. On a vu que cette subtilité ne posait pas de problème dans le cas gaussien, car les distributions conditionnelle et inconditionnelle sont de même nature, mais est importante dans les autres cas de figure. En effet, on propose pour illustrer ce propos de prendre $\Sigma_{XY} = \mathbf{0}_{\mathbb{R}^d}$. Le prédicteur $q_\alpha^{RQ}(Y|\mathbf{X} = \mathbf{x})$ renvoie le quantile "indépendant" $\sqrt{\Sigma_Y} \Phi_Y^{-1}(\alpha)$, lorsque le quantile théorique vaut le quantile "décorrélé" $\sqrt{\Sigma_Y} \Phi_{Y|\mathbf{x}}^{-1}(\alpha)$. C'est

la différence entre indépendance et non-corrélation, inexistante dans le cas gaussien, qui pose alors problème dans le cas général. Pour donner un aperçu visuel, on propose de comparer les fonctions $\Phi_Y^{-1}(\alpha)$ et $\Phi_{Y|\mathbf{x}}^{-1}(\alpha)$ sur deux exemples. Prenons dans un premier temps la distribution de Laplace de paramètre $\lambda > 0$. Dans ce cas, on obtient en utilisant la densité de la Table [II.2](#):

$$\Phi_Y^{-1}(\alpha) = \sqrt{\frac{\lambda}{2}} \ln(2\alpha), \quad \alpha < \frac{1}{2} \text{ et } \Phi_Y^{-1}(\alpha) = \sqrt{\frac{\lambda}{2}} \ln\left(\frac{1}{2(1-\alpha)}\right), \quad \alpha \geq \frac{1}{2}.$$

La fonction $\Phi_{Y|\mathbf{x}}^{-1}(\alpha)$ est quant à elle obtenue numériquement. Le second exemple sera une loi de Student de paramètre $\nu > 0$. On sait grâce à l'Equation [\(7\)](#) que l'on peut fournir les deux fonctions de manière explicite.

$$\begin{cases} \Phi_Y^{-1}(\alpha) &= \Phi_{\nu}^{-1}(\alpha) \\ \Phi_{Y|\mathbf{x}}^{-1}(\alpha) &= \sqrt{\frac{\nu+M(\mathbf{x})}{\nu+d}} \Phi_{\nu+d}^{-1}(\alpha) \end{cases},$$

où Φ_m est la fonction de répartition d'une loi de Student à m degrés de liberté. La Figure [III.1](#) donne un aperçu de l'écart entre ces deux fonctions, et donc du biais que peut amener une telle approche. Dans cet exemple, on a $M(\mathbf{x}) = 1$, $d = 5$, $\nu = 3$ et $\lambda = 2$.

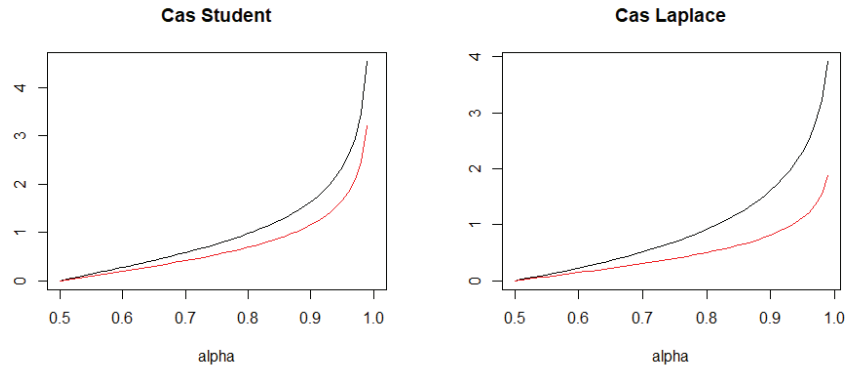


FIGURE III.1 – Comparaison des fonctions $\Phi_Y^{-1}(\alpha)$ (en noir) et $\Phi_{Y|\mathbf{x}}^{-1}(\alpha)$ (en rouge) pour $\alpha \in]\frac{1}{2}, 1[$ dans les cas Student et Laplace.

Le modèle que l'on a introduit possède certaines limites, que l'on vient d'énoncer. Nous avons vu qu'il était parfaitement adapté au cas $\alpha = 1/2$ (il est d'ailleurs utilisé en krigeage pour prédire la moyenne conditionnelle d'un champ aléatoire partiellement observé). Cependant, le prédicteur fait une erreur en dehors ce cas là, et cette erreur semble, de manière assez logique, croissante avec $|\alpha - 1/2|$. En effet, les niveaux de quantile

proches de 0 ou 1 faisant intervenir de très grandes valeurs, il paraît clair que le modèle de régression sera de moins en moins performant. En particulier, nous verrons que dans le cas des distributions à queue lourde, les lois conditionnelle $Y|(\mathbf{X} = \mathbf{x})$ et inconditionnelle Y n'ont pas le même indice de queue, ce qui mène à une approximation extrêmement mauvaise des quantiles conditionnels extrêmes. Pour cela, on propose dans la section suivante de corriger le prédicteur de régression quantile.

III.3. Quantiles extrêmes

L'estimation des quantiles extrêmes est très répandue dans la littérature. [Weissman, 1978] est l'un des premiers papiers considérant des niveaux de quantile tendant vers 1, puis [Dekkers and De Haan, 1989] s'est par la suite intéressé au cas où cette convergence est particulièrement rapide. Plus récemment, [Gardes and Girard, 2005] se sont intéressés aux lois à queue Weibull, tandis que [El Methni et al., 2012] ont proposé une approche pour des distributions à queue lourde et légère. La différence avec notre approche est que nous nous intéressons à des quantiles extrêmes conditionnels. Dans ce cadre, [Wang et al., 2012] ont obtenus des résultats dans un cadre de régression, que nous cherchons ici à contourner au vu de la section précédente.

Dans cette section, nous proposons de nouveaux prédicteurs de quantile adaptés aux niveaux extrêmes. Nous allons dans un premier temps proposer une relation asymptotique entre les quantiles $\Phi_{Y|\mathbf{x}}^{-1}$ et Φ_Y^{-1} . Par la suite, nous proposerons des estimateurs adaptés pour pouvoir estimer de tels quantiles à partir d'un jeu de données.

III.3.a. Approche méthodologique

Le but ici est de fournir une équivalence entre $\Phi_{Y|\mathbf{x}}^{-1}$ et Φ_Y^{-1} . Dans le cadre des distributions elliptiques, étant données les formules de densité données dans le Théorème 2 il paraît assez simple de fournir une équivalence asymptotique entre les fonctions de répartition $\Phi_{Y|\mathbf{x}}$ et Φ_Y . Cependant, le passage à l'inverse nécessite quelques outils analytiques que l'on peut notamment trouver dans [Djurčić and Torgašev, 2001]. La définition suivante en est d'ailleurs issue.

Définition 11 (φ -fonction). *Une fonction f est une φ -fonction si $f : [0, \infty) \rightarrow [0, \infty)$, $f(0) = 0$, f est continue, croissante sur $[0, \infty)$, et $f(t) \rightarrow \infty$ lorsque $t \rightarrow \infty$. De plus, on note K_c l'ensemble des φ -fonctions f qui vérifient*

$$\lim_{(t,\lambda) \rightarrow (\infty,1)} \frac{f(\lambda t)}{f(t)} = 1. \quad (\text{III.9})$$

On remarque que la condition de l'Equation (11) est plus générale que celle des fonctions à variation régulière au sens de la Définition 6. En effet, il est évident qu'une fonction à variation régulière est de classe K_c . En revanche, certaines fonctions comme la fonction de quantile gaussienne (nous le montrerons) sont K_c sans être à variation régulière. La Définition 11 est indispensable pour énoncer le lemme suivant, dû à [Djurčić and Torgašev, 2001], qui donne la relation entre l'équivalence de deux fonctions et l'équivalence de leurs inverses.

Lemme 10 (Equivalence des inverses). *Supposons que f et g soient deux φ -fonctions strictement croissantes, qu'au moins une des fonctions f^{-1} , g^{-1} appartienne à la classe K_c , et que $f(t) \underset{t \rightarrow \infty}{\sim} g(t)$. Alors $f^{-1}(t) \underset{t \rightarrow \infty}{\sim} g^{-1}(t)$.*

Rappelons que notre but est de donner une équivalence entre $\Phi_{Y|x}^{-1}$ et Φ_Y^{-1} . Pour cela, il faudra nécessairement proposer une équivalence entre $1 - \Phi_{Y|x}$ et $1 - \Phi_Y$. Cependant, ces deux fonctions de survie ne sont clairement pas des φ -fonctions, et ce pour plusieurs raisons :

$$1 - \Phi_{Y|x}(0) = 1 - \Phi_Y(0) = 1/2, \quad \lim_{t \rightarrow +\infty} 1 - \Phi_{Y|x}(t) = \lim_{t \rightarrow +\infty} 1 - \Phi_Y(t) = 1.$$

On doit alors opérer une transformation sur ces fonctions, pour pouvoir ensuite, en utilisant le Lemme 10, donner une équivalence de leurs inverses. Dans cette optique, on propose l'hypothèse suivante.

Hypothèse 1. *Il existe $0 < \ell(\mathbf{x}) < \infty$ et $\eta \in \mathbb{R}$ tels que*

$$\lim_{t \rightarrow \infty} \frac{\bar{\Phi}_{Y|x}(t)}{\bar{\Phi}_Y(t^\eta)} = \ell(\mathbf{x}), \quad (\text{III.10})$$

où $\bar{\Phi} = 1 - \Phi$ est la fonction de survie associée à Φ . De plus, on définit Ψ et $\Psi_{\mathbf{x}}$ comme

$$\Psi(t) = \frac{1}{\ell(\mathbf{x})\bar{\Phi}_Y(t^\eta)} - 2/\ell(\mathbf{x}), \quad \Psi_{\mathbf{x}}(t) = \frac{1}{\bar{\Phi}_{Y|x}(t)} - 2.$$

Alors Ψ^{-1} ou $\Psi_{\mathbf{x}}^{-1}$ appartient à la classe K_c .

On rappelle, en utilisant le Théorème 2 et la Proposition 5 que :

$$\Phi_Y(t) = \int_{-\infty}^t c_1 g_1(t^2) dt \text{ et } \Phi_{Y|x}(t) = \int_{-\infty}^t c_{\mathbf{x}} g_{d+1}(M(\mathbf{x}) + t^2) dt \text{ avec } c_{\mathbf{x}} = \frac{c_{d+1}}{c_d g_d(M(\mathbf{x}))}$$

où c_{d+1} et c_d sont les coefficients de normalisation de la distribution elliptique consistante de variable d'échelle ξ , en dimensions $d+1$ et d . Dès lors, la règle de l'Hôpital appliquée à l'Equation (1) permet d'exprimer les coefficients η et $\ell(\mathbf{x})$ en fonction des densités.

$$\lim_{t \rightarrow \infty} \frac{c_{\mathbf{x}} g_{d+1}(M(\mathbf{x}) + t^2)}{c_1 \eta t^{\eta-1} g_1(t^{2\eta})} = \ell(\mathbf{x}).$$

Basée sur l'Hypothèse [1](#) puis s'aidant du Lemme [10](#) la proposition suivante fournit une équivalence entre $\Phi_{Y|\mathbf{x}}^{-1}$ et Φ_Y^{-1} .

Proposition 11 (Equivalence quantiles conditionnel et inconditionnel). *Sous l'Hypothèse [1](#), on a*

$$\Phi_{Y|\mathbf{x}}^{-1}(\alpha) \underset{\alpha \rightarrow 1}{\sim} \left[\Phi_Y^{-1} \left(1 - \frac{1}{\frac{\ell(\mathbf{x})}{1-\alpha} + 2(1-\ell(\mathbf{x}))} \right) \right]^{1/\eta}. \quad (\text{III.11})$$

Preuve : Clairement, $\Psi_{\mathbf{x}}$ et Ψ sont deux φ -fonctions strictement croissantes, et $\Psi_{\mathbf{x}}(t) \underset{t \rightarrow \infty}{\sim} \Psi(t)$. De plus, si Ψ^{-1} ou $\Psi_{\mathbf{x}}^{-1}$ appartient à la classe K_c , alors le Lemme [10](#) implique

$$\Psi_{\mathbf{x}}^{-1}(t) \underset{t \rightarrow \infty}{\sim} \Psi^{-1}(t)$$

En d'autres termes,

$$\Phi_{Y|\mathbf{x}}^{-1} \left(1 - \frac{1}{t+2} \right) \underset{t \rightarrow \infty}{\sim} \left[\Phi_Y^{-1} \left(1 - \frac{1}{\ell(\mathbf{x})t+2} \right) \right]^{1/\eta}.$$

On peut également réécrire cela

$$\Phi_{Y|\mathbf{x}}^{-1}(\alpha) \underset{\alpha \rightarrow 1}{\sim} \left[\Phi_Y^{-1} \left(1 - \frac{1}{\frac{\ell(\mathbf{x})}{1-\alpha} + 2(1-\ell(\mathbf{x}))} \right) \right]^{1/\eta}$$

avec $\alpha = 1 - 1/(t+2)$, d'où le résultat. $\square$

Grâce à la Proposition [11](#) on est maintenant capable de fournir des prédicteurs de quantile conditionnel adaptés aux niveaux extrêmes. Nous en proposons alors deux, respectivement pour des niveaux proches de 1 et 0.

Définition 12 (Prédicteurs de quantile extrêmes). *On définit*

$$\begin{cases} q_{\alpha\uparrow}^E(Y|\mathbf{X} = \mathbf{x}) = \mu_{Y|\mathbf{x}} + \sqrt{\Sigma_{Y|\mathbf{X}}} \left[\Phi_Y^{-1} \left(1 - \frac{1}{\frac{\ell(\mathbf{x})}{1-\alpha} + 2(1-\ell(\mathbf{x}))} \right) \right]^{1/\eta}, \\ q_{\alpha\downarrow}^E(Y|\mathbf{X} = \mathbf{x}) = \mu_{Y|\mathbf{x}} - \sqrt{\Sigma_{Y|\mathbf{X}}} \left[\Phi_Y^{-1} \left(1 - \frac{1}{\frac{\ell(\mathbf{x})}{\alpha} + 2(1-\ell(\mathbf{x}))} \right) \right]^{1/\eta}. \end{cases} \quad (\text{III.12})$$

La proposition suivante montre que les prédicteurs introduits dans la Définition [12](#) sont asymptotiquement équivalents au quantile théorique $q_{\alpha}(Y|\mathbf{X} = \mathbf{x})$, respectivement pour $\alpha \rightarrow 1$ et $\alpha \rightarrow 0$.

Théorème 12 (Equivalence des prédicteurs). *Soit $\mathbf{Z} = (\mathbf{X}, Y)^{\top} \sim \mathcal{EC}_{d+1}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \xi)$. Sous l'Hypothèse [1](#),*

$$q_{\alpha\uparrow}^E(Y|\mathbf{X} = \mathbf{x}) \underset{\alpha \rightarrow 1}{\sim} q_{\alpha}(Y|\mathbf{X} = \mathbf{x}) \text{ et } q_{\alpha\downarrow}^E(Y|\mathbf{X} = \mathbf{x}) \underset{\alpha \rightarrow 0}{\sim} q_{\alpha}(Y|\mathbf{X} = \mathbf{x}).$$

Preuve : On rappelle que l'Equation (III.1.a.) donne $q_\alpha(Y|\mathbf{X} = \mathbf{x}) = \mu_{Y|\mathbf{x}} + \sqrt{\Sigma_{Y|\mathbf{X}}} \Phi_{Y|\mathbf{x}}^{-1}(\alpha)$. Des équations (11) et (12), on déduit immédiatement la première partie du résultat, à savoir

$$q_\alpha(Y|\mathbf{X} = \mathbf{x}) \underset{\alpha \rightarrow 1}{\sim} \mu_{Y|\mathbf{x}} + \sqrt{\Sigma_{Y|\mathbf{X}}} \left[\Phi_Y^{-1} \left(1 - \frac{1}{\frac{\ell(\mathbf{x})}{1-\alpha} + 2(1 - \ell(\mathbf{x}))} \right) \right]^{1/\eta} = q_{\alpha^\uparrow}^E(Y|\mathbf{X} = \mathbf{x}).$$

L'équivalence pour $\alpha \rightarrow 0$ est obtenue de la même manière, en utilisant les propriétés de symétrie des distributions elliptiques. $\square$

Maintenant que l'objectif de cette section est atteint, à savoir fournir des prédicteurs de quantile conditionnel adaptés aux niveaux extrêmes, il paraît intéressant de vérifier quelles distributions elliptiques satisfont l'Hypothèse 1. On donne pour cela la proposition suivante. En résumé, parmi les distributions que nous avons introduites précédemment, seule la distribution de Laplace ne satisfait pas cette hypothèse.

Proposition 13 (Exemples). *Les distributions Normale, Normale contaminée, Student, et Slash satisfont l'Hypothèse 1 avec des coefficients η et $\ell(\mathbf{x})$ donnés dans la Table III.1*

Loi	η	$\ell(\mathbf{x})$
Normale	1	1
Normale contaminée	1	$\frac{\max(\theta_1, \dots, \theta_m)^{-d} \exp\left\{-\frac{M(\mathbf{x})}{2 \max(\theta_1, \dots, \theta_m)^2}\right\}}{\sum_{k=1}^m \pi_k \theta_k^{-d} \exp\left(-\frac{M(\mathbf{x})}{2\theta_k^2}\right)}$
Student, $\nu > 0$	$1 + \frac{d}{\nu}$	$\frac{\Gamma(\frac{\nu+d+1}{2})\Gamma(\frac{\nu}{2})}{\Gamma(\frac{\nu+d}{2})\Gamma(\frac{\nu+1}{2})} \left(1 + \frac{M(\mathbf{x})}{\nu}\right)^{\frac{d+\nu}{2}} \frac{\nu^{\frac{d}{2}+1}}{\nu+d}$
Slash, $a > 0$	$1 + \frac{d}{a}$	$\frac{\Gamma(\frac{d+1+a}{2})M(\mathbf{x})^{\frac{d+a}{2}} 2^{1-\frac{a}{2}}}{\Gamma(\frac{d+a}{2})(d+a)\chi_{d+a}^2(M(\mathbf{x}))\Gamma(\frac{1+a}{2})}$
Laplace, $\lambda > 0$	$\times$	$\times$

TABLE III.1 – Coefficients η et $\ell(\mathbf{x})$ pour certaines distributions elliptiques classiques, avec $M(\mathbf{x}) = (\mathbf{x} - \boldsymbol{\mu}_X)^\top \boldsymbol{\Sigma}_X^{-1}(\mathbf{x} - \boldsymbol{\mu}_X)$ (voir Proposition 5)

Preuve : Loi normale : Premièrement, calculons $\ell(\mathbf{x})$ et η de l'Equation (1). Ici $\eta = 1$ mène à la limite

$$\lim_{t \rightarrow \infty} \frac{c_x g_{d+1}(M(\mathbf{x}) + t^2)}{c_1 g_1(t^2)}.$$

Les valeurs c_1 , $g_{d+1}(M(\mathbf{x}) + t^2)$, et $g_1(t^2)$ peuvent être déduites directement de la Table I.2. Soit $c_x = c_{d+1} / \{c_d g_d(M(\mathbf{x}))\}$,

$$\ell(\mathbf{x}) = \lim_{t \rightarrow \infty} \frac{\exp\left(-\frac{M(\mathbf{x}) + t^2}{2}\right)}{\exp(-M(\mathbf{x})/2) \exp(-t^2/2)} = 1.$$

Il reste à vérifier si la limite suivante existe (on remplace $\delta = \lambda + 1$ et $y = 1/t$) :

$$\epsilon_{\ell(x)} = \lim_{\substack{t \rightarrow \infty \\ \lambda \rightarrow 1}} \frac{\Psi^{-1}(\lambda t)}{\Psi^{-1}(t)} = \lim_{\substack{t \rightarrow \infty \\ \lambda \rightarrow 1}} \frac{\Phi^{-1}\left(1 - \frac{1}{\lambda t + 2}\right)}{\Phi^{-1}\left(1 - \frac{1}{t + 2}\right)} = \lim_{(\delta, y) \rightarrow (0, 0)} \frac{\Phi^{-1}\left(1 - \frac{y}{2y + \delta + 1}\right)}{\Phi^{-1}\left(1 - \frac{y}{2y + 1}\right)}.$$

Passons maintenant au coordonnées polaires, i.e. prenons $\delta = r \cos(\theta)$ et $y = r \sin(\theta)$, et considérons la limite lorsque $r \rightarrow 0$. Si cette limite ne dépend pas de θ , alors la limite quand $(\delta, y) \rightarrow (0, 0)$ existe. Considérons

$$\lim_{r \rightarrow 0} \frac{\Phi^{-1}\left(1 - \frac{r \sin(\theta)}{2r \sin(\theta) + r \cos(\theta) + 1}\right)}{\Phi^{-1}\left(1 - \frac{r \sin(\theta)}{2r \sin(\theta) + 1}\right)} = f(\theta).$$

Clairement, le numérateur et le dénominateur tendent vers $+\infty$, on utilise alors la règle de l'Hôpital pour obtenir

$$f(\theta) = \lim_{r \rightarrow 0} Q(r, \theta) \frac{\phi\left[\Phi^{-1}\left(1 - \frac{r \sin(\theta)}{2r \sin(\theta) + r \cos(\theta) + 1}\right)\right]}{\phi\left[\Phi^{-1}\left(1 - \frac{r \sin(\theta)}{2r \sin(\theta) + 1}\right)\right]},$$

où $Q(r, \theta)$ est un ratio qui tend clairement vers 1 lorsque $r \rightarrow 0$. Sur la droite, on rappelle que ϕ est la densité de la loi normale centrée réduite, i.e., $\phi(x) = (2\pi)^{-1}e^{-x^2/2}$. Alors

$$\epsilon_{\ell(x)} = \lim_{r \rightarrow 0} \frac{\phi\left[\Phi^{-1}\left(1 - \frac{r \sin(\theta)}{2r \sin(\theta) + r \cos(\theta) + 1}\right)\right]}{\phi\left[\Phi^{-1}\left(1 - \frac{r \sin(\theta)}{2r \sin(\theta) + r \cos(\theta) + 1}\right)\right]}.$$

Les deux termes de la fraction tendent vers 0. En utilisant une fois de plus la règle de l'Hôpital, alors

$$f(\theta) = \lim_{r \rightarrow 0} \frac{1}{Q(r, \theta)} \frac{\phi\left[\Phi^{-1}\left(1 - \frac{r \sin(\theta)}{2r \sin(\theta) + r \cos(\theta) + 1}\right)\right]}{\phi\left[\Phi^{-1}\left(1 - \frac{r \sin(\theta)}{2r \sin(\theta) + 1}\right)\right]} \frac{\phi'\left[\Phi^{-1}\left(1 - \frac{r \sin(\theta)}{2r \sin(\theta) + r \cos(\theta) + 1}\right)\right]}{\phi'\left[\Phi^{-1}\left(1 - \frac{r \sin(\theta)}{2r \sin(\theta) + r \cos(\theta) + 1}\right)\right]},$$

où $Q(r, \theta)$ est le ratio évoqué précédemment, qui tend vers 1. De plus, $\phi'(x) = -x\phi(x)$. Finalement,

$$f(\theta) = \lim_{r \rightarrow 0} \frac{\Phi^{-1}\left(1 - \frac{r \sin(\theta)}{2r \sin(\theta) + 1}\right)}{\Phi^{-1}\left(1 - \frac{r \sin(\theta)}{2r \sin(\theta) + r \cos(\theta) + 1}\right)} = \frac{1}{f(\theta)} = 1.$$

Loi Student : On doit calculer la limite

$$\lim_{t \rightarrow \infty} \frac{c_{\mathbf{x}} g_{d+1}(M(\mathbf{x}) + t^2)}{c_1 \eta t^{\eta-1} g_1(t^{2\eta})},$$

où c_d et g_d sont donnés dans la Table II.2 et $c_{\mathbf{x}}$ en est déduit. La limite précédente devient alors

$$\lim_{t \rightarrow \infty} \frac{\Gamma\left(\frac{d+1+\nu}{2}\right) \Gamma\left(\frac{\nu}{2}\right)}{\Gamma\left(\frac{1+\nu}{2}\right) \Gamma\left(\frac{d+\nu}{2}\right)} \left(1 + \frac{M(\mathbf{x})}{\nu}\right)^{\frac{d+\nu}{2}} \frac{\left(1 + \frac{M(\mathbf{x})+t^2}{\nu}\right)^{-\frac{d+1+\nu}{2}}}{\eta t^{\eta-1} \left(1 + \frac{t^{2\eta}}{\nu}\right)^{-\frac{1+\nu}{2}}}.$$

A partir de maintenant, on considère $\eta = (d + \nu)/\nu$ et on s'intéresse au ratio sur la droite. La limite est simplement $\nu^{d/2}$. Finalement,

$$\ell(\mathbf{x}) = \frac{\Gamma\left(\frac{d+1+\nu}{2}\right) \Gamma\left(\frac{\nu}{2}\right)}{\Gamma\left(\frac{1+\nu}{2}\right) \Gamma\left(\frac{d+\nu}{2}\right)} \left(1 + \frac{M(\mathbf{x})}{\nu}\right)^{\frac{d+\nu}{2}} \frac{\nu^{\frac{d}{2}+1}}{d + \nu}.$$

Concernant la seconde partie de l'Hypothèse I la preuve est similaire à celle du cas gaussien, mais nous avons $\eta \neq 1$ et $\ell(\mathbf{x}) \neq 1$ ici. Nous devons alors vérifier l'existence et la valeur de la limite

$$\lim_{\substack{t \rightarrow \infty \\ \lambda \rightarrow 1}} \frac{\Phi_{\nu}^{-1}\left(1 - \frac{1}{\ell(\mathbf{x})\lambda t + 2}\right)^{1/\eta}}{\Phi_{\nu}^{-1}\left(1 - \frac{1}{\ell(\mathbf{x})t + 2}\right)^{1/\eta}}.$$

Comme pour le cas gaussien, on utilise les coordonnées polaires pour obtenir le résultat. Les détails du calculs ne sont pas spécifiés ici.

Loi normale contaminée : On prend $\eta = 1$, et considère

$$\ell(\mathbf{x}) = \lim_{t \rightarrow \infty} \frac{c_{\mathbf{x}} g_{d+1}(M(\mathbf{x}) + t^2)}{c_1 g_1(t^2)},$$

où c_d et g_d sont donnés dans la Table II.2 et $c_{\mathbf{x}}$ en est déduit, d'où la limite

$$\ell(\mathbf{x}) = \lim_{t \rightarrow \infty} \frac{\sum_{k=1}^m \pi_k \theta_k^{-d-1} \exp(-M(\mathbf{x})/(2\theta_k^2)) \exp(-t^2/(2\theta_k^2))}{\left(\sum_{k=1}^m \pi_k \theta_k^{-d} \exp(-M(\mathbf{x})/(2\theta_k^2))\right) \left(\sum_{k=1}^m \pi_k \theta_k^{-1} \exp(-t^2/(2\theta_k^2))\right)}.$$

Soit k^* tel que $\theta_{k^*} = \max(\theta_1, \dots, \theta_m)$. Après simplifications, il reste

$$\ell(\mathbf{x}) = \frac{\theta_{k^*}^{-d} \exp(-M(\mathbf{x})/(2\theta_{k^*}^2))}{\sum_{k=1}^m \pi_k \theta_k^{-d} \exp(-M(\mathbf{x})/(2\theta_k^2))}.$$

Nous devons dès lors vérifier la deuxième condition de l'Hypothèse [1](#). Le même type de calculs que ceux du cas gaussien permettent de conclure.

Loi Slash : Soit $\eta = d/a + 1$, et calculons la limite $\ell(\mathbf{x})$ de l'Equation [\(1\)](#). Ici, c_d et g_d sont donnés dans la Table [II.2](#), et $c_{\mathbf{x}}$ en est déduit, d'où la limite

$$\ell(\mathbf{x}) = \lim_{t \rightarrow \infty} \frac{\Gamma\left(\frac{d+1+a}{2}\right) M(\mathbf{x})^{\frac{d+a}{2}} 2^{1-a/2}}{\Gamma\left(\frac{d+a}{2}\right) \Gamma\left(\frac{1+a}{2}\right) \chi_{d+a}^2(M(\mathbf{x}))} \frac{\chi_{d+1+a}^2(M(\mathbf{x}) + t^2)}{\chi_{1+a}^2(t^{2(d/a+1)})} \frac{t^{(1+a)(d/a+1)}}{(M(\mathbf{x}) + t^2)^{\frac{d+1+a}{2}}} \frac{1}{(d+a)t^{d/a}},$$

qui conduit à

$$\ell(\mathbf{x}) = \frac{\Gamma\left(\frac{d+1+a}{2}\right) M(\mathbf{x})^{\frac{d+a}{2}} 2^{1-\frac{a}{2}}}{\Gamma\left(\frac{d+a}{2}\right) (d+a) \chi_{d+a}^2(M(\mathbf{x})) \Gamma\left(\frac{1+a}{2}\right)}.$$

Pour vérifier la seconde partie de l'Hypothèse [1](#) comme pour les autres cas, on doit calculer la limite

$$\lim_{\substack{t \rightarrow \infty \\ \lambda \rightarrow 1}} \frac{\Phi_Y^{-1}\left(1 - \frac{1}{\ell(\mathbf{x})\lambda t + 2}\right)}{\Phi_Y^{-1}\left(1 - \frac{1}{\ell(\mathbf{x})t + 2}\right)} = \lim_{r \rightarrow 0} \frac{\phi_Y\left[\Phi_Y^{-1}\left(1 - \frac{r \sin(\theta)}{2r \sin(\theta) + \ell(\mathbf{x})}\right)\right]}{\phi_Y\left[\Phi_Y^{-1}\left(1 - \frac{r \sin(\theta)}{2r \sin(\theta) + \ell(\mathbf{x})r \cos(\theta) + \ell(\mathbf{x})}\right)\right]} = f(\theta).$$

On obtient une fois encore $f(\theta) = 1$. □

Dans cette section, nous avons introduit des prédictors basés uniquement sur la répartition non-conditionnelle Φ_Y . Ces prédictors sont censés améliorer le prédictor de régression pour des niveaux de quantile conditionnel extrêmes. Dans la section suivante, nous proposons des méthodes pour estimer ces prédictors à partir d'un jeu de données.

III.3.b. Estimation des paramètres extrémaux

La Définition [12](#) donne des prédictors du quantile théorique $q_\alpha(Y|\mathbf{X} = \mathbf{x})$ qui dépendent d'abord des moments conditionnels $\mu_{Y|\mathbf{x}}$, $\Sigma_{Y|\mathbf{X}}$, du quantile non-conditionnel Φ_Y^{-1} , puis de deux paramètres η et $\ell(\mathbf{x})$ que nous appellerons paramètres extrémaux. Il faut donc être capable d'estimer toutes ces quantités pour en déduire une estimation des prédictors $q_{\alpha\uparrow}^E(Y|\mathbf{X} = \mathbf{x})$ et $q_{\alpha\downarrow}^E(Y|\mathbf{X} = \mathbf{x})$, sans pour autant connaître la nature de la distribution. Pour cela, on va dans un premier temps s'autoriser une petite simplification. En effet, les quantités $\mu_{Y|\mathbf{x}}$ et $\Sigma_{Y|\mathbf{X}}$ peuvent être estimées très simplement, soit par la méthode des moments (si ces derniers existent), soit par les M-estimateurs (voir [\[Maronna, 1976\]](#) ou [\[Tyler, 1987\]](#)). De plus, la vitesse de convergence de ces méthodes (en $\sqrt{n}$) est plus rapide que celles des estimateurs que nous allons proposer par la suite. C'est pourquoi, dans la suite, nous considérons $\mu_{Y|\mathbf{x}}$ et $\Sigma_{Y|\mathbf{X}}$ connus, sans que cela ne change les résultats asymptotiques présentés. Il nous reste donc à estimer les quantités η , $\ell(\mathbf{x})$ et Φ_Y^{-1} , à partir d'observations $\mathbf{X}_1, \dots, \mathbf{X}_n$ indépendantes et identiquement distribuées de loi identique à $\mathbf{X} \sim \mathcal{EC}_d(\boldsymbol{\mu}_{\mathbf{X}}, \Sigma_{\mathbf{X}}, \xi)$. Notons que les quantités $\mu_{Y|\mathbf{x}}$ et

$\Sigma_{Y|\mathbf{X}}$ étant connues, on ne considère que des répétitions de la covariable. En pratique, leur estimation nécessite d'avoir des répétitions $\mathbf{Z}_1, \dots, \mathbf{Z}_n$ indépendantes et identiquement distribuées du vecteur complet $\mathbf{Z} = (\mathbf{X}, Y) \sim \mathcal{EC}_{d+1}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \xi)$. On notera également $M(\mathbf{x}) = (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{X}})^\top \boldsymbol{\Sigma}_{\mathbf{X}}^{-1}(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{X}})$.

Dans cette optique, nous allons dans un premier temps faire une hypothèse sur la fonction quantile Φ_Y^{-1} . Cette hypothèse, que nous pouvons qualifier de condition de variation régulière de second ordre au sens de la Définition 7 est classique dans la littérature des quantiles extrêmes (voir par exemple [Resnick and Stărică, 1997] ou plus récemment [Daouia et al., 2017a]).

Hypothèse 2 (Variation régulière de second ordre). *On suppose qu'il existe une fonction de signe constant A telle que $A(t) \rightarrow 0$ quand $t \rightarrow +\infty$, et*

$$\lim_{t \rightarrow +\infty} \frac{\frac{\Phi_Y^{-1}(1 - \frac{1}{\omega t})}{\Phi_Y^{-1}(1 - \frac{1}{t})} - \omega^\gamma}{A(t)} = \omega^\gamma \frac{\omega^\rho - 1}{\rho}, \forall \omega > 0$$

où $\gamma > 0$ et $\rho < 0$.

Une première conséquence de l'Hypothèse 2 est que Φ_Y est une distribution à variation régulière d'indice γ , au regard de la Définition 6. De plus, on a $\Phi_Y^{-1}(1 - 1/t) \sim k_1 t^\gamma$, ou de manière équivalente $\bar{\Phi}_Y(t) \sim k_2 t^{-1/\gamma}$ lorsque $t \rightarrow +\infty$ (voir [de Haan and Ferreira, 2006]). Par exemple, les distributions Student et Slash généralisée satisfont l'Hypothèse 2. Le lemme suivant donne quelques équivalences qui découlent immédiatement de cette hypothèse.

Lemme 14 (Quelques équivalences). *Sous l'Hypothèse 2, on a les relations asymptotiques suivantes :*

(i) *La variable d'échelle ξ satisfait*

$$\bar{F}_\xi(t) \underset{t \rightarrow +\infty}{\sim} \lambda t^{-\frac{1}{\gamma}}, \lambda \in \mathbb{R}.$$

(ii) *Pour tout $d \in \mathbb{N}^*$, le rayon $R_d = \sqrt{\chi_d^2} \xi$ est à variation régulière d'indice γ , et*

$$\bar{F}_{R_d}(t) \underset{t \rightarrow +\infty}{\sim} 2^{\frac{1}{2\gamma}} \frac{\Gamma\left(\frac{d+\gamma^{-1}}{2}\right)}{\Gamma\left(\frac{d}{2}\right)} \bar{F}_\xi(t) \underset{t \rightarrow +\infty}{\sim} 2^{\frac{1}{2\gamma}} \frac{\Gamma\left(\frac{d+\gamma^{-1}}{2}\right)}{\Gamma\left(\frac{d}{2}\right)} \lambda t^{-\frac{1}{\gamma}}, \lambda \in \mathbb{R}. \quad (\text{III.13})$$

(iii) *Pour tout $\eta > 0$, $d \in \mathbb{N}^*$,*

$$\frac{f_{R_d}(t)}{f_{R_1}(t^\eta)} \underset{t \rightarrow +\infty}{\sim} \frac{\sqrt{\pi} \Gamma\left(\frac{d+\gamma^{-1}}{2}\right)}{\Gamma\left(\frac{d}{2}\right) \Gamma\left(\frac{1+\gamma^{-1}}{2}\right)} t^{(\eta-1)(\gamma^{-1}+1)}. \quad (\text{III.14})$$

Preuve :

- (i) On a $R_1 \stackrel{d}{=} \chi_1 \xi$, où χ_1 a pour densité $\sqrt{\frac{2}{\pi}} e^{-\frac{t^2}{2}}$. Conformément au Lemme 4.3 de [Jessen and Mikosch, 2006], ξ satisfait $\bar{F}_\xi(t\omega)/\bar{F}_\xi(t) \rightarrow \omega^{-\frac{1}{\gamma}}$ quand $t \rightarrow +\infty$. De plus, le Lemme 4.2 dans [Jessen and Mikosch, 2006] implique

$$\mathbb{P}(\xi > t) \underset{t \rightarrow +\infty}{\sim} \mathbb{E} \left[\chi_1^{\frac{1}{\gamma}} \right]^{-1} \mathbb{P}(R_1 > t)$$

L'Hypothèse [2] assure $\mathbb{P}(R_1 > t) \sim \lambda t^{-\frac{1}{\gamma}}$, d'où le résultat.

- (ii) En utilisant le Lemme 4.2 de [Jessen and Mikosch, 2006] pour $R_d \stackrel{d}{=} \chi_d \xi$, il vient immédiatement

$$\mathbb{P}(R_d > t) \underset{t \rightarrow +\infty}{\sim} \mathbb{E} \left[\chi_d^{\frac{1}{\gamma}} \right] \mathbb{P}(\xi > t)$$

Les formules de moments d'une loi du χ^2 donnent $\mathbb{E} \left[\chi_d^{\frac{1}{\gamma}} \right] = 2^{\frac{1}{\gamma}} \frac{\Gamma\left(\frac{d+\gamma-1}{2}\right)}{\Gamma\left(\frac{d}{2}\right)}$.

- (iii) De (ii), on a, pour tout $d \in \mathbb{N}$, $f_{R_d}(t) \underset{t \rightarrow +\infty}{\sim} 2^{\frac{1}{\gamma}} \frac{\Gamma\left(\frac{d+\gamma-1}{2}\right)}{\Gamma\left(\frac{d}{2}\right)} \lambda' t^{-\frac{1}{\gamma}-1}$, où $\lambda' \in \mathbb{R}$ est indépendant de d . Le résultat est immédiat avec cette expression. $\square$

Dans ce paragraphe, on propose des méthodes d'estimation des paramètres η et $\ell(\mathbf{x})$ sous l'Hypothèse [2]. Nous avons vu dans l'approche méthodologique que l'existence de ces paramètres n'était pas assurée pour toutes les distributions. On va donc dans un premier temps montrer que l'Hypothèse [2] implique l'existence de η et $\ell(\mathbf{x})$, puis nous donnerons une forme explicite de ces paramètres, ce qui nous permettra par la suite de les estimer.

Proposition 15 (Existence des paramètres extrémaux). *Sous l'Hypothèse [2], les paramètres η et $\ell(\mathbf{x})$ existent, et sont exprimés comme suit :*

$$\begin{cases} \eta = 1 + \gamma d \\ \ell(\mathbf{x}) = \frac{\Gamma\left(\frac{d+\gamma-1}{2}\right)}{\Gamma\left(\frac{\gamma-1}{2}\right)} \frac{\gamma^{-1} \pi^{-\frac{d}{2}}}{(d+\gamma-1) c_d g_d(M(\mathbf{x}))}. \end{cases} \quad (\text{III.15})$$

Preuve : La formule de densité conditionnelle de la Proposition [5] mène à :

$$\lim_{t \rightarrow \infty} \frac{\bar{\Phi}_{Y|\mathbf{x}}(t)}{\bar{\Phi}_Y(t^\eta)} = \lim_{t \rightarrow \infty} \frac{c_{d+1} g_{d+1}(M(\mathbf{x}) + t^2)}{c_d g_d(M(\mathbf{x})) \eta t^{\eta-1} c_1 g_1(t^{2\eta})} = \lim_{t \rightarrow \infty} \frac{\Gamma\left(\frac{d+1}{2}\right) (M(\mathbf{x}) + t^2)^{-\frac{d}{2}} f_{R_{d+1}}\left(\sqrt{M(\mathbf{x}) + t^2}\right)}{\pi^{\frac{d+1}{2}} c_d g_d(M(\mathbf{x})) \eta t^{\eta-1} f_{R_1}(t^\eta)}.$$

En utilisant l'Equation (iii) du Lemme 14 il vient

$$\frac{\overline{\Phi}_{Y|\mathbf{x}}(t)}{\overline{\Phi}_Y(t^\eta)} \underset{t \rightarrow +\infty}{\sim} \frac{1}{\pi^{\frac{d}{2}} c_d g_d(M(\mathbf{x})) \eta} \frac{\Gamma\left(\frac{d+1+\gamma^{-1}}{2}\right)}{\Gamma\left(\frac{1+\gamma^{-1}}{2}\right)} t^{(\eta-1)(\gamma^{-1}+1)+1-\eta-d}.$$

Evidemment, on impose $0 < \ell(\mathbf{x}) < +\infty$, donc $1 - \eta - d + (\eta - 1)(\gamma^{-1} + 1) = 0$, d'où $\eta = d\gamma + 1$. En remplaçant η dans l'équation précédente, $\ell(\mathbf{x})$ facilement déduit

$$\ell(\mathbf{x}) = \frac{1}{\pi^{\frac{d}{2}} c_d g_d(M(\mathbf{x})) \eta} \frac{\Gamma\left(\frac{d+1+\gamma^{-1}}{2}\right)}{\Gamma\left(\frac{1+\gamma^{-1}}{2}\right)}.$$

□

On va maintenant chercher des estimateurs $\hat{\eta}$ et $\hat{\ell}(\mathbf{x})$. On voit que η dépend uniquement de l'indice de queue γ , pour lequel il existe de nombreux estimateurs dans la littérature. En revanche, $\ell(\mathbf{x})$ dépend à la fois de γ et de la fonction $c_d g_d$, que l'on doit également estimer.

Estimation de η Il existe en effet de nombreux estimateurs de l'indice γ . [Pickands, 1975], [Schultze and Steinebach, 1996] ou [Kratz and Resnick, 1996] en sont de bons exemples. Nous utiliserons pour notre part l'estimateur de Hill, introduit dans [Hill, 1975] :

$$\hat{\gamma}_{k_n} = \frac{1}{k_n} \sum_{i=1}^{k_n} \ln \left(\frac{W_{[i]}}{W_{[k_n+1]}} \right) \quad (\text{III.16})$$

où $W_{[1]} \geq \dots \geq W_{[k_n+1]} \geq \dots \geq W_{[n]}$ et $k_n = o(n)$ tel que $k_n \rightarrow +\infty$ quand $n \rightarrow +\infty$. Dans ce contexte, la statistique W peut être :

- La première (ou indifféremment n'importe laquelle) composante du vecteur centré réduit $\Lambda_{\mathbf{X}}^{-1}(\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}})$, où $\Lambda_{\mathbf{X}}^{\top} \Lambda_{\mathbf{X}} = \Sigma_{\mathbf{X}}$. Cette approche marche bien, mais n'utilise pas toutes les données disponibles.
- La norme de Mahalanobis $\sqrt{(\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}})^{\top} \Sigma_{\mathbf{X}}^{-1} (\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}})}$. Cette approche a l'avantage d'utiliser toutes les données disponibles.

Dans la suite, nous utiliserons plutôt la première approche à une composante, l'Hypothèse 2 étant appliquée sur la répartition univariée et donc plus simple à vérifier. De plus, quelques comparaisons numériques n'ont pas montré d'amélioration vraiment significative de la second approche par rapport à la première. En utilisant la Proposition 15 il est alors aisé de fournir un estimateur de η .

Définition 13 (Estimateur de η). *On définit $\hat{\eta}_{k_n}$ comme*

$$\hat{\eta}_{k_n} = \frac{d}{k_n} \sum_{i=1}^{k_n} \ln \left(\frac{W_{[i]}}{W_{[k_n+1]}} \right) + 1 \quad (\text{III.17})$$

Les propriétés principales de $\hat{\gamma}_{k_n}$ peuvent être trouvées dans [de Haan and Resnick, 1998]. Sous la condition de second ordre de l'Hypothèse 2 [de Haan and Ferreira, 2006] montre la normalité asymptotique de $\hat{\gamma}_{k_n}$.

$$\sqrt{k_n}(\hat{\gamma}_{k_n} - \gamma) \xrightarrow{n \rightarrow +\infty} \mathcal{N}\left(\frac{\lambda}{1 - \rho}, \gamma^2\right), \quad (\text{III.18})$$

où $\lambda = \lim_{n \rightarrow +\infty} \sqrt{k_n} A\left(\frac{n}{k_n}\right)$ et $k_n = o(n)$ tel que $k_n \rightarrow +\infty$ quand $n \rightarrow +\infty$. L'estimateur $\hat{\eta}_{k_n}$ étant une simple transformation affine de $\hat{\gamma}_{k_n}$, sa normalité asymptotique est alors immédiate.

Proposition 16 (Normalité asymptotique de $\hat{\eta}_{k_n}$). *Sous l'Hypothèse 2, et si $\lim_{n \rightarrow +\infty} \sqrt{k_n} A\left(\frac{n}{k_n}\right) = 0$, alors*

$$\sqrt{k_n}(\hat{\eta}_{k_n} - \eta) \xrightarrow{n \rightarrow +\infty} \mathcal{N}(0, d^2 \gamma^2)$$

Estimation de $\ell(\mathbf{x})$ La forme de $\ell(\mathbf{x})$, donnée dans la Proposition 15, donne lieu à une estimation plus compliquée que la précédente. En effet, $\ell(\mathbf{x})$ dépend non seulement de γ , mais aussi de la quantité $c_d g_d(M(\mathbf{x}))$. Nous garderons l'estimateur de Hill (III.3.b.) pour γ . Concernant $c_d g_d(M(\mathbf{x}))$, nous proposons un estimateur à noyau. La classe des estimateurs à noyau, introduite dans [Rosenblatt, 1956] puis [Parzen, 1962], permet d'estimer notamment des densités de probabilité. Dans cet esprit, le lemme suivant nous sera utile pour la construction d'un estimateur adapté. Il vient de la page 108 de [Johnson, 1987].

Lemme 17. *La distance de Mahalanobis $(\mathbf{X} - \boldsymbol{\mu}_\mathbf{X})^\top \boldsymbol{\Sigma}_\mathbf{X}^{-1} (\mathbf{X} - \boldsymbol{\mu}_\mathbf{X})$ a pour densité :*

$$\frac{\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}\right)} t^{\frac{d}{2}-1} c_d g_d(t).$$

En utilisant le Lemme 17 on définit un estimateur à noyau $\hat{g}_{h_n}(M(\mathbf{x}))$ pour $c_d g_d(M(\mathbf{x}))$.

Définition 14 (Estimateur du générateur). *On définit $\hat{g}_{h_n}(M(\mathbf{x}))$ comme*

$$\hat{g}_{h_n}(M(\mathbf{x})) = \frac{M(\mathbf{x})^{1-\frac{d}{2}} \Gamma\left(\frac{d}{2}\right)}{\pi^{\frac{d}{2}} n h_n} \sum_{i=1}^n K\left(\frac{M(\mathbf{x}) - (\mathbf{X}_i - \boldsymbol{\mu}_\mathbf{X})^\top \boldsymbol{\Sigma}_\mathbf{X}^{-1} (\mathbf{X}_i - \boldsymbol{\mu}_\mathbf{X})}{h_n}\right) \quad (\text{III.19})$$

où le noyau K vérifie certaines conditions données dans [Parzen, 1962] et la fenêtre h_n vérifie $h_n \rightarrow 0$ et $nh_n \rightarrow +\infty$ quand $n \rightarrow +\infty$.

Les conditions sur K sont fournies dans le Théorème 61 de l'Annexe B. Sous ces dernières, [Parzen, 1962] a montré la normalité asymptotique des estimateurs à noyau. Dès lors, la Définition 14 ainsi que le Lemme 17 entraînent la distribution asymptotique suivante pour $\hat{g}_{h_n}(M(\mathbf{x}))$.

Proposition 18 (Normalité asymptotique de $\hat{g}_{h_n}(M(\mathbf{x}))$). *Si g_d et K remplissent les conditions (i) – (iv) du Théorème 61 de l'Annexe B, l'estimateur $\hat{g}_{h_n}(M(\mathbf{x}))$ vérifie la normalité asymptotique suivante :*

$$\sqrt{nh_n}(\hat{g}_{h_n}(M(\mathbf{x})) - c_d g_d(M(\mathbf{x}))) \xrightarrow{n \rightarrow +\infty} \mathcal{N}\left(0, \frac{M(\mathbf{x})^{1-\frac{d}{2}} \Gamma\left(\frac{d}{2}\right)}{\pi^{\frac{d}{2}}} c_d g_d(M(\mathbf{x})) \int K(u)^2 du\right).$$

En remplaçant γ par $\hat{\gamma}_{k_n}$ et $c_d g_d(M(\mathbf{x}))$ par $\hat{g}_{h_n}(M(\mathbf{x}))$ dans l'Equation (15), on est maintenant capable de proposer un estimateur $\hat{\ell}_{k_n, h_n}(\mathbf{x})$ pour $\ell(\mathbf{x})$, dans la définition suivante. De plus, sous l'Hypothèse 2 on donne la normalité asymptotique de $\hat{\ell}_{k_n, h_n}(\mathbf{x})$.

Définition 15 (Estimateur de $\ell(\mathbf{x})$). *On définit $\hat{\ell}_{k_n, h_n}(\mathbf{x})$ comme :*

$$\hat{\ell}_{k_n, h_n}(\mathbf{x}) = \frac{\Gamma\left(\frac{d+\hat{\gamma}_{k_n}^{-1}+1}{2}\right)}{\Gamma\left(\frac{\hat{\gamma}_{k_n}^{-1}+1}{2}\right)} \frac{\hat{\gamma}_{k_n}^{-1} \pi^{-\frac{d}{2}}}{(d + \hat{\gamma}_{k_n}^{-1}) \hat{g}_{h_n}(M(\mathbf{x}))} \quad (\text{III.20})$$

où $\hat{\gamma}_{k_n}$ et $\hat{g}_{h_n}(M(\mathbf{x}))$ sont respectivement donnés dans les Equations (III.3.b.) et (14).

Proposition 19 (Normalité asymptotique de $\hat{\ell}_{k_n, h_n}(\mathbf{x})$). *Si g_d et K remplissent les conditions (i) – (iii) du Théorème 61 de l'Annexe B, sous l'Hypothèse 2 et si de plus*

$\lim_{n \rightarrow +\infty} \sqrt{k_n} A\left(\frac{n}{k_n}\right) = 0$, *on a les relations asymptotiques suivantes :*

(i) *Si $nh_n/k_n \xrightarrow{n \rightarrow +\infty} +\infty$, et $\sqrt{k_n} h_n^2 \xrightarrow{n \rightarrow +\infty} 0$, alors*

$$\sqrt{k_n}(\hat{\ell}_{k_n, h_n}(\mathbf{x}) - \ell(\mathbf{x})) \xrightarrow{n \rightarrow +\infty} \mathcal{N}(0, V_1(\gamma, c_d g_d(M(\mathbf{x})))) \quad (\text{III.21})$$

(ii) *Si $nh_n/k_n \xrightarrow{n \rightarrow +\infty} 0$, et $nh_n^5 \xrightarrow{n \rightarrow +\infty} 0$, alors*

$$\sqrt{nh_n}(\hat{\ell}_{k_n, h_n}(\mathbf{x}) - \ell(\mathbf{x})) \xrightarrow{n \rightarrow +\infty} \mathcal{N}(0, V_2(\gamma, c_d g_d(M(\mathbf{x})))) ,$$

où Ψ est la fonction digamma (voir p.258 de [Abramowitz et al., 1966] et Définition 36), et :

$$\begin{cases} V_1(\gamma, c_d g_d(M(\mathbf{x}))) = \frac{\pi^{-d} \gamma^2}{c_d^2 g_d(M(\mathbf{x}))^2} \frac{\Gamma\left(\frac{d+\gamma^{-1}+1}{2}\right)^2}{\Gamma\left(\frac{\gamma^{-1}+1}{2}\right)^2} \left[\frac{\Psi\left(\frac{\gamma^{-1}+1}{2}\right) - \Psi\left(\frac{d+\gamma^{-1}+1}{2}\right)}{2\gamma^2(d\gamma+1)} - \frac{d}{(d\gamma+1)^2} \right]^2 \\ V_2(\gamma, c_d g_d(M(\mathbf{x}))) = \frac{\Gamma\left(\frac{d}{2}\right)}{M(\mathbf{x})^{\frac{d}{2}-1} \pi^{\frac{d}{2}}} c_d g_d(M(\mathbf{x})) \int K(u)^2 du \left[\frac{\Gamma\left(\frac{d+\gamma^{-1}+1}{2}\right)}{\Gamma\left(\frac{\gamma^{-1}+1}{2}\right)} \frac{\gamma^{-1} \pi^{-\frac{d}{2}}}{(d+\gamma^{-1}) c_d^2 g_d(M(\mathbf{x}))^2} \right]^2 \end{cases} \quad (\text{III.22})$$

Preuve : Le Théorème 61 apporte $(\hat{g}_{h_n}(M(\mathbf{x})) - c_d g_d(M(\mathbf{x}))) = O_{\mathbb{P}}(h_n^2)$ et $\text{Var} [\hat{g}_{h_n}(M(\mathbf{x}))] = O_{\mathbb{P}}((nh_n)^{-1})$. Si $k_n = o(nh_n)$ et $\sqrt{k_n} h_n^2 \xrightarrow{n \rightarrow +\infty} 0$, il vient $\sqrt{k_n} (\hat{g}_{h_n}(M(\mathbf{x})) - c_d g_d(M(\mathbf{x}))) \xrightarrow[n \rightarrow +\infty]{L_2} 0$. On obtient alors la normalité asymptotique suivante :

$$\sqrt{k_n} \begin{pmatrix} \hat{g}_{h_n}(M(\mathbf{x})) - c_d g_d(M(\mathbf{x})) \\ \hat{\gamma}_{k_n} - \gamma \end{pmatrix} \xrightarrow{n \rightarrow +\infty} \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \gamma^2 & 0 \\ 0 & 0 \end{pmatrix} \right).$$

Comme $\ell(\mathbf{x}) = u(\gamma)$, la méthode delta implique

$$\sqrt{k_n} (\hat{\ell}_{k_n, h_n}(\mathbf{x}) - \ell(\mathbf{x})) \xrightarrow{n \rightarrow +\infty} \mathcal{N} (0, u'(\gamma)^2 \gamma^2).$$

Calculons alors u' , en utilisant l'Equation (15). On a

$$u'(\gamma) = \frac{\pi^{-d/2}}{c_d g_d(M(\mathbf{x}))} \left[\frac{\Gamma\left(\frac{d+\gamma^{-1}+1}{2}\right)}{\Gamma\left(\frac{\gamma^{-1}+1}{2}\right)} \right]' \left[\frac{1}{d\gamma+1} \right] + \frac{\pi^{-d/2}}{c_d g_d(M(\mathbf{x}))} \left[\frac{\Gamma\left(\frac{d+\gamma^{-1}+1}{2}\right)}{\Gamma\left(\frac{\gamma^{-1}+1}{2}\right)} \right] \left[\frac{1}{d\gamma+1} \right]'$$

La dérivée du ratio des fonctions gamma donne :

$$\left[\frac{\Gamma\left(\frac{d+\gamma^{-1}+1}{2}\right)}{\Gamma\left(\frac{\gamma^{-1}+1}{2}\right)} \right]' = \frac{1}{2\gamma^2} \frac{\Gamma'\left(\frac{\gamma^{-1}+1}{2}\right) \Gamma\left(\frac{d+\gamma^{-1}+1}{2}\right) - \Gamma'\left(\frac{d+\gamma^{-1}+1}{2}\right) \Gamma\left(\frac{\gamma^{-1}+1}{2}\right)}{\Gamma\left(\frac{\gamma^{-1}+1}{2}\right)^2}.$$

En utilisant la relation $\Gamma'(z) = \Gamma(z)\Psi(z)$, on réécrit alors :

$$\left[\frac{\Gamma\left(\frac{d+\gamma^{-1}+1}{2}\right)}{\Gamma\left(\frac{\gamma^{-1}+1}{2}\right)} \right]' = \frac{1}{2\gamma^2} \frac{\Gamma\left(\frac{d+\gamma^{-1}+1}{2}\right)}{\Gamma\left(\frac{\gamma^{-1}+1}{2}\right)} \left(\Psi\left(\frac{\gamma^{-1}+1}{2}\right) - \Psi\left(\frac{d+\gamma^{-1}+1}{2}\right) \right).$$

En observant également que $[(d\gamma+1)^{-1}]' = -d(d\gamma+1)^{-2}$, on obtient le résultat ((i)).

La seconde partie de la preuve est similaire. En effet, si $nh_n = o(k_n)$ et $nh_n^5 \rightarrow 0$ quand $n \rightarrow +\infty$, alors

$$\sqrt{nh_n} \begin{pmatrix} \hat{\gamma}_{k_n} - \gamma \\ \hat{g}_{h_n}(M(\mathbf{x})) - c_d g_d(M(\mathbf{x})) \end{pmatrix} \xrightarrow{n \rightarrow +\infty} \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & \frac{M(\mathbf{x})^{1-\frac{d}{2}} \Gamma(\frac{d}{2})}{\pi^{\frac{d}{2}}} c_d g_d(M(\mathbf{x})) \int K(u)^2 du \end{pmatrix} \right)$$

En notant que $\ell(\mathbf{x}) = v(c_d g_d(M(\mathbf{x})))$, et que

$$v'(c_d g_d(M(\mathbf{x}))) = -\frac{1}{c_d^2 g_d(M(\mathbf{x}))^2} \ell(\mathbf{x}),$$

la méthode delta complète la preuve. $\square$

Nous avons désormais une idée du taux de convergence de nos estimateurs $\hat{\eta}_{k_n}$ et $\hat{\ell}_{k_n, h_n}(\mathbf{x})$. On peut également s'intéresser à leur loi jointe.

Proposition 20 (Normalité asymptotique de $(\hat{\eta}_{k_n}, \hat{\ell}_{k_n, h_n}(\mathbf{x}))$). *Si g_d et K remplissent les conditions (i)–(iii) du Théorème 61 de l'Annexe B, sous l'Hypothèse 2, et si $\sqrt{k_n} A\left(\frac{n}{k_n}\right) \rightarrow 0$ quand $n \rightarrow +\infty$, alors on a les relations asymptotiques suivantes :*

(i) *Si $nh_n/k_n \xrightarrow{n \rightarrow +\infty} 0$ et $nh_n^5 \xrightarrow{n \rightarrow +\infty} 0$, alors*

$$\sqrt{nh_n} \begin{pmatrix} \hat{\ell}_{k_n, h_n}(\mathbf{x}) - \ell(\mathbf{x}) \\ \hat{\eta}_{k_n} - \eta \end{pmatrix} \xrightarrow{n \rightarrow +\infty} \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} V_2(\gamma, c_d g_d(M(\mathbf{x}))) & 0 \\ 0 & 0 \end{pmatrix} \right), \quad (\text{III.23})$$

où $V_2(\gamma, c_d g_d(M(\mathbf{x})))$ est donné dans l'Equation (19).

(ii) *Si $nh_n/k_n \xrightarrow{n \rightarrow +\infty} +\infty$ et $\sqrt{k_n} h_n^2 \xrightarrow{n \rightarrow +\infty} 0$, alors*

$$\sqrt{k_n} \begin{pmatrix} \hat{\ell}_{k_n, h_n}(\mathbf{x}) - \ell(\mathbf{x}) \\ \hat{\eta}_{k_n} - \eta \end{pmatrix} \xrightarrow{n \rightarrow +\infty} \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} V_1(\gamma, c_d g_d(M(\mathbf{x}))) & -d\gamma \sqrt{V_1(\gamma, c_d g_d(M(\mathbf{x})))} \\ -d\gamma \sqrt{V_1(\gamma, c_d g_d(M(\mathbf{x})))} & d^2 \gamma^2 \end{pmatrix} \right), \quad (\text{III.24})$$

où $V_1(\gamma, c_d g_d(M(\mathbf{x})))$ est donné dans l'Equation (19).

Preuve : La preuve de cette proposition est proche de celle de la Proposition 19.

- (i) Ce cas est immédiat. On utilise d'abord le résultat de l'Equation (i). De plus, l'hypothèse $nh_n = o(k_n)$ donne directement le résultat (i).
- (ii) En s'inspirant de la preuve de la Proposition 19 on rappelle la normalité asymptotique :

$$\sqrt{k_n} \begin{pmatrix} \hat{\gamma}_{k_n} - \gamma \\ \hat{g}_{h_n}(M(\mathbf{x})) - c_d g_d(M(\mathbf{x})) \end{pmatrix} \xrightarrow{n \rightarrow +\infty} \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \gamma^2 & 0 \\ 0 & 0 \end{pmatrix} \right).$$

On peut alors écrire le couple d'estimateurs comme fonction de $\hat{\gamma}_{k_n}$.

$$(\hat{\ell}_{k_n, h_n}(M(\mathbf{x})), \hat{\eta}_{k_n}) = f(\hat{\gamma}_{k_n}) = (f_1(\hat{\gamma}_{k_n}), f_2(\hat{\gamma}_{k_n}))$$

La méthode delta multivariée donne alors la normalité asymptotique, avec une matrice de covariance

$$\gamma^2 \begin{pmatrix} (f'_1(\gamma))^2 & f'_1(\gamma)f'_2(\gamma) \\ f'_1(\gamma)f'_2(\gamma) & (f'_2(\gamma))^2 \end{pmatrix}.$$

Avec les expressions de $\ell(M(\mathbf{x}))$ et η de l'Equation (15), on a $f'_2(\gamma) = d$. De plus, les calculs effectués dans la Proposition 19 donnent $f'_1(\gamma) = \sqrt{V_1(\gamma, c_d g_d(M(\mathbf{x})))}/\gamma$, d'où le résultat (ii).

□

En se basant sur les estimateurs introduits dans cette section, nous proposons dans la section suivante des estimateurs pour les quantiles conditionnels extrêmes $q_\alpha(Y|\mathbf{X} = \mathbf{x})$.

III.3.c. Estimation des quantiles extrêmes

On rappelle que les quantiles conditionnels théoriques $q_\alpha(Y|\mathbf{X} = \mathbf{x})$, donnés par l'Equation (III.1.a.), dépendent de la fonction quantile conditionnelle $\Phi_{Y|\mathbf{x}}^{-1}$. On a pour notre part fourni des prédicteurs $q_{\alpha\uparrow}^E(Y|\mathbf{X} = \mathbf{x})$ et $q_{\alpha\downarrow}^E(Y|\mathbf{X} = \mathbf{x})$ dans la Définition 12, dépendant de la fonction quantile inconditionnelle Φ_Y^{-1} , ainsi que deux paramètres η et $\ell(\mathbf{x})$ que l'on sait estimer grâce à la section précédente. L'objectif de cette section est donc de fournir des estimateurs $\hat{q}_{\alpha\uparrow}^E(Y|\mathbf{X} = \mathbf{x})$ et $\hat{q}_{\alpha\downarrow}^E(Y|\mathbf{X} = \mathbf{x})$ pour des suites $\alpha_n \rightarrow 1$ lorsque $n \rightarrow +\infty$. On pourrait également considérer les cas où $\alpha_n \rightarrow 0$. Cependant, on décide de se consacrer au cas où le niveau de quantile tend vers 1, le cas opposé pouvant être facilement déduit des propriétés de symétrie des distributions elliptiques. Pour cela, on divise l'étude en deux cas :

- Quantiles intermédiaires, i.e. on suppose $n(1 - \alpha_n) \rightarrow +\infty$ quand $n \rightarrow +\infty$. Cela signifie que l'estimation du quantile de niveau α_n conduit à une interpolation sur les données observées.
- Hauts quantiles. A l'instar de [de Haan and Rootzén, 1993], on suppose $n(1 - \alpha_n) \rightarrow 0$ lorsque $n \rightarrow +\infty$, i.e. on a besoin d'extrapoler notre échantillon.

Avant de traiter ces deux cas de figure distinctement, on introduit dans le paragraphe suivant certaines conditions qui nous serviront dans cette section.

III.3.c.1. Conditions

On présente ici quelques conditions sur les suites k_n , h_n et α_n . On rappelle que $k_n \rightarrow +\infty$, $k_n = o(n)$, $h_n \rightarrow 0$, $nh_n \rightarrow +\infty$ quand $n \rightarrow +\infty$, et définit les conditions C , C_{int} et C_{high} comme suit :

- (C) : $k_n = o(nh_n)$, $h_n \rightarrow 0$, $nh_n \rightarrow +\infty$, $\sqrt{k_n}h_n^2 \rightarrow 0$ et $\sqrt{k_n}A\left(\frac{n}{k_n}\right) \rightarrow 0$ quand $n \rightarrow +\infty$. On ajoute également que g_d et K satisfont les conditions (i)-(iii) du Théorème 61 de l'Annexe B

- $(C_{int}) : \alpha_n \rightarrow 1, \ln(1 - \alpha_n) = o(\sqrt{k_n})$ et $\frac{\sqrt{k_n}}{\ln(1 - \alpha_n)} = o\left(\sqrt{n(1 - \alpha_n)}\right)$ quand $n \rightarrow +\infty$.
- $(C_{high}) : n(1 - \alpha_n) \rightarrow 0, \ln(n(1 - \alpha_n)) = o(\sqrt{k_n})$ et $\frac{\ln(1 - \alpha_n)}{\ln\left(\frac{n}{k_n}(1 - \alpha_n)\right)} \rightarrow \theta \in [0, +\infty[$ quand $n \rightarrow +\infty$.

La condition (C) est commune aux deux approches, et est nécessaire pour que l'estimateur de Hill soit asymptotiquement non-biaisé, conformément à l'Equation (III.3.b.). De plus, $k_n = o(nh_n)$ et $\sqrt{k_n}h_n^2$ impliquent que $\hat{\gamma}_{k_n}$ (respectivement $\hat{\eta}_{k_n}$) converge vers γ (respectivement η) plus vite que $\hat{\ell}_{k_n, h_n}(\mathbf{x})$ vers $\ell(\mathbf{x})$. En pratique, cette condition semble appropriée, car k_n ne doit pas être trop grand pour que l'estimateur de Hill soit asymptotiquement non-biaisé, et nh_n doit être suffisamment grand pour fournir une bonne estimation de $\ell(\mathbf{x})$. Enfin, les conditions sur la densité et le noyau K sont nécessaires à la normalité asymptotique de l'estimateur (14).

Les conditions (C_{int}) et (C_{high}) vont être respectivement utilisées pour les quantiles intermédiaires et hauts, et dépendent uniquement des suites k_n et α_n . Dans le but de rendre ces conditions plus intuitives, on propose un exemple simple : on choisit nos suites sous des formes polynomiales $k_n = n^b$, $0 < b < 1$ et $\alpha_n = 1 - n^{-a}$, $a > 0$. Il est facile de voir que $\ln(1 - \alpha_n) = o(\sqrt{k_n})$ et $\ln(n(1 - \alpha_n)) = o(\sqrt{k_n})$, $\forall a > 0, 0 < b < 1$. Cependant, $\frac{\sqrt{k_n}}{\ln(1 - \alpha_n)} = o\left(\sqrt{n(1 - \alpha_n)}\right)$ si et seulement si $a < 1 - b$, i.e. $n(1 - \alpha_n) \rightarrow +\infty$ quand $n \rightarrow +\infty$. Autrement, si $a > 1$ (ou de manière équivalente $n(1 - \alpha_n) \rightarrow 0$ quand $n \rightarrow +\infty$), alors (C_{high}) est remplie avec un θ particulier que nous donnons plus tard. Nous ajoutons également des conditions plus fortes (C_{int}^{HG}) et (C_{high}^{HG}) utilisées plus tard pour des résultats de normalité asymptotique :

- $(C_{int}^{HG}) : (C_{int})$ est remplie. De plus, $\sqrt{k_n}(1 - \alpha_n) = o(\ln(1 - \alpha_n))$ et

$$\lim_{n \rightarrow +\infty} \frac{\sqrt{k_n}}{\ln(1 - \alpha_n)} (1 - \alpha_n)^{\frac{\min(-\rho, 2\gamma)}{\gamma d + 1}} = 0.$$

- $(C_{high}^{HG}) : (C_{high})$ est remplie. De plus,

$$\lim_{n \rightarrow +\infty} \frac{\sqrt{k_n}}{\ln\left(\frac{k_n}{n(1 - \alpha_n)}\right)} (1 - \alpha_n)^{\frac{\min(-\rho, 2\gamma)}{\gamma d + 1}} = 0.$$

Ces conditions, dépendant des suites k_n , α_n et des indices de valeurs extrêmes γ et ρ introduits dans l'Hypothèse 2 ne sont pas nécessaires pour les résultats de consistance. En revanche, elles le sont pour les résultats de normalité asymptotique.

III.3.c.2. Quantiles intermédiaires

On considère le cas où $\alpha_n \rightarrow 1$ et $n(1 - \alpha_n) \rightarrow +\infty$ quand $n \rightarrow +\infty$. En se basant sur le Théorème 2.4.1 de [de Haan and Ferreira, 2006], on définit l'estimateur $\hat{q}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})$ pour $q_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})$. Rappelons pour cette définition la notation $\mathbf{\Lambda}_{\mathbf{X}} \mathbf{\Lambda}_{\mathbf{X}}^\top = \mathbf{\Sigma}_{\mathbf{X}}$.

Définition 16 (Estimateur de quantiles intermédiaires). *On définit $(\hat{q}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x}))_{n \in \mathbb{N}}$ comme :*

$$\hat{q}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x}) = \mu_{Y|\mathbf{x}} + \sqrt{\Sigma_{Y|\mathbf{x}}} (W_{[n\tilde{v}_n+1]})^{\frac{1}{\hat{\eta}_{k_n}}} \quad (\text{III.25})$$

où $\tilde{v}_n = \left(2 + \hat{\ell}_{k_n, h_n}(\mathbf{x}) \left(\frac{1}{1-\alpha_n} - 2\right)\right)^{-1}$ et W est la première (ou indifféremment n'importe laquelle) composante du vecteur $\Lambda_{\mathbf{X}}^{-1}(\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}})$.

On estime donc dans un premier temps un quantile de niveau $1 - v_n$, où $v_n = \left(2 + \ell(\mathbf{x}) \left(\frac{1}{1-\alpha_n} - 2\right)\right)^{-1}$. Sous quelques conditions, il est possible d'obtenir la normalité asymptotique de $\hat{q}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})$ par rapport au prédicteur $q_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})$, et par conséquent la convergence vers le quantile théorique $q_{\alpha_n}(Y|\mathbf{X} = \mathbf{x})$.

Théorème 21 (Consistance de $\hat{q}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})$). *Sous l'Hypothèse [2] et les conditions (C), (C_{int}) :*

$$\frac{\sqrt{k_n}}{\ln(1 - \alpha_n)} \left(\frac{\hat{q}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})}{q_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})} - 1 \right) \xrightarrow{n \rightarrow +\infty} \mathcal{N}\left(0, \frac{d^2 \gamma^4}{(\gamma d + 1)^4}\right). \quad (\text{III.26})$$

Par conséquent :

$$\frac{\hat{q}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})}{q_{\alpha_n}(Y|\mathbf{X} = \mathbf{x})} \xrightarrow{\mathbb{P}} 1 \text{ lorsque } n \rightarrow +\infty. \quad (\text{III.27})$$

Preuve : Dans un premier temps, on remarque que $\tilde{v}_n$ dépend de $\hat{\ell}_{k_n, h_n}(\mathbf{x})$. Selon la Proposition [19] (i) implique que l'on peut remplacer $\tilde{v}_n$ par v_n dans l'Equation [21]. De plus, on décompose :

$$\begin{aligned} & \frac{\sqrt{k_n}}{\ln(1 - \alpha_n)} \left(\frac{\hat{q}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})}{q_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})} - 1 \right) \xrightarrow{n \rightarrow +\infty} \frac{\sqrt{k_n}}{\ln(1 - \alpha_n)} \left(\frac{(W_{[n\tilde{v}_n+1]})^{1/\hat{\eta}_{k_n}}}{\Phi_Y^{-1}(1 - v_n)^{1/\eta}} - 1 \right) = \\ & \frac{\sqrt{k_n}}{\ln(1 - \alpha_n)} \left(\frac{(W_{[nv_n+1]})^{\frac{1}{\hat{\eta}_{k_n}}}}{\Phi_Y^{-1}(1 - v_n)^{\frac{1}{\hat{\eta}_{k_n}}}} - 1 \right) \Phi_Y^{-1}(1 - v_n)^{\frac{1}{\hat{\eta}_{k_n}} - \frac{1}{\eta}} + \frac{\sqrt{k_n}}{\ln(1 - \alpha_n)} \left(\Phi_Y^{-1}(1 - v_n)^{\frac{1}{\hat{\eta}_{k_n}} - \frac{1}{\eta}} - 1 \right) \end{aligned}$$

Sous l'Hypothèse [2] et selon la Proposition [16] et le Théorème 2.4.1 dans [de Haan and Ferreira, 2006] (avec (C)), on a :

$$\left\{ \begin{array}{l} \sqrt{k_n} \left(\frac{1}{\hat{\eta}_{k_n}} - \frac{1}{\eta} \right) \xrightarrow{n \rightarrow +\infty} \mathcal{N}\left(0, \frac{d^2 \gamma^2}{(\gamma d + 1)^4}\right) \\ \sqrt{nv_n} \left(\frac{(W_{[nv_n+1]})}{\Phi_Y^{-1}(1 - v_n)} - 1 \right) \xrightarrow{n \rightarrow +\infty} \mathcal{N}(0, \gamma^2) \end{array} \right. ,$$

En notant que v_n est équivalent à $\ell(\mathbf{x})^{-1}(1 - \alpha_n)$ quand $n \rightarrow +\infty$, et grâce à (C_{int}) , il vient

$$\frac{\sqrt{k_n}}{\ln(1 - \alpha_n)} \left(\frac{(W_{[nv_n+1]})^{\frac{1}{\eta_{k_n}}}}{\Phi_Y^{-1}(1 - v_n)^{\frac{1}{\eta_{k_n}}}} - 1 \right) \rightarrow 0 \text{ lorsque } n \rightarrow +\infty.$$

De plus, sous l'Hypothèse [2](#) $\ln(\Phi_Y^{-1}(1 - v_n))$ est clairement équivalent à $-\gamma \ln(v_n)$, ou $-\gamma \ln(1 - \alpha_n)$. Alors (C_{int}) implique $\Phi_Y^{-1}(1 - v_n)^{\frac{1}{\eta_{k_n}} - \frac{1}{\eta}} \rightarrow 1$ quand $n \rightarrow +\infty$, et par conséquent le premier terme de la décomposition tend vers 0. Il reste encore à calculer la limite du second terme. Il n'est pas compliqué de remarquer que

$$\frac{\sqrt{k_n}}{\ln(\Phi_Y^{-1}(1 - v_n))} \left(\Phi_Y^{-1}(1 - v_n)^{\frac{1}{\eta_{k_n}} - \frac{1}{\eta}} - 1 \right) \xrightarrow{n \rightarrow +\infty} \mathcal{N}\left(0, \frac{d^2 \gamma^2}{(\gamma d + 1)^4}\right)$$

En utilisant l'équivalence $\ln(\Phi_Y^{-1}(1 - v_n)) \sim -\gamma \ln(v_n) \sim -\gamma \ln(1 - \alpha_n)$, on obtient le résultat [\(21\)](#). Avec la relation [\(11\)](#), la consistance [\(21\)](#) est évidente. $\square$

La même normalité asymptotique que [\(21\)](#) peut être obtenue en remplaçant $q_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})$ par le quantile théorique $q_{\alpha_n}(Y|\mathbf{X} = \mathbf{x})$ sous la condition

$$\lim_{n \rightarrow +\infty} \frac{\sqrt{k_n}}{\ln(1 - \alpha_n)} \ln \left(\frac{\Phi_Y^{-1}(1 - v_n)^{\frac{1}{\eta}}}{\Phi_{Y|\mathbf{x}}^{-1}(\alpha_n)} \right) = 0.$$

Cette condition est difficile à vérifier dans le cas général. Nous avons en effet besoin d'une expansion de second ordre de l'Equation [\(11\)](#). Mais les propriétés de second ordre du quantile inconditionnel Φ_Y^{-1} (données dans l'Hypothèse [2](#)) ne sont pas nécessairement les mêmes que celles du quantile conditionnel $\Phi_{Y|\mathbf{x}}^{-1}$, ce qui rend l'étude compliquée. Cependant, dans certains cas simples, on est capable de surmonter cette difficulté. C'est pourquoi on présente une nouvelle hypothèse, plus forte que l'Hypothèse [2](#). Dans la suite, on se référera à cette hypothèse pour certains résultats de normalité asymptotique.

Hypothèse 3. $\forall d \in \mathbb{N}^*$, il existe $\lambda_1, \lambda_2 \in \mathbb{R}$ tels que :

$$c_d g_d(t) = \lambda_1 t^{-\frac{d+\gamma-1}{2}} \left[1 + \lambda_2 t^{\frac{\rho}{2\gamma}} + o\left(t^{\frac{\rho}{2\gamma}}\right) \right]. \quad (\text{III.28})$$

Il est évident que l'Hypothèse [3](#) implique l'Hypothèse [2](#). En effet, conformément à [\[Hua and Joe, 2011\]](#), l'Equation [\(3\)](#) signifie que $c_1 g_1(t^2)$ est à variation régulière de second ordre d'indices $-1 - 1/\gamma$, ρ/γ et une fonction auxiliaire proportionnelle à $t^{\rho/\gamma}$. La Proposition 6 de [\[Hua and Joe, 2011\]](#) implique alors que $\bar{\Phi}_Y$ est à variation régulière de second ordre d'indices $-1/\gamma$, ρ/γ et le même type de fonction auxiliaire. Finalement, c'est exactement (voir [\[de Haan and Ferreira, 2006\]](#)) l'Hypothèse [2](#) avec γ , ρ et une fonction auxiliaire proportionnelle à t^ρ . Par exemple, la loi Student satisfait l'Hypothèse [3](#) avec des coefficients que l'on donne dans la Section [III.4](#).

Proposition 22 (Normalité asymptotique de $\hat{q}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})$). *Supposons que l'Hypothèse [3](#) et les conditions (C), (C_{int}^{HG}) soient satisfaites. Alors :*

$$\frac{\sqrt{k_n}}{\ln(1 - \alpha_n)} \left(\frac{\hat{q}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})}{q_{\alpha_n}(Y|\mathbf{X} = \mathbf{x})} - 1 \right) \xrightarrow{n \rightarrow +\infty} \mathcal{N} \left(0, \frac{d^2 \gamma^4}{(\gamma d + 1)^4} \right) \quad (\text{III.29})$$

Preuve : On rappelle que la densité de $\Phi_{Y|\mathbf{x}}$ est proportionnelle à $c_{d+1}g_{d+1}(M(\mathbf{x}) + t^2)$, et, par l'Hypothèse [3](#) il existe $\lambda_1, \lambda_2 \in \mathbb{R}$ tels que :

$$c_{d+1}g_{d+1}(M(\mathbf{x}) + t^2) = \lambda_1 (M(\mathbf{x}) + t^2)^{-\frac{d+1+\gamma^{-1}}{2}} \left[1 + \lambda_2 (M(\mathbf{x}) + t^2)^{\frac{\rho}{2\gamma}} + o\left(t^{\frac{\rho}{\gamma}}\right) \right]$$

L'expression précédente peut être réécrite comme suit, où $\lambda_1, \lambda_2, \lambda_3 \in \mathbb{R}$:

$$c_{d+1}g_{d+1}(M(\mathbf{x}) + t^2) = \lambda_1 t^{-(d+1+\gamma^{-1})} \left[1 + \lambda_2 (M(\mathbf{x}) + t^2)^{\frac{\rho}{2\gamma}} + \lambda_3 t^{-2} + o\left(t^{\frac{\rho}{\gamma}}\right) \right]$$

Dans le but d'aérer la preuve, on ne détaille pas les valeurs des constantes λ_i , elles ne sont en effet pas indispensables. Dans le cas $\rho/\gamma \leq -2$, on obtient

$$c_{d+1}g_{d+1}(M(\mathbf{x}) + t^2) = \lambda_1 t^{-(d+1+\gamma^{-1})} \left[1 + \lambda_2 t^{-2} + o(t^{-2}) \right], \lambda_1, \lambda_2 \in \mathbb{R}$$

En d'autres termes, $c_{d+1}g_{d+1}(M(\mathbf{x}) + t^2)$ est à variation régulière de second ordre d'indices $(d+1+\gamma^{-1})^{-1}$, -2 , et une fonction auxiliaire proportionnelle à t^{-2} . Selon la Proposition 6 de [\[Hua and Joe, 2011\]](#), la fonction de survie $\bar{\Phi}_{Y|\mathbf{x}}(t) = \int_t^{+\infty} c_{d+1}g_{d+1}(M(\mathbf{x}) + u^2) du \in 2RV\left((d+\gamma^{-1})^{-1}, -2\right)$ avec une fonction auxiliaire proportionnelle à t^{-2} . De manière équivalente, il existe $\lambda_1, \lambda_2 \in \mathbb{R}$ tels que

$$\Phi_{Y|\mathbf{x}}^{-1}\left(1 - \frac{1}{t}\right) = \lambda_1 t^{\frac{\gamma}{\gamma d+1}} \left[1 + \lambda_2 t^{-\frac{2\gamma}{\gamma d+1}} + o\left(t^{-\frac{2\gamma}{\gamma d+1}}\right) \right]$$

Les Hypothèses [2](#) et [3](#) donnant $\Phi_Y^{-1}(1 - 1/t) = \lambda_3 t^\rho [1 + \lambda_4 t^\rho + o(t^\rho)]$, il vient

$$\frac{\Phi_Y^{-1}(1 - v_n)^{\frac{1}{\eta}}}{\Phi_{Y|\mathbf{x}}^{-1}(\alpha_n)} = \ell(\mathbf{x})^{-\frac{\gamma}{\gamma d+1}} \left(\frac{1 - \alpha_n}{v_n} \right)^{\frac{\gamma}{\gamma d+1}} \frac{1 + \lambda_1 v_n^{-\rho} + o(v_n^{-\rho})}{1 + \lambda_2 (1 - \alpha_n)^{\frac{2\gamma}{\gamma d+1}} + o\left((1 - \alpha_n)^{\frac{2\gamma}{\gamma d+1}}\right)}$$

pour certaines constantes $\lambda_1, \lambda_2 \in \mathbb{R}$. Dans ce cas, on a considéré $\rho \leq -2\gamma$, d'où $-\rho > 2\gamma/(\gamma d + 1)$. On en déduit donc l'expansion suivante :

$$\frac{\Phi_Y^{-1}(1 - v_n)^{\frac{1}{\eta}}}{\Phi_{Y|\mathbf{x}}^{-1}(\alpha_n)} = \ell(\mathbf{x})^{-\frac{\gamma}{\gamma d+1}} \left(\frac{1 - \alpha_n}{v_n} \right)^{\frac{\gamma}{\gamma d+1}} \left[1 + \lambda(1 - \alpha_n)^{\frac{2\gamma}{\gamma d+1}} + o\left((1 - \alpha_n)^{\frac{2\gamma}{\gamma d+1}}\right) \right]$$

pour une certaine constante $\lambda \in \mathbb{R}$. On peut remarquer que $(1 - \alpha_n)/v_n = 2(1 - \ell(\mathbf{x}))(1 - \alpha_n) + \ell(\mathbf{x})$, et intéressons nous désormais à la limite :

$$\begin{aligned} \lim_{n \rightarrow +\infty} \frac{\sqrt{k_n}}{\ln(1 - \alpha_n)} \ln \left(\frac{\Phi_Y^{-1}(1 - v_n)^{\frac{1}{\eta}}}{\Phi_{Y|\mathbf{x}}^{-1}(\alpha_n)} \right) &= \frac{\gamma}{\gamma d + 1} \lim_{n \rightarrow +\infty} \frac{\sqrt{k_n}}{\ln(1 - \alpha_n)} \ln \left(2 \frac{1 - \ell(\mathbf{x})}{\ell(\mathbf{x})} (1 - \alpha_n) + 1 \right) \\ &+ \lim_{n \rightarrow +\infty} \frac{\sqrt{k_n}}{\ln(1 - \alpha_n)} \ln \left(1 + \lambda(1 - \alpha_n)^{\frac{2\gamma}{\gamma d + 1}} + o \left((1 - \alpha_n)^{\frac{2\gamma}{\gamma d + 1}} \right) \right) \end{aligned}$$

Le premier terme est simple à calculer. En effet, étant donné que $\sqrt{k_n}(1 - \alpha_n)/\ln(1 - \alpha_n) \rightarrow 0$ quand $n \rightarrow +\infty$, on déduit

$$\lim_{n \rightarrow +\infty} \frac{\sqrt{k_n}}{\ln(1 - \alpha_n)} \ln \left(2 \frac{1 - \ell(\mathbf{x})}{\ell(\mathbf{x})} (1 - \alpha_n) + 1 \right) = 2 \frac{1 - \ell(\mathbf{x})}{\ell(\mathbf{x})} \lim_{n \rightarrow +\infty} \frac{\sqrt{k_n}}{\ln(1 - \alpha_n)} (1 - \alpha_n) = 0$$

Par un calcul similaire, le second terme tend également vers 0, en supposant (C_{int}^{HG}) . On en déduit alors, en utilisant le Théorème [21](#) :

$$\begin{aligned} \frac{\sqrt{k_n}}{\ln(1 - \alpha_n)} \left(\frac{\hat{q}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})}{q_{\alpha_n}(Y|\mathbf{X} = \mathbf{x})} - 1 \right) &\underset{n \rightarrow +\infty}{\sim} \frac{\sqrt{k_n}}{\ln(1 - \alpha_n)} \left(\frac{(W_{[n\tilde{v}_n+1]})^{1/\hat{\eta}_{k_n}}}{\Phi_{Y|\mathbf{x}}^{-1}(\alpha_n)^{1/\eta}} - 1 \right) = \\ \frac{\sqrt{k_n}}{\ln(1 - \alpha_n)} \left(\frac{(W_{[n\tilde{v}_n+1]})^{1/\hat{\eta}_{k_n}}}{\Phi_Y^{-1}(1 - v_n)^{\frac{1}{\eta}}} - 1 \right) &\frac{\Phi_Y^{-1}(1 - v_n)^{\frac{1}{\eta}}}{\Phi_{Y|\mathbf{x}}^{-1}(\alpha_n)} + \frac{\sqrt{k_n}}{\ln(1 - \alpha_n)} \left(\frac{\Phi_Y^{-1}(1 - v_n)^{\frac{1}{\eta}}}{\Phi_{Y|\mathbf{x}}^{-1}(\alpha_n)} - 1 \right) \\ &\xrightarrow{n \rightarrow +\infty} \mathcal{N} \left(0, \frac{d^2 \gamma^4}{(\gamma d + 1)^4} \right) \end{aligned}$$

Concentrons nous maintenant sur le cas $\rho/\gamma > -2$. La preuve est exactement la même, avec

$$c_{d+1}g_{d+1} (M(\mathbf{x}) + t^2) = \lambda_1 t^{-(d+1+\gamma^{-1})} \left[1 + \lambda_2 t^{\frac{\rho}{\gamma}} + o \left(t^{\frac{\rho}{\gamma}} \right) \right], \lambda_1, \lambda_2 \in \mathbb{R}$$

Les mêmes calculs et la condition de $(C_{int}^{HG}) \lim_{n \rightarrow +\infty} \frac{\sqrt{k_n}}{\ln(1 - \alpha_n)} (1 - \alpha_n)^{-\frac{\rho}{\gamma d + 1}} = 0$ conduisent au résultat. $\square$

On peut noter que la variance asymptotique de l'Equation [21](#) tend vers 0 lorsque le nombre d de covariables tend vers $+\infty$. En effet, on observe une convergence rapide de $\hat{q}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})$ vers $\Phi_Y^{-1}(1 - v_n)^{\frac{1}{\eta}}$ lorsque d est grand. Cependant, la condition (C_{high}^{HG}) n'est clairement pas satisfaite si d est trop grand. La normalité asymptotique [22](#) n'est alors plus vérifiée. Cela s'explique par le fait que plus d est grand, plus $\Phi_{Y|\mathbf{x}}^{-1}(\alpha_n)/\Phi_Y^{-1}(1 - v_n)$ (voir Equation [11](#)) tend vers 1 lentement.

III.3.c.3. Hauts quantiles

On considère maintenant $n(1 - \alpha_n) \rightarrow 0$ quand $n \rightarrow +\infty$. Dans la définition qui suit, on introduit un nouvel estimateur $\hat{q}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})$ pour $q_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})$, et donc pour

$q_{\alpha_n}(Y|\mathbf{X} = \mathbf{x})$. L'approche, inspirée par [de Haan and Ferreira, 2006] ou [Weissman, 1978], consiste à prendre une suite intermédiaire k_n , et d'utiliser les propriétés de variation régulière.

Définition 17 (Estimateur de haut quantile). *On définit $(\hat{q}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x}))_{n \in \mathbb{N}}$ comme :*

$$\hat{q}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x}) = \mu_{Y|\mathbf{x}} + \sqrt{\Sigma_{Y|\mathbf{X}}} \left[W_{[k_n+1]} \left(\frac{k_n}{n} \left(2 + \hat{\ell}_{k_n, h_n}(\mathbf{x}) \left(\frac{1}{1 - \alpha_n} - 2 \right) \right) \right)^{\hat{\gamma}_{k_n}} \right]^{\frac{1}{\hat{\eta}_{k_n}}}. \quad (\text{III.30})$$

Le but est maintenant d'étudier les propriétés asymptotiques de $\hat{q}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})$. Comme pour les quantiles intermédiaires, on propose un résultat de normalité asymptotique, que l'on affinera par la suite sous l'Hypothèse [3](#)

Théorème 23 (Consistance de $\hat{q}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})$). *On note $v_n = (2 + \ell(\mathbf{x})((1 - \alpha_n)^{-1} - 2))^{-1}$ et $\tilde{v}_n = (2 + \hat{\ell}_{k_n, h_n}(\mathbf{x})((1 - \alpha_n)^{-1} - 2))^{-1}$. Sous l'Hypothèse [2](#), et les conditions (C), (C_{high}) :*

$$\frac{\sqrt{k_n}}{\ln\left(\frac{k_n}{n(1-\alpha_n)}\right)} \left(\frac{\hat{q}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})}{q_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})} - 1 \right) \xrightarrow{n \rightarrow +\infty} \mathcal{N}\left(0, \left(\frac{\gamma}{\gamma d + 1} - \theta \frac{d\gamma^2}{(\gamma d + 1)^2} \right)^2\right). \quad (\text{III.31})$$

Par conséquent :

$$\frac{\hat{q}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})}{q_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})} \xrightarrow{\mathbb{P}} 1 \text{ quand } n \rightarrow +\infty.$$

Preuve : Premièrement, on peut remarquer que

$$\frac{\sqrt{k_n}}{\ln\left(\frac{k_n}{n(1-\alpha_n)}\right)} \left(\frac{\hat{q}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})}{q_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})} - 1 \right) \underset{n \rightarrow +\infty}{\sim} \frac{\sqrt{k_n}}{\ln\left(\frac{k_n}{n(1-\alpha_n)}\right)} \left(\frac{\left[W_{[k_n+1]} \left(\frac{k_n}{n\tilde{v}_n} \right)^{\hat{\gamma}_{k_n}} \right]^{\frac{1}{\hat{\eta}_{k_n}}}}{\Phi_Y^{-1}(1 - v_n)^{\frac{1}{\eta}}} - 1 \right)$$

De plus,

$$\frac{\left[W_{[k_n+1]} \left(\frac{k_n}{n\tilde{v}_n} \right)^{\hat{\gamma}_{k_n}} \right]^{\frac{1}{\hat{\eta}_{k_n}}}}{\Phi_Y^{-1}(1 - v_n)^{\frac{1}{\eta}}} - 1 = \left(\frac{\left[W_{[k_n+1]} \left(\frac{k_n}{n\tilde{v}_n} \right)^{\hat{\gamma}_{k_n}} \right]^{\frac{1}{\hat{\eta}_{k_n}}}}{\Phi_Y^{-1}(1 - v_n)^{\frac{1}{\eta}}} - 1 \right) \left(\left(\frac{v_n}{\tilde{v}_n} \right)^{\frac{\hat{\gamma}_{k_n}}{\hat{\eta}_{k_n}}} + \left(\frac{v_n}{\tilde{v}_n} \right)^{\frac{\hat{\gamma}_{k_n}}{\hat{\eta}_{k_n}}} - 1 \right).$$

Comme $k_n = o(nh_n)$, on déduit

$$\frac{\sqrt{k_n}}{\ln\left(\frac{k_n}{nv_n}\right)} \left(\frac{\left[W_{[k_n+1]} \left(\frac{k_n}{nv_n} \right)^{\hat{\gamma}_{k_n}} \right]^{\frac{1}{\hat{\eta}_{k_n}}}}{\Phi_Y^{-1}(1-v_n)^{\frac{1}{\eta}}} - 1 \right) \underset{n \rightarrow +\infty}{\sim} \frac{\sqrt{k_n}}{\ln\left(\frac{k_n}{nv_n}\right)} \left(\frac{\left[W_{[k_n+1]} \left(\frac{k_n}{nv_n} \right)^{\hat{\gamma}_{k_n}} \right]^{\frac{1}{\hat{\eta}_{k_n}}}}{\Phi_Y^{-1}(1-v_n)^{\frac{1}{\eta}}} - 1 \right).$$

De plus, conformément au Théorème 4.3.8 de [de Haan and Ferreira, 2006], (C) et (C_{high}) mènent à

$$\frac{\sqrt{k_n}}{\ln\left(\frac{k_n}{nv_n}\right)} \left(\frac{W_{[k_n+1]} \left(\frac{k_n}{nv_n} \right)^{\hat{\gamma}_{k_n}}}{\Phi_Y^{-1}(1-v_n)} - 1 \right) \underset{n \rightarrow +\infty}{\sim} \frac{\sqrt{k_n}}{\ln\left(\frac{k_n}{nv_n}\right)} \left(\left(\frac{k_n}{nv_n} \right)^{\hat{\gamma}_{k_n} - \gamma} - 1 \right).$$

Avec l'Hypothèse 2 il est aisé de montrer que $\ln(\Phi_Y^{-1}(1-v_n)) / \ln(k_n/(nv_n))$ est asymptotiquement équivalent à $\gamma \ln(1-\alpha_n) / \ln(n(1-\alpha_n)/k_n)$. Si on considère le second terme, il vient, en utilisant la limite donnée dans (C_{high}) :

$$\frac{\sqrt{k_n}}{\ln\left(\frac{k_n}{nv_n}\right)} \left(\frac{\left(\frac{k_n}{nv_n} \right)^{\hat{\gamma}_{k_n} - \gamma} - 1}{\Phi_Y^{-1}(1-v_n)^{\frac{1}{\hat{\eta}_{k_n}} - \frac{1}{\eta}} - 1} \right) \underset{n \rightarrow +\infty}{\rightarrow} \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \gamma^2 & -\theta \frac{d\gamma^3}{(\gamma d+1)^2} \\ -\theta \frac{d\gamma^3}{(\gamma d+1)^2} & \theta^2 \frac{d^2\gamma^4}{(\gamma d+1)^4} \end{pmatrix} \right).$$

Finalement,

$$\begin{aligned} \frac{\sqrt{k_n}}{\ln\left(\frac{k_n}{nv_n}\right)} \left(\frac{\left[W_{[k_n+1]} \left(\frac{k_n}{nv_n} \right)^{\hat{\gamma}_{k_n}} \right]^{\frac{1}{\hat{\eta}_{k_n}}}}{\Phi_Y^{-1}(1-v_n)^{\frac{1}{\eta}}} - 1 \right) &= \frac{\sqrt{k_n}}{\ln\left(\frac{k_n}{nv_n}\right)} \left(\Phi_Y^{-1}(1-v_n)^{\frac{1}{\hat{\eta}_{k_n}} - \frac{1}{\eta}} - 1 \right) \\ &+ \frac{\sqrt{k_n}}{\ln\left(\frac{k_n}{nv_n}\right)} \left(\frac{\left[W_{[k_n+1]} \left(\frac{k_n}{nv_n} \right)^{\hat{\gamma}_{k_n}} \right]^{\frac{1}{\hat{\eta}_{k_n}}}}{\Phi_Y^{-1}(1-v_n)^{\frac{1}{\hat{\eta}_{k_n}}}} - 1 \right) \Phi_Y^{-1}(1-v_n)^{\frac{1}{\hat{\eta}_{k_n}} - \frac{1}{\eta}} \end{aligned}$$

Quand $n \rightarrow \infty$, cette expression est la somme de la loi normale bivariée suivante :

$$\frac{\sqrt{k_n}}{\ln\left(\frac{k_n}{nv_n}\right)} \left(\frac{\left[W_{[k_n+1]} \left(\frac{k_n}{nv_n} \right)^{\hat{\gamma}_{k_n}} \right]^{\frac{1}{\hat{\eta}_{k_n}}}}{\Phi_Y^{-1}(1-v_n)^{\frac{1}{\hat{\eta}_{k_n}}}} - 1 \right) \underset{n \rightarrow +\infty}{\rightarrow} \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{\gamma^2}{(\gamma d+1)^2} & -\theta \frac{d\gamma^3}{(\gamma d+1)^3} \\ -\theta \frac{d\gamma^3}{(\gamma d+1)^3} & \theta^2 \frac{d^2\gamma^4}{(\gamma d+1)^4} \end{pmatrix} \right),$$

Pour conclure, $\ln\left(\frac{k_n}{nv_n}\right) \sim \ln\left(\frac{k_n}{n(1-\alpha_n)}\right)$ quand $n \rightarrow +\infty$, d'où le résultat. La consistance est immédiate. $\square$

Nous pouvons souligner que la condition (C_{high}) est satisfaite dans la plupart des cas. En effet, les exemples simples qui ne satisfont pas (ii) sont de la forme $\alpha_n = 1 - n^{-1} \ln(n)^{-\kappa}$, $\kappa > 0$ et $k_n = \ln(n)$. Mais un tel choix de suites pourrait mener à une mauvaise estimation de $\hat{\gamma}_{k_n}$ et $\hat{\eta}_{k_n}$, car $k_n \rightarrow +\infty$ très lentement, et en plus une mauvaise estimation du quantile, le niveau α_n tendant vers 1 lentement. Ces suites ne sont donc pas recommandées en pratique. Le corollaire suivant donne la valeur de θ lorsque les suites k_n et α_n ont une forme polynomiale.

Corollaire 1. *Sous l'Hypothèse $\textcircled{2}$, les conditions (C) , (C_{high}) , et en prenant $k_n = n^b$, $0 < b < 1$ et $\alpha_n = 1 - n^{-a}$, $a > 1$, la relation $(\textcircled{23})$ est vérifiée avec $\theta = \frac{a}{a+b-1}$.*

Comme pour les quantiles intermédiaires, la normalité asymptotique $(\textcircled{23})$ peut être améliorée sous la condition

$$\lim_{n \rightarrow +\infty} \frac{\sqrt{k_n}}{\ln\left(\frac{k_n}{n(1-\alpha_n)}\right)} \ln\left(\frac{\Phi_Y^{-1}(1-v_n)^{\frac{1}{\eta}}}{\Phi_{Y|\mathbf{x}}^{-1}(\alpha_n)}\right) = 0.$$

L'Hypothèse $\textcircled{3}$ nous place dans un cadre où il est facile de vérifier cette condition, d'où le résultat suivant.

Proposition 24 (Normalité asymptotique de $\hat{q}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})$). *Supposons que l'Hypothèse $\textcircled{3}$ et les conditions (C) , (C_{high}^{HG}) soient satisfaites. Alors :*

$$\frac{\sqrt{k_n}}{\ln\left(\frac{k_n}{n(1-\alpha_n)}\right)} \left(\frac{\hat{q}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})}{q_{\alpha_n}(Y|\mathbf{X} = \mathbf{x})} - 1 \right) \xrightarrow{n \rightarrow +\infty} \mathcal{N}\left(0, \left(\frac{\gamma}{\gamma d + 1} - \theta \frac{d\gamma^2}{(\gamma d + 1)^2}\right)^2\right). \quad (\text{III.32})$$

Preuve : La preuve est similaire à celle de la Proposition $\textcircled{22}$. En effet, nous avons donné dans le cas $\rho/\gamma \leq -2$:

$$\frac{\Phi_Y^{-1}(1-v_n)^{\frac{1}{\eta}}}{\Phi_{Y|\mathbf{x}}^{-1}(\alpha_n)} = \ell(\mathbf{x})^{-\frac{\gamma}{\gamma d + 1}} (2(1 - \ell(\mathbf{x}))(1 - \alpha_n) + \ell(\mathbf{x}))^{\frac{\gamma}{\gamma d + 1}} \left[1 + \lambda(1 - \alpha_n)^{\frac{2\gamma}{\gamma d + 1}} + o\left((1 - \alpha_n)^{\frac{2\gamma}{\gamma d + 1}}\right) \right]$$

pour une certaine constante $\lambda \in \mathbb{R}$. Il reste alors à calculer

$$\begin{aligned} \lim_{n \rightarrow +\infty} \frac{\sqrt{k_n}}{\ln\left(\frac{k_n}{n(1-\alpha_n)}\right)} \ln\left(\frac{\Phi_Y^{-1}(1-v_n)^{\frac{1}{\eta}}}{\Phi_{Y|\mathbf{x}}^{-1}(\alpha_n)}\right) &= \frac{\gamma}{\gamma d + 1} \lim_{n \rightarrow +\infty} \frac{\sqrt{k_n}}{\ln\left(\frac{k_n}{n(1-\alpha_n)}\right)} \ln\left(2 \frac{1 - \ell(\mathbf{x})}{\ell(\mathbf{x})} (1 - \alpha_n) + 1\right) \\ &\quad + \lim_{n \rightarrow +\infty} \frac{\sqrt{k_n}}{\ln\left(\frac{k_n}{n(1-\alpha_n)}\right)} \ln\left(1 + \lambda(1 - \alpha_n)^{\frac{2\gamma}{\gamma d + 1}} + o\left((1 - \alpha_n)^{\frac{2\gamma}{\gamma d + 1}}\right)\right) \end{aligned}$$

Le premier terme est simple à calculer. En effet, comme $n(1 - \alpha_n) \rightarrow 0$ et $k_n = o(n)$ quand $n \rightarrow +\infty$, on déduit

$$\lim_{n \rightarrow +\infty} \frac{\sqrt{k_n}}{\ln\left(\frac{k_n}{n(1-\alpha_n)}\right)} \ln\left(2 \frac{1 - \ell(\mathbf{x})}{\ell(\mathbf{x})} (1 - \alpha_n) + 1\right) = 2 \frac{1 - \ell(\mathbf{x})}{\ell(\mathbf{x})} \lim_{n \rightarrow +\infty} \frac{\sqrt{k_n}}{\ln\left(\frac{k_n}{n(1-\alpha_n)}\right)} (1 - \alpha_n) = 0$$

Par un calcul similaire, le second terme tend aussi vers 0, en supposant $\frac{\sqrt{k_n}}{\ln\left(\frac{k_n}{n(1-\alpha_n)}\right)} (1 - \alpha_n)^{\frac{2\gamma}{\gamma d + 1}} \rightarrow 0$ quand $n \rightarrow +\infty$. On déduit alors, à l'aide de la Proposition [23](#) :

$$\begin{aligned} & \frac{\sqrt{k_n}}{\ln\left(\frac{k_n}{n(1-\alpha_n)}\right)} \left(\frac{\hat{q}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})}{q_{\alpha_n}(Y|\mathbf{X} = \mathbf{x})} - 1 \right) \underset{n \rightarrow +\infty}{\sim} \frac{\sqrt{k_n}}{\ln\left(\frac{k_n}{n(1-\alpha_n)}\right)} \left(\frac{\left[W_{[k_n+1]} \left(\frac{k_n}{n\hat{v}_n} \right)^{\hat{\gamma}_{k_n}} \right]^{\frac{1}{\hat{\eta}_{k_n}}}}{\Phi_{Y|\mathbf{x}}^{-1}(\alpha_n)} - 1 \right) \\ &= \frac{\sqrt{k_n}}{\ln\left(\frac{k_n}{n(1-\alpha_n)}\right)} \left(\frac{\left[W_{[k_n+1]} \left(\frac{k_n}{n\hat{v}_n} \right)^{\hat{\gamma}_{k_n}} \right]^{\frac{1}{\hat{\eta}_{k_n}}}}{\Phi_Y^{-1}(1 - v_n)^{\frac{1}{\eta}}} - 1 \right) \frac{\Phi_Y^{-1}(1 - v_n)^{\frac{1}{\eta}}}{\Phi_{Y|\mathbf{x}}^{-1}(\alpha_n)} + \frac{\sqrt{k_n}}{\ln\left(\frac{k_n}{n(1-\alpha_n)}\right)} \left(\frac{\Phi_Y^{-1}(1 - v_n)^{\frac{1}{\eta}}}{\Phi_{Y|\mathbf{x}}^{-1}(\alpha_n)} - 1 \right) \\ &\quad \xrightarrow{n \rightarrow +\infty} \mathcal{N}\left(0, \frac{\gamma^2}{(\gamma d + 1)^2} - 2\theta \frac{d\gamma^3}{(\gamma d + 1)^3} + \theta^2 \frac{d^2\gamma^4}{(\gamma d + 1)^4}\right). \end{aligned}$$

Intéressons nous maintenant au cas $\rho/\gamma > -2$. La preuve est exactement la même, avec

$$\frac{\Phi_Y^{-1}(1 - v_n)^{\frac{1}{\eta}}}{\Phi_{Y|\mathbf{x}}^{-1}(\alpha_n)} = \ell(\mathbf{x})^{-\frac{\gamma}{\gamma d + 1}} (2(1 - \ell(\mathbf{x}))(1 - \alpha_n) + \ell(\mathbf{x}))^{\frac{\gamma}{\gamma d + 1}} \left[1 + \lambda(1 - \alpha_n)^{\frac{\rho}{\gamma d + 1}} + o\left((1 - \alpha_n)^{\frac{\rho}{\gamma d + 1}}\right) \right], \lambda \in \mathbb{R}$$

En usant des mêmes calculs, et supposant $\lim_{n \rightarrow +\infty} \frac{\sqrt{k_n}}{\ln\left(\frac{k_n}{n(1-\alpha_n)}\right)} (1 - \alpha_n)^{-\frac{\rho}{\gamma d + 1}} = 0$, on arrive au résultat. $\square$

On peut faire exactement la même remarque que celle de la Proposition [22](#) concernant le comportement de la variance asymptotique lorsque le nombre de covariables d devient grand.

La Section [III.4](#) propose une application numérique avec une distribution pour laquelle on est capable de calculer les quantiles conditionnels théoriques. De plus, on propose également un exemple sur des données réelles.

III.4. Etude numérique

On propose ici de tester les estimateurs définis dans ce chapitre sur un jeu de données simulées. Pour cela, on choisit de simuler des répétitions de loi de Student, car on sait

dans ce cas là fournir la formule théorique des quantiles conditionnels (7). De plus, il est mentionné dans [de Haan and Ferreira, 2006] que loi de Student à ν degrés de liberté satisfait l'Hypothèse 2 avec des indices $\gamma = 1/\nu$, $\rho = -2/\nu$ et une fonction auxiliaire $A(t)$ proportionnelle à $t^{-2/\nu}$. Enfin, la distribution de Student satisfait l'Hypothèse 3 ce qui en fait un parfait exemple à étudier. On pourra alors avoir une idée précise de l'efficacité de nos estimateurs.

III.4.a. Quantiles intermédiaires

On simule donc 1 000 échantillons $\mathbf{Z}_1 = (\mathbf{X}_1, Y_1), \dots, \mathbf{Z}_n = (\mathbf{X}_n, Y_n)$ de n simulations d'une loi de Student à $\nu = 1.5$ degrés de liberté, de dimension $d + 1 = 4$. Pour simplifier, on prendra $\boldsymbol{\mu} = \mathbf{0}_{\mathbb{R}^4}$ et $\boldsymbol{\Sigma} = \mathbf{I}_4$. On suppose également que la valeur de la covariable observée est telle que $M(\mathbf{x}) = 1$. Dans ce cadre, l'Equation (7) donne

$$q_{\alpha_n}(Y|\mathbf{X} = \mathbf{x}) = \sqrt{\frac{5}{9}} \Phi_{4.5}^{-1}(\alpha_n).$$

On propose alors de comparer les comportements asymptotiques de nos estimateurs $\hat{\eta}_{k_n}$, $\hat{\ell}_{k_n, h_n}(\mathbf{x})$ et $\hat{q}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})$. On procède dans un premier temps au choix de nos suites k_n , h_n et α_n . Pour cela, on se concentrera sur les suites de la forme $k_n = n^b$, $b < 1$, $h_n = n^{-c}$, $c < 1$ et $\alpha_n = 1 - n^{-a}$, $a < 1$. On doit choisir judicieusement nos paramètres a , b et c de manière à ce qu'ils satisfassent (C) , (C_{int}) et (C_{int}^{HG}) . Il vient que la condition (C) impose $b < 1 - c$, $b < 4c$ et $b < \frac{4}{\nu+4}$. De plus, (C_{int}) ajoute $b \leq 1 - a$. Enfin, (C_{int}^{HG}) ajoute $b \leq 2a$ et $b \leq \frac{4a}{d+\nu}$.

Le choix $k_n = n^{0.45}$, $\alpha_n = 1 - n^{-0.55}$ et $h_n = n^{-0.15}$ semble être un bon compromis, et est fidèle à toutes les hypothèses. On choisit également la densité gaussienne comme noyau K dans l'estimateur de $\hat{\ell}_{k_n, h_n}(\mathbf{x})$. On mesure alors empiriquement, dans la Table III.2 les variances asymptotiques de nos estimateurs. Dans la dernière ligne sont données les variances asymptotiques théoriques calculées dans les Equations ((ii)) et (21). Rappelons que notre approximation $q_{\alpha_n \uparrow}^E$ vaut dans ce cas de figure :

$$q_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x}) = \Phi_{1.5}^{-1} \left(1 - 1 / \left(\ell(\mathbf{x})(1 - \alpha_n)^{-1} + 2(1 - \ell(\mathbf{x})) \right) \right)^{1/3}.$$

La dernière colonne de la Table III.2 donne alors le comportement asymptotique de $\hat{q}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})$ par rapport à $q_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})$. Pour simplifier, on introduit les notations suivantes :

$$\begin{cases} \hat{v}_{\ell(\mathbf{x}), \eta} &= \widehat{\text{Var}} \left[\sqrt{k_n} \left(\hat{\ell}_{k_n, h_n}(\mathbf{x}) - \ell(\mathbf{x}), \hat{\eta}_{k_n} - \eta \right)^\top \right], \\ \hat{v}_{q^E} &= \widehat{\text{Var}} \left[\frac{\sqrt{k_n}}{\ln(1 - \alpha_n)} \left(\frac{\hat{q}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})}{q_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})} - 1 \right) \right], \\ \hat{v}_q &= \widehat{\text{Var}} \left[\frac{\sqrt{k_n}}{\ln(1 - \alpha_n)} \left(\frac{\hat{q}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})}{q_{\alpha_n}(Y|\mathbf{X} = \mathbf{x})} - 1 \right) \right]. \end{cases}$$

Notons que l'écriture $\widehat{\mathbb{V}ar}$ représente la variance mesurée sur l'échantillon de taille 1 000.

n	α_n	$\hat{v}_{\ell(\mathbf{x}),\eta}$	$\hat{v}_{q^E}$	$\hat{v}_q$
1 000	0.9776128	$\begin{pmatrix} 39.89251 & -11.30352 \\ -11.30352 & 3.877192 \end{pmatrix}$	0.03214032	0.04701728
10 000	0.9936904	$\begin{pmatrix} 34.93154 & -11.45070 \\ -11.45070 & 4.010496 \end{pmatrix}$	0.02982159	0.03661058
100 000	0.9982217	$\begin{pmatrix} 29.65537 & -10.69910 \\ -10.69910 & 3.947848 \end{pmatrix}$	0.02863619	0.03207807
1 000 000	0.9994988	$\begin{pmatrix} 30.94410 & -11.35545 \\ -11.35545 & 4.206882 \end{pmatrix}$	0.02844920	0.03031777
2 000 000	0.9996577	$\begin{pmatrix} 28.75429 & -10.54041 \\ -10.54041 & 3.889293 \end{pmatrix}$	0.02575056	0.02716628
$+\infty$	1	$\begin{pmatrix} 28.98503 & -10.76755 \\ -10.76755 & 4 \end{pmatrix}$	0.02194787	0.02194787

TABLE III.2 – Variances asymptotiques mesurées et théorique (pour $n = +\infty$) de nos estimateurs

Concentrons nous maintenant sur le cas $n = 2\,000\,000$. La Figure III.4.a. représente les boxplots des 1 000 répétitions de $\hat{\eta}_{k_n}$, $\hat{\ell}_{k_n, h_n}(\mathbf{x})$ et $\hat{q}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})$ respectivement.

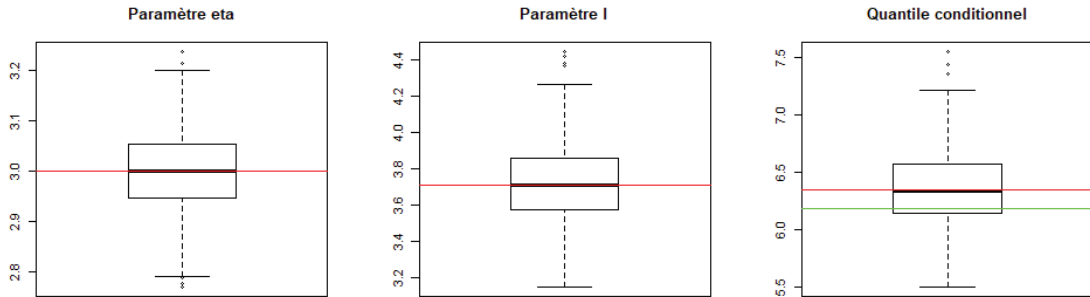


FIGURE III.2 – Boxplots de 1 000 répétitions de $\hat{\eta}_{k_n}$, $\hat{\ell}_{k_n, h_n}(\mathbf{x})$ et $\hat{q}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})$ respectivement. Les valeurs de η , $\ell(\mathbf{x})$ et $q_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})$ apparaissent en rouge. $q_{\alpha_n}(Y|\mathbf{X} = \mathbf{x})$ est en vert.

On remarque que les estimateurs $\hat{\eta}_{k_n}$, $\hat{\ell}_{k_n, h_n}(\mathbf{x})$ présentent une bonne efficacité. En revanche, si $\hat{q}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})$ paraît fournir une bonne estimation de $q_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x}) =$

6.345426, il n'est pas forcément une estimation optimale du quantile conditionnel théorique $q_{\alpha_n}(Y|\mathbf{X} = \mathbf{x}) = 6.177874$. En effet, au regard de la relation asymptotique (11), le fait de considérer une suite tendant vers 1 assez lentement conduit à une convergence assez lente de notre prédicteur. Les résultats sont alors meilleurs en prenant une suite α_n plus rapide, i.e. en passant au cas des hauts quantiles.

III.4.b. Hauts quantiles

Plaçons nous dans des conditions similaires, c'est à dire que l'on considère une distribution Student avec $d = 3$, $\nu = 1.5$ et $M(\mathbf{x}) = 1$. Notre estimateur, introduit dans la Définition 17, sera donc de la forme :

$$\hat{q}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x}) = \mu_{Y|\mathbf{x}} + \sqrt{\Sigma_{Y|\mathbf{x}}} \left[W_{[k_n+1]} \left(\frac{k_n}{n} \left(2 + \hat{\ell}_{k_n, h_n}(\mathbf{x}) \left(\frac{1}{1 - \alpha_n} - 2 \right) \right) \right)^{\hat{\gamma}_{k_n}} \right]^{\frac{1}{\hat{\eta}_{k_n}}}.$$

Comme pour les quantiles intermédiaires, prenons des suites $k_n = n^b$, $b < 1$, $h_n = n^{-c}$, $c < 1$ et $\alpha_n = 1 - n^{-a}$, $a > 1$, et regardons les conditions sur les paramètres a , b et c pour que (C) , (C_{high}) et (C_{high}^{HG}) soient remplies. Commune au cas des quantiles intermédiaires, (C) impose $b < 1 - c$, $b < 4c$ et $b < \frac{4}{\nu+4}$. (C_{high}) est satisfaite avec $\theta = \frac{a}{a+b-1}$, et (C_{high}^{HG}) implique $b \leq 2a$ et $b \leq \frac{4a}{d+\nu}$. En regardant alors de plus près les normalités asymptotiques (23) et (24), on s'aperçoit que le choix $a = (1 - b)(\gamma d + 1)$ mène à une variance nulle ! Dans ce cas, on prendra donc $b = 0.6$ ainsi que $a = 0.4 \times (2 + 1) = 1.2$. En pratique, il faut bien sûr estimer γ . On peut alors fixer b puis prendre $a = (1 - b)(\hat{\gamma}_{k_n} d + 1)$. On choisit également $h_n = n^{-0.20}$, ainsi que la densité gaussienne comme noyau K . Regardons alors dans la Table III.3 le comportement de nos estimateurs pour plusieurs valeurs de n .

n	α_n	$\hat{v}_{\ell(\mathbf{x}), \eta}$	$\hat{v}_{q^E}$	$\hat{v}_q$
1 000	0.9997488	$\begin{pmatrix} 30.18146 & -10.19442 \\ -10.19442 & 4.24693 \end{pmatrix}$	0.00270203	0.002830734
10 000	0.9999842	$\begin{pmatrix} 33.48424 & -11.26003 \\ -11.26003 & 4.22301 \end{pmatrix}$	0.001280309	0.001297676
100 000	0.999999	$\begin{pmatrix} 29.47296 & -10.42533 \\ -10.42533 & 3.97305 \end{pmatrix}$	0.0007163233	0.0007191455
1 000 000	0.9999999	$\begin{pmatrix} 29.60700 & -10.42303 \\ -10.42303 & 3.837567 \end{pmatrix}$	0.0005030306	0.0005036096
10 000 000	0.999999996	$\begin{pmatrix} 29.73965 & -10.72249 \\ -10.72249 & 3.982351 \end{pmatrix}$	0.0003761021	0.0003762288
$+\infty$	1	$\begin{pmatrix} 28.98503 & -10.76755 \\ -10.76755 & 4 \end{pmatrix}$	0	0

TABLE III.3 – Variances asymptotiques mesurées et théorique (pour $n = +\infty$) de nos estimateurs

Les quantités $\hat{v}_{q^E}$ et $\hat{v}_q$ sont définies comme suit :

$$\begin{cases} \hat{v}_{q^E} = \widehat{\text{Var}} \left[\frac{\sqrt{k_n}}{\ln\left(\frac{k_n}{n(1-\alpha_n)}\right)} \left(\frac{\hat{q}_{\alpha_n \uparrow}^E(Y|\mathbf{X}=\mathbf{x})}{q_{\alpha_n \uparrow}^E(Y|\mathbf{X}=\mathbf{x})} - 1 \right) \right], \\ \hat{v}_q = \widehat{\text{Var}} \left[\frac{\sqrt{k_n}}{\ln\left(\frac{k_n}{n(1-\alpha_n)}\right)} \left(\frac{\hat{q}_{\alpha_n \uparrow}^E(Y|\mathbf{X}=\mathbf{x})}{q_{\alpha_n}(Y|\mathbf{X}=\mathbf{x})} - 1 \right) \right]. \end{cases}$$

Prenons plus particulièrement le cas $n = 10\,000\,000$. Les boxplots des 1 000 estimations de $\hat{\eta}_{k_n}$, $\hat{\ell}_{k_n, h_n}(\mathbf{x})$ et $\hat{q}_{\alpha_n \uparrow}^E(Y|\mathbf{X}=\mathbf{x})$ sont illustrés dans la Figure III.4.b.

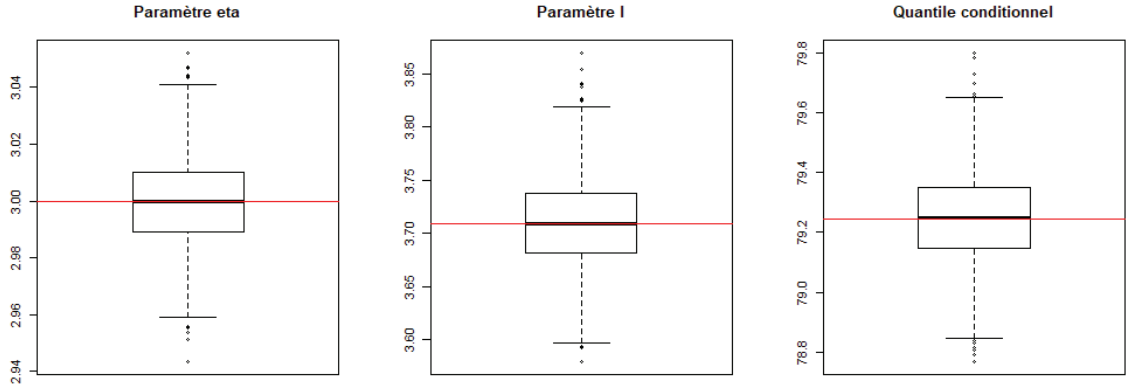


FIGURE III.3 – Boxplots de 1 000 répétitions de $\hat{\eta}_{k_n}$, $\hat{\ell}_{k_n, h_n}(\mathbf{x})$ et $\hat{q}_{\alpha_n \uparrow}^E(Y|\mathbf{X}=\mathbf{x})$ respectivement. Les valeurs de η , $\ell(\mathbf{x})$ et $q_{\alpha_n \uparrow}^E(Y|\mathbf{X}=\mathbf{x})$ apparaissent en rouge. $q_{\alpha_n}(Y|\mathbf{X}=\mathbf{x})$ est en vert.

Si on compare aux quantiles intermédiaires, on voit que notre approximation $q_{\alpha_n \uparrow}^E(Y|\mathbf{X}=\mathbf{x}) = 79.25944$ est très proche de la valeur théorique $q_{\alpha_n}(Y|\mathbf{X}=\mathbf{x}) = 79.2461$. Par conséquent, notre estimateur $\hat{q}_{\alpha_n \uparrow}^E(Y|\mathbf{X}=\mathbf{x})$ est beaucoup plus efficace.

III.4.c. Un exemple de données réelles

En guise d'application, on propose d'utiliser les rendements journaliers d'actifs financiers de 2006 à 2016, disponibles à <http://stanford.edu/class/ee103/portfolio.html>. Intéressons nous par exemple aux 4 premiers actifs, à savoir iShares Core U.S. Aggregate Bond ETF, PowerShares DB Commodity Index Tracking Fund, WisdomTree Europe SmallCap Dividend Fund et SPDR Dow Jones Industrial Average ETF. La Figure III.4 représente les rendements quotidiens de ces 4 actifs.

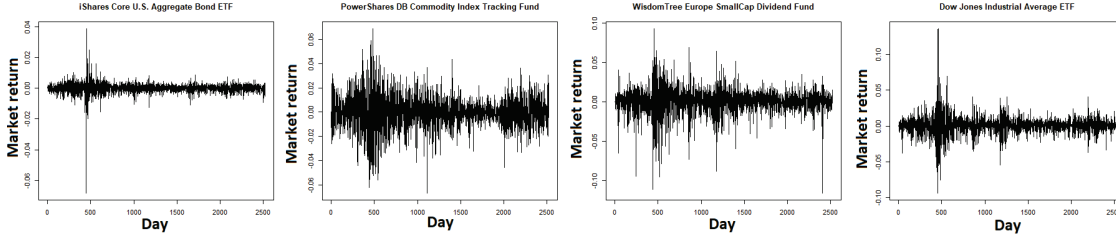


FIGURE III.4 – Rendements journaliers de 4 actifs financiers.

On pourrait considérer, par exemple, que ces rendements sont les premiers disponibles chaque jour. Le but serait alors d'estimer un quantile sur le rendement d'un autre actif disponible sur un autre marché ouvrant plus tard. On considère alors les rendements d'un 5e actif, qui sera WisdomTree Japan Hedged Equity Fund. La taille de l'échantillon est ici de 2520. On considère alors les 2519 premiers jours (du 03/01/2007 au 05/12/2016) comme un échantillon d'apprentissage, et on se place au 2520e jour (soit le 06/12/2016), où la covariable $\mathbf{X}$ vaut $\mathbf{x} = (-0.0185\%, -0.4464\%, 0.9614\%, 0.1405\%)$. En attendant l'ouverture du second marché, estimons donc le quantile du 5e rendement Y sachant $\mathbf{X} = \mathbf{x}$.

Après une brève étude des fonctions d'autocorrélation, on s'autorise à considérer que les répétitions sont indépendantes. Concernant la forme de la distribution, les histogrammes des marginales semblent symétriques. De plus, les indices de valeurs extrêmes mesurés sont sensiblement les mêmes pour les 4 actifs. Avec tous ces éléments, on se permet alors de considérer que l'on dispose de répétitions indépendantes et identiquement distribuées de vecteurs elliptiques en dimension 4. Après avoir estimé $\boldsymbol{\mu}$ et $\boldsymbol{\Sigma}$ par la méthode des moments, on obtient $M(\mathbf{x}) = 1.072952$. On applique alors nos estimateurs $\hat{\eta}_{k_n}$ et $\hat{\ell}_{k_n, h_n}(\mathbf{x})$ donnés dans les Equations (13) et (15). On prend pour cela les suites $k_n = n^{0.6}$ ($b = 0.6$) et $h_n = n^{-0.2}$ ($c = 0.2$), et comme noyau K la densité gaussienne, d'où l'on déduit les intervalles de confiance de l'Equation (i). Notre coefficient $\ell(\mathbf{x})$ se situerait entre 5.27907 et 7.60761 avec une probabilité proche de 0.95. De même, η serait entre 2.315297 et 2.920395. Si on se place dans le cadre des hauts quantiles, et que l'on cherche à minimiser la variance asymptotique de l'Equation (24), on prendra alors $a = (1 - b)(\hat{\gamma}_{k_n} + 1) = 1.047146$. En appliquant l'estimateur (17), on obtient un quantile de niveau 0.9997256 proche de 3.744985% pour Y sachant $\mathbf{X} = \mathbf{x}$. En d'autres termes, avant l'ouverture du second marché, on estime qu'étant donnés les rendements de nos 4 premiers actifs, celui de WisdomTree Japan Hedged Equity Fund a une probabilité 0.9997256 d'être inférieur à 3.744985%. A titre d'information, le rendement mesuré était en réalité de 0.7141%.

EXPECTILES ET AUTRES MESURES CONDITIONNELLES

Récemment, le quantile, ou Value at Risk, comme il est appelé dans la littérature financière (voir [Linsmeier and Pearson, 2000]), a subi quelques critiques sur sa capacité à être une mesure de risque acceptable dans un cadre financier ou actuariel. Certains auteurs, comme [Artzner et al., 1999], lui reprochent de ne pas être une mesure de risque cohérente, au sens de la Définition 3 car il ne vérifie pas la propriété de sous-additivité. Cette dernière assure que le risque de la somme de deux portefeuilles ne doit pas dépasser la somme des risques des portefeuilles. Si, dans le cas des distributions elliptiques, [McNeil et al., 2015] montre que le quantile est bien sous-additif pour $\alpha \geq 0.5$, ce dernier n'est pas sensible à l'ampleur des événements extrêmes. De manière à corriger ce défaut, l'expectile, que l'on peut trouver dans [Newey and Powell, 1987], est une option. Nous verrons qu'il est construit de manière similaire au quantile, et qu'il a l'avantage d'être une mesure de risque cohérente, comme montré dans [Bellini et al., 2014]. Il possède donc un nombre croissant d'applications en sciences actuarielle et financière. [Cai and Weng, 2016] et [Taylor, 2008] en sont de bons exemples. Dans ce chapitre, nous proposons une approche similaire à celle effectuée pour le quantile au Chapitre III appliquée cette fois à l'expectile. Nous donnons alors dans un premier temps une formalisation mathématique de l'expectile dans le cas général, puis plus particulièrement elliptique. Nous testons ensuite le modèle de régression pour approximer des expectiles conditionnels, puis après avoir constaté les limites de ce modèle, on propose une approche méthodologique pour les approximer puis les estimer. Une étude numérique est proposée pour conclure. Une partie de ce chapitre est publiée dans [Maume-Deschamps et al., 2018].

IV.1. Introduction

On donne d'abord la définition d'un expectile, et quelques propriétés dans le cas des lois elliptiques. Puis, l'expectile ne disposant pas de formule fermée, on propose ensuite quelques méthodes numériques pour approximer des expectiles elliptiques. Enfin, à l'aide de ces algorithmes, on donne quelques expectiles elliptiques conditionnels.

IV.1.a. La notion d'expectile

Les expectiles, introduits dans [Newey and Powell, 1987], peuvent être vus comme une généralisation des quantiles. Dans cette section d'introduction, on définit ces quantités

avec la notion d'élicitabilité (voir par exemple [Ziegel, 2016]). En effet, si on considère Z une variable aléatoire réelle, le quantile de niveau $\alpha \in]0, 1[$ $q_\alpha(Z)$ de Z est donné par le problème de minimisation suivant :

$$q_\alpha(Z) = \arg \min_{q \in \mathbb{R}} \mathbb{E} [\mathcal{S}_{1,\alpha}(Z - q)] \quad (\text{IV.1})$$

avec la fonction asymétrique $\mathcal{S}_{1,\alpha}(t) = (1-\alpha)|t^-| + \alpha|t^+|$ introduite dans l'Equation (III.2.a.). La résolution de cette équation mène au résultat classique $\mathbb{P}(Z \leq q) = \alpha$ pour des distributions continues. En s'inspirant de cette approche, on peut définir l'expectile de niveau $\alpha \in]0, 1[$ de manière similaire en modifiant la fonction de scoring.

Définition 18 (Expectiles). *Soit Z une variable aléatoire réelle admettant un moment d'ordre 1. L'expectile de niveau $\alpha \in]0, 1[$ de Z est donné par :*

$$e_\alpha(Z) = \arg \min_{e \in \mathbb{R}} \mathbb{E} [\mathcal{S}_{2,\alpha}(Z - e)] \quad (\text{IV.2})$$

où la fonction $\mathcal{S}_{2,\alpha}$ est définie de la manière suivante :

$$\mathcal{S}_{2,\alpha}(t) = (1-\alpha)|t^-|^2 + \alpha|t^+|^2. \quad (\text{IV.3})$$

Le problème (18) semble nécessiter un moment d'ordre 2. En réalité, la condition sur le gradient fait intervenir uniquement des moments d'ordre 1 :

$$\alpha \mathbb{E} [(Z - e_\alpha(Z))^+] = (1-\alpha) \mathbb{E} [(e_\alpha(Z) - Z)^+].$$

La nouvelle fonction de scoring de l'expectile est finalement une version quadratique de celle du quantile. La Figure IV.1 propose un aperçu visuel des deux fonctions de scoring $\mathcal{S}_{1,\alpha}$ et $\mathcal{S}_{2,\alpha}$.

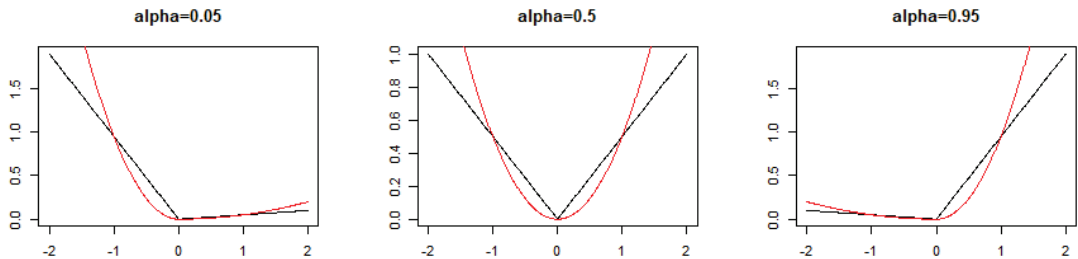


FIGURE IV.1 – Fonctions $\mathcal{S}_{1,\alpha}$ en noir, et $\mathcal{S}_{2,\alpha}$ en rouge pour des niveaux α respectivement de 0.05, 0.5 et 0.95.

Contrairement au quantile, où la solution du problème (IV.1.a.) est assez évidente et intuitive, exprimer l'expectile simplement est moins aisé. C'est ce que nous essaierons de faire dans un premier temps, dans notre contexte de distributions elliptiques. Pour cela, on effectue quelques rappels des sections précédentes. On considère notre vecteur $\mathbf{Z} = (\mathbf{X}, Y) \sim \mathcal{EC}_{d+1}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \xi)$ avec $\mathbf{X} \in \mathbb{R}^d$ et $Y \in \mathbb{R}$. On rappelle les notations suivantes :

$$\begin{cases} \Phi_Y(t) = \mathbb{P}\left(\frac{Y - \mu_Y}{\sqrt{\Sigma_Y}} \leq t\right) \\ \Phi_{Y|\mathbf{x}}(t) = \mathbb{P}\left(\frac{Y|(\mathbf{X}=\mathbf{x}) - \mu_{Y|\mathbf{x}}}{\sqrt{\Sigma_{Y|\mathbf{x}}}} \leq t\right). \end{cases}$$

Les quantiles de Y et $Y|(\mathbf{X} = \mathbf{x})$ pouvant facilement être exprimés à partir de ces fonctions, on va alors introduire une nouvelle fonction Ψ pour faire de même avec les expectiles.

Définition 19. Soit $\Psi_Y : \mathbb{R}^* \rightarrow]-\infty, 0[\cup]1, +\infty[$ telle que :

$$\Psi_Y(t) = \Phi_Y(t) + \frac{1}{t} \int_t^{+\infty} u c_1 g_1(u^2) du. \quad (\text{IV.4})$$

De la même manière, on peut définir la fonction $\Psi_{Y|\mathbf{x}}$ relative à la distribution $Y|(\mathbf{X} = \mathbf{x})$. A l'instar de la fonction de répartition Φ , la fonction Ψ possède quelques propriétés évidentes que l'on énonce ci-dessous.

Lemme 25 (Propriétés de Ψ_Y). Ψ_Y satisfait les propriétés suivantes :

- $\Psi_Y(-t) = 1 - \Psi_Y(t), \forall t \in \mathbb{R}^*$.
- $\Psi_Y : \mathbb{R}^* \rightarrow]-\infty, 0[\cup]1, +\infty[$ est bijective et décroissante.
- $\Psi_Y^{-1}(1 - \alpha) = -\Psi_Y^{-1}(\alpha), \forall \alpha \in \mathbb{R}^*$.

La preuve du Lemme 25 étant suffisamment évidente, elle n'est pas fournie. On va pouvoir, à l'aide des définitions et propriétés précédentes, donner une forme générale des expectiles dans le cas des distributions elliptiques. Pour cela, on a besoin de l'expression d'un moment tronqué.

Lemme 26. Si $Y \sim \mathcal{EC}_1(\mu_Y, \Sigma_Y, \xi)$, on a la relation suivante :

$$\mathbb{E}[Y^+] = \mu_Y \Phi_Y\left(\frac{\mu_Y}{\sqrt{\Sigma_Y}}\right) + \sqrt{\Sigma_Y} \int_{-\frac{\mu_Y}{\sqrt{\Sigma_Y}}}^{+\infty} u c_1 g_1(u^2) du. \quad (\text{IV.5})$$

Preuve : On rappelle que la densité de Y est donnée par :

$$f_Y(t) = \frac{c_1}{\sqrt{\Sigma_Y}} g_1 \left(\frac{(t - \mu_Y)^2}{\Sigma_Y} \right).$$

Alors :

$$\mathbb{E}[Y^+] = \int_0^{+\infty} u \frac{c_1}{\sqrt{\Sigma_Y}} g_1 \left(\frac{(u - \mu_Y)^2}{\Sigma_Y} \right) du = \mu_Y \int_{-\frac{\mu_Y}{\sqrt{\Sigma_Y}}}^{+\infty} c_1 g_1(u^2) du + \sqrt{\Sigma_Y} \int_{-\frac{\mu_Y}{\sqrt{\Sigma_Y}}}^{+\infty} u c_1 g_1(u^2) du.$$

Comme $\Phi_Y(t) = \int_{-\infty}^t c_1 g_1(u^2) du$ et $1 - \Phi_Y(x) = \Phi_Y(-x)$ (par les propriétés de symétrie des distributions elliptiques), on obtient l'Equation (26). $\square$

Donnons maintenant une expression générale pour l'expectile de niveau $\alpha \in]0, 1[$ d'une loi elliptique univariée. Cette expression rappelle notamment l'approche de [Jones, 1994], où les expectiles sont vus comme les quantiles d'une autre distribution.

Proposition 27 (Expectile elliptique). *Soit $Y \sim \mathcal{EC}_1(\mu_Y, \Sigma_Y, \xi)$ avec $\mathbb{E}[\xi] < +\infty$. L'expectile de niveau $\alpha \in]0, 1[$ noté e_α de Y est donné par :*

$$e_\alpha(Y) = \begin{cases} \mu_Y + \sqrt{\Sigma_Y} \Psi_Y^{-1} \left(\frac{\alpha}{2\alpha-1} \right) & , \alpha \neq \frac{1}{2} \\ \mu_Y & , \alpha = \frac{1}{2} \end{cases}. \quad (\text{IV.6})$$

Preuve : On doit résoudre le problème de minimisation suivant :

$$e_\alpha(Y) = \arg \min_{e \in \mathbb{R}} \mathbb{E}[\mathcal{S}_\alpha(Y - e)].$$

De manière à rendre la preuve plus agréable, on notera par la suite e_α au lieu de $e_\alpha(Y)$. En dérivant la fonction de scoring, on obtient :

$$(1 - \alpha)\mathbb{E}[Y - e_\alpha] + (2\alpha - 1)\mathbb{E}[(Y - e_\alpha)^+] = 0.$$

En utilisant le Lemme 26 on a la relation :

$$\mathbb{E}[(Y - e_\alpha)^+] = (\mu_Y - e_\alpha)\Phi_Y \left(\frac{\mu_Y - e_\alpha}{\sqrt{\Sigma_Y}} \right) + \sqrt{\Sigma_Y} \int_{-\frac{\mu_Y - e_\alpha}{\sqrt{\Sigma_Y}}}^{+\infty} u c_1 g_1(u^2) du.$$

En l'incorporant dans l'égalité précédente, on obtient :

$$(1 - \alpha)(\mu_Y - e_\alpha) + (2\alpha - 1) \left[(\mu_Y - e_\alpha) \Phi_Y \left(\frac{\mu_Y - e_\alpha}{\sqrt{\Sigma_Y}} \right) + \sqrt{\Sigma_Y} \int_{-\frac{\mu_Y - e_\alpha}{\sqrt{\Sigma_Y}}}^{+\infty} u c_1 g_1(u^2) du \right] = 0.$$

Un niveau $\alpha = \frac{1}{2}$ mène à la solution évidente $e_\alpha = \mu_Y$. On considère maintenant $\alpha \neq \frac{1}{2}$ (d'où $e_\alpha \neq \mu_Y$), l'équation peut alors être réécrite :

$$\Phi_Y \left(\frac{e_\alpha - \mu_Y}{\sqrt{\Sigma_Y}} \right) + \frac{\sqrt{\Sigma_Y}}{e_\alpha - \mu_Y} \int_{\frac{e_\alpha - \mu_Y}{\sqrt{\Sigma_Y}}}^{+\infty} u c_1 g_1(u^2) du = \frac{\alpha}{2\alpha - 1}.$$

D'où $\Psi_Y \left(\frac{e_\alpha - \mu_Y}{\sqrt{\Sigma_Y}} \right) = \frac{\alpha}{2\alpha - 1}$, et $e_\alpha = \mu_Y + \sqrt{\Sigma_Y} \Psi_Y^{-1} \left(\frac{\alpha}{2\alpha - 1} \right)$. $\square$

On vient de définir une expression générale pour les expectiles de distributions elliptiques. Comme pour le quantile, on s'intéresse aux expectiles conditionnels de Y connaissant la covariable $\mathbf{X} = \mathbf{x}$. Avant cela, on propose quelques méthodes numériques pour obtenir des expectiles elliptiques.

IV.1.b. Algorithmes

La forme de la fonction Ψ , introduite dans la Définition [19](#) laisse à penser qu'il n'est pas immédiat d'obtenir des formules explicites pour les expectiles. Pour cela, on propose ici deux algorithmes pour le calcul du terme $\Psi^{-1} \left(\frac{\alpha}{2\alpha - 1} \right)$ donné dans la formule des expectiles [\(27\)](#). Le premier d'entre eux est l'algorithme MM (voir [\[Hunter and Lange, 2004\]](#)), le second est un algorithme de type point-fixe. On donne également les vitesses de convergence de ces deux algorithmes.

Dans cette optique, introduisons d'abord quelques définitions et propriétés que l'on peut notamment trouver dans [\[Frontini and Sormani, 2003\]](#). Soit une suite $(e^{(k)})_{k \in \mathbb{N}}$ et une valeur $e^* \in \mathbb{R}$. On dira que la suite $(e^{(k)})_{k \in \mathbb{N}}$ converge vers e^* d'ordre $p \in [1, \infty[$ si pour tout k , $|e^{(k)} - e^*| \leq \epsilon_k$ où $(\epsilon_k)_{k \in \mathbb{N}}$ est une suite positive qui satisfait

$$\exists c > 0 : \lim_{k \rightarrow +\infty} \frac{\epsilon_{k+1}}{\epsilon_k^p} = c. \quad (\text{IV.7})$$

Le coefficient c est appelé facteur asymptotique. En particulier, on dira que la convergence est linéaire si $p = 1$, et quadratique si $p = 2$.

Considérons maintenant l'algorithme de point fixe $e^{(k+1)} = f(e^{(k)})$, où $e^{(k+1)}, e^{(k)} \in \mathbb{R}$, et $f : \mathbb{R} \rightarrow \mathbb{R}$ est une fonction dérivable. La convergence de l'algorithme est alors assurée par $|f'(e)| < 1, \forall e \in \mathbb{R}$. Si de plus f est p fois dérivable avec, $\forall m < p$, $f^{(m)}(e^*) = 0$ et $f^{(p)}(e^*) \neq 0$, alors la convergence est d'ordre p avec un facteur asymptotique $c = f^{(p)}(e^*)/p!$.

IV.1.b.1. Algorithme MM

On présente dans un premier temps une approche du calcul de $\Psi^{-1}\left(\frac{\alpha}{2\alpha-1}\right)$ basée sur l'algorithme MM. Une approche similaire a été proposée dans [Yang et al., 2018], et quelques résultats feront référence à ce papier.

Lemme 28 ([Yang et al., 2018]). *La fonction $\mathcal{S}'_{2,\alpha}$ est Lipschitzienne :*

$$|\mathcal{S}'_{2,\alpha}(a) - \mathcal{S}'_{2,\alpha}(b)| \leq 2 \max\{1 - \alpha, \alpha\} |a - b|, \forall a, b \in \mathbb{R}.$$

D'où la borne suivante :

$$\mathcal{S}_{2,\alpha}(a) \leq \mathcal{S}_{2,\alpha}(b) + \mathcal{S}'_{2,\alpha}(b)(a - b) + \max\{1 - \alpha, \alpha\}(a - b)^2, \forall a, b \in \mathbb{R}.$$

En utilisant le Lemme [28] on propose notre algorithme, détaillé ci-dessous.

Proposition 29 (Algorithme MM). *Soit $Y \sim \mathcal{EC}_1(\mu_Y, \Sigma_Y, \xi)$ avec $\mathbb{E}[\xi] < +\infty$. La suite suivante $(e^{(k)})_{k \in \mathbb{N}}$ converge vers $\Psi_Y^{-1}\left(\frac{\alpha}{2\alpha-1}\right)$:*

$$\begin{cases} e^{(0)} &= 0 \\ e^{(k+1)} &= e^{(k)} - \frac{e^{(k)}}{\max\{1-\alpha, \alpha\}} [\alpha - (2\alpha - 1)\Phi_Y(e^{(k)})] + \frac{2\alpha-1}{\max\{1-\alpha, \alpha\}} \int_{e^{(k)}}^{+\infty} u c_1 g_1(u^2) du \end{cases} \quad (\text{IV.8})$$

De plus, la convergence est linéaire au sens de l'Equation ([IV.1.b.]) avec un facteur asymptotique

$$c = 1 - \frac{1}{\max\{1 - \alpha, \alpha\}} \left[\alpha - (2\alpha - 1)\Phi_Y\left(\Psi_Y^{-1}\left(\frac{\alpha}{2\alpha - 1}\right)\right) \right]. \quad (\text{IV.9})$$

Preuve : Soit $Y_0 = (Y - \mu_Y)/\sqrt{\Sigma_Y} \sim \mathcal{EC}_1(0, 1, \xi)$. A la k^e itération, on nomme $R^{(k)} = Y_0 - e^{(k)}$. Il vient que pour tout $e \in \mathbb{R}$:

$$Y_0 - e = R^{(k)} - (e - e^{(k)}).$$

Grâce au Lemme [28] on a l'inégalité suivante :

$$\mathbb{E}[\mathcal{S}_{2,\alpha}(R^{(k)} - (e - e^{(k)}))] \leq \mathbb{E}[\mathcal{S}_{2,\alpha}(R^{(k)})] - (e - e^{(k)})\mathbb{E}[\mathcal{S}'_{2,\alpha}(R^{(k)})] + \max\{1 - \alpha, \alpha\}(e - e^{(k)})^2.$$

On note $Q(e, e^{(k)}) = \mathbb{E}[\mathcal{S}_{2,\alpha}(R^{(k)})] - (e - e^{(k)})\mathbb{E}[\mathcal{S}'_{2,\alpha}(R^{(k)})] + \max\{1 - \alpha, \alpha\}(e - e^{(k)})^2$. Afin de minimiser $\mathbb{E}[\mathcal{S}_{2,\alpha}(R^{(k)} - (e - e^{(k)}))]$, le principe de l'algorithme MM est de minimiser sa borne supérieure, et de choisir $e^{(k+1)} = \arg \min_{e \in \mathbb{R}} Q(e, e^{(k)})$. On obtient alors facilement :

$$e^{(k+1)} = e^{(k)} + \frac{1}{2\max\{1-\alpha, \alpha\}} \mathbb{E} [\mathcal{S}'_{2,\alpha} (Y_0 - e^{(k)})].$$

Il reste à calculer $\mathbb{E} [\mathcal{S}'_{2,\alpha} (Y_0 - e^{(k)})]$.

$$\mathbb{E} [\mathcal{S}'_{2,\alpha} (Y_0 - e^{(k)})] = 2(1-\alpha)\mathbb{E} [Y_0 - e^{(k)}] + 2(2\alpha-1)\mathbb{E} [(Y_0 - e^{(k)})^+].$$

Le Lemme [26](#) mène à l'Equation [\(29\)](#). Prouvons maintenant la convergence de l'algorithme. Comme mentionné précédemment, il est suffisant d'avoir $|f'(t)| < 1, \forall t$, avec $e^{(k+1)} = f(e^{(k)})$.

$$f'(t) = 1 - \frac{\alpha - (2\alpha - 1)\Phi_Y(t)}{\max\{1 - \alpha, \alpha\}}.$$

Clairement, comme $0 < \Phi_Y(t) < 1, \forall t \in \mathbb{R}$, $0 < f'(t) < 1$. D'où la convergence. De plus, comme $f'(t) \neq 0, \forall t \in \mathbb{R}$, la convergence est linéaire, avec facteur asymptotique $f'(\Psi_Y^{-1}(\frac{\alpha}{2\alpha-1}))$, d'où l'Equation [\(29\)](#). $\square$

Logiquement, $c \rightarrow 1$ quand $\alpha \rightarrow 1$. En effet, lorsque α est proche de 1, l'expectile diverge, et son calcul requiert plus d'itérations. D'un autre côté, un niveau $\alpha = 1/2$ donne $c = 0$, car la suite de l'algorithme est constante, avec $e^{(k)} = 0$ pour tout k .

IV.1.b.2. Algorithme de point fixe

On peut montrer aisément que $e_\alpha(Y)$ satisfait une relation de point fixe dans le cas des distributions elliptiques. Dans ce paragraphe, nous explicitons cette relation et l'algorithme de point fixe correspondant.

Proposition 30 (Algorithme de point fixe). *Soit $Y \sim \mathcal{EC}_1(\mu_Y, \Sigma_Y, \xi)$ avec $\mathbb{E}[\xi] < +\infty$. La suite suivante $(e^{(k)})_{k \in \mathbb{N}}$ converge vers $\Psi_Y^{-1}(\frac{\alpha}{2\alpha-1})$:*

$$\begin{cases} e^{(0)} = 0 \\ e^{(k+1)} = \frac{(1-2\alpha) \int_{e^{(k)}}^{+\infty} u c_1 g_1(u^2) dy}{(2\alpha-1)\Phi_Y(e^{(k)}) - \alpha} \end{cases} \quad (\text{IV.10})$$

De plus, la convergence est quadratique au sens de l'Equation [\(IV.1.b.\)](#) avec un facteur asymptotique

$$c = \frac{(2\alpha-1)c_1 g_1\left(\Psi_Y^{-1}\left(\frac{\alpha}{2\alpha-1}\right)^2\right)}{2[(2\alpha-1)\Phi_Y\left(\Psi_Y^{-1}\left(\frac{\alpha}{2\alpha-1}\right)\right) - \alpha]}. \quad (\text{IV.11})$$

Preuve : Soit $Y_0 = (Y - \mu_Y)/\sqrt{\Sigma_Y} \sim \mathcal{EC}_1(0, 1, \xi)$. On a alors :

$$\Psi_Y^{-1}\left(\frac{\alpha}{2\alpha-1}\right) = \arg \min_{e \in \mathbb{R}} \mathbb{E}[\mathcal{S}_{2,\alpha}(Y_0 - e)].$$

Si pour la suite de la preuve on note $e_\alpha = \Psi_Y^{-1}\left(\frac{\alpha}{2\alpha-1}\right)$, alors e_α est tel que $\mathbb{E}[\mathcal{S}'_{2,\alpha}(Y_0 - e_\alpha)] = 0$, i.e. :

$$2(1-\alpha)\mathbb{E}[(Y_0 - e_\alpha)] + 2(2\alpha-1)\mathbb{E}[(Y_0 - e_\alpha)^+] = 0.$$

Avec l'équation précédente, et la formule du Lemme 26 on obtient la relation de point fixe suivante :

$$e_\alpha = \frac{(1-2\alpha) \int_{e_\alpha}^{+\infty} u c_1 g_1(u^2) du}{(2\alpha-1)\Phi_Y(e_\alpha) - \alpha}.$$

Comme dans le cas de l'algorithme MM donné précédemment, on a un algorithme récursif $e^{(k+1)} = f(e^{(k)})$. La convergence est alors assurée par $|f'(e_\alpha)| < 1$. La convergence quadratique est montrée si $f'(e_\alpha) = 0$ et $f''(e_\alpha) \neq 0$. On a

$$f'(t) = (2\alpha-1)c_1 g_1(t^2) \frac{t(2\alpha-1)\Phi_Y(t) - \alpha t + (2\alpha-1) \int_t^{+\infty} u c_1 g_1(u^2) du}{[(2\alpha-1)\Phi_Y(t) - \alpha]^2}.$$

L'Equation (19) implique $f'(e_\alpha) = 0$. Avec ce résultat, on montre non seulement que l'algorithme converge, mais sa convergence est, de plus, au moins quadratique. Après quelques calculs simples, on obtient :

$$f''(e_\alpha) = \frac{(2\alpha-1)c_1 g_1(e_\alpha^2)}{[(2\alpha-1)\Phi_Y(e_\alpha) - \alpha]}.$$

Si $\alpha \neq \frac{1}{2}$, $f''(e_\alpha) \neq 0$ et la convergence est quadratique. De plus, le facteur asymptotique c est facilement déduit, à l'aide de la relation $c = \frac{f''(e_\alpha)}{2}$, d'où l'Equation (30). $\square$

Si on se base sur le taux de convergence, cet algorithme de point fixe est plus efficace que l'algorithme MM défini précédemment. C'est pourquoi dans la suite, nous utiliserons uniquement cet algorithme.

IV.1.c. Quelques formes théoriques d'expectiles conditionnels

Après cette parenthèse numérique sur le calcul des expectiles, revenons à la problématique de ce chapitre, à savoir les expectiles conditionnels de $Y|(\mathbf{X} = \mathbf{x})$. Si on a vu que le cas des quantiles n'avait rien d'immédiat, il en va de même pour celui des expectiles. On

peut cependant écrire une forme générale de l'expectile de niveau $\alpha \in]0, 1[$ de la variable aléatoire $Y|(\mathbf{X} = \mathbf{x})$.

$$e_\alpha(Y|\mathbf{X} = \mathbf{x}) = \mu_{Y|\mathbf{x}} + \sqrt{\Sigma_{Y|\mathbf{x}}} \Psi_{Y|\mathbf{x}}^{-1} \left(\frac{\alpha}{2\alpha - 1} \right) \quad (\text{IV.12})$$

où on rappelle que les quantités $\mu_{Y|\mathbf{x}}, \Sigma_{Y|\mathbf{x}}$ sont données dans l'Equation (5), et la fonction Ψ dans l'Equation (19). Le problème est alors sensiblement le même que celui du chapitre précédent, c'est à dire que l'utilisation de $\Psi_{Y|\mathbf{x}}^{-1} \left(\frac{\alpha}{2\alpha - 1} \right)$ nécessite la connaissance de la distribution conditionnelle $Y|(\mathbf{X} = \mathbf{x})$. On propose dans ce paragraphe quelques résultats d'expectiles conditionnels théoriques, qui nous permettront par la suite de comparer nos résultats avec ces valeurs.

IV.1.c.1. Loi Normale

Ce cas est comme souvent le plus simple. En effet, si le vecteur $\mathbf{Z} = (\mathbf{X}, Y)$ avec $\mathbf{X} \in \mathbb{R}^d$ et $Y \in \mathbb{R}$ est gaussien, la distribution conditionnelle $Y|(\mathbf{X} = \mathbf{x})$ reste gaussienne. La Proposition 30 assure alors que l'algorithme suivant converge vers $\Psi_{Y|\mathbf{x}}^{-1} \left(\frac{\alpha}{2\alpha - 1} \right)$.

$$\begin{cases} e^{(0)} = 0 \\ e^{(k+1)} = \frac{(1-2\alpha)\varphi(e^{(k)})}{(2\alpha-1)\Phi(e^{(k)})-\alpha} \end{cases},$$

où Φ et φ sont respectivement la fonction de répartition et la densité de probabilité de la loi normale centrée réduite.

IV.1.c.2. Loi Student

Le cas Student à $\nu > 0$ degrés de liberté demande plus de calculs, mais, comme pour les quantiles, aboutit à une formule plus ou moins explicite. L'algorithme de point fixe pour $\Psi_{Y|\mathbf{x}}^{-1} \left(\frac{\alpha}{2\alpha - 1} \right)$ est alors

$$\begin{cases} e^{(0)} = 0 \\ e^{(k+1)} = \frac{(2\alpha-1) \frac{\Gamma(\frac{\nu+d+1}{2})}{\Gamma(\frac{\nu+d}{2})\sqrt{\pi}} \frac{\sqrt{\nu+M(\mathbf{x})}}{1-\nu-d} \left(1 + \frac{1}{\nu} \frac{(e^{(k)})^2}{(1+\frac{1}{\nu}M(\mathbf{x}))} \right)^{\frac{1-\nu-d}{2}}}{(2\alpha-1)\Phi_{\nu+d}(e^{(k)})\sqrt{\frac{\nu+d}{\nu+M(\mathbf{x})}}-\alpha} \end{cases}, \quad (\text{IV.13})$$

où Φ_m est la fonction de répartition d'une loi de Student à m degrés de liberté, et $M(\mathbf{x}) = (\mathbf{x} - \boldsymbol{\mu}_\mathbf{X})^\top \Sigma_\mathbf{X}^{-1} (\mathbf{x} - \boldsymbol{\mu}_\mathbf{X})$.

IV.1.c.3. Loi Normale contaminée

En usant de calculs assez similaires au cas gaussien, on arrive également à une expression fermée de notre algorithme. $\Psi_{Y|\mathbf{x}}^{-1}\left(\frac{\alpha}{2\alpha-1}\right)$ est alors obtenu à l'aide des itérations :

$$\begin{cases} e^{(0)} = 0 \\ e^{(k+1)} = \frac{(1-2\alpha) \sum_{i=1}^m \pi_i \theta_i^{1-d} \exp\left(-\frac{M(\mathbf{x})}{2\theta_i^2}\right) \varphi(\theta_i^{-1} e^{(k)})}{(2\alpha-1) \sum_{i=1}^m \pi_i \theta_i^{-d} \exp\left(-\frac{M(\mathbf{x})}{2\theta_i^2}\right) \Phi(\theta_i^{-1} e^{(k)}) - \alpha \sum_{i=1}^m \pi_i \theta_i^{-d} \exp\left(-\frac{M(\mathbf{x})}{2\theta_i^2}\right)} \end{cases}, \quad (\text{IV.14})$$

où Φ et φ sont respectivement la fonction de répartition et la densité de probabilité de la loi normale centrée réduite, et $M(\mathbf{x}) = (\mathbf{x} - \boldsymbol{\mu}_X)^\top \boldsymbol{\Sigma}_X^{-1} (\mathbf{x} - \boldsymbol{\mu}_X)$.

Le but de ce chapitre sera alors de proposer des approches pour estimer $\Psi_{Y|\mathbf{x}}^{-1}\left(\frac{\alpha}{2\alpha-1}\right)$. A l'instar du quantile, on proposera dans un premier temps le modèle de régression, qui s'avérera inefficace particulièrement dans le cas des niveaux extrêmes, ce qui nous poussera à proposer d'autres méthodes.

IV.2. Le modèle de régression

On considère ici un modèle analogue à la régression quantile, adapté aux expectiles. Ce modèle, simplement appelé régression expectile, présente de nombreuses similitudes, tant au niveau de la définition que des résultats.

IV.2.a. Présentation du modèle

Inspirée par la régression quantile (voir Section III.2.), la régression expectile (introduite dans [Newey and Powell, 1987]) utilise la même approche, à savoir l'approximation de l'expectile de niveau $\alpha \in]0, 1[$ de $Y|(\mathbf{X} = \mathbf{x})$ par une combinaison affine de la covariable observée $\mathbf{x}$.

$$e_\alpha^{RE}(Y|\mathbf{X} = \mathbf{x}) = \boldsymbol{\beta}^{*\top} \mathbf{x} + \beta_0^*.$$

La quantité $e_\alpha^{RE}(Y|\mathbf{X} = \mathbf{x})$ sera appelée prédicteur de régression expectile. Les coefficients $\boldsymbol{\beta}^*$ et β_0^* sont choisis de la même manière que dans l'Equation (III.2.a.), en utilisant la fonction de scoring relative à l'expectile, et introduite dans l'Equation (18).

$$(\boldsymbol{\beta}^*, \beta_0^*) = \arg \min_{\boldsymbol{\beta} \in \mathbb{R}^d, \beta_0 \in \mathbb{R}} \mathbb{E} [\mathcal{S}_{2,\alpha}(Y - \boldsymbol{\beta}^\top \mathbf{X} - \beta_0)]. \quad (\text{IV.15})$$

Dans ce cas, le problème nécessite des moments d'ordre 2, ou de manière équivalente $\mathbb{E}[\xi^2] < +\infty$ dans le cas des MEGC. Evidemment, le cas $\alpha = 1/2$ mène à la méthode

des moindres carrés, et donne $\beta^* = (\mathbb{E}[\mathbf{X}\mathbf{X}^\top] - \mathbb{E}[\mathbf{X}]\mathbb{E}[\mathbf{X}]^\top)^{-1}(\mathbb{E}[\mathbf{X}Y] - \mathbb{E}[\mathbf{X}]\mathbb{E}[Y])$ et $\beta_0^* = \mathbb{E}[Y] - \beta^\top \mathbb{E}[\mathbf{X}]$. Ces résultats sont exactement ceux que l'on peut trouver dans les approches de krigeage (voir par exemple [Ligas and Kulczycki, 2010]). Le cas $\alpha \neq 1/2$ mène à des calculs plus compliqués, et il paraît difficile d'obtenir une expression générale. Heureusement, dans notre cadre de distributions elliptiques avec la propriété de consistance, il est possible de résoudre ce problème.

IV.2.b. Résultats dans le cadre elliptique consistant

Un rapide coup d'oeil à l'Equation (IV.2.a.) laisse penser que le calcul de certains moments tronqués seront indispensable pour venir à bout du problème de minimisation. Dans cette optique, le lemme suivant est un résultat intermédiaire.

Lemme 31. Soit $(X_1, X_2) \sim \mathcal{EC}_2(\mu, \Sigma, \xi)$ un MEGC bivarié avec des paramètres :

$$\mu = (\mu_1, \mu_2)^\top, \quad \Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

Alors

$$\mathbb{E}[X_1 X_2^+] = \left(\mu_1 \mu_2 - \rho \frac{\sigma_1}{\sigma_2} \mu_2^2 \right) \Phi_{X_1} \left(\frac{\mu_2}{\sigma_2} \right) + (\mu_1 \sigma_2 + \rho \sigma_1 \mu_2) \int_{-\frac{\mu_2}{\sigma_2}}^{+\infty} u c_1 g_1(u^2) du + \rho \sigma_1 \sigma_2 \int_{-\frac{\mu_2}{\sigma_2}}^{+\infty} u^2 c_1 g_1(u^2) du. \quad (\text{IV.16})$$

Preuve : On écrit :

$$\mathbb{E}[X_1 X_2^+] = \int_0^{+\infty} x_2 f_{X_2}(x_2) \left(\int_{-\infty}^{+\infty} x_1 \frac{f_{X_1, X_2}(x_1, x_2)}{f_{X_2}(x_2)} dx_1 \right) dx_2.$$

On a $\int_{-\infty}^{+\infty} x_1 \frac{f_{X_1, X_2}(x_1, x_2)}{f_{X_2}(x_2)} dx_1 = \mathbb{E}[X_1 | X_2]$. En utilisant l'Equation (5), on obtient :

$$\mathbb{E}[X_1 X_2^+] = \int_0^{+\infty} x_2 f_{X_2}(x_2) \left(\mu_1 + \rho \frac{\sigma_1}{\sigma_2} (x_2 - \mu_2) \right) dx_2.$$

On rappelle que la densité de X_2 est $f_{X_2}(x_2) = \frac{c_1}{\sigma_2} g_1 \left(\frac{(x_2 - \mu_2)^2}{\sigma_2^2} \right)$. L'Equation (31) est obtenue en utilisant le Lemme 26 $\square$

On est maintenant capable de fournir une expression explicite du vecteur (β^*, β_0^*) , et on s'aperçoit que β^* est exactement le même que celui calculé pour le quantile.

Théorème 32 (Prédicteur de régression expectile). *Soit $\mathbf{Z} = (\mathbf{X}, Y)^\top \sim \mathcal{EC}_{d+1}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \xi)$ avec $\mathbb{E}[\xi^2] < +\infty$. Le vecteur de régression expectile $(\boldsymbol{\beta}^*, \beta_0^*)$ de $(Y|\mathbf{X} = \mathbf{x})$, qui satisfait l'Equation (IV.2.a.), est donné par :*

$$\begin{cases} \boldsymbol{\beta}^* = \\ \beta_0^* = \end{cases} \mu_Y - \boldsymbol{\Sigma}_{Y\mathbf{X}} \boldsymbol{\Sigma}_{\mathbf{X}}^{-1} \boldsymbol{\mu}_{\mathbf{X}} + \sqrt{\boldsymbol{\Sigma}_{Y|\mathbf{X}}} \Psi_Y^{-1} \left(\frac{\alpha}{2\alpha-1} \right).$$

Le prédicteur de régression expectile de niveau $\alpha \in [0, 1]$ est donné par :

$$e_\alpha^{RE}(Y|\mathbf{X} = \mathbf{x}) = \mu_{Y|\mathbf{x}} + \sqrt{\boldsymbol{\Sigma}_{Y|\mathbf{X}}} \Psi_Y^{-1} \left(\frac{\alpha}{2\alpha-1} \right).$$

De plus,

$$e_\alpha^{RE}(Y|\mathbf{X}) \sim \mathcal{EC}_1 \left(\mu_Y + \sqrt{\boldsymbol{\Sigma}_{Y|\mathbf{X}}} \Psi_Y^{-1} \left(\frac{\alpha}{2\alpha-1} \right), \boldsymbol{\Sigma}_{Y\mathbf{X}} \boldsymbol{\Sigma}_{\mathbf{X}}^{-1} \boldsymbol{\Sigma}_{\mathbf{X}Y}, \xi \right).$$

Preuve : Dans un premier temps, introduisons quelques notations : soit ρ_j le coefficient de corrélation entre X_j et la variable aléatoire $Y - \boldsymbol{\beta}^{*\top} \mathbf{X} - \beta_0^*$. En effet, $X_j \sim \mathcal{EC}_1(\mu_j, \sigma_j^2, \xi)$. De plus, $Y - \boldsymbol{\beta}^{*\top} \mathbf{X} - \beta_0^* \sim \mathcal{EC}_1(\mu_Y - \boldsymbol{\beta}^{*\top} \boldsymbol{\mu}_{\mathbf{X}} - \beta_0^*, (-\boldsymbol{\beta}^*, 1)^\top \boldsymbol{\Sigma} (-\boldsymbol{\beta}^{*\top}, 1), \xi)$. On a alors, pour tout $j = 1, \dots, d$, $(X_j, Y - \boldsymbol{\beta}^{*\top} \mathbf{X} - \beta_0^*) \sim \mathcal{EC}_2(\mathbf{m}_j, \boldsymbol{\Omega}_j, \xi)$ où :

$$\boldsymbol{\Omega}_j = \begin{pmatrix} \sigma_j^2 & \rho_j \sigma_j \sqrt{(-\boldsymbol{\beta}^*, 1)^\top \boldsymbol{\Sigma} (-\boldsymbol{\beta}^{*\top}, 1)} \\ \rho_j \sigma_j \sqrt{(-\boldsymbol{\beta}^*, 1)^\top \boldsymbol{\Sigma} (-\boldsymbol{\beta}^{*\top}, 1)} & (-\boldsymbol{\beta}^*, 1)^\top \boldsymbol{\Sigma} (-\boldsymbol{\beta}^{*\top}, 1) \end{pmatrix}.$$

On rappelle le problème de minimisation que vérifie $(\boldsymbol{\beta}^*, \beta_0^*)$:

$$\arg \min_{\boldsymbol{\beta} \in \mathbb{R}^d, \beta_0 \in \mathbb{R}} (1-\alpha) \mathbb{E}[(Y - \boldsymbol{\beta}^\top \mathbf{X} - \beta_0)^2] - (1-2\alpha) \mathbb{E}[(Y - \boldsymbol{\beta}^\top \mathbf{X} - \beta_0)^+]^2.$$

Le calcul du gradient mène au système d'équations suivant :

$$\begin{cases} (1-\alpha) \mathbb{E}[\mathbf{X}(Y - \boldsymbol{\beta}^{*\top} \mathbf{X} - \beta_0^*)] + (1-2\alpha) \mathbb{E}[\mathbf{X}(Y - \boldsymbol{\beta}^{*\top} \mathbf{X} - \beta_0^*)^+] = 0 \\ (1-\alpha) \mathbb{E}[Y - \boldsymbol{\beta}^{*\top} \mathbf{X} - \beta_0^*] + (1-2\alpha) \mathbb{E}[(Y - \boldsymbol{\beta}^{*\top} \mathbf{X} - \beta_0^*)^+] = 0 \end{cases}.$$

On sait que $Y - \boldsymbol{\beta}^{*\top} \mathbf{X} - \beta_0^* \sim \mathcal{EC}_1(\bar{\mu}, \bar{\sigma}^2, \xi)$ avec $\bar{\mu} = \mu_Y - \boldsymbol{\beta}^{*\top} \boldsymbol{\mu}_{\mathbf{X}} - \beta_0^*$ et $\bar{\sigma}^2 = (-\boldsymbol{\beta}^*, 1)^\top \boldsymbol{\Sigma} (-\boldsymbol{\beta}^*, 1) = \boldsymbol{\Sigma}_Y - 2\boldsymbol{\beta}^{*\top} \boldsymbol{\Sigma}_{Y\mathbf{X}} + \boldsymbol{\beta}^{*\top} \boldsymbol{\Sigma}_{\mathbf{X}} \boldsymbol{\beta}^*$. En utilisant les Lemmes 26 et 31 on obtient le système suivant, $\forall j \in \{1, \dots, d\}$

$$\begin{cases} (1-\alpha) \bar{\mu} + (1-2\alpha) \bar{\mu} \Phi_Y \left(\frac{\bar{\mu}}{\bar{\sigma}} \right) + (1/2 - \alpha) \bar{\sigma} \int_{\bar{\mu}/\bar{\sigma}}^{+\infty} u c_1 g_1(u^2) du = 0 \\ (1-2\alpha) \left(\mu_j \bar{\mu} - \rho_j \frac{\sigma_j}{\bar{\sigma}} \bar{\mu}^2 \right) \Phi_Y \left(\frac{\bar{\mu}}{\bar{\sigma}} \right) + (1/2 - \alpha) \left(\mu_j \bar{\sigma} - \rho_j \sigma_j \bar{\sigma} \right) \int_{\bar{\mu}/\bar{\sigma}}^{+\infty} u c_1 g_1(u^2) du \\ \quad + (1-2\alpha) \rho_j \frac{\sigma_j}{\bar{\sigma}} \int_0^{+\infty} u^2 f_Y(u) du = 0 \end{cases}.$$

En combinant ces deux équations, il reste

$$\begin{aligned} & - (1 - 2\alpha)\rho_j \frac{\sigma_j}{\bar{\sigma}} \bar{\mu}^2 \Phi_Y \left(\frac{\bar{\mu}}{\bar{\sigma}} \right) - (1/2 - \alpha)\rho_j \sigma_j \bar{\sigma} \int_{\bar{\mu}/\bar{\sigma}}^{+\infty} u c_1 g_1(u^2) du \\ & + (1 - 2\alpha)\rho_j \frac{\sigma_j}{\bar{\sigma}} \int_0^{+\infty} u^2 f_Y(u) du + (1 - \alpha)\rho_j \sigma_j \bar{\sigma} = 0. \end{aligned}$$

Il paraît évident que $\rho_j = 0$. On a déjà vu, dans la preuve du Théorème 9 que $\rho_j = 0, \forall j$ était équivalent à $\beta^* = \Sigma_X^{-1} \Sigma_{XY}$.

β_0^* satisfait :

$$\beta_0^* = \arg \min_{\beta_0 \in \mathbb{R}} \mathbb{E} [\mathcal{S}_{2,\alpha}(Y - \beta^{*\top} \mathbf{X} - \beta_0)].$$

Finalement, β_0^* est l'expectile de niveau α de la variable aléatoire $Y - \beta^{*\top} \mathbf{X}$. Toujours dans la preuve du Théorème 9 on a vu que $Y - \beta^{*\top} \mathbf{X} \sim \mathcal{EC}_1(\mu_Y - \Sigma_{YX} \Sigma_X^{-1} \mu_X, \Sigma_{Y|X}, \xi)$. Dès lors, en utilisant la formule de l'expectile de l'Equation (27), on obtient :

$$\beta_0^* = \mu_Y - \Sigma_{YX} \Sigma_X^{-1} \mu_X + \sqrt{\Sigma_{Y|X}} \Psi_Y^{-1} \left(\frac{\alpha}{2\alpha - 1} \right).$$

On peut maintenant exprimer notre prédicteur de régression expectile de Y sachant $\mathbf{X} = \mathbf{x}$:

$$e_\alpha^{RE}(Y|\mathbf{X} = \mathbf{x}) = \beta^{*\top} \mathbf{x} + \beta_0^* = \mu_Y + \Sigma_{YX} \Sigma_X^{-1} (\mathbf{x} - \mu_X) + \sqrt{\Sigma_{Y|X}} \Psi_Y^{-1} \left(\frac{\alpha}{2\alpha - 1} \right).$$

On reconnait sur la gauche l'expression de $\mu_{Y|\mathbf{x}}$ donnée dans l'Equation (5).

Comme $e_\alpha^{RE}(Y|\mathbf{X}) = \mu_Y + \Sigma_{YX} \Sigma_X^{-1} (\mathbf{X} - \mu_X) + \sqrt{\Sigma_{Y|X}} \Psi_Y^{-1} \left(\frac{\alpha}{2\alpha - 1} \right)$, avec $\mathbf{X} \sim \mathcal{E}_d(\mu_X, \Sigma_X, \xi)$. On obtient, à l'aide de la Proposition 4

$$e_\alpha^{RE}(Y|\mathbf{X}) \sim \mathcal{EC}_1 \left(\mu_Y + \sqrt{\Sigma_{Y|X}} \Psi_Y^{-1} \left(\frac{\alpha}{2\alpha - 1} \right), (\Sigma_{YX} \Sigma_X^{-1}) \Sigma_X (\Sigma_X^{-1} \Sigma_{XY}), \xi \right).$$

D'où le résultat. □

IV.2.c. Commentaires

Les conclusions à tirer du modèle de régression dans le cas de l'expectile sont assez similaires à celles du quantile. Considérons alors premièrement le modèle linéaire avec $\beta_0^* = 0$. On voit alors clairement que $e_\alpha^{RE}(Y|\mathbf{X}) = \beta^{*\top} \mathbf{X}$ est indépendant de α . En dehors du cas $\alpha = 1/2$, le modèle n'est donc pas recommandé. De plus, si $\Sigma_{XY} = \mathbf{0}_{\mathbb{R}^d}$, le

prédicteur $e_\alpha^{RE}(Y|\mathbf{X})$ vaut 0. On ajoute donc le terme β_0^* et passe au modèle affine. Comme pour le quantile, on s'aperçoit que l'expectile conditionnel théorique $e_\alpha(Y|\mathbf{X} = \mathbf{x})$ de l'Equation (IV.1.c.) et le prédicteur $e_\alpha^{RE}(Y|\mathbf{X} = \mathbf{x})$ ont une expression similaire qui ne diffère que par l'expectile conditionnel $\Psi_{Y|\mathbf{x}}^{-1}$ remplacé par l'inconditionnel Ψ_Y^{-1} . Le modèle de régression est donc toujours tout à fait adapté dans le cas gaussien, mais peut mener à de mauvais résultats dans les autres cas. L'illustration faite pour le quantile est aussi parlante dans ce cas : prenons $\Sigma_{\mathbf{X}Y} = \mathbf{0}_{\mathbb{R}^d}$. Le prédicteur $e_\alpha^{RE}(Y|\mathbf{X} = \mathbf{x})$ renvoie l'expectile "indépendant" $\sqrt{\Sigma_Y}\Psi_Y^{-1}(\alpha)$, lorsque l'expectile théorique vaut l'expectile "décorrélé" $\sqrt{\Sigma_Y}\Psi_{Y|\mathbf{x}}^{-1}(\alpha)$. La différence entre indépendance et non-corrélation une fois quitté le cadre gaussien pose donc également problème pour l'expectile.

Comme nous l'avons fait dans la Section III.2.c. pour le quantile, comparons numériquement nos fonctions Ψ_Y^{-1} et $\Psi_{Y|\mathbf{x}}^{-1}$. Pour cela, intéressons nous aux distributions Normale contaminée et Student. Les algorithmes (IV.1.c.3.) et (IV.1.c.2.) nous permettent de calculer les fonctions $\Psi_{Y|\mathbf{x}}^{-1}$, tandis que l'algorithme (30) donne Ψ_Y^{-1} pour chaque exemple. En choisissant comme paramètres $M(\mathbf{x}) = 1$, $d = 5$, $\nu = 3$, $m = 2$, $\theta_1 = 4$, $\theta_2 = 1$, $p_1 = 0.5$ et $p_2 = 0.5$, on obtient les graphiques de la Figure IV.2. On voit clairement que si dans le cas de la loi Normale contaminée, les deux fonctions ont l'air assez proches, ce n'est pas le cas de la loi de Student. Rappelons que la queue de distribution est légère pour le premier exemple, et lourde pour le second.

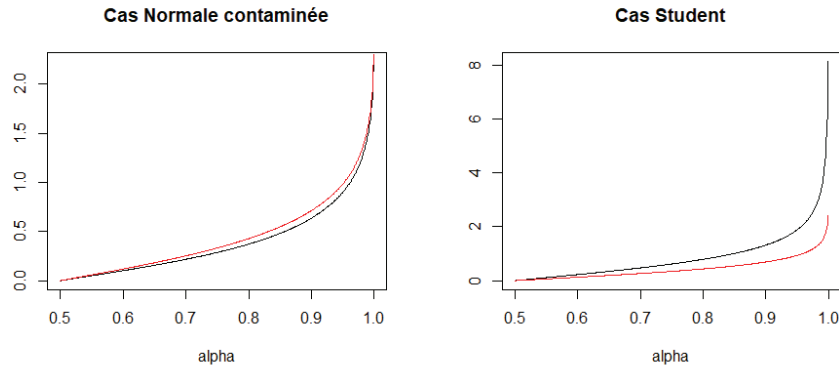


FIGURE IV.2 – Comparaison des fonctions $\Psi_Y^{-1}(\alpha)$ (en noir) et $\Psi_{Y|\mathbf{x}}^{-1}(\alpha)$ (en rouge) pour $\alpha \in]\frac{1}{2}, 1[$ dans les cas Normale contaminée et Student.

À l'instar de la régression quantile, la régression expectile présente donc de nombreuses limites. Si le cas $\alpha = 1/2$, équivalent alors à la régression quantile (dans le cas elliptique), est performant et d'ailleurs utilisé dans les problématiques de krigeage, les autres cas laissent à désirer, particulièrement lorsque α devient proche de 0 ou de 1. On propose alors dans la section suivante des prédicteurs adaptés aux niveaux extrêmes, en corri-

geant notamment la différence d'indice de queue de distribution des lois conditionnelle et inconditionnelle dans un cadre de variation régulière.

IV.3. Expectiles extrêmes

Cette section est inspirée de la Section III.3. en remplaçant le quantile par l'expectile. La structure est donc la même, à savoir une approche méthodologique puis statistique des expectiles extrêmes.

IV.3.a. Approche méthodologique

On propose dans cette section une relation asymptotique entre les fonctions $\Psi_Y^{-1}(\frac{\alpha}{2\alpha-1})$ et $\Psi_{Y|\mathbf{x}}^{-1}(\frac{\alpha}{2\alpha-1})$. L'approche utilisée est alors la même que pour les fonctions $\Phi_Y^{-1}(\alpha)$ et $\Phi_{Y|\mathbf{x}}^{-1}(\alpha)$ du chapitre sur les quantiles. Il faut alors donner dans un premier temps une équivalence entre $1 - \Psi_Y$ et $1 - \Psi_{Y|\mathbf{x}}$. Là encore, ces deux fonctions ne sont pas des φ -fonctions au sens de la Définition I.1. On fait alors l'hypothèse suivante sur une transformation de ces fonctions.

Hypothèse 4. *Il existe $0 < \ell^e(\mathbf{x}) < +\infty$ et $\eta^e \in \mathbb{R}$ tels que :*

$$\lim_{t \rightarrow +\infty} \frac{\bar{\Psi}_{Y|\mathbf{x}}(t)}{\bar{\Psi}_Y(t^{\eta^e})} = \ell^e(\mathbf{x}), \quad (\text{IV.17})$$

où $\bar{\Psi} = 1 - \Psi$. De plus, on définit Ω et $\Omega_{\mathbf{x}}$ comme

$$\begin{cases} \Omega(t) &= -\frac{1}{\ell^e(\mathbf{x})\bar{\Psi}_Y(t^{\eta^e})} \\ \Omega_{\mathbf{x}}(t) &= -\frac{1}{\bar{\Psi}_{Y|\mathbf{x}}(t)} \end{cases}.$$

Alors Ω^{-1} ou $\Omega_{\mathbf{x}}^{-1}$ appartient à la classe K_c .

On rappelle

$$\begin{cases} \Psi_Y(t) = & \Phi_Y(x) + \frac{1}{t} \int_t^{+\infty} u c_1 g_1(u^2) du, \\ \Psi_{Y|\mathbf{x}}(t) = & \Phi_{Y|\mathbf{x}}(t) + \frac{1}{t} \int_t^{+\infty} u c_{\mathbf{x}} g_{d+1}(M(\mathbf{x}) + u^2) du, \end{cases}$$

avec $c_{\mathbf{x}} = \frac{c_{d+1}}{c_d g_d(M(\mathbf{x}))}$ (voir l'Equation (5)). Alors les coefficients η^e et $\ell^e(\mathbf{x})$ satisfont :

$$\lim_{t \rightarrow +\infty} \frac{t^{\eta^e - 1} \int_t^{+\infty} u c_{\mathbf{x}} g_{d+1}(M(\mathbf{x}) + u^2) du}{\gamma \int_{t^{\eta^e}}^{+\infty} u c_1 g_1(u^2) du} = \ell^e(\mathbf{x}). \quad (\text{IV.18})$$

Avec l'Equation (4), on a, sous l'Hypothèse 4 l'équivalence suivante :

$$\bar{\Psi}_{Y|\mathbf{x}}(t) \underset{t \rightarrow +\infty}{\sim} \ell^e(\mathbf{x}) \bar{\Psi}_Y(t^{\eta^e}). \quad (\text{IV.19})$$

En utilisant alors la Définition 11 et l'Hypothèse 4 on est capable d'inverser l'équivalence (IV.3.a.), d'où la proposition suivante.

Proposition 33 (Equivalence des expectiles conditionnel et inconditionnel). *Sous l'Hypothèse 4, on a :*

$$\Psi_{Y|\mathbf{x}}^{-1} \left(\frac{\alpha}{2\alpha - 1} \right) \underset{\alpha \rightarrow 1}{\sim} \left[\Psi_Y^{-1} \left(1 - \frac{\alpha - 1}{(2\alpha - 1)\ell^e(\mathbf{x})} \right) \right]^{\frac{1}{\eta^e}}. \quad (\text{IV.20})$$

Preuve : Des calculs rapides donnent $\Omega^{-1}(t) = \left[\Psi_Y^{-1} \left(1 + \frac{1}{\ell^e(\mathbf{x})t} \right) \right]^{\frac{1}{\eta^e}}$ et $\Omega_{\mathbf{x}}^{-1}(t) = \Psi_{Y|\mathbf{x}}^{-1} \left(1 + \frac{1}{t} \right)$. Grâce à l'Hypothèse 4 on s'aperçoit que $\Omega(t) \underset{t \rightarrow \infty}{\sim} \Omega_{\mathbf{x}}(t)$, avec Ω et $\Omega_{\mathbf{x}}$ étant des φ -fonctions. De plus, Ω^{-1} ou $\Omega_{\mathbf{x}}^{-1}$ appartient à la classe K_c . On peut alors appliquer le Lemme 10. D'où l'équivalence $\Omega^{-1}(t) \underset{t \rightarrow \infty}{\sim} \Omega_{\mathbf{x}}^{-1}(t)$. En d'autres termes,

$$\Psi_{Y|\mathbf{x}}^{-1} \left(1 + \frac{1}{t} \right) \underset{t \rightarrow \infty}{\sim} \left[\Psi_Y^{-1} \left(1 + \frac{1}{\ell^e(\mathbf{x})t} \right) \right]^{\frac{1}{\eta^e}}.$$

En effectuant le changement de variable $1 + \frac{1}{t} = \frac{\alpha}{2\alpha - 1}$, on obtient le résultat (33). $\square$

Notons que si $\Psi^{-1} \left(\frac{\alpha}{2\alpha - 1} \right)$ correspond à un expectile de niveau α , $\Psi^{-1} \left(1 - \frac{\alpha - 1}{(2\alpha - 1)\ell^e(\mathbf{x})} \right)$ est un expectile de niveau $\frac{(2\alpha - 1)\ell^e(\mathbf{x}) + 1 - \alpha}{(2\alpha - 1)\ell^e(\mathbf{x}) + 2(1 - \alpha)}$. Evidemment, si $\ell^e(\mathbf{x}) = 1$, ce niveau est α . On définit alors deux prédicteurs $e_{\alpha\uparrow}^E(Y|\mathbf{X} = \mathbf{x})$ et $e_{\alpha\downarrow}^E(Y|\mathbf{X} = \mathbf{x})$. On montre par la suite qu'ils sont équivalents à l'expectile théorique $e_{\alpha}(Y|\mathbf{X} = \mathbf{x})$ lorsque $\alpha \rightarrow 1$ et $\alpha \rightarrow 0$, respectivement.

Définition 20 (Prédicteurs d'expectiles extrêmes). *On définit :*

$$\begin{cases} e_{\alpha\uparrow}^E(Y|\mathbf{X} = \mathbf{x}) &= \mu_{Y|\mathbf{x}} + \sqrt{\Sigma_{Y|\mathbf{X}}} \left[\Psi_Y^{-1} \left(1 - \frac{\alpha-1}{(2\alpha-1)\ell^e(\mathbf{x})} \right) \right]^{\frac{1}{\eta^e}} \\ e_{\alpha\downarrow}^E(Y|\mathbf{X} = \mathbf{x}) &= \mu_{Y|\mathbf{x}} - \sqrt{\Sigma_{Y|\mathbf{X}}} \left[\Psi_Y^{-1} \left(1 - \frac{\alpha}{(2\alpha-1)\ell^e(\mathbf{x})} \right) \right]^{\frac{1}{\eta^e}} \end{cases}.$$

Théorème 34 (Equivalence des prédicteurs). *Soit $\mathbf{Z} = (\mathbf{X}, Y)^\top \sim \mathcal{EC}_{d+1}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \xi)$. Sous l'Hypothèse [4](#),*

$$\begin{cases} e_{\alpha\uparrow}^E(Y|\mathbf{X} = \mathbf{x}) &\underset{\alpha \rightarrow 1}{\sim} e_\alpha(Y|\mathbf{X} = \mathbf{x}) \\ e_{\alpha\downarrow}^E(Y|\mathbf{X} = \mathbf{x}) &\underset{\alpha \rightarrow 0}{\sim} e_\alpha(Y|\mathbf{X} = \mathbf{x}) \end{cases}.$$

Preuve : On rappelle la relation $e_\alpha(Y|\mathbf{X} = \mathbf{x}) = \mu_{Y|\mathbf{x}} + \sqrt{\Sigma_{Y|\mathbf{X}}} \Psi_{Y|\mathbf{x}}^{-1} \left(\frac{\alpha}{2\alpha-1} \right)$. Avec l'Equation [33](#), la première équivalence est immédiate. Concernant la seconde, i.e. quand $\alpha \rightarrow 0$, on utilise les propriétés de symétrie des distributions elliptiques : on sait que $e_\alpha(Y|\mathbf{X} = \mathbf{x}) = \mu_{Y|\mathbf{x}} - \sqrt{\Sigma_{Y|\mathbf{X}}} \Psi_{Y|\mathbf{x}}^{-1} \left(\frac{1-\alpha}{1-2\alpha} \right)$. Ou, si $\alpha' = 1 - \alpha$, on a $e_\alpha(Y|\mathbf{X} = \mathbf{x}) = \mu_{Y|\mathbf{x}} - \sqrt{\Sigma_{Y|\mathbf{X}}} \Psi_{Y|\mathbf{x}}^{-1} \left(\frac{\alpha'}{2\alpha'-1} \right)$, avec $\alpha' \rightarrow 1$. Avec l'Equation [33](#), on obtient

$$e_\alpha(Y|\mathbf{X} = \mathbf{x}) \underset{\alpha' \rightarrow 1}{\sim} \mu_{Y|\mathbf{x}} - \sqrt{\Sigma_{Y|\mathbf{X}}} \left[\Psi_Y^{-1} \left(1 - \frac{\alpha' - 1}{(2\alpha' - 1)\ell^e(\mathbf{x})} \right) \right]^{\frac{1}{\eta^e}}.$$

Remplaçons α' par $1 - \alpha$ pour la seconde équivalence. □

Comme pour le quantile, on a fourni des prédicteurs adaptés aux niveaux extrêmes. On vérifie maintenant quelles distributions satisfont l'Hypothèse [4](#) parmi celles que nous avons introduites dans le Chapitre [II](#). Une fois encore, la distribution de Laplace est la seule qui ne la vérifie pas. On donne en revanche les coefficients associés pour les autres.

Proposition 35 (Exemples). *Les distributions Normale, Normale contaminée, Student, et Slash satisfont l'Hypothèse [4](#) avec des coefficients η^e et $\ell^e(\mathbf{x})$ donnés dans la Table [IV.1](#).*

Distribution	η^e	$\ell^e(\mathbf{x})$
Normale	1	1
Normale contaminée	1	$\frac{\max(\theta_1, \dots, \theta_m)^{-d} \exp\left(-\frac{M(\mathbf{x})}{2 \max(\theta_1, \dots, \theta_m)^2}\right)}{\sum_{k=1}^m \pi_k \theta_k^{-d} \exp\left(-\frac{M(\mathbf{x})}{2 \theta_k^2}\right)}$
Student, $\nu > 0$	$\frac{d}{\nu} + 1$	$\frac{\Gamma\left(\frac{\nu+d+1}{2}\right)\Gamma\left(\frac{\nu}{2}\right)}{\Gamma\left(\frac{\nu+d}{2}\right)\Gamma\left(\frac{\nu+1}{2}\right)} \left(1 + \frac{M(\mathbf{x})}{\nu}\right)^{\frac{d+\nu}{2}} \frac{\nu^{\frac{d}{2}+1}}{\nu+d} \frac{\nu-1}{\nu+d-1}$
Slash, $a > 0$	$\frac{d}{a} + 1$	$\frac{2^{1-\frac{a}{2}}(a-1)\Gamma\left(\frac{d+1+a}{2}\right)M(\mathbf{x})^{\frac{d+a}{2}}}{a\left(\frac{d}{a}+1\right)(d+a-1)\Gamma\left(\frac{d+a}{2}\right)\Gamma\left(\frac{1+a}{2}\right)\chi_{d+a}^2(M(\mathbf{x}))}$
Laplace, $\lambda > 0$	$\times$	$\times$

TABLE IV.1 – Coefficients η^e et $\ell^e(\mathbf{x})$ pour quelques exemples de distributions elliptiques, où $M(\mathbf{x}) = (x - \boldsymbol{\mu}_X)^\top \boldsymbol{\Sigma}_X^{-1}(\mathbf{x} - \boldsymbol{\mu}_X)$.

Preuve : Loi Normale : Premièrement, on doit calculer la limite de l'Equation (IV.3.a.), avec $g_n(t) = \exp\left(-\frac{t}{2}\right)$, $\forall n \in \mathbb{N}$, $c_1 = \frac{1}{\sqrt{2\pi}}$ et $c_x = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{M(\mathbf{x})}{2}\right)$. On considère alors

$$\lim_{t \rightarrow +\infty} \frac{t^{\eta^e-1} \int_t^{+\infty} u \exp\left(-\frac{u^2}{2}\right) du}{\gamma \int_{t^{\eta^e}}^{+\infty} u \exp\left(-\frac{u^2}{2}\right) du}.$$

Si on prend $\eta^e = 1$, on obtient directement la limite $\ell^e(\mathbf{x})$ égale à 1.

Maintenant, on doit montrer que Ω^{-1} appartient à la classe K_c . Pour cela, on s'intéresse à la limite

$$\lim_{\substack{t \rightarrow +\infty \\ \lambda \rightarrow 1}} \frac{\Omega^{-1}(\lambda t)}{\Omega^{-1}(t)}.$$

On a vu précédemment que $\Omega^{-1}(t) = \left[\Psi_Y^{-1}\left(1 + \frac{1}{\ell^e(\mathbf{x})t}\right)\right]^{\frac{1}{\eta^e}}$. On doit alors considérer la limite de $\frac{\Psi_Y^{-1}\left(1 + \frac{1}{\ell^e(\mathbf{x})\lambda t}\right)}{\Psi_Y^{-1}\left(1 + \frac{1}{\ell^e(\mathbf{x})t}\right)}$. Avec $\delta = \lambda - 1$ et $y = \frac{1}{t}$, cette limite devient

$$\lim_{(y, \delta) \rightarrow (0, 0)} \frac{\Psi_Y^{-1}\left(1 + \frac{y}{\ell^e(\mathbf{x})(\delta+1)}\right)}{\Psi_Y^{-1}\left(1 + \frac{y}{\ell^e(\mathbf{x})}\right)}.$$

Dans le but de calculer cette limite, on passe en coordonnées polaires, i.e. on prend $\delta = r \cos(\theta)$, $y = r \sin(\theta)$, et vérifie si la limite quand $r \rightarrow 0$ existe et ne dépend pas de θ . D'où

$$f(\theta) = \lim_{r \rightarrow 0} \frac{\Psi_Y^{-1}\left(1 + \frac{r \sin(\theta)}{\ell^e(\mathbf{x})(r \cos(\theta)+1)}\right)}{\Psi_Y^{-1}\left(1 + \frac{r \sin(\theta)}{\ell^e(\mathbf{x})}\right)}.$$

Les deux termes tendent vers 0, on applique alors la règle de l'Hôpital, et

$$f(\theta) = \lim_{r \rightarrow 0} \frac{\frac{\sin(\theta)(r \cos(\theta)+1) - r \sin(\theta) \cos(\theta)}{(r \cos(\theta)+1)^2}}{\sin(\theta)} \frac{\Psi'_Y \left(\Psi_Y^{-1} \left(1 + \frac{r \sin(\theta)}{\ell^e(\mathbf{x})} \right) \right)}{\Psi'_Y \left(\Psi_Y^{-1} \left(1 + \frac{r \sin(\theta)}{\ell^e(\mathbf{x})(r \cos(\theta)+1)} \right) \right)}.$$

Le ratio sur la droite tend vers 1. De plus, un rapide calcul donne $\Psi'_Y(t) = -\frac{1}{2t^2} \int_t^{+\infty} u f_Y(|u|) du$.

Dans le cas gaussien, cette valeur vaut $-\frac{1}{t^2} \varphi(t)$, et $\ell^e(\mathbf{x}) = 1$. On obtient alors

$$f(\theta) = \lim_{r \rightarrow 0} \frac{\left[\Psi_Y^{-1} \left(1 + \frac{r \sin(\theta)}{r \cos(\theta)+1} \right) \right]^2}{\left[\Psi_Y^{-1} (1 + r \sin(\theta)) \right]^2} \frac{\varphi(\Psi_Y^{-1}(1 + r \sin(\theta)))}{\varphi\left(\Psi_Y^{-1} \left(1 + \frac{r \sin(\theta)}{r \cos(\theta)+1} \right)\right)}.$$

Le ratio sur la gauche tend clairement vers $f(\theta)^2$. Concernant celui sur la droite, on applique une fois encore la règle de l'Hôpital. Comme $\varphi'(t) = -t\varphi(t)$, cela conduit à

$$f(\theta) = f(\theta)^2 \lim_{r \rightarrow 0} \frac{\left[\Psi_Y^{-1} (1 + r \sin(\theta)) \right]^3}{\left[\Psi_Y^{-1} \left(1 + \frac{r \sin(\theta)}{r \cos(\theta)+1} \right) \right]^3} = \frac{1}{f(\theta)} = 1.$$

Loi Student : Dans un premier temps, on calcule la limite de l'Equation (IV.3.a.) avec

$g_n(t) = \left(1 + \frac{t}{\nu}\right)^{-\frac{n+\nu}{2}}$, $\forall n \in \mathbb{N}$, $c_1 = \frac{\Gamma(\frac{1+\nu}{2})}{\Gamma(\frac{\nu}{2})} \frac{1}{\sqrt{\nu\pi}}$ et $c_{\mathbf{x}} = \frac{\Gamma(\frac{d+1+\nu}{2})}{\Gamma(\frac{d+\nu}{2})} \frac{1}{\sqrt{\nu\pi}} \left(1 + \frac{M(\mathbf{x})}{\nu}\right)^{\frac{d+\nu}{2}}$. Il reste

$$\lim_{t \rightarrow +\infty} \frac{t^{\eta^e-1} \int_t^{+\infty} u c_{\mathbf{x}} \left(1 + \frac{M(\mathbf{x})+u^2}{\nu}\right)^{-\frac{d+1+\nu}{2}} du}{\eta^e \int_{t^{\eta^e}}^{+\infty} u c_1 \left(1 + \frac{u^2}{\nu}\right)^{-\frac{1+\nu}{2}} du} = \ell^e(\mathbf{x}).$$

Les deux fonctions sous les intégrales sont polynomiales, et donc simples à intégrer, d'où

$$\lim_{t \rightarrow +\infty} \frac{t^{\eta^e-1} c_{\mathbf{x}} \frac{\nu}{\nu+d-1} \left(1 + \frac{M(\mathbf{x})+t^2}{\nu}\right)^{\frac{1-d-\nu}{2}}}{\eta^e c_1 \frac{\nu}{\nu-1} \left(1 + \frac{t^{2\eta^e}}{\nu}\right)^{\frac{1-\nu}{2}}} = \ell^e(\mathbf{x}).$$

Si on prend alors $\eta^e = \frac{d+\nu}{\nu}$, et remplace c_1 et $c_{\mathbf{x}}$ par leurs valeurs, on obtient

$$\frac{\Gamma\left(\frac{\nu+d+1}{2}\right) \Gamma\left(\frac{\nu}{2}\right)}{\Gamma\left(\frac{\nu+d}{2}\right) \Gamma\left(\frac{\nu+1}{2}\right)} \left(1 + \frac{M(\mathbf{x})}{\nu}\right)^{\frac{d+\nu}{2}} \frac{\nu^{\frac{d}{2}+1}}{\nu+d} \frac{\nu-1}{\nu+d-1} = \ell^e(\mathbf{x}).$$

On doit maintenant vérifier la seconde partie de l'Hypothèse 4 i.e. si la fonction Ω^{-1} appartient à la classe K_c . Calculons la limite

$$\lim_{\substack{t \rightarrow +\infty \\ \lambda \rightarrow 1}} \frac{\Omega^{-1}(\lambda t)}{\Omega^{-1}(t)}.$$

En opérant le même changement de variable que dans le cas gaussien, on considère

$$f(\theta) = \lim_{r \rightarrow 0} \frac{\Psi'_Y \left(\Psi_Y^{-1} \left(1 + \frac{r \sin(\theta)}{\ell^e(\mathbf{x})} \right) \right)}{\Psi'_Y \left(\Psi_Y^{-1} \left(1 + \frac{r \sin(\theta)}{\ell^e(\mathbf{x})(r \cos(\theta) + 1)} \right) \right)}.$$

Dans le cas Student, $\Psi_Y(t)' = -\frac{1}{t^2} \frac{\Gamma(\frac{1+\nu}{2})}{\Gamma(\frac{\nu}{2})} \frac{1}{\sqrt{\nu\pi}} \frac{\nu}{\nu-1} \left(1 + \frac{t^2}{\nu} \right)^{\frac{1-\nu}{2}}$. Il reste alors uniquement

$$f(\theta) = \lim_{r \rightarrow 0} \frac{\Psi_Y^{-1} \left(1 + \frac{r \sin(\theta)}{\ell^e(\mathbf{x})(r \cos(\theta) + 1)} \right)^2}{\Psi_Y^{-1} \left(1 + \frac{r \sin(\theta)}{\ell^e(\mathbf{x})} \right)^2} \left(\frac{1 + \frac{\Psi_Y^{-1} \left(1 + \frac{r \sin(\theta)}{\ell^e(\mathbf{x})} \right)^2}{\nu}}{1 + \frac{\Psi_Y^{-1} \left(1 + \frac{r \sin(\theta)}{\ell^e(\mathbf{x})(r \cos(\theta) + 1)} \right)^2}{\nu}} \right)^{\frac{1-\nu}{2}}.$$

Sur la gauche, le ratio tend vers $f(\theta)^2$. Le terme sur la droite vaut asymptotiquement

$$\lim_{r \rightarrow 0} \left(\frac{\Psi_Y^{-1} \left(1 + \frac{r \sin(\theta)}{\ell^e(\mathbf{x})} \right)^2}{\Psi_Y^{-1} \left(1 + \frac{r \sin(\theta)}{\ell^e(\mathbf{x})(r \cos(\theta) + 1)} \right)^2} \right)^{\frac{1-\nu}{2}} = f(\theta)^{\nu-1}. \text{ Finalement, on a la relation}$$

$$f(\theta) = f(\theta)^{\nu+1}, \nu > 0.$$

D'où $f(\theta) = 1$.

Loi Normale contaminée : Calculons la limite de l'Equation (IV.3.a.). Dans ce cas, $g_d(t) = \sum_{k=1}^m \pi_k \theta_k^{-d} \exp\left(-\frac{t}{2\theta_k^2}\right)$, $c_1 = \frac{1}{\sqrt{2\pi}}$ et $c_{\mathbf{x}} = \frac{1}{\sqrt{2\pi} \sum_{k=1}^m \pi_k \theta_k^{-d} \exp\left(-\frac{M(\mathbf{x})}{2\theta_k^2}\right)}$. Le calcul de la limite donne

$$\lim_{t \rightarrow +\infty} \frac{t^{\eta^e-1} \sum_{k=1}^m \theta_k^{1-d} c_{\mathbf{x}} \exp\left(-\frac{M(\mathbf{x})}{2\theta_k^2}\right) \exp\left(-\frac{t^2}{2\theta_k^2}\right)}{\eta^e \sum_{k=1}^m \pi_k \theta_k c_1 \exp\left(-\frac{t^2 \eta^e}{2\theta_k^2}\right)}.$$

A partir de maintenant, considérons que $\eta^e = 1$. Asymptotiquement, on s'intéresse uniquement aux termes $\exp\left(-\frac{t^2}{2\theta_k^2}\right)$ avec le plus grand coefficient $-\frac{1}{2\theta_k^2}$, i.e. le plus grand θ_k . Soit k^* tel que $\theta_{k^*} = \max\{\theta_1, \dots, \theta_m\}$. On a

$$\ell^e(\mathbf{x}) = \lim_{t \rightarrow +\infty} \frac{c_{\mathbf{x}} \pi_{k^*} \theta_{k^*}^{1-d} \exp\left(-\frac{M(\mathbf{x})}{2\theta_{k^*}^2}\right) \exp\left(-\frac{t^2}{2\theta_{k^*}^2}\right)}{c_1 \pi_{k^*} \theta_{k^*} \exp\left(-\frac{t^2}{2\theta_{k^*}^2}\right)}.$$

Après simplifications, et en remplaçant c_1 et $c_{\mathbf{x}}$ par leurs valeurs, on obtient

$$\ell^e(\mathbf{x}) = \frac{\theta_{k^*}^{-d} \exp\left(-\frac{M(\mathbf{x})}{2\theta_{k^*}^2}\right)}{\sum_{k=1}^m \pi_k \theta_k^{-d} \exp\left(-\frac{M(\mathbf{x})}{2\theta_k^2}\right)}.$$

On vient de vérifier la première partie de l'Hypothèse [4](#). Vérifions la seconde en considérant encore une fois la limite

$$f(\theta) = \lim_{r \rightarrow 0} \frac{\Psi'_Y \left(\Psi_Y^{-1} \left(1 + \frac{r \sin(\theta)}{\ell^e(\mathbf{x})} \right) \right)}{\Psi'_Y \left(\Psi_Y^{-1} \left(1 + \frac{r \sin(\theta)}{\ell^e(\mathbf{x})(r \cos(\theta)+1)} \right) \right)}.$$

Dans ce cas, on a la relation $\Psi'_Y(t) = -t^{-2} \sum_{k=1}^m \pi_k \theta_k \varphi(x/\theta_k)$. En remplaçant Ψ'_Y dans la limite précédente, on obtient

$$f(\theta) = \lim_{r \rightarrow 0} \frac{\Psi_Y^{-1} \left(1 + \frac{r \sin(\theta)}{\ell^e(\mathbf{x})(r \cos(\theta)+1)} \right)^2}{\Psi_Y^{-1} \left(1 + \frac{r \sin(\theta)}{\ell^e(\mathbf{x})} \right)^2} \frac{\sum_{k=1}^m \pi_k \theta_k \varphi \left(\theta_k^{-1} \Psi_Y^{-1} \left(1 + \frac{r \sin(\theta)}{\ell^e(\mathbf{x})} \right) \right)}{\sum_{k=1}^m \pi_k \theta_k \varphi \left(\theta_k^{-1} \Psi_Y^{-1} \left(1 + \frac{r \sin(\theta)}{\ell^e(\mathbf{x})(r \cos(\theta)+1)} \right) \right)}.$$

Le terme sur la gauche tend vers $f(\theta)^2$. Pour calculer la limite du second ratio, on s'intéresse uniquement aux termes dominants, i.e. les termes avec les plus grands $-\frac{1}{2\theta_k^2}$, soit les plus grands θ_k . Soit k^* tel que $\theta_{k^*} = \max\{\theta_1, \dots, \theta_m\}$. Il reste

$$f(\theta) = f(\theta)^2 \lim_{r \rightarrow 0} \frac{\varphi \left(\theta_{k^*}^{-1} \Psi_Y^{-1} \left(1 + \frac{r \sin(\theta)}{\ell^e(\mathbf{x})} \right) \right)}{\varphi \left(\theta_{k^*}^{-1} \Psi_Y^{-1} \left(1 + \frac{r \sin(\theta)}{\ell^e(\mathbf{x})(r \cos(\theta)+1)} \right) \right)}.$$

Dans le cas gaussien, on a montré de manière similaire que $f(\theta)$ était égal à 1.

Loi Slash : On considère la limite de l'Equation [\[IV.3.a.\]](#), avec $g_n(t) = \frac{\chi_{n+a}^2(t)}{t^{\frac{n+a}{2}}}$, $\forall n \in \mathbb{N}^*$,

$c_{\mathbf{x}} = \frac{1}{\sqrt{\pi}} \frac{\Gamma(\frac{d+1+a}{2})}{\Gamma(\frac{d+a}{2})} \frac{M(\mathbf{x})^{\frac{d+a}{2}}}{\chi_{d+a}^2(M(\mathbf{x}))}$ et $c_1 = \frac{2^{\frac{a}{2}-1} a \Gamma(\frac{1+a}{2})}{\sqrt{\pi}}$. Par intégration par partie, on obtient

$$\ell^e(\mathbf{x}) = \lim_{t \rightarrow +\infty} \frac{t^{\eta^e-1} \frac{c_{\mathbf{x}}}{d+a-1} \left[(M(\mathbf{x}) + t^2)^{\frac{1-d-a}{2}} \chi_{d+1+a}^2(M(\mathbf{x}) + t^2) + \frac{2^{\frac{1-d-a}{2}} \exp\left(-\frac{M(\mathbf{x})+t^2}{2}\right)}{\Gamma\left(\frac{d+1+a}{2}\right)} \right]}{\eta^e \frac{c_1}{a-1} \left[t^{\eta^e(1-a)} \chi_{1+a}^2(t^{2\eta^e}) + \frac{2^{\frac{1-a}{2}} \exp\left(-\frac{t^2}{2}\right)}{\Gamma\left(\frac{1+a}{2}\right)} \right]}.$$

Asymptotiquement, il reste

$$\ell^e(\mathbf{x}) = \lim_{t \rightarrow +\infty} \frac{(a-1)t^{\eta^e-1} c_{\mathbf{x}} t^{1-d-a}}{\eta^e c_1 (d+a-1) t^{\eta^e(1-a)}}.$$

Avec $\eta^e = d/a + 1$, et les valeurs de $c_{\mathbf{x}}$ et c_1 , on obtient directement la valeur de $\ell^e(\mathbf{x})$ donnée dans la Table [IV.1](#). Concernant l'appartenance à la classe K_c , on pose

$$f(\theta) = \lim_{r \rightarrow 0} \frac{\Psi'_Y \left(\Psi_Y^{-1} \left(1 + \frac{r \sin(\theta)}{\ell^e(\mathbf{x})} \right) \right)}{\Psi'_Y \left(\Psi_Y^{-1} \left(1 + \frac{r \sin(\theta)}{\ell^e(\mathbf{x})(r \cos(\theta)+1)} \right) \right)}.$$

Dans le cas Slash, $\Psi'_Y(t) = -\frac{1}{t^2} 2^{\frac{a}{2}-1} \frac{a}{a-1} \frac{\Gamma(\frac{1+a}{2})}{\sqrt{\pi}} \left(\frac{\chi_{1+a}^2(t^2)}{|t|^{a-1}} + \frac{2^{\frac{1-a}{2}}}{\Gamma(\frac{1+a}{2})} \sqrt{2\pi} \varphi(t) \right)$. D'où la limite

$$f(\theta) = \lim_{r \rightarrow 0} \frac{\Psi_Y^{-1} \left(1 + \frac{r \sin(\theta)}{\ell^e(\mathbf{x})(r \cos(\theta)+1)} \right)^2}{\Psi_Y^{-1} \left(1 + \frac{r \sin(\theta)}{\ell^e(\mathbf{x})} \right)^2} \frac{\frac{\chi_{1+a}^2 \left(\Psi_Y^{-1} \left(1 + \frac{r \sin(\theta)}{\ell^e(\mathbf{x})} \right)^2 \right)}{|\Psi_Y^{-1} \left(1 + \frac{r \sin(\theta)}{\ell^e(\mathbf{x})} \right)|^{a-1}} + \frac{2^{\frac{1-a}{2}}}{\Gamma(\frac{1+a}{2})} \sqrt{2\pi} \varphi \left(\Psi_Y^{-1} \left(1 + \frac{r \sin(\theta)}{\ell^e(\mathbf{x})} \right) \right)}{\frac{\chi_{1+a}^2 \left(\Psi_Y^{-1} \left(1 + \frac{r \sin(\theta)}{\ell^e(\mathbf{x})(r \cos(\theta)+1)} \right)^2 \right)}{|\Psi_Y^{-1} \left(1 + \frac{r \sin(\theta)}{\ell^e(\mathbf{x})(r \cos(\theta)+1)} \right)|^{a-1}} + \frac{2^{\frac{1-a}{2}}}{\Gamma(\frac{1+a}{2})} \sqrt{2\pi} \varphi \left(\Psi_Y^{-1} \left(1 + \frac{r \sin(\theta)}{\ell^e(\mathbf{x})(r \cos(\theta)+1)} \right) \right)}.$$

Le ratio sur la gauche tend clairement vers $f(\theta)^2$. Concernant celui sur la droite $\frac{\chi_{1+a}^2(t)}{t}$ tend vers 0 quand $t \rightarrow \infty$. La limite peut alors être obtenue facilement

$$f(\theta) = f(\theta)^2 \lim_{r \rightarrow 0} \frac{\varphi \left(\Psi_Y^{-1} \left(1 + \frac{r \sin(\theta)}{\ell^e(\mathbf{x})} \right) \right)}{\varphi \left(\Psi_Y^{-1} \left(1 + \frac{r \sin(\theta)}{\ell^e(\mathbf{x})(r \cos(\theta)+1)} \right) \right)}$$

On a déjà calculé cette limite dans le cas gaussien, et vu la relation $f(\theta) = \frac{1}{f(\theta)} = 1$, d'où le résultat. $\square$

On a alors défini des prédicteurs d'expectile basés uniquement sur les expectiles non-conditionnels, asymptotiquement équivalents aux expectiles conditionnels pour des niveaux d'expectile proches de 0 et 1. A l'instar du quantile, ces prédicteurs dépendent de deux paramètres η^e et $\ell^e(\mathbf{x})$. La section suivante est dédiée à l'estimation de ces paramètres appelés extrémaux.

IV.3.b. Estimation des paramètres extrémaux

On cherche ici à estimer les paramètres η^e et $\ell^e(\mathbf{x})$ de la Définition [20](#). On voit que les prédicteurs qui en découlent dépendent également des quantités $\mu_{Y|\mathbf{x}}$, $\Sigma_{Y|\mathbf{x}}$ et Ψ_Y^{-1} . On va alors supposer, pour les mêmes raisons que celles expliquées dans la Section [III.3.b.](#), que $\mu_{Y|\mathbf{x}}$ et $\Sigma_{Y|\mathbf{x}}$ sont connus. Il ne nous reste donc plus qu'à estimer η^e et $\ell^e(\mathbf{x})$ dans cette section, puis l'expectile dans la suivante. Nous verrons d'ailleurs que les paramètres extrémaux η^e et $\ell^e(\mathbf{x})$ relatifs à l'expectile sont assez proches de η et $\ell(\mathbf{x})$ obtenus pour le quantile.

Prenons alors $\mathbf{Z} = (\mathbf{X}, Y) \sim \mathcal{EC}_{d+1}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \xi)$, et supposons que l'on dispose d'un jeu de données $\mathbf{X}_1, \dots, \mathbf{X}_n$ indépendantes et identiquement distribuées de loi identique à $\mathbf{X} \sim \mathcal{EC}_d(\boldsymbol{\mu}_{\mathbf{X}}, \boldsymbol{\Sigma}_{\mathbf{X}}, \xi)$. On rappelle qu'en pratique, on a également besoin de répétitions

$Y_1, \dots, Y_n$. Cependant, $\mu_{Y|\mathbf{x}}$ et $\Sigma_{Y|\mathbf{X}}$ étant supposés connus, nous ne l'utiliserons pas dans cette section. On a également besoin de faire une hypothèse concernant les variations régulières de second ordre de notre distribution. Pour cela, on se place dans le cadre de l'Hypothèse [2] avec $\gamma < 1$, ce détail garantissant l'existence de l'expectile.

Proposition 36 (Existence des paramètres extrémaux). *Sous l'Hypothèse [2] avec $\gamma < 1$, les paramètres η^e et $\ell^e(\mathbf{x})$ existent, et sont exprimés comme suit :*

$$\begin{cases} \eta^e = 1 + \gamma d \\ \ell^e(\mathbf{x}) = \frac{\Gamma\left(\frac{d+\gamma^{-1}+1}{2}\right)}{\Gamma\left(\frac{\gamma^{-1}+1}{2}\right)} \frac{\pi^{-\frac{d}{2}}}{(\gamma d+1)c_d g_d(M(\mathbf{x}))} \frac{1-\gamma}{1+\gamma d-\gamma} \end{cases}.$$

Pour démontrer cette proposition, on a besoin de quelques indications sur les propriétés de variation régulière de la distribution $Y|(\mathbf{X} = \mathbf{x})$. On donne alors le lemme suivant, qui montre que $Y|(\mathbf{X} = \mathbf{x})$ est à variation régulière, puis démontre la Proposition [36].

Lemme 37 (Indice de variation régulière de $Y|(\mathbf{X} = \mathbf{x})$). *Sous l'Hypothèse [2], la distribution $Y|(\mathbf{X} = \mathbf{x})$ est à variation régulière d'indice $(\gamma^{-1} + d)^{-1}$, c'est à dire :*

$$\lim_{t \rightarrow +\infty} \frac{\overline{\Phi}_{Y|\mathbf{x}}(t\omega)}{\overline{\Phi}_{Y|\mathbf{x}}(t)} = \omega^{-\frac{1}{\gamma}-d}, \quad \forall \omega > 0.$$

Preuve du Lemme: La densité de $Y|(\mathbf{X} = \mathbf{x})$ est donnée par

$$c_{d+1}g_{d+1} \left(M(\mathbf{x}) + (t - \mu_{Y|\mathbf{x}})^2 \Sigma_{Y|\mathbf{X}}^{-1} \right) (c_d g_d(M(\mathbf{x})))^{-1},$$

où $M(\mathbf{x}) = (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{X}})^\top \Sigma_{\mathbf{X}}^{-1} (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{X}})$. Pour simplifier, on considère le cas centré réduit, i.e. $\mu_{Y|\mathbf{x}} = 0$ et $\Sigma_{Y|\mathbf{X}} = 1$. Un calcul rapide donne

$$\lim_{t \rightarrow +\infty} \frac{\overline{\Phi}_{Y|\mathbf{x}}(\omega t)}{\overline{\Phi}_{Y|\mathbf{x}}(t)} = \omega \lim_{t \rightarrow +\infty} \frac{g_{d+1}(M(\mathbf{x}) + \omega^2 t^2)}{g_{d+1}(M(\mathbf{x}) + t^2)} = \omega \lim_{t \rightarrow +\infty} \frac{(M(\mathbf{x}) + \omega^2 t^2)^{-\frac{d}{2}} f_{R_{d+1}} \left(\sqrt{M(\mathbf{x}) + \omega^2 t^2} \right)}{(M(\mathbf{x}) + t^2)^{-\frac{d}{2}} f_{R_{d+1}} \left(\sqrt{M(\mathbf{x}) + t^2} \right)}.$$

L'Equation ([ii]) mène à

$$\lim_{t \rightarrow +\infty} \frac{\overline{\Phi}_{Y|\mathbf{x}}(\omega t)}{\overline{\Phi}_{Y|\mathbf{x}}(t)} = \omega \omega^{-d} \omega^{-\frac{1}{\gamma}-1} = \omega^{-\frac{1}{\gamma}-d}.$$

□

Preuve de la Proposition: On rappelle

$$\ell^e(\mathbf{x}) = \lim_{t \rightarrow +\infty} \frac{\overline{\Psi}_{Y|\mathbf{x}}(t)}{\overline{\Psi}_Y(t^{\eta^e})}.$$

Il n'est pas très compliqué de montrer, en utilisant l'Equation (19) puis la théorie de Karamata sur les intégrales de fonction à variation régulière, ou bien directement le résultat de [Bellini et al., 2014], que $\overline{\Psi}_Y(t) \sim (1 - \gamma^{-1})^{-1} \overline{\Phi}_Y(t)$ quand $t \rightarrow +\infty$. De la même manière, on a $\overline{\Psi}_{Y|\mathbf{x}}(t) \sim (1 - \gamma^{-1} - d)^{-1} \overline{\Phi}_{Y|\mathbf{x}}(t)$ à l'aide du Lemme 37. Le calcul de $\ell^e(\mathbf{x})$ se réduit alors à

$$\ell^e(\mathbf{x}) = \lim_{t \rightarrow +\infty} \frac{(1 - \gamma^{-1} - d)^{-1} \overline{\Phi}_{Y|\mathbf{x}}(t)}{(1 - \gamma^{-1})^{-1} \overline{\Phi}_Y(t^{\eta^e})},$$

soit, à un coefficient près, au calcul de $\ell(\mathbf{x})$ dans le cas du quantile, donné dans l'Equation (15). Dès lors, il suffit de prendre $\eta^e = \eta$, et on obtient $\ell^e(\mathbf{x}) = \ell(\mathbf{x})(1 - \gamma)/(1 + \gamma d - \gamma)$. $\square$

On voit clairement que $\eta^e = \eta$ et $\ell^e(\mathbf{x}) = \ell(\mathbf{x})(1 - \gamma)/(1 + \gamma d - \gamma)$, où η et $\ell(\mathbf{x})$ sont donnés dans l'Equation (15). Par commodité, on notera alors η au lieu de η^e par la suite. L'étude similaire effectuée dans la Section III.3.b. nous permet de déduire assez facilement des estimateurs pour ces paramètres, ainsi que leurs distributions asymptotiques. L'estimateur pour η étant déjà introduit dans la Définition 13 on propose alors un estimateur pour $\ell^e(\mathbf{x})$ proche de celui de la Définition 15 relatif à $\ell(\mathbf{x})$.

Définition 21 (Estimateur de $\ell^e(\mathbf{x})$). On définit $\hat{\ell}_{k_n, h_n}^e(\mathbf{x})$ comme :

$$\hat{\ell}_{k_n, h_n}^e(\mathbf{x}) = \frac{\Gamma\left(\frac{d + \hat{\gamma}_{k_n}^{-1} + 1}{2}\right)}{\Gamma\left(\frac{\hat{\gamma}_{k_n}^{-1} + 1}{2}\right)} \frac{\pi^{-\frac{d}{2}}}{(d\hat{\gamma}_{k_n} + 1) \hat{g}_{h_n}(M(\mathbf{x}))} \frac{1 - \hat{\gamma}_{k_n}}{1 + d\hat{\gamma}_{k_n} - \hat{\gamma}_{k_n}},$$

où $\hat{\gamma}_{k_n}$ et $\hat{g}_{h_n}(M(\mathbf{x}))$ sont respectivement donnés dans les Equations (III.3.b.) et (14).

La loi asymptotique jointe du couple $(\hat{\eta}_{k_n}, \hat{\ell}_{k_n, h_n}^e(\mathbf{x}))$ se déduit également aisément des résultats de la Section III.3.b.

Proposition 38 (Normalité asymptotique de $(\hat{\eta}_{k_n}, \hat{\ell}_{k_n, h_n}^e(\mathbf{x}))$). Si g_d et K satisfont les conditions (i) – (iii) du Théorème 61 de l'Annexe B sous l'Hypothèse 2 avec $\gamma < 1$, et si $\sqrt{k_n} A\left(\frac{n}{k_n}\right) \rightarrow 0$ quand $n \rightarrow +\infty$, alors on a les relations asymptotiques suivantes :

(i) Si $nh_n/k_n \xrightarrow{n \rightarrow +\infty} 0$ et $nh_n^5 \xrightarrow{n \rightarrow +\infty} 0$, alors

$$\sqrt{nh_n} \begin{pmatrix} \hat{\ell}_{k_n, h_n}^e(\mathbf{x}) - \ell^e(\mathbf{x}) \\ \hat{\eta}_{k_n} - \eta \end{pmatrix} \xrightarrow{n \rightarrow +\infty} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} V_2^e(\gamma, c_d g_d(M(\mathbf{x}))) & 0 \\ 0 & 0 \end{pmatrix}\right).$$

(ii) Si $nh_n/k_n \xrightarrow{n \rightarrow +\infty} +\infty$ et $\sqrt{k_n}h_n^2 \xrightarrow{n \rightarrow +\infty} 0$, alors

$$\sqrt{k_n} \left(\frac{\hat{\ell}_{k_n, h_n}^e(\mathbf{x}) - \ell^e(\mathbf{x})}{\hat{\eta}_{k_n} - \eta} \right) \xrightarrow{n \rightarrow +\infty} \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} V_1^e(\gamma, c_d g_d(M(\mathbf{x}))) & -d\gamma \sqrt{V_1^e(\gamma, c_d g_d(M(\mathbf{x})))} \\ -d\gamma \sqrt{V_1^e(\gamma, c_d g_d(M(\mathbf{x})))} & d^2 \gamma^2 \end{pmatrix} \right),$$

où Ψ est la fonction digamma (voir p.258 de [\[Abramowitz et al., 1966\]](#)), et :

$$\begin{cases} V_1^e(\gamma, c_d g_d(M(\mathbf{x}))) = \frac{\pi^{-d} \gamma^2}{c_d^2 g_d(M(\mathbf{x}))^2} \frac{\Gamma\left(\frac{d+\gamma^{-1}+1}{2}\right)^2}{\Gamma\left(\frac{\gamma^{-1}+1}{2}\right)^2} \left[\frac{1-\gamma}{2\gamma^2} \frac{\Psi\left(\frac{\gamma^{-1}+1}{2}\right) - \Psi\left(\frac{d+\gamma^{-1}+1}{2}\right)}{\gamma^2 d(d-1) + \gamma(2d-1)+1} + \frac{\gamma^2 d(d-1) - 2\gamma d(d-1) - 2d}{(\gamma^2 d(d-1) + \gamma(2d-1)+1)^2} \right]^2 \\ V_2^e(\gamma, c_d g_d(M(\mathbf{x}))) = \frac{\Gamma\left(\frac{d}{2}\right)}{M(\mathbf{x})^{\frac{d}{2}-1} \pi^{\frac{d}{2}}} c_d g_d(M(\mathbf{x})) \int K(u)^2 du \left[\frac{\Gamma\left(\frac{d+\gamma^{-1}+1}{2}\right)}{\Gamma\left(\frac{\gamma^{-1}+1}{2}\right)} \frac{\pi^{-\frac{d}{2}}}{(\gamma d+1) c_d^2 g_d(M(\mathbf{x}))^2} \frac{1-\gamma}{1+\gamma d-\gamma} \right]^2 \end{cases}.$$

Preuve :

(i) On a montré dans la preuve de la Proposition [20](#) que

$$\sqrt{nh_n} \left(\frac{\hat{\gamma}_{k_n} - \gamma}{\hat{g}_{h_n}(M(\mathbf{x})) - c_d g_d(M(\mathbf{x}))} \right) \xrightarrow{n \rightarrow +\infty} \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & \frac{M(\mathbf{x})^{1-\frac{d}{2}} \Gamma\left(\frac{d}{2}\right)}{\pi^{\frac{d}{2}}} c_d g_d(M(\mathbf{x})) \int K(u)^2 du \end{pmatrix} \right).$$

La méthode delta multivariée donne alors le résultat.

(ii) Même raisonnement avec

$$\sqrt{k_n} \left(\frac{\hat{\gamma}_{k_n} - \gamma}{\hat{g}_{h_n}(M(\mathbf{x})) - c_d g_d(M(\mathbf{x}))} \right) \xrightarrow{n \rightarrow +\infty} \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \gamma^2 & 0 \\ 0 & 0 \end{pmatrix} \right).$$

□

En se basant sur les résultats de cette section, on propose dans la section suivante des estimateurs pour les expectiles conditionnels extrêmes $e_\alpha(Y|\mathbf{X} = \mathbf{x})$.

IV.3.c. Estimation des expectiles extrêmes

Les estimateurs relatifs aux paramètres η et $\ell^e(\mathbf{x})$ de la section précédente vont maintenant nous être utiles pour estimer des expectiles de niveau α_n avec $\alpha_n \rightarrow 1$ quand $n \rightarrow +\infty$. Comme dans la Section [III.3.c.](#), on va diviser l'étude aux cas $n(1 - \alpha_n) \rightarrow +\infty$ et $n(1 - \alpha_n) \rightarrow 0$. On fera d'ailleurs référence aux conditions (C) , (C_{int}) , (C_{high}) , (C_{int}^{HG}) et (C_{high}^{HG}) .

IV.3.c.1. Expectiles intermédiaires

Considérons alors $n(1 - \alpha_n) \rightarrow +\infty$. Si on peut se permettre d'interpoler notre échantillon dans le cas des quantiles, l'idée est la même dans celui des expectiles. On va donc chercher, de manière assez semblable, à l'aide de l'Equation (23), un expectile de niveau u_n , qui sera la solution d'un problème de minimisation construit avec la fonction de scoring (18) (voir [Daouia et al., 2017a]), puis l'élever à la puissance $1/\eta$.

Définition 22 (Estimateur d'expectiles intermédiaires). *Sous l'Hypothèse 2 avec $\gamma < \frac{1}{2}$, on définit $(\hat{e}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x}))_{n \in \mathbb{N}}$ comme :*

$$\hat{e}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x}) = \mu_{Y|\mathbf{x}} + \sqrt{\Sigma_{Y|\mathbf{x}}} \bar{e}_{\tilde{u}_n}^{\frac{1}{\eta_{k_n}}}, \quad (\text{IV.21})$$

où

$$\forall \alpha \in]0, 1[, \bar{e}_\alpha = \arg \min_{e \in \mathbb{R}} \frac{1}{n} \sum_{i=1}^n S_{2,\alpha}(W_i - e),$$

$\tilde{u}_n = 1 - (1 - \alpha_n) \left((2\alpha_n - 1) \hat{\ell}_{k_n, h_n}^e(\mathbf{x}) + 2(1 - \alpha_n) \right)^{-1}$ et W est la première (ou indifféremment n'importe laquelle) composante du vecteur $\Lambda_{\mathbf{X}}^{-1}(\mathbf{X} - \mu_{\mathbf{X}})$.

On remarque ici que la condition $\gamma < 1/2$ est nécessaire. En effet, l'estimation de l'expectile de niveau $u_n = 1 - (1 - \alpha_n) ((2\alpha_n - 1) \ell^e(\mathbf{x}) + 2(1 - \alpha_n))^{-1}$ par le problème de minimisation requiert l'existence de moments d'ordre 2. En utilisant le même raisonnement que celui utilisé pour la preuve du Théorème 21 et les mêmes conditions, on trouve un résultat de normalité asymptotique similaire, et a fortiori de consistance.

Théorème 39 (Propriétés de $\hat{e}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})$). *Sous l'Hypothèse 2 avec $\gamma < \frac{1}{2}$, et les conditions (C), (C_{int}) :*

$$\frac{\sqrt{k_n}}{\ln(1 - \alpha_n)} \left(\frac{\hat{e}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})}{e_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})} - 1 \right) \xrightarrow{n \rightarrow +\infty} \mathcal{N} \left(0, \frac{d^2 \gamma^4}{(\gamma d + 1)^4} \right). \quad (\text{IV.22})$$

Par conséquent :

$$\frac{\hat{e}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})}{e_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})} \xrightarrow{\mathbb{P}} 1 \text{ lorsque } n \rightarrow +\infty. \quad (\text{IV.23})$$

Enfin, en remplaçant l'Hypothèse 2 par l'Hypothèse 3 et en ajoutant la condition (C_{int}^{HG}), on obtient :

$$\frac{\sqrt{k_n}}{\ln(1 - \alpha_n)} \left(\frac{\hat{e}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})}{e_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})} - 1 \right) \xrightarrow{n \rightarrow +\infty} \mathcal{N} \left(0, \frac{d^2 \gamma^4}{(\gamma d + 1)^4} \right). \quad (\text{IV.24})$$

Preuve : La preuve est assez similaire à celle du Théorème 21. On a d'abord

$$\frac{\sqrt{k_n}}{\ln(1 - \alpha_n)} \left(\frac{\hat{e}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})}{e_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})} - 1 \right) \underset{n \rightarrow +\infty}{\sim} \frac{\sqrt{k_n}}{\ln(1 - \alpha_n)} \left(\frac{\bar{e}_{u_n}^{1/\hat{\eta}_{k_n}}}{\Psi_Y^{-1}\left(\frac{u_n}{2u_n-1}\right)^{1/\eta}} - 1 \right).$$

Si on s'intéresse à ce dernier terme, on le décompose de la manière suivante

$$\frac{\sqrt{k_n}}{\ln(1 - \alpha_n)} \left(\frac{\bar{e}_{u_n}^{1/\hat{\eta}_{k_n}}}{\Psi_Y^{-1}\left(\frac{u_n}{2u_n-1}\right)^{1/\hat{\eta}_{k_n}}} - 1 \right) \Psi_Y^{-1}\left(\frac{u_n}{2u_n-1}\right)^{1/\hat{\eta}_{k_n}-1/\eta} + \frac{\sqrt{k_n}}{\ln(1 - \alpha_n)} \left(\Psi_Y^{-1}\left(\frac{u_n}{2u_n-1}\right)^{1/\hat{\eta}_{k_n}-1/\eta} - 1 \right)$$

Concentrons nous sur le premier terme. On a l'estimateur $\hat{\eta}_{k_n}$ qui converge en $\sqrt{k_n}$, quand $\bar{e}_{u_n}$ converge en $\sqrt{n(1 - u_n)}$ (voir [Daouia et al., 2017a]). On a d'une part l'assertion évidente $\sqrt{k_n}/\ln(1 - \alpha_n) = o(\sqrt{k_n})$. De l'autre, la condition (C_{int}) assure $\sqrt{k_n}/\ln(1 - \alpha_n) = o(\sqrt{n(1 - \alpha_n)})$. Or on a $\sqrt{n(1 - u_n)} \sim \ell^e(\mathbf{x})^{-1/2} \sqrt{n(1 - \alpha_n)}$, d'où le premier terme tend vers 0 quand $n \rightarrow +\infty$. Prenons alors le second terme. On s'aperçoit, comme pour le Théorème 21 que

$$\frac{\sqrt{k_n}}{\ln\left(\Psi_Y^{-1}\left(\frac{u_n}{2u_n-1}\right)\right)} \left(\Psi_Y^{-1}\left(\frac{u_n}{2u_n-1}\right)^{1/\hat{\eta}_{k_n}-1/\eta} - 1 \right) \underset{n \rightarrow +\infty}{\rightarrow} \mathcal{N}\left(0, \frac{d^2 \gamma^2}{(d\gamma + 1)^4}\right).$$

Enfin, il est montré dans [Bellini et al., 2014] ou [Daouia et al., 2017a] que les quantiles sont asymptotiquement proportionnels aux expectiles. D'où l'équivalence $\ln\left(\Psi_Y^{-1}\left(\frac{u_n}{2u_n-1}\right)\right) \sim -\gamma \ln(1 - \alpha_n)$ quand $n \rightarrow +\infty$, et le résultat (39).

Enfin, pour montrer la dernière normalité asymptotique, il faut prouver que

$$\lim_{n \rightarrow +\infty} \frac{\sqrt{k_n}}{\ln(1 - \alpha_n)} \ln \left(\frac{\Psi_Y^{-1}\left(\frac{u_n}{2u_n-1}\right)^{\frac{1}{\eta}}}{\Psi_{Y|\mathbf{x}}^{-1}\left(\frac{\alpha_n}{2\alpha_n-1}\right)} \right) = 0.$$

On utilise encore et toujours la proportionnalité asymptotique entre quantile et expectile, d'où la relation suivante sous l'Hypothèse 3:

$$\ln \left(\frac{\Psi_Y^{-1}\left(\frac{u_n}{2u_n-1}\right)^{\frac{1}{\eta}}}{\Psi_{Y|\mathbf{x}}^{-1}\left(\frac{\alpha_n}{2\alpha_n-1}\right)} \right) = \ln \left(\frac{\Phi_Y^{-1}(u_n)^{\frac{1}{\eta}}}{\Phi_{Y|\mathbf{x}}^{-1}(\alpha_n)} \lambda_1 \frac{1 + \lambda_2(1 - u_n)^{-\rho}}{1 + \lambda_3(1 - \alpha_n)^{\frac{\min\{-\rho, 2\gamma\}}{\gamma d + 1}}} \right), \lambda_i \in \mathbb{R}.$$

En décomposant en une somme de deux logarithmes, on a vu dans la preuve du Théorème 22 que le premier tendait vers 0 sous (C_{int}^{HG}) . Dans cette même preuve, on montre en particulier que

$$\frac{\sqrt{k_n}}{\ln(1 - \alpha_n)} \ln \left(\frac{1 + \lambda_2(1 - u_n)^{-\rho}}{1 + \lambda_3(1 - \alpha_n)^{\frac{\min\{-\rho, 2\gamma\}}{\gamma d + 1}}} \right)$$

tend vers 0, d'où le résultat (39). $\square$

Comme pour l'estimateur de quantile $\hat{q}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})$, l'ajout de l'Hypothèse 3 permet d'obtenir la normalité asymptotique (39) avec $e_{\alpha_n}^E(Y|\mathbf{X} = \mathbf{x})$ au lieu de $e_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})$.

IV.3.c.2. Hauts Expectiles

Prenons $n(1 - \alpha_n) \rightarrow +\infty$. Le suite α_n tendant vers 1 trop rapidement pour appliquer directement l'approche précédente, on va alors choisir une suite dite intermédiaire pour laquelle on va estimer un expectile avec l'estimateur $\bar{e}$, puis utiliser les propriétés de variation régulière pour construire un estimateur de notre expectile extrême. Cette idée, que l'on a déjà utilisée dans la Section III.3.c. est basée sur les approches de [de Haan and Ferreira, 2006] ou [Weissman, 1978] dans le cadre des quantiles. Elle a également été appliquée aux expectiles dans [Daouia et al., 2017a].

Définition 23 (Estimateur de hauts expectiles). *Sous l'Hypothèse 2 avec $\gamma < \frac{1}{2}$, on définit $\left(\hat{e}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})\right)_{n \in \mathbb{N}}$ comme :*

$$\hat{e}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x}) = \mu_{Y|\mathbf{x}} + \sqrt{\Sigma_{Y|\mathbf{x}}} \left[\left(\frac{k_n}{n(1 - \tilde{u}_n)} \right)^{\hat{\gamma}_{k_n}} \bar{e}_{1-k_n/n} \right]^{\frac{1}{\hat{\eta}_{k_n}}}, \quad (\text{IV.25})$$

où les quantités $\tilde{u}_n$, $\bar{e}$ et W sont rappelées dans la Définition 22.

Comme mentionné dans [Daouia et al., 2017a], un tel estimateur a la même loi asymptotique que l'estimateur de quantile extrême défini dans le chapitre précédent. On déduit alors sans difficulté le résultat suivant.

Théorème 40 (Propriétés de $\hat{e}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})$). *Sous l'Hypothèse 2 avec $\gamma < \frac{1}{2}$, les conditions (C), (C_{high}) ainsi que $\frac{k_n^{\gamma+1/2}}{n^\gamma} \rightarrow \lambda$ quand $n \rightarrow +\infty$, on a :*

$$\frac{\sqrt{k_n}}{\ln\left(\frac{k_n}{n(1-\alpha_n)}\right)} \left(\frac{\hat{e}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})}{e_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})} - 1 \right) \xrightarrow{n \rightarrow +\infty} \mathcal{N}\left(0, \left(\frac{\gamma}{\gamma d + 1} - \theta \frac{d\gamma^2}{(\gamma d + 1)^2} \right)^2\right).$$

Par conséquent :

$$\frac{\hat{e}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})}{e_{\alpha_n}^E(Y|\mathbf{X} = \mathbf{x})} \xrightarrow{\mathbb{P}} 1 \text{ lorsque } n \rightarrow +\infty.$$

Enfin, en remplaçant l'Hypothèse 2 par l'Hypothèse 3 et en ajoutant la condition (C_{high}^{HG}) , on obtient :

$$\frac{\sqrt{k_n}}{\ln\left(\frac{k_n}{n(1-\alpha_n)}\right)} \left(\frac{\hat{e}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})}{e_{\alpha_n}^E(Y|\mathbf{X} = \mathbf{x})} - 1 \right) \xrightarrow{n \rightarrow +\infty} \mathcal{N}\left(0, \left(\frac{\gamma}{\gamma d + 1} - \theta \frac{d\gamma^2}{(\gamma d + 1)^2} \right)^2\right). \quad (\text{IV.26})$$

Preuve : On s'aperçoit rapidement que la loi limite de (23) est celle de

$$\frac{\sqrt{k_n}}{\ln\left(\frac{k_n}{n(1-\alpha_n)}\right)} \left(\frac{\left[\left(\frac{k_n}{n(1-u_n)}\right)^{\hat{\gamma}_{k_n}} \bar{e}_{1-k_n/n}\right]^{\frac{1}{\hat{\eta}_{k_n}}}}{\Psi_Y^{-1}\left(\frac{u_n}{2u_n-1}\right)^{\frac{1}{\hat{\eta}_{k_n}}}} - 1 \right) \Psi_Y^{-1}\left(\frac{u_n}{2u_n-1}\right)^{\frac{1}{\hat{\eta}_{k_n}} - \frac{1}{\eta}} + \frac{\sqrt{k_n}}{\ln\left(\frac{k_n}{n(1-\alpha_n)}\right)} \left(\Psi_Y^{-1}\left(\frac{u_n}{2u_n-1}\right)^{\frac{1}{\hat{\eta}_{k_n}} - \frac{1}{\eta}} - 1 \right)$$

Comme mentionné plus haut, on peut s'inspirer du Théorème 3 de [Daouia et al., 2017a], et remplacer l'estimateur d'expectile par celui du quantile. Les deux termes de cette somme ont donc exactement comme distribution asymptotique la loi bivariée du Théorème 23 d'où le résultat. La normalité asymptotique (40) s'obtient de la même manière que dans la preuve du Théorème 39. $\square$

IV.3.d. Liens avec les quantiles extrêmes

L'approche que nous avons proposée jusqu'à maintenant pour estimer des expectiles conditionnels n'utilise pas vraiment les relations asymptotiques qui existent entre les quantiles et expectiles. En effet, [Bellini et al., 2014] ont établi une équivalence entre les quantiles et expectiles, lorsque le niveau α tendait vers 1. De tels résultats nous permettraient de définir de nouveaux estimateurs basés sur les estimateurs de quantiles extrêmes présentés dans le Chapitre III. La proposition suivante met en évidence cette équivalence. Si elle a été dans un premier temps introduite dans [Bellini et al., 2014], elle s'avère être plus parlante sous l'Hypothèse 2. Elle est notamment utilisée dans [Daouia et al., 2017a].

Proposition 41 (Equivalence quantile-expectile). *Soit la variable $Y \sim \mathcal{EC}_1(\mu_Y, \Sigma_Y, \xi)$. On suppose que l'Hypothèse 2 est satisfaite avec $\gamma < 1$, et on note $Y' = (Y - \mu_Y)\Sigma_Y^{-1/2}$. Si Φ_Y est strictement croissante, alors*

$$\frac{e_\alpha(Y')}{q_\alpha(Y')} = (\gamma^{-1} - 1)^{-\gamma} (1 + r(\alpha)) \quad (\text{IV.27})$$

avec $r(\alpha) = \left(\frac{(\gamma^{-1}-1)^{-\rho}}{1-\rho-\gamma} + \frac{(\gamma^{-1}-1)^{-\rho}-1}{\rho} + o(1) \right) A((1-\alpha)^{-1})$ lorsque $\alpha \rightarrow 1$.

Notons que l'expression du reste $r(\alpha)$ ci-dessus utilise le fait que $\mathbb{E}[Y'] = 0$. A partir des quantiles, il est donc assez simple d'en déduire des expectiles pour des niveaux extrêmes. Ce résultat nous permet alors de définir des estimateurs pour les expectiles extrêmes différents de ceux de la Section IV.3.c. En effet, on combine les estimateurs de la Section III.3.c. avec la Proposition 41 pour définir les estimateurs suivants.

Définition 24 (D'autres estimateurs). *Sous l'Hypothèse 2, on définit les estimateurs*

$(\tilde{e}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x}))_{n \in \mathbb{N}}$, $(\tilde{\tilde{e}}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x}))_{n \in \mathbb{N}}$ et $(\tilde{\tilde{\tilde{e}}}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x}))_{n \in \mathbb{N}}$ comme :

$$\begin{cases} \tilde{e}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x}) = & \mu_{Y|\mathbf{x}} + \sqrt{\Sigma_{Y|\mathbf{X}}} (W_{[n\tilde{v}_n+1]})^{\frac{1}{\tilde{\eta}_{k_n}}} (\hat{\gamma}_{k_n}^{-1} + d - 1)^{-(\hat{\gamma}_{k_n}^{-1} + d)^{-1}} \\ \tilde{\tilde{e}}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x}) = & \mu_{Y|\mathbf{x}} + \sqrt{\Sigma_{Y|\mathbf{X}}} \left(\left(\frac{k_n}{n(1-\tilde{u}_n)} \right)^{\hat{\gamma}_{k_n}} W_{[k_n+1]} (\hat{\gamma}_{k_n}^{-1} - 1)^{-\hat{\gamma}_{k_n}} \right)^{\frac{1}{\tilde{\eta}_{k_n}}} \\ \tilde{\tilde{\tilde{e}}}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x}) = & \mu_{Y|\mathbf{x}} + \sqrt{\Sigma_{Y|\mathbf{X}}} \left[W_{[k_n+1]} \left(\frac{k_n}{n\tilde{v}_n} \right)^{\hat{\gamma}_{k_n}} \right]^{\frac{1}{\tilde{\eta}_{k_n}}} (\hat{\gamma}_{k_n}^{-1} + d - 1)^{-(\hat{\gamma}_{k_n}^{-1} + d)^{-1}} \end{cases} \quad (\text{IV.28})$$

Le premier estimateur $\tilde{e}$ est adapté aux expectiles intermédiaires. On voit qu'il estime d'abord un quantile de niveau intermédiaire, puis utilise la relation asymptotique (41). Une telle approche est une alternative intéressante à la Définition 22 car elle permet de se libérer de la condition $\gamma < 1/2$. Le second $\tilde{\tilde{e}}$ est applicable dans le cas des hauts expectiles, et est exactement le même que (23) à un détail près : l'expectile $\bar{e}_{1-k_n/n}$ estimé par les moindres carrés asymétriques est remplacé par une statistique d'ordre estimant le quantile de niveau $1 - k_n/n$, que l'on multiplie par le coefficient de l'Equation (23). Enfin, le dernier estimateur, introduit pour les hauts expectiles, estime dans un premier temps le haut quantile à la manière de la Définition 17 avant de multiplier par le coefficient asymptotique. Dans un premier temps, donnons la distribution asymptotique de $\tilde{\tilde{e}}$.

Théorème 42 (Propriétés de $\tilde{\tilde{e}}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})$). *Sous l'Hypothèse 2 avec $\gamma < \frac{1}{2}$, les conditions (C), (C_{high}) ainsi que $\frac{k_n^{\gamma+1/2}}{n^\gamma} \rightarrow \lambda < +\infty$ quand $n \rightarrow +\infty$, on a :*

$$\frac{\sqrt{k_n}}{\ln\left(\frac{k_n}{n(1-\alpha_n)}\right)} \left(\frac{\tilde{\tilde{e}}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})}{e_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})} - 1 \right) \xrightarrow{n \rightarrow +\infty} \mathcal{N}\left(0, \left(\frac{\gamma}{\gamma d + 1} - \theta \frac{d\gamma^2}{(\gamma d + 1)^2} \right)^2\right).$$

Par conséquent :

$$\frac{\tilde{\tilde{e}}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})}{e_{\alpha_n}(Y|\mathbf{X} = \mathbf{x})} \xrightarrow{\mathbb{P}} 1 \text{ lorsque } n \rightarrow +\infty.$$

Enfin, en remplaçant l'Hypothèse 2 par l'Hypothèse 3 et en ajoutant la condition (C_{high}^{HG}) , on obtient :

$$\frac{\sqrt{k_n}}{\ln\left(\frac{k_n}{n(1-\alpha_n)}\right)} \left(\frac{\tilde{\tilde{e}}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})}{e_{\alpha_n}(Y|\mathbf{X} = \mathbf{x})} - 1 \right) \xrightarrow{n \rightarrow +\infty} \mathcal{N}\left(0, \left(\frac{\gamma}{\gamma d + 1} - \theta \frac{d\gamma^2}{(\gamma d + 1)^2} \right)^2\right).$$

Preuve : La démarche est exactement la même que celle du Théorème 40 et découle du Théorème 3 de [Daouia et al., 2017a]. $\square$

Dans la section suivante, on propose des estimateurs pour une famille de mesures de risques englobant les quantiles et expectiles. Des résultats asymptotiques généraux sont alors donnés, à partir desquels on pourra déduire ceux des estimateurs $\tilde{e}$ et $\tilde{\tilde{e}}$ de la Définition [24]. Cependant, ces résultats nécessitent des conditions plus fortes. Le cas de l'expectile mérite donc un intérêt particulier énoncé ci-dessous.

Théorème 43 (Propriétés de $\tilde{e}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})$ et $\tilde{\tilde{e}}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})$). *Supposons que l'Hypothèse [2] et la condition (C) soient satisfaites. Sous les conditions supplémentaires (C_{int}) et (C_{high}) respectivement, on a :*

$$\begin{cases} \frac{\tilde{e}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})}{e_{\alpha_n}(Y|\mathbf{X} = \mathbf{x})} \xrightarrow{\mathbb{P}} 1 \\ \frac{\tilde{\tilde{e}}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})}{e_{\alpha_n}(Y|\mathbf{X} = \mathbf{x})} \xrightarrow{\mathbb{P}} 1 \end{cases} \quad \text{lorsque } n \rightarrow +\infty.$$

De plus, en remplaçant l'Hypothèse [2] par l'Hypothèse [3] et en ajoutant les conditions (C_{int}^{HG}) et (C_{high}^{HG}) respectivement, on obtient :

$$\begin{cases} \frac{\sqrt{k_n}}{\ln(1-\alpha_n)} \left(\frac{\tilde{e}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})}{e_{\alpha_n}(Y|\mathbf{X} = \mathbf{x})} - 1 \right) \xrightarrow{n \rightarrow +\infty} \mathcal{N}\left(0, \frac{d^2 \gamma^4}{(\gamma d + 1)^4}\right) \\ \frac{\sqrt{k_n}}{\ln\left(\frac{k_n}{n(1-\alpha_n)}\right)} \left(\frac{\tilde{\tilde{e}}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})}{e_{\alpha_n}(Y|\mathbf{X} = \mathbf{x})} - 1 \right) \xrightarrow{n \rightarrow +\infty} \mathcal{N}\left(0, \left(\frac{\gamma}{\gamma d + 1} - \theta \frac{d \gamma^2}{(\gamma d + 1)^2}\right)^2\right) \end{cases}.$$

Preuve : Au vu des Théorèmes [21] et [23] les consistances sont immédiates. Dans cette preuve, on note $Y_0|\mathbf{x} = (Y|(\mathbf{X} = \mathbf{x}) - \mu_{Y|\mathbf{x}}) \Sigma_{Y|\mathbf{X}}^{-1/2}$ la variable centrée réduite relative à $Y|(\mathbf{X} = \mathbf{x})$. Avec cette notation, on a $e_{\alpha}(Y|\mathbf{X} = \mathbf{x}) = \mu_{Y|\mathbf{x}} + \sqrt{\Sigma_{Y|\mathbf{X}}} e_{\alpha}(Y_0|\mathbf{x})$. De plus, si $\bar{q}_{\alpha_n}(Y|\mathbf{X} = \mathbf{x})$ est un estimateur de quantile conditionnel ($\hat{q}$ ou $\hat{\hat{q}}$), on a la décomposition suivante :

$$\begin{aligned} \frac{\bar{q}_{\alpha_n}(Y_0|\mathbf{X} = \mathbf{x}) (\hat{\gamma}_{k_n}^{-1} + d - 1)^{-(\hat{\gamma}_{k_n}^{-1} + d)^{-1}}}{e_{\alpha_n}(Y_0|\mathbf{x})} - 1 &= \frac{q_{\alpha_n}(Y_0|\mathbf{x}) (\gamma^{-1} + d - 1)^{-(\gamma^{-1} + d)^{-1}}}{e_{\alpha_n}(Y_0|\mathbf{x})} - 1 \\ &+ \left(\frac{\bar{q}_{\alpha_n}(Y_0|\mathbf{X} = \mathbf{x}) (\hat{\gamma}_{k_n}^{-1} + d - 1)^{-(\hat{\gamma}_{k_n}^{-1} + d)^{-1}}}{q_{\alpha_n}(Y_0|\mathbf{x}) (\gamma^{-1} + d - 1)^{-(\gamma^{-1} + d)^{-1}}} - 1 \right) \frac{q_{\alpha_n}(Y_0|\mathbf{x}) (\gamma^{-1} + d - 1)^{-(\gamma^{-1} + d)^{-1}}}{e_{\alpha_n}(Y_0|\mathbf{x})}. \end{aligned}$$

En utilisant la relation ([41]), il existe $\lambda \in \mathbb{R}$ tel que

$$\frac{q_{\alpha_n}(Y_0|\mathbf{x}) (\gamma^{-1} + d - 1)^{-(\gamma^{-1} + d)^{-1}}}{e_{\alpha_n}(Y_0|\mathbf{x})} - 1 = (\lambda + o(1)) A^* \left(\frac{1}{1 - \alpha_n} \right) \quad \text{quand } n \rightarrow +\infty.$$

On a vu dans la preuve de la Proposition [22] que sous l'Hypothèse [3] ainsi que les conditions (C_{int}^{HG}) et (C_{high}^{HG}) , on avait

$$\begin{cases} \frac{\sqrt{k_n}}{\ln(1-\alpha_n)} A^* \left(\frac{1}{1-\alpha_n} \right) \rightarrow 0 \\ \frac{\sqrt{k_n}}{\ln\left(\frac{k_n}{n(1-\alpha_n)}\right)} A^* \left(\frac{1}{1-\alpha_n} \right) \rightarrow 0 \end{cases} \quad \text{quand } n \rightarrow +\infty.$$

Quel que soit $\bar{q}$, le premier terme de la précédente décomposition n'interviendra donc pas dans la loi asymptotique. Intéressons alors au second terme. Les deux taux de convergence étant négligeables devant $\sqrt{k_n}$, il est évident que le terme entre parenthèses se comportera asymptotiquement comme l'estimateur de quantile $\bar{q}$. D'où les normalités asymptotiques identiques à celles des Equations (22) et (24). $\square$

IV.4. Autres mesures de risque extrêmes basées sur le quantile

Comme expliqué dans l'introduction de ce chapitre, les limites du quantile, appelé parfois Value at Risk (voir [Linsmeier and Pearson, 2000]), ont poussé de nombreux auteurs à proposer d'autres mesures de risque comme alternative au quantile. Si nous avons exploré jusqu'à maintenant le cas de l'expectile, on présente dans cette section deux familles de mesures de risque. De plus, ces deux classes sont asymptotiquement liées au quantile, ce qui nous permettra de construire des estimateurs basés sur ceux de la Section III.3.c.

IV.4.a. L_p -quantiles

On vient de voir en détail les notions de quantile et d'expectile. Finalement, ces deux mesures appartiennent à une même famille appelée L_p -quantiles. Une des premières apparitions de cette notion peut être trouvée dans [Chen, 1996].

Définition 25 (L_p -quantiles). *Soit Y une variable aléatoire réelle admettant un moment d'ordre $p - 1$. Le L_p -quantile de niveau $\alpha \in]0, 1[$ de Y est donné par :*

$$q_{p,\alpha}(Y) = \arg \min_{y \in \mathbb{R}} \mathbb{E} [\mathcal{S}_{p,\alpha}(Y - y)],$$

où la fonction $\mathcal{S}_{p,\alpha}$ est définie de la manière suivante :

$$\mathcal{S}_{p,\alpha}(t) = (1 - \alpha)|t^-|^p + \alpha|t^+|^p.$$

Avec cette définition, il est évident que le cas $p = 1$ mène au quantile (voir Equation (IV.1.a.)), tandis que $p = 2$ conduit à l'expectile (Equation (18)). La Figure IV.3 permet de visualiser l'allure de cette famille de fonction de scoring.

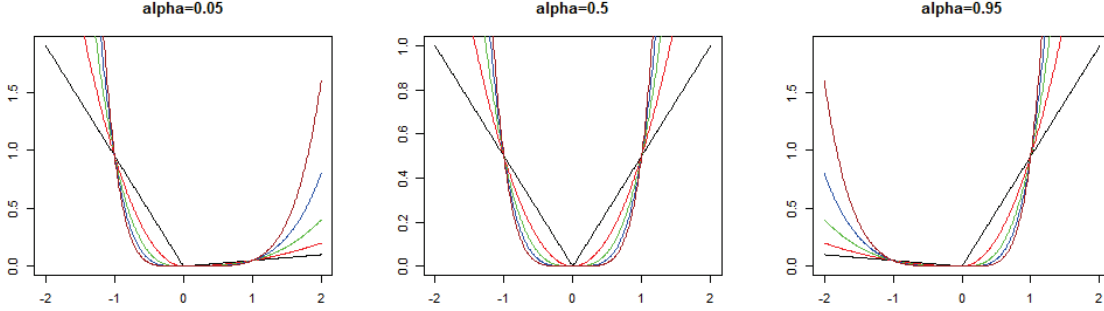


FIGURE IV.3 – Fonctions $S_{1,\alpha}$ en noir, $S_{2,\alpha}$ en rouge, $S_{3,\alpha}$ en vert, $S_{4,\alpha}$ en bleu et $S_{5,\alpha}$ en marron pour des niveaux α respectivement de 0.05, 0.5 et 0.95.

Les propriétés de ces mesures de risque sont notamment discutées dans [Bellini et al., 2014]. En particulier, il est montré que le L_p –quantile est une mesure de risque cohérente si et seulement si $p = 2$, c’est à dire dans le cas de l’expectile. Cependant, les autres cas de figure possèdent les propriétés d’homogénéité et d’invariance par translation. Dans notre contexte, cela signifie que si $Y \sim \mathcal{EC}_1(\mu_Y, \Sigma_Y, \xi)$, alors $q_{p,\alpha}(Y) = \mu_Y + \sqrt{\Sigma_Y} q_{p,\alpha}(Y')$, avec $Y' \sim \mathcal{EC}_1(0, 1, \xi)$ la loi centrée réduite relative à Y . On va alors s’intéresser aux niveaux de L_p –quantiles proches de 1. On introduit pour cela la proposition suivante, donnée dans [Daouia et al., 2017b].

Proposition 44 (Equivalence L_p –quantiles). *Si Y vérifie l’Hypothèse $\mathcal{Q}$ avec $\gamma < \frac{1}{p-1}$, alors :*

$$\frac{q_{p,\alpha}(Y)}{q_\alpha(Y)} \underset{\alpha \rightarrow 1}{\sim} \left[\frac{\gamma}{B(p, \gamma^{-1} - p + 1)} \right]^{-\gamma} = f_L(\gamma, p). \quad (\text{IV.29})$$

On s’aperçoit que les L_p –quantiles sont asymptotiquement proportionnels aux quantiles (ou L_1 –quantiles). Le cas $p = 2$ mène bien au résultat de la Proposition 41. De manière à généraliser nos estimateurs de quantiles en estimateurs de L_p –quantiles extrêmes, on introduit deux conditions supplémentaires, qui nous seront utiles plus loin.

- $(C_{int}^{L_p})$: (C_{int}) est remplie. De plus, $\sqrt{k_n}(1 - \alpha_n) = o(\ln(1 - \alpha_n))$ quand $n \rightarrow +\infty$, et :

$$\lim_{n \rightarrow +\infty} \frac{\sqrt{k_n}}{\ln(1 - \alpha_n)} (1 - \alpha_n)^{\frac{\min\{-\rho, \gamma\}}{\gamma d + 1}} = 0.$$

- $(C_{high}^{L_p})$: (C_{high}) est remplie. De plus,

$$\lim_{n \rightarrow +\infty} \frac{\sqrt{k_n}}{\ln\left(\frac{k_n}{n(1 - \alpha_n)}\right)} (1 - \alpha_n)^{\frac{\min\{-\rho, \gamma\}}{\gamma d + 1}} = 0.$$

Finalement, parmi toutes les conditions que l'on a établies dans la Section III.3.c. puis ici, on peut déduire l'ordre suivant :

$$\left\{ \begin{array}{l} (C_{int}^{L_p}) \Rightarrow (C_{int}^{HG}) \Rightarrow (C_{int}) \\ (C_{high}^{L_p}) \Rightarrow (C_{high}^{HG}) \Rightarrow (C_{high}) \end{array} \right. .$$

Grâce à la Proposition 44 donnons maintenant des estimateurs de L_p -quantiles extrêmes de $Y|(\mathbf{X} = \mathbf{x})$.

Définition 26 (Prédicteurs de L_p -quantiles extrêmes). *Soit $(\alpha_n)_{n \in \mathbb{N}}$ une suite telle que $\alpha_n \rightarrow 1$ quand $n \rightarrow +\infty$. Si $p \leq d+1$ ou $\gamma < \frac{1}{p-d-1}$, on définit :*

$$\left\{ \begin{array}{l} \hat{q}_{p,\alpha_n}(Y|\mathbf{X} = \mathbf{x}) = \mu_{Y|\mathbf{x}} + \sqrt{\Sigma_{Y|\mathbf{X}}} (W_{[n\tilde{v}_n+1]})^{\frac{1}{\hat{\eta}_{k_n}}} f_L \left((\hat{\gamma}_{k_n}^{-1} + d)^{-1}, p \right) \\ \hat{\hat{q}}_{p,\alpha_n}(Y|\mathbf{X} = \mathbf{x}) = \mu_{Y|\mathbf{x}} + \sqrt{\Sigma_{Y|\mathbf{X}}} \left[W_{[k_n+1]} \left(\frac{k_n}{n\tilde{v}_n} \right)^{\hat{\gamma}_{k_n}} \right]^{\frac{1}{\hat{\eta}_{k_n}}} f_L \left((\hat{\gamma}_{k_n}^{-1} + d)^{-1}, p \right) \end{array} \right. ,$$

où les quantités $\mu_{Y|\mathbf{x}}$, $\Sigma_{Y|\mathbf{X}}$ sont définies dans l'Equation (5), la suite $\tilde{v}_n$ dans la Définition 16 et les estimateurs $\hat{\gamma}_{k_n}$, $\hat{\ell}_{k_n, h_n}(\mathbf{x})$ dans les Equations (III.3.b.) et (15). Enfin, W est la première (ou indifféremment n'importe laquelle) composante du vecteur $\Lambda_{\mathbf{X}}^{-1}(\mathbf{X} - \mu_{\mathbf{X}})$.

Si on prend $p = 1$, on obtient bien sûr les estimateurs de quantile (16) et (17), quand $p = 2$ mène aux estimateurs d'expectile $\tilde{e}$ et $\tilde{\tilde{e}}$ de l'Equation (24). On a déjà prouvé la consistance de nos estimateurs de quantiles extrêmes dans la Section III.3.c. De la même manière, on peut en déduire facilement celle de nos estimateurs de L_p -quantiles extrêmes. La preuve n'est pas explicitée. Notons que nous n'avons pour l'instant pas besoin des conditions $(C_{int}^{L_p})$ et $(C_{high}^{L_p})$.

Proposition 45 (Consistances de $\hat{q}_{p,\alpha_n}(Y|\mathbf{X} = \mathbf{x})$ et $\hat{\hat{q}}_{p,\alpha_n}(Y|\mathbf{X} = \mathbf{x})$). *Supposons que l'Hypothèse 2 avec $p \leq d+1$ ou $\gamma < \frac{1}{p-d-1}$ et la condition (C) soient satisfaites. Sous les conditions supplémentaires (C_{int}) et (C_{high}) respectivement, $\hat{q}_{p,\alpha_n}(Y|\mathbf{X} = \mathbf{x})$ et $\hat{\hat{q}}_{p,\alpha_n}(Y|\mathbf{X} = \mathbf{x})$ sont consistants, i.e. :*

$$\left\{ \begin{array}{l} \frac{\hat{q}_{p,\alpha_n}(Y|\mathbf{X}=\mathbf{x})}{q_{p,\alpha_n}(Y|\mathbf{X}=\mathbf{x})} \xrightarrow{\mathbb{P}} 1 \\ \frac{\hat{\hat{q}}_{p,\alpha_n}(Y|\mathbf{X}=\mathbf{x})}{q_{p,\alpha_n}(Y|\mathbf{X}=\mathbf{x})} \xrightarrow{\mathbb{P}} 1 \end{array} \right. .$$

Utilisons maintenant nos conditions $(C_{int}^{L_p})$ et $(C_{high}^{L_p})$. Combinées à une expansion de second ordre de l'équivalence (44) que l'on peut trouver dans [Daouia et al., 2017b], on peut généraliser les normalités asymptotiques des Propositions 22 et 24. Ce résultat est bien sûr valable sous l'Hypothèse 3

IV.4.. AUTRES MESURES DE RISQUE EXTRÊMES BASÉES SUR LE QUANTILE 91

Proposition 46 (Normalités asymptotiques de $\hat{q}_{p,\alpha_n}(Y|\mathbf{X}=\mathbf{x})$ et $\hat{\hat{q}}_{p,\alpha_n}(Y|\mathbf{X}=\mathbf{x})$). Supposons que l'Hypothèse [3](#) avec $p \leq d+1$ ou $\gamma < \frac{1}{p-d-1}$ et la condition (C) soient satisfaites. Sous les conditions supplémentaires (C_{int}^{Lp}) et (C_{high}^{Lp}) respectivement, et si $p > 1$, alors :

$$\left\{ \begin{array}{l} \frac{\sqrt{k_n}}{\ln(1-\alpha_n)} \left(\frac{\hat{q}_{p,\alpha_n}(Y|\mathbf{X}=\mathbf{x})}{q_{p,\alpha_n}(Y|\mathbf{X}=\mathbf{x})} - 1 \right) \xrightarrow{n \rightarrow +\infty} \mathcal{N}\left(0, \frac{d^2 \gamma^4}{(\gamma d+1)^4}\right) \\ \frac{\sqrt{k_n}}{\ln\left(\frac{k_n}{n(1-\alpha_n)}\right)} \left(\frac{\hat{\hat{q}}_{p,\alpha_n}(Y|\mathbf{X}=\mathbf{x})}{q_{p,\alpha_n}(Y|\mathbf{X}=\mathbf{x})} - 1 \right) \xrightarrow{n \rightarrow +\infty} \mathcal{N}\left(0, \left(\frac{\gamma}{\gamma d+1} - \theta \frac{d\gamma^2}{(\gamma d+1)^2}\right)^2\right) \end{array} \right.$$

Preuve : Dans cette preuve, on note $Y_0|\mathbf{x} = (Y|(\mathbf{X}=\mathbf{x}) - \mu_{Y|\mathbf{x}}) \Sigma_{Y|\mathbf{x}}^{-1/2}$ la variable centrée réduite relative à $Y|(\mathbf{X}=\mathbf{x})$. Avec cette notation, on a $q_{p,\alpha}(Y|\mathbf{X}=\mathbf{x}) = \mu_{Y|\mathbf{x}} + \sqrt{\Sigma_{Y|\mathbf{x}}} q_{p,\alpha}(Y_0|\mathbf{x})$. On rappelle ensuite que (C_{int}^{Lp}) entraîne (C_{int}^{HG}) . De plus, on a la décomposition suivante :

$$\begin{aligned} & \frac{\sqrt{k_n}}{\ln(1-\alpha_n)} \left(\frac{\hat{q}_{p,\alpha_n}(Y|\mathbf{X}=\mathbf{x})}{q_{p,\alpha_n}(Y|\mathbf{X}=\mathbf{x})} - 1 \right) \xrightarrow{n \rightarrow +\infty} \frac{\sqrt{k_n}}{\ln(1-\alpha_n)} \left(\frac{(W_{[n\bar{v}_n+1]})^{\frac{1}{\eta_{k_n}}}}{q_{p,\alpha_n}(Y_0|\mathbf{x})} f_L\left(\left(\hat{\gamma}_{k_n}^{-1} + d\right)^{-1}, p\right) - 1 \right) \\ &= \frac{\sqrt{k_n}}{\ln(1-\alpha_n)} \left(\frac{f_L\left(\left(\hat{\gamma}_{k_n}^{-1} + d\right)^{-1}, p\right)}{f_L\left(\left(\gamma^{-1} + d\right)^{-1}, p\right)} - 1 \right) \frac{(W_{[n\bar{v}_n+1]})^{\frac{1}{\eta_{k_n}}}}{\Phi_{Y|\mathbf{x}}^{-1}(\alpha_n)} \frac{f_L\left(\left(\gamma^{-1} + d\right)^{-1}, p\right) \Phi_{Y|\mathbf{x}}^{-1}(\alpha_n)}{q_{p,\alpha_n}(Y_0|\mathbf{x})} \\ &+ \frac{\sqrt{k_n}}{\ln(1-\alpha_n)} \left(\frac{(W_{[n\bar{v}_n+1]})^{\frac{1}{\eta_{k_n}}}}{\Phi_{Y|\mathbf{x}}^{-1}(\alpha_n)} - 1 \right) \frac{f_L\left(\left(\gamma^{-1} + d\right)^{-1}, p\right) \Phi_{Y|\mathbf{x}}^{-1}(\alpha_n)}{q_{p,\alpha_n}(Y_0|\mathbf{x})} + \frac{\sqrt{k_n}}{\ln(1-\alpha_n)} \left(\frac{f_L\left(\left(\gamma^{-1} + d\right)^{-1}, p\right) \Phi_{Y|\mathbf{x}}^{-1}(\alpha_n)}{q_{p,\alpha_n}(Y_0|\mathbf{x})} - 1 \right) \end{aligned}$$

On sait que $f_L\left(\left(\hat{\gamma}_{k_n}^{-1} + d\right)^{-1}, p\right)$, étant une fonction de $\hat{\gamma}_{k_n}$, est asymptotiquement normale avec un taux de convergence $\sqrt{k_n}$ (voir l'Equation [\(III.3.b.\)](#)). Le premier terme tend alors clairement vers 0 quand $n \rightarrow +\infty$. Avec la Proposition [22](#) le second terme tend vers la loi normale en [\(22\)](#). Finalement, il nous reste à vérifier que le troisième et dernier terme tend vers 0. Pour cela, on a recours à l'expansion de second ordre donnée dans [\[Daouia et al., 2017b\]](#) :

$$\frac{q_{p,\alpha_n}(Y_0|\mathbf{x})}{f_L\left(\left(\gamma^{-1} + d\right)^{-1}, p\right) \Phi_{Y|\mathbf{x}}^{-1}(\alpha_n)} = 1 - (\gamma^{-1} + d)^{-1} r(\alpha_n, p) + (\lambda + o(1)) A^*\left(\frac{1}{1-\alpha_n}\right),$$

où $r(\alpha_n, p) = \lambda_1 \frac{1}{\Phi_{Y|\mathbf{x}}^{-1}(\alpha_n)} (\mathbb{E}[Y_0|\mathbf{x}] + o(1)) + \lambda_2 A^*\left(\frac{1}{1-\alpha_n}\right) (1 + o(1))$, $\lambda, \lambda_1, \lambda_2 \in \mathbb{R}$ étant indépendants de n et $A^*(t)$ est la fonction auxiliaire de $\Phi_{Y|\mathbf{x}}\left(1 - \frac{1}{t}\right)$. On rappelle que $Y|(\mathbf{X}=\mathbf{x})$ est à variation régulière d'indice $\gamma^{-1} + d > 1$, d'où $\mathbb{E}[Y_0|\mathbf{x}]$ existe et, $Y_0|\mathbf{x}$

étant symétrique, vaut 0. Une condition suffisante pour la normalité asymptotique peut alors être

$$\begin{cases} \lim_{n \rightarrow +\infty} \frac{\sqrt{k_n}}{\ln(1-\alpha_n)\Phi_{Y|\mathbf{x}}^{-1}(\alpha_n)} = 0 \\ \lim_{n \rightarrow +\infty} \frac{\sqrt{k_n}}{\ln(1-\alpha_n)} A^*\left(\frac{1}{1-\alpha_n}\right) = 0 \end{cases}.$$

On sait, avec l'Hypothèse [3](#) et la preuve de la Proposition [24](#) que $\Phi_{Y|\mathbf{x}}^{-1}(\alpha_n)$ est asymptotiquement proportionnel à $(1-\alpha_n)^{-\frac{\gamma}{\gamma d+1}}$, tandis que $A^*\left(\frac{1}{1-\alpha_n}\right)$ est lui asymptotiquement proportionnel à $(1-\alpha_n)^{-\frac{\rho}{\gamma d+1}}$ si $\rho > -2\gamma$ et à $(1-\alpha_n)^{\frac{2\gamma}{\gamma d+1}}$ sinon. Finalement, il n'est pas compliqué de voir que $\left(C_{int}^{L_p}\right)$ mène à la nullité des deux limites, et par conséquent du dernier terme de la décomposition, d'où le premier résultat. La preuve est exactement la même pour le second, en remplaçant $\hat{q}_{p,\alpha_n}(Y|\mathbf{X}=\mathbf{x})$ par $\hat{q}_{p,\alpha_n}(Y|\mathbf{X}=\mathbf{x})$, $\ln(1-\alpha_n)$ par $\ln\left(\frac{k_n}{n(1-\alpha_n)}\right)$ et en se servant de la Proposition [24](#) plutôt que [22](#) $\square$

IV.4.b. Mesures d'Haezendonck-Goovaerts

D'abord introduites comme les mesures d'Orlicz dans [\[Haezendonck and Goovaerts, 1982\]](#), puis d'Haezendonck dans [\[Goovaerts et al., 2004\]](#), et enfin d'Haezendonck-Goovaerts dans [\[Tang and Yang, 2012\]](#), cette famille de mesures de risque est vue par certains auteurs comme une alternative au quantile. Définissons d'abord cette notion.

Définition 27 (Mesures de risque d'Haezendonck-Goovaerts). *Soit Y une variable aléatoire réelle, et ϕ une fonction convexe positive telle $\phi(0) = 0$, $\phi(1) = 1$ et $\phi(+\infty) = +\infty$. La mesure d'Haezendonck-Goovaerts de Y de niveau $\alpha \in]0, 1[$ associée à ϕ est donnée par l'équation suivante :*

$$H_\alpha(Y) = \inf_{y \in \mathbb{R}} \{y + H_\alpha(Y, y)\},$$

où $H_\alpha(Y, y)$ est l'unique solution h de l'équation :

$$\mathbb{E} \left[\phi \left(\frac{(Y - y)_+}{h} \right) \right] = 1 - \alpha.$$

ϕ est appelée fonction de Young.

Le cas particulier $\phi(t) = t$ mène à la Tail Value at Risk de niveau α $\text{TVaR}_\alpha(Y)$, présentée dans la Définition [4](#). Cette mesure a l'avantage d'être sous-additive et connaît de nombreuses applications, notamment en finance et actuariat (voir par exemple [\[Bargès et al., 2009\]](#)). D'ailleurs, [\[Landsman and Valdez, 2003\]](#) s'est intéressé à la TVaR pour les distributions elliptiques. Plus généralement, [\[Bellini and Rosazza Gianin, 2008\]](#) ont montré que les mesures d'Haezendonck-Goovaerts étaient des mesures de risque cohérentes. Cela implique pour nous que si $Y \sim \mathcal{EC}_1(\mu_Y, \Sigma_Y, \xi)$, alors $H_\alpha(Y) = \mu_Y + \sqrt{\Sigma_Y} H_\alpha(Y')$, avec $Y' \sim$

$\mathcal{EC}_1(0, 1, \xi)$ la loi centrée réduite relative à Y . On va alors s'intéresser aux niveaux extrêmes, c'est à dire lorsque α tend vers 1.

Pour une certaine classe de fonctions de Young, [Tang and Yang, 2012] ont montré qu'à l'image des L_p -quantiles, les mesures d'Haezendonck-Goovaerts sont asymptotiquement proportionnelles aux quantiles, d'où la proposition ci-dessous.

Proposition 47 (Equivalence des mesures d'HG). *Si Y vérifie l'Hypothèse 2 avec $\gamma < \frac{1}{p}$, et en prenant une fonction de Young $\phi(t) = t^p$, alors on a la relation suivante :*

$$\frac{H_{p,\alpha}(Y)}{q_\alpha(Y)} \underset{\alpha \rightarrow 1}{\sim} \frac{\gamma^{-1}(\gamma^{-1} - p)^{p\gamma-1}}{p^{\gamma(p-1)}} B(\gamma^{-1} - p, p)^\gamma = f_H(\gamma, p). \quad (\text{IV.30})$$

L'exemple trivial de la TVaR donne $\text{TVaR}_\alpha(Y) \sim (1 - \gamma)^{-1} q_\alpha(Y)$ quand $\alpha \rightarrow 1$. En utilisant le Lemme 37 les estimateurs de quantile extrêmes des Définitions 16, 17 et la Proposition 47 on peut déduire des estimateurs extrêmes de mesure d'Haezendonck-Goovaerts $H_{p,\alpha}(Y|\mathbf{X} = \mathbf{x})$ (avec fonction de Young $\phi(t) = t^p, p \geq 1$) de la loi conditionnelle $Y|\mathbf{X} = \mathbf{x}$.

Définition 28 (Prédicteurs de mesures d'HG extrêmes). *Soit $(\alpha_n)_{n \in \mathbb{N}}$ une suite telle que $\alpha_n \rightarrow 1$ quand $n \rightarrow +\infty$. Si soit $p \leq d$, soit $\gamma < \frac{1}{p-d}$, on définit :*

$$\begin{cases} \hat{H}_{p,\alpha_n}(Y|\mathbf{X} = \mathbf{x}) = \mu_{Y|\mathbf{x}} + \sqrt{\Sigma_{Y|\mathbf{X}}}(W_{[n\tilde{v}_n+1]})^{\frac{1}{\hat{\eta}_{k_n}}} f_H\left((\hat{\gamma}_{k_n}^{-1} + d)^{-1}, p\right) \\ \hat{\hat{H}}_{p,\alpha_n}(Y|\mathbf{X} = \mathbf{x}) = \mu_{Y|\mathbf{x}} + \sqrt{\Sigma_{Y|\mathbf{X}}}\left[W_{[k_n+1]}\left(\frac{k_n}{n\tilde{v}_n}\right)^{\hat{\gamma}_{k_n}}\right]^{\frac{1}{\hat{\eta}_{k_n}}} f_H\left((\hat{\gamma}_{k_n}^{-1} + d)^{-1}, p\right) \end{cases}, \quad (\text{IV.31})$$

où les quantités $\mu_{Y|\mathbf{x}}, \Sigma_{Y|\mathbf{X}}$ sont définies dans l'Equation 5, la suite $\tilde{v}_n$ dans la Définition 16 et les estimateurs $\hat{\gamma}_{k_n}, \hat{\ell}_{k_n, h_n}(\mathbf{x})$ dans les Equations (III.3.b.) et (15). Enfin, W est la première (ou indifféremment n'importe laquelle) composante du vecteur $\Lambda_{\mathbf{X}}^{-1}(\mathbf{X} - \mu_{\mathbf{X}})$.

La condition $p \leq d$ ou $\gamma < \frac{1}{p-d}$ assure simplement l'existence de la mesure de risque conditionnelle. En utilisant les résultats de consistance des Théorèmes 21 et 23, la consistance de ces estimateurs est immédiate. La preuve étant assez évidente, elle n'est pas fournie.

Proposition 48 (Consistance de $\hat{H}_{p,\alpha_n}(Y|\mathbf{X} = \mathbf{x})$ et $\hat{\hat{H}}_{p,\alpha_n}(Y|\mathbf{X} = \mathbf{x})$). *Supposons que l'Hypothèse 2 et la condition (C) soient satisfaites avec $p \leq d$ ou $\gamma < \frac{1}{p-d}$. Sous les conditions (C_{int}) et (C_{high}) respectivement, $\hat{H}_{p,\alpha_n}(Y|\mathbf{X} = \mathbf{x})$ et $\hat{\hat{H}}_{p,\alpha_n}(Y|\mathbf{X} = \mathbf{x})$ sont consistants, i.e. :*

$$\begin{cases} \frac{\hat{H}_{p,\alpha_n}(Y|\mathbf{X}=\mathbf{x})}{H_{p,\alpha_n}(Y|\mathbf{X}=\mathbf{x})} \xrightarrow{\mathbb{P}} 1 \\ \frac{\hat{\hat{H}}_{p,\alpha_n}(Y|\mathbf{X}=\mathbf{x})}{H_{p,\alpha_n}(Y|\mathbf{X}=\mathbf{x})} \xrightarrow{\mathbb{P}} 1 \end{cases}.$$

En renforçant l'Hypothèse [2] on peut également généraliser la normalités asymptotiques [22] et [24]. Pour cela, on a besoin d'une expansion de second ordre de l'Equation [47]. Cette dernière est donnée dans le Théorème 4.5 de [Mao and Hu, 2012]. Avec tous ces éléments, il vient facilement la proposition ci-dessous.

Proposition 49 (Normalités asymptotiques de $\hat{H}_{p,\alpha_n}(Y|\mathbf{X}=\mathbf{x})$ et $\hat{H}_{p,\alpha_n}(Y|\mathbf{X}=\mathbf{x})$). *Supposons que l'Hypothèse [3] et la condition (C) soient satisfaites avec $p \leq d$ ou $\gamma < \frac{1}{p-d}$. Sous les conditions supplémentaires (C_{int}^{HG}) et (C_{high}^{HG}) respectivement, on a :*

$$\left\{ \begin{array}{l} \frac{\sqrt{k_n}}{\ln(1-\alpha_n)} \left(\frac{\hat{H}_{p,\alpha_n}(Y|\mathbf{X}=\mathbf{x})}{H_{p,\alpha_n}(Y|\mathbf{X}=\mathbf{x})} - 1 \right) \xrightarrow{n \rightarrow +\infty} \mathcal{N}\left(0, \frac{d^2 \gamma^4}{(\gamma d + 1)^4}\right) \\ \frac{\sqrt{k_n}}{\ln\left(\frac{k_n}{n(1-\alpha_n)}\right)} \left(\frac{\hat{H}_{p,\alpha_n}(Y|\mathbf{X}=\mathbf{x})}{H_{p,\alpha_n}(Y|\mathbf{X}=\mathbf{x})} - 1 \right) \xrightarrow{n \rightarrow +\infty} \mathcal{N}\left(0, \left(\frac{\gamma}{\gamma d + 1} - \theta \frac{d \gamma^2}{(\gamma d + 1)^2}\right)^2\right) \end{array} \right. \quad (\text{IV.32})$$

Preuve : Pour cette preuve, on note $Y'|\mathbf{X}=\mathbf{x} \stackrel{\mathcal{L}}{=} (Y|\mathbf{X}=\mathbf{x}) - \mu_{Y|\mathbf{x}} \Sigma_{Y|\mathbf{x}}^{-1/2}$ la variable aléatoire centrée réduite relative à $Y|\mathbf{X}=\mathbf{x}$. D'où les relations :

$$\left\{ \begin{array}{l} \hat{H}_{p,\alpha_n}(Y'|\mathbf{X}=\mathbf{x}) = f_H\left((\hat{\gamma}_{k_n}^{-1} + d)^{-1}, p\right) (W_{[n\tilde{v}_n+1]})^{\frac{1}{\tilde{\eta}_{k_n}}} \\ H_{p,\alpha_n}(Y'|\mathbf{X}=\mathbf{x}) \underset{n \rightarrow +\infty}{\sim} f_H\left((\gamma^{-1} + d)^{-1}, p\right) \Phi_{Y|\mathbf{x}}^{-1}(\alpha_n) \end{array} \right.$$

On a alors la décomposition suivante :

$$\begin{aligned} & \frac{\sqrt{k_n}}{\ln(1-\alpha_n)} \left(\frac{\hat{H}_{p,\alpha_n}(Y|\mathbf{X}=\mathbf{x})}{H_{p,\alpha_n}(Y|\mathbf{X}=\mathbf{x})} - 1 \right) \underset{n \rightarrow +\infty}{\sim} \frac{\sqrt{k_n}}{\ln(1-\alpha_n)} \left(\frac{\hat{H}_{p,\alpha_n}(Y'|\mathbf{X}=\mathbf{x})}{H_{p,\alpha_n}(Y'|\mathbf{X}=\mathbf{x})} - 1 \right) \\ &= \frac{\sqrt{k_n}}{\ln(1-\alpha_n)} \left(\frac{f_H\left((\hat{\gamma}_{k_n}^{-1} + d)^{-1}, p\right)}{f_H\left((\gamma^{-1} + d)^{-1}, p\right)} - 1 \right) \frac{(W_{[n\tilde{v}_n+1]})^{\frac{1}{\tilde{\eta}_{k_n}}}}{\Phi_{Y|\mathbf{x}}^{-1}(\alpha_n)} \frac{f_H\left((\gamma^{-1} + d)^{-1}, p\right) \Phi_{Y|\mathbf{x}}^{-1}(\alpha_n)}{H_{p,\alpha_n}(Y'|\mathbf{X}=\mathbf{x})} \\ &+ \frac{\sqrt{k_n}}{\ln(1-\alpha_n)} \left(\frac{(W_{[n\tilde{v}_n+1]})^{\frac{1}{\tilde{\eta}_{k_n}}}}{\Phi_{Y|\mathbf{x}}^{-1}(\alpha_n)} - 1 \right) \frac{f_H\left((\gamma^{-1} + d)^{-1}, p\right) \Phi_{Y|\mathbf{x}}^{-1}(\alpha_n)}{H_{p,\alpha_n}(Y'|\mathbf{X}=\mathbf{x})} + \frac{\sqrt{k_n}}{\ln(1-\alpha_n)} \left(\frac{f_H\left((\gamma^{-1} + d)^{-1}, p\right) \Phi_{Y|\mathbf{x}}^{-1}(\alpha_n)}{H_{p,\alpha_n}(Y'|\mathbf{X}=\mathbf{x})} - 1 \right). \end{aligned}$$

On sait que $f_H\left((\hat{\gamma}_{k_n}^{-1} + d)^{-1}, p\right)$, étant une fonction de $\hat{\gamma}_{k_n}$, est asymptotiquement normal avec un taux de convergence $\sqrt{k_n}$ (voir l'Equation [III.3.b.]). Le premier terme tend alors vers 0 quand $n \rightarrow +\infty$. En se servant de la Proposition [22] le second terme tend vers la loi normale donnée dans [22]. Finalement, il ne reste plus qu'à vérifier que le troisième terme tend vers 0. Pour cela, on utilise le résultat du Théorème 4.5 de [Mao and Hu, 2012], qui montre qu'il existe $\lambda \in \mathbb{R}$ tel que :

$$\frac{H_{p,\alpha_n}(Y'|\mathbf{X}=\mathbf{x})}{f_H\left((\gamma^{-1} + d)^{-1}, p\right) \Phi_{Y|\mathbf{x}}^{-1}(\alpha_n)} = 1 + \lambda A^* \left(\frac{1}{1 - \alpha_n} \right) (1 + o(1)),$$

où A^* est la fonction auxiliaire de $\Phi_{Y|\mathbf{x}}^{-1}(1 - \frac{1}{t})$. Dans la preuve de la Proposition [22] on a vu que $A^*(t)$ était proportionnel soit à $t^{-\frac{2\gamma}{\gamma d+1}}$ si $\rho \leq -2\gamma$ soit à $t^{\frac{\rho}{\gamma d+1}}$ autrement. Les conditions de (C_{int}^{HG}) assurent

$$\lim_{n \rightarrow +\infty} \frac{\sqrt{k_n}}{\ln(1 - \alpha_n)} \ln \left(\frac{H_{p, \alpha_n}(Y'|\mathbf{X} = \mathbf{x})}{f_H((\gamma^{-1} + d)^{-1}, p) \Phi_{Y|\mathbf{x}}^{-1}(\alpha_n)} \right) = 0.$$

D'où le troisième terme tend vers 0, et le premier résultat de ([49]) est prouvé. La preuve est exactement la même pour le second, avec un taux $\frac{\sqrt{k_n}}{\ln(\frac{k_n}{n(1-\alpha_n)})}$ au lieu de $\frac{\sqrt{k_n}}{\ln(1-\alpha_n)}$. Alors les conditions de (C_{high}^{HG}) mènent au résultat attendu. $\square$

On propose dans la Section [IV.5.] un exemple numérique avec $p = 1$, c'est à dire dans le cas de la Tail Value at Risk.

Dans cette section, nous avons donc introduit plusieurs prédicteurs d'expectile, aux résultats asymptotiques sensiblement égaux. Cependant, il existe quelques différences au niveau des conditions. Par exemple, $\hat{e}$ nécessite la condition $\gamma < 1/2$.

Nous avons également introduit des estimateurs d'autres mesures de risque que sont les L_p -quantiles et les mesures d'Haezendonck-Goovaerts. On propose dans la Section [IV.5.] une étude numérique pour tester tous ces estimateurs.

IV.5. Etude numérique

Prenons, comme pour l'étude numérique relative aux quantiles extrêmes, un jeu de données simulées de la loi de Student centrée réduite à $\nu = 2.5$ degrés de liberté en dimension $d + 1 = 4$. On peut facilement calculer les expectiles conditionnels dans ce cas à l'aide de l'Equation ([IV.1.c.2.]). Notons que l'on prendra également $M(\mathbf{x}) = 1$.

IV.5.a. Expectiles intermédiaires

Pour estimer des expectiles intermédiaires, nous disposons de deux estimateurs $\hat{e}_{\alpha_n \uparrow}^E$ et $\tilde{e}_{\alpha_n \uparrow}^E$, donnés respectivement dans les Equations ([22]) et ([24]). Les résultats asymptotiques de ces deux estimateurs requièrent des conditions très proches. On donne alors le même jeu de suites α_n , k_n et h_n pour les deux.

$\hat{e}_{\alpha_n \uparrow}^E$ comme $\tilde{e}_{\alpha_n \uparrow}^E$ ont besoin des conditions (C) , (C_{int}) et (C_{int}^{HG}) pour les résultats de normalité asymptotiques. Si on se réfère à la Section [III.4.] du Chapitre [III] on peut alors choisir nos suites $\alpha_n = 1 - n^{-0.6}$, $k_n = n^{0.4}$ et $h_n = n^{-0.15}$. Le noyau K sera une fois de plus la densité gaussienne.

La Table IV.2 donne les variances mesurées, avec les notations suivantes :

$$\begin{cases} \hat{v}_e = \widehat{\mathbb{V}ar} \left[\frac{\sqrt{k_n}}{\ln(1-\alpha_n)} \left(\frac{\hat{e}_{\alpha_n \uparrow}^E(Y|\mathbf{X}=\mathbf{x})}{e_{\alpha_n}(Y|\mathbf{X}=\mathbf{x})} - 1 \right) \right], \\ \tilde{v}_e = \widehat{\mathbb{V}ar} \left[\frac{\sqrt{k_n}}{\ln(1-\alpha_n)} \left(\frac{\tilde{e}_{\alpha_n \uparrow}^E(Y|\mathbf{X}=\mathbf{x})}{e_{\alpha_n}(Y|\mathbf{X}=\mathbf{x})} - 1 \right) \right]. \end{cases}$$

n	α_n	$\hat{v}_e$	$\tilde{v}_e$
1 000	0.9841511	0.035606589	0.024919204
10 000	0.9960189	0.023860291	0.020516479
100 000	0.999	0.021496576	0.017257715
1 000 000	0.9997488	0.018309767	0.015327578
10 000 000	0.9999369	0.01497986	0.014316220
$+\infty$	1	0.009835394	0.009835394

TABLE IV.2 – Variances asymptotiques mesurées et théorique (pour $n = +\infty$) de nos estimateurs

La Figure IV.4 représente les boxplots de 1 000 estimations de $\hat{e}_{\alpha_n \uparrow}^E(Y|\mathbf{X}=\mathbf{x})$ et $\tilde{e}_{\alpha_n \uparrow}^E(Y|\mathbf{X}=\mathbf{x})$ pour $n = 10\,000\,000$. On voit que nos deux estimateurs ont des performances assez similaires. Notons cependant que $\tilde{e}_{\alpha_n \uparrow}^E(Y|\mathbf{X}=\mathbf{x})$ a l'avantage de ne pas nécessiter la condition $\gamma < \frac{1}{2}$.

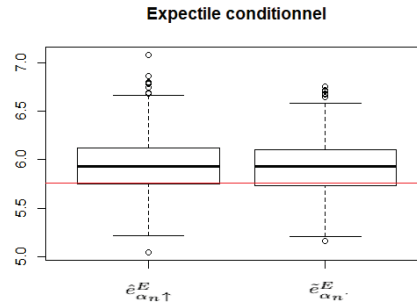


FIGURE IV.4 – Boxplots de 1 000 répétitions de $\hat{e}_{\alpha_n \uparrow}^E(Y|\mathbf{X}=\mathbf{x})$ et $\tilde{e}_{\alpha_n \uparrow}^E(Y|\mathbf{X}=\mathbf{x})$ respectivement, pour $n = 10\,000\,000$. La valeur de $e_{\alpha_n}(Y|\mathbf{X}=\mathbf{x})$ apparaît en rouge.

Comme pour le quantile, le prédicteur $e_{\alpha_n \uparrow}^E(Y|\mathbf{X}=\mathbf{x})$ est relativement éloigné de l'expectile théorique $e_{\alpha_n}(Y|\mathbf{X}=\mathbf{x})$, α_n tendant vers 1 plus lentement pour les expectiles intermédiaires.

IV.5.b. Hauts expectiles

En ce qui concerne les hauts expectiles, nous avons précédemment introduit 3 estimateurs différents : $\hat{e}_{\alpha_n \uparrow}^E$ dans la Définition 23, puis $\tilde{e}_{\alpha_n \uparrow}^E$ et $\tilde{\tilde{e}}_{\alpha_n \uparrow}^E$ dans la Définition 24. On a également observé au travers de leurs résultats de consistance et de normalité asymptotique que nos trois estimateurs ne nécessitaient pas d'hypothèses supplémentaires à celles des hauts quantiles. Nous choisirons alors nos suites $k_n = n^{0.5}$ en conséquence. Nous prendrons la suite $\alpha_n = 1 - n^{-1.1}$ et $h_n = n^{-0.15}$. Le noyau K sera la densité gaussienne. On introduit les notations suivantes :

$$\begin{cases} \hat{v}_e &= \widehat{\text{Var}} \left[\frac{\sqrt{k_n}}{\ln\left(\frac{k_n}{n(1-\alpha_n)}\right)} \left(\frac{\hat{e}_{\alpha_n \uparrow}^E(Y|\mathbf{X}=\mathbf{x})}{e_{\alpha_n}(Y|\mathbf{X}=\mathbf{x})} - 1 \right) \right], \\ \tilde{v}_e &= \widehat{\text{Var}} \left[\frac{\sqrt{k_n}}{\ln\left(\frac{k_n}{n(1-\alpha_n)}\right)} \left(\frac{\tilde{e}_{\alpha_n \uparrow}^E(Y|\mathbf{X}=\mathbf{x})}{e_{\alpha_n}(Y|\mathbf{X}=\mathbf{x})} - 1 \right) \right], \\ \tilde{\tilde{v}}_e &= \widehat{\text{Var}} \left[\frac{\sqrt{k_n}}{\ln\left(\frac{k_n}{n(1-\alpha_n)}\right)} \left(\frac{\tilde{\tilde{e}}_{\alpha_n \uparrow}^E(Y|\mathbf{X}=\mathbf{x})}{e_{\alpha_n}(Y|\mathbf{X}=\mathbf{x})} - 1 \right) \right]. \end{cases}$$

En reprenant ces notations, la Table IV.3 donne un aperçu des variances mesurées - que l'on peut comparer aux théoriques dans la ligne $+\infty$ - de nos 3 estimateurs.

n	α_n	$\hat{v}_e$	$\tilde{v}_e$	$\tilde{\tilde{v}}_e$
1 000	0.9994988	0.00672838	0.00377646	0.003521414
10 000	0.9999602	0.002641393	0.001595622	0.001603951
100 000	0.9999968	0.001507326	0.0009715758	0.001033282
1 000 000	0.9999997	0.001304404	0.0006238518	0.0006808723
10 000 000	0.99999998004	0.0007775403	0.0004313537	0.000452559
$+\infty$	1	0	0	0

TABLE IV.3 – Variances asymptotiques mesurées et théorique (pour $n = +\infty$) de nos estimateurs

Là encore, nos trois estimateurs ont des performances semblables, conformément à leur distribution asymptotique. Les boxplots de la Figure IV.5 mettent bien en évidence ce phénomène. Ajoutons que $\hat{e}_{\alpha_n \uparrow}^E(Y|\mathbf{X}=\mathbf{x})$ a l'inconvénient de nécessiter $\gamma < \frac{1}{2}$.

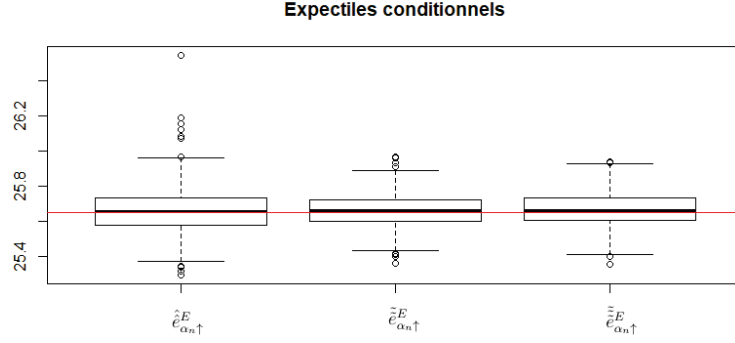


FIGURE IV.5 – Boxplots de 1 000 répétitions de $\hat{e}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})$, $\tilde{e}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})$ et $\tilde{\tilde{e}}_{\alpha_n \uparrow}^E(Y|\mathbf{X} = \mathbf{x})$ respectivement, pour $n = 10\,000\,000$. La valeur de $e_{\alpha_n}(Y|\mathbf{X} = \mathbf{x})$ apparaît en rouge.

Dans le cas des hauts expectiles, notre prédicteur semble donc bien plus efficace que dans le cas des expectiles intermédiaires. Cependant, l'estimateur $\hat{e}$ semble légèrement plus dispersé que les deux autres.

En guise d'exemple de mesure d'Haezendonck-Goovaerts, on propose dans le paragraphe suivant une estimation de Tail Value at Risk.

IV.5.c. TVaR intermédiaires

Comme exemple pour les mesures d'Haezendonck-Goovaerts, on propose le cas trivial $p = 1$ correspondant à la Tail Value at Risk. Toujours dans le cas Student étudié précédemment, il est possible d'obtenir des formules explicites pour la TVaR de $Y|\mathbf{X} = \mathbf{x}$.

$$\text{TVaR}_\alpha(Y|\mathbf{X} = \mathbf{x}) = H_{1,\alpha}(Y|\mathbf{X} = \mathbf{x}) = \frac{1}{1-\alpha} \frac{\Gamma\left(\frac{d+1+\nu}{2}\right)}{\Gamma\left(\frac{d+\nu}{2}\right)} \frac{\sqrt{\nu + M(\mathbf{x})}}{\sqrt{\pi}(\nu + d - 1)} \left(1 + \frac{\Phi_{\nu+d}^{-1}(\alpha)^2}{\nu + d}\right)^{\frac{1-d-\nu}{2}}.$$

Pour l'estimation, on prendra notre estimateur $\hat{H}_{1,\alpha_n}$ de l'Equation (28) avec les suites $h_n = n^{-0.15}$, $k_n = n^{0.45}$ et $\alpha_n = 1 - n^{-0.55}$. La Table IV.4 donne les variances mesurées, avec la notation :

$$\hat{v}_H = \widehat{\mathbb{V}ar} \left[\frac{\sqrt{k_n}}{\ln(1 - \alpha_n)} \left(\frac{\hat{H}_{1,\alpha_n}(Y|\mathbf{X} = \mathbf{x})}{H_{1,\alpha_n}(Y|\mathbf{X} = \mathbf{x})} - 1 \right) \right].$$

n	α_n	$\hat{v}_{\ell(\mathbf{x}),\eta}$	$\hat{v}_H$
1 000	0.9776128	$\begin{pmatrix} 130.0403 & -13.28885 \\ -13.28885 & 1.592599 \end{pmatrix}$	0.015557875
10 000	0.9936904	$\begin{pmatrix} 121.1456 & -12.74622 \\ -12.74622 & 1.436565 \end{pmatrix}$	0.013988046
100 000	0.9982217	$\begin{pmatrix} 120.8767 & -13.04611 \\ -13.04611 & 1.448935 \end{pmatrix}$	0.013671486
1 000 000	0.9994988	$\begin{pmatrix} 119.9793 & -13.53169 \\ -13.53169 & 1.537808 \end{pmatrix}$	0.012132873
10 000 000	0.9998587	$\begin{pmatrix} 121.7649 & -13.68233 \\ -13.68233 & 1.54204 \end{pmatrix}$	0.012281728
$+\infty$	1	$\begin{pmatrix} 114.0569 & -12.81569 \\ -12.81569 & 1.44 \end{pmatrix}$	0.009835394

TABLE IV.4 – Variances asymptotiques mesurées et théorique (pour $n = +\infty$) de nos estimateurs.

Comparons les performances de nos estimateurs dans le cas $n = 10\,000\,000$. La Figure IV.5.c. représente les boxplots de nos estimateurs dans ce cas.

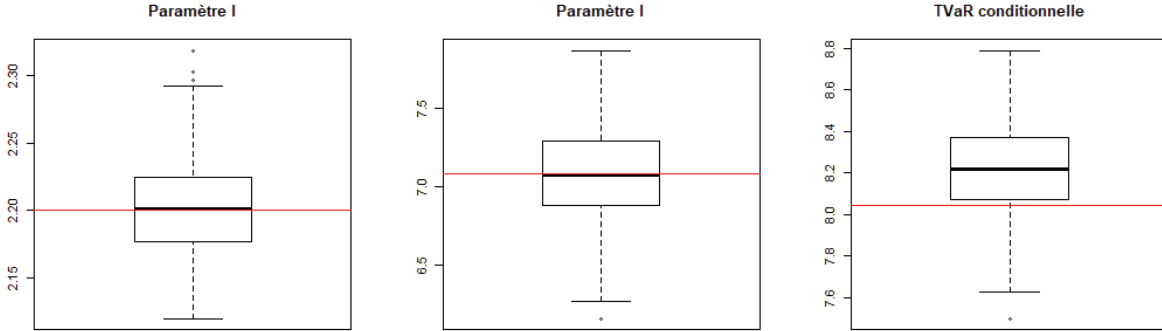


FIGURE IV.6 – Boxplots de 1 000 répétitions de $\hat{\eta}_{k_n}$, $\hat{\ell}_{k_n, h_n}(\mathbf{x})$ et $\hat{H}_{1, \alpha_n}(Y|\mathbf{X} = \mathbf{x})$ respectivement. Les valeurs de η , $\ell(\mathbf{x})$ et $H_{1, \alpha_n}(Y|\mathbf{X} = \mathbf{x})$ apparaissent en rouge.

Il est évident que l'estimation de la TVaR conditionnelle $H_{1, \alpha_n}(Y|\mathbf{X} = \mathbf{x})$ souffre également de la lenteur de la convergence de α_n vers 1. De plus, l'estimation est particulièrement variable. Ces problèmes sont censés s'atténuer pour les TVaR hautes.

IV.5.d. TVaR hautes

Dans le cas des hautes Tail Value at Risk, les conditions étant les mêmes que celles des quantiles extrêmes, on se base sur la Section III.4. pour définir les suites $h_n = n^{-0.15}$, $k_n = n^{0.5}$ et $\alpha_n = 1 - n^{-1.1}$. Un tel jeu de paramètres annule la normalité asymptotique de l'Equation (49). Si on introduit la notation :

$$\hat{v}_H = \widehat{\mathbb{V}ar} \left[\frac{\sqrt{k_n}}{\ln \left(\frac{k_n}{n(1-\alpha_n)} \right)} \left(\frac{\hat{H}_{1,\alpha_n}(Y|\mathbf{X} = \mathbf{x})}{H_{1,\alpha_n}(Y|\mathbf{X} = \mathbf{x})} - 1 \right) \right],$$

la Table IV.5 donne les variances mesurées. On observe alors bien la décroissance de $\hat{v}_H$ avec n .

n	α_n	$\hat{v}_{\ell(\mathbf{x}),\eta}$	$\hat{v}_H$
1 000	0.9994988	$\begin{pmatrix} 115.9570 & -13.17105 \\ -13.17105 & 1.762918 \end{pmatrix}$	0.00292465
10 000	0.9999602	$\begin{pmatrix} 116.4151 & -13.13737 \\ -13.13737 & 1.558384 \end{pmatrix}$	0.001152373
100 000	0.9999968	$\begin{pmatrix} 108.5100 & -12.32441 \\ -12.32441 & 1.42867 \end{pmatrix}$	0.0008032991
1 000 000	0.9999997	$\begin{pmatrix} 115.1339 & -12.95916 \\ -12.95916 & 1.470622 \end{pmatrix}$	0.0005234553
10 000 000	0.99999998004	$\begin{pmatrix} 110.6031 & -12.51083 \\ -12.51083 & 1.419509 \end{pmatrix}$	0.000340293
$+\infty$	1	$\begin{pmatrix} 114.0569 & -12.81569 \\ -12.81569 & 1.44 \end{pmatrix}$	0

TABLE IV.5 – Variances asymptotiques mesurées et théorique (pour $n = +\infty$) de nos estimateurs

Enfin, il n'est pas étonnant d'observer, comme pour les quantiles et expectiles, que les estimateurs sont très efficaces lorsque $n(1 - \alpha_n) \rightarrow 0$. La Figure IV.5.d., qui représente les boxplots des estimateurs dans le cas $n = 10\,000\,000$, en est une bonne illustration.

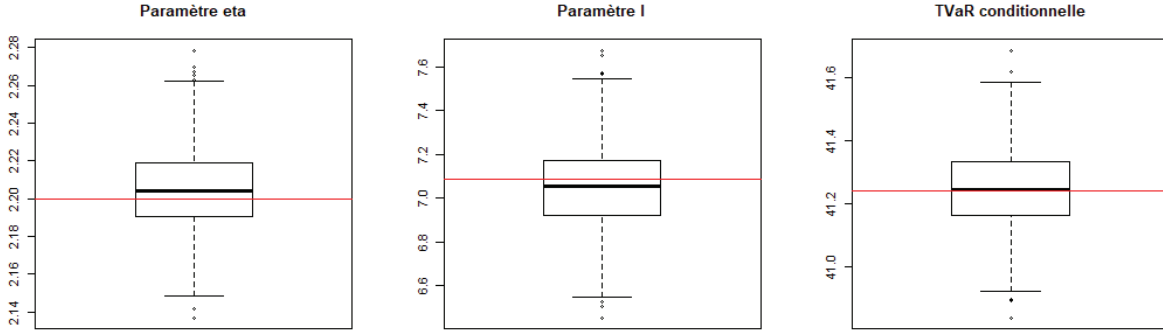


FIGURE IV.7 – Boxplots de 1 000 répétitions de $\hat{\eta}_{k_n}$, $\hat{\ell}_{k_n, h_n}(\mathbf{x})$ et $\hat{H}_{1, \alpha_n}(Y|\mathbf{X} = \mathbf{x})$ respectivement. Les valeurs de η , $\ell(\mathbf{x})$ et $H_{1, \alpha_n}(Y|\mathbf{X} = \mathbf{x})$ apparaissent en rouge.

Dans cette section, nous avons donc pu mettre en évidence l'efficacité de nos estimateurs. En revanche, ces mêmes estimateurs sont beaucoup moins bons dans le cas de la grande dimension. En effet, si on regarde les conditions (C_{int}^{HG}) et (C_{high}^{HG}) , elles ne sont plus satisfaites si d est trop grand. Les résultats asymptotiques de nos estimateurs ne sont alors plus valables. Il faut donc considérer une autre approche dans ce cas. C'est le propos du prochain chapitre.

LE CAS DE LA GRANDE DIMENSION

Lors des deux chapitres précédents, nous avons pu observer que les approches proposées pour l'estimation de mesures de risque extrêmes conditionnelles étaient difficilement applicables lorsque le nombre de covariables d était élevé. En effet, les relations asymptotiques (11) et (33), au coeur de nos approches, peuvent s'avérer très lentes, et donc inutilisables dans ce cas de figure. Dans ce chapitre, nous proposons alors une alternative à ces approches dans le cas où d tend vers $+\infty$. Un tel cadre mène à plusieurs problèmes très présents dans la littérature, le premier d'entre eux étant l'estimation de la matrice de covariance Σ .

La structure de ce chapitre est la suivante : dans un premier temps, nous montrons que si le nombre de covariables tend vers l'infini, alors la distribution conditionnelle tend vers une loi gaussienne. L'estimation du quantile devient alors aisée, encore faut-il être capable d'estimer les paramètres μ et Σ de notre distribution. C'est pourquoi nous détaillons par la suite les différentes méthodes d'estimation des paramètres en grande dimension. Une attention toute particulière est donnée à l'estimation de la matrice de covariance inverse Σ^{-1} . Une partie de ces méthodes est directement empruntée à la littérature existante, et nous en proposons quelques unes adaptées à notre problème. Enfin, un exemple numérique est proposé, toujours dans le cas commode de la distribution de Student pour lequel les valeurs théoriques sont explicites.

V.1. Approximation gaussienne

Nous proposons dans ce chapitre d'approximer la loi conditionnelle par une loi gaussienne dans le cas de la grande dimension. Une approche similaire peut notamment être trouvée dans des papiers tels que [Hashorva and Jaworski, 2012] ou [Hashorva, 2006a], mais pas en grande dimension.

Comme pour les chapitres précédents, on considère un vecteur $\mathbf{Z} = (\mathbf{X}, Y) \sim \mathcal{EC}_{d+1}(\mu, \Sigma, \xi)$ avec $\mathbf{X} \in \mathbb{R}^d$, $Y \in \mathbb{R}$ tels que

$$\Sigma = \begin{pmatrix} \Sigma_{\mathbf{X}} & \Sigma_{\mathbf{X}Y} \\ \Sigma_{Y\mathbf{X}} & \Sigma_Y \end{pmatrix}, \mu = \begin{pmatrix} \mu_{\mathbf{X}} \\ \mu_Y \end{pmatrix},$$

et on s'intéressera à la loi conditionnelle $Y|\mathbf{X} = \mathbf{x}$, $\mathbf{x} \in \mathbb{R}^d$. La différence ici est que d tend vers l'infini lorsque la taille n de l'échantillon d'apprentissage $\mathbf{Z}_n, \dots, \mathbf{Z}_n$ tend vers $+\infty$. On notera alors plutôt d_n . On rappelle, en utilisant l'Equation (5), que la loi conditionnelle $Y|\mathbf{X} = \mathbf{x}$ étant elliptique, on a la représentation stochastique suivante :

$$Y|(\mathbf{X} = \mathbf{x}) \stackrel{\mathcal{L}}{=} \mu_{Y|\mathbf{x}} + \sqrt{\Sigma_{Y|\mathbf{X}}} R^* U^{(1)},$$

avec

$$R^* \stackrel{\mathcal{L}}{=} R\sqrt{1-\beta} \left\{ R\sqrt{\beta} \mathbf{U}^{(d_n)} = \mathbf{C}_{\mathbf{X}}^{-1}(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{X}}) \right\}.$$

On va donc s'intéresser au comportement de la loi symétrique $R^*U^{(1)}$ lorsque $d_n \rightarrow +\infty$. On notera alors φ_{R^*} sa densité. Avant de calculer la limite de cette densité lorsque n tend vers $+\infty$, il paraît important d'ajouter quelques précisions. Si jusqu'à maintenant la quantité $M(\mathbf{x}) = (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{X}})^\top \boldsymbol{\Sigma}_{\mathbf{X}}^{-1}(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{X}})$ était considérée déterministe (d était fixé), on voit maintenant cette distance comme une quantité aléatoire qui évolue avec n , et d_n . On donne le résultat intermédiaire suivant sur les propriétés asymptotiques de la distance de Mahalanobis $M(\mathbf{x})$.

Lemme 50 (Propriétés de $M(\mathbf{x})$). *On a les résultats suivants lorsque $n \rightarrow +\infty$:*

- (i) $M(\mathbf{x}) \rightarrow +\infty$ presque sûrement.
- (ii) $\frac{M(\mathbf{x})}{d_n} \rightarrow \xi^2$ presque sûrement.

Preuve :

- (i) Il n'est pas compliqué de remarquer que $M(\mathbf{x}) \stackrel{\mathcal{L}}{=} \xi^2 \chi_{d_n}^2$. De plus, un rapide calcul donne

$$\mathbb{E}[\chi_1^2 \wedge 1] = \frac{2^{5/2} \Gamma(\frac{5}{2})}{\sqrt{2\pi}} \mathbb{P}(\chi_1^2 \leq 1) + \mathbb{P}(\chi_1^2 \geq 1) > 0.$$

Etant donné que $\chi_{d_n}^2$ peut être vu comme une somme de d_n loi du χ^2 indépendantes à 1 degré de liberté, et que $\sum_n^{+\infty} \mathbb{E}[\chi_1^2 \wedge 1] = +\infty$, le Théorème 3.17 ainsi que le Corollaire 2.14 de [Kallenberg, 2006] indiquent que $\chi_{d_n}^2$, et par conséquent $M(\mathbf{x})$, divergent presque sûrement.

- (ii) La loi forte des grands nombres implique une convergence presque sûre de $\chi_{d_n}^2/d_n$ vers 1, ce qui donne directement le résultat attendu.

□

A l'aide du lemme précédent, intéressons nous à la loi limite de φ_{R^*} . On rappelle que φ_{R^*} dépendant de $M(\mathbf{x})$, c'est également une variable aléatoire. Si il est connu que la loi de Student à ν degrés de liberté tend vers une loi gaussienne lorsque ν tend vers $+\infty$, la loi Slash connaît le même phénomène avec son paramètre a . Dès lors, il paraît intéressant de se demander si cette limite gaussienne est commune à tous les Mélanges d'Echelle Gaussiens à queue lourde. La proposition suivante apporte une première réponse.

Proposition 51 (Loi limite). *Sous l'Hypothèse [2](#), on a $\forall t \in \mathbb{R}$:*

$$\varphi_{R^*}(t) \rightarrow \frac{1}{\sqrt{2\pi}\xi} e^{-\frac{t^2}{2\xi^2}} \text{ presque sûrement quand } n \rightarrow +\infty. \quad (\text{V.1})$$

Preuve : Un simple développement de la représentation stochastique (5) mène à

$$f_{R^*}(t) = \frac{2f_{R^2}(M(\mathbf{x}) + t^2)}{f_{R^2\beta}(M(\mathbf{x})) B\left(\frac{d_n}{2}, \frac{1}{2}\right) M(\mathbf{x})} \left(1 + \frac{t^2}{M(\mathbf{x})}\right)^{-1-d_n/2} \left(1 - \frac{M(\mathbf{x})}{M(\mathbf{x}) + t^2}\right)^{-1/2}.$$

A l'aide du Lemme 50, on a $\left(1 - \frac{M(\mathbf{x})}{M(\mathbf{x}) + t^2}\right)^{-1/2} \sim \sqrt{M(\mathbf{x})}/t$ quand $n \rightarrow +\infty$ presque sûrement. De plus, on s'aperçoit que quand $n \rightarrow +\infty$,

$$\left(1 + \frac{d_n}{2}\right) \ln \left(1 + \frac{t^2}{M(\mathbf{x})}\right) \sim \frac{d_n}{M(\mathbf{x})} \frac{t^2}{2} \rightarrow \frac{t^2}{2\xi^2} \text{ presque sûrement,}$$

d'où $\left(1 + \frac{t^2}{M(\mathbf{x})}\right)^{-1-d_n/2} \rightarrow \exp\left(-\frac{t^2}{2\xi^2}\right)$ presque sûrement quand $n \rightarrow +\infty$. Enfin, on cherche une équivalence entre la densité de R^2 et celle de sa contraction $R^2\beta$. Pour cela, on utilise le résultat (4.20) de [Hashorva and Pakes, 2010], et on obtient sous l'Hypothèse 2:

$$f_{R^2\beta}(t) \underset{t \rightarrow +\infty}{\sim} \frac{\Gamma\left(\frac{d_n+1}{2}\right) \Gamma\left(\frac{d_n}{2} + \frac{1}{2\gamma}\right)}{\Gamma\left(\frac{d_n}{2}\right) \Gamma\left(\frac{d_n+1}{2} + \frac{1}{2\gamma}\right)} f_{R^2}(t).$$

On arrive alors à

$$f_{R^*}(t) \sim \frac{2}{\Gamma\left(\frac{1}{2}\right)} \frac{f_{R^2}(M(\mathbf{x}) + t^2)}{f_{R^2}(M(\mathbf{x}))} \frac{\Gamma\left(\frac{d_n+1}{2} + \frac{1}{2\gamma}\right)}{\Gamma\left(\frac{d_n}{2} + \frac{1}{2\gamma}\right)} M(\mathbf{x})^{-1/2} e^{-\frac{t^2}{2\xi^2}}, \forall t \in \mathbb{R} \text{ p.s. quand } n \rightarrow +\infty.$$

Une conséquence immédiate de l'Hypothèse 2 est que f_{R^2} conserve les équivalences, d'où le ratio des deux densités tend clairement vers 1. De plus, il est facile de voir que le ratio des deux fonctions gamma est asymptotiquement équivalent à $\sqrt{d_n}/2$. En utilisant le résultat (ii) du Lemme 50, ainsi que la relation $\varphi_{R^*} = f_{R^*}/2$, on trouve asymptotiquement le résultat (51). $\square$

Ce résultat est assez intéressant, et ce pour plusieurs raisons. En effet, on s'aperçoit que conditionnellement à la variable d'échelle ξ , la loi de $Y|\mathbf{X} = \mathbf{x}$ est une loi normale de paramètres $\mu_{Y|\mathbf{x}}$ et $\xi^2 \Sigma_{Y|\mathbf{X}}$. Cette loi conditionnelle limite dépend finalement uniquement de la variable d'échelle de la loi elliptique inconditionnelle. Si on a alors suffisamment d'observations pour estimer la valeur de la variable d'échelle $\xi_{\mathbf{x}}$ issue de notre vecteur $\mathbf{x}$ (ce qui est le cas avec $\mathbf{x}$ en grande dimension), on sera capable de fournir une bonne approximation de la loi conditionnelle, et par conséquent du quantile conditionnel.

Dès lors, en remarquant que $(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{X}})^\top \boldsymbol{\Sigma}_{\mathbf{X}}^{-1} (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{X}}) | (\xi = \xi_{\mathbf{x}}) \stackrel{\mathcal{L}}{\sim} \chi_{d_n}^2 \xi_{\mathbf{x}}^2$, ainsi que $\chi_{d_n}^2/d_n \xrightarrow{p.s.} 1$, on peut proposer les estimateurs suivants pour $\xi_{\mathbf{x}}$ ainsi que $q_\alpha(Y|\mathbf{X} = \mathbf{x})$.

Définition 29 (Approximation du quantile conditionnel). *On introduit l'estimateur suivant :*

$$\hat{\xi}_{\mathbf{x}, d_n} = \sqrt{\frac{(\mathbf{x} - \hat{\boldsymbol{\mu}}_{\mathbf{X}})^\top \hat{\boldsymbol{\Sigma}}_{\mathbf{X}}^{-1} (\mathbf{x} - \hat{\boldsymbol{\mu}}_{\mathbf{X}})}{d_n}}$$

où $\hat{\boldsymbol{\mu}}_{\mathbf{X}}$ et $\hat{\boldsymbol{\Sigma}}_{\mathbf{X}}$ sont des estimateurs appropriés de $\boldsymbol{\mu}_{\mathbf{X}}$ et $\boldsymbol{\Sigma}_{\mathbf{X}}$. On en déduit l'estimateur de quantile conditionnel :

$$\bar{q}_{\alpha}(Y|\mathbf{X} = \mathbf{x}) = \hat{\mu}_{Y|\mathbf{x}} + \sqrt{\hat{\Sigma}_{Y|\mathbf{X}} \hat{\xi}_{\mathbf{x}, d_n}} \Phi^{-1}(\alpha) \quad (\text{V.2})$$

où Φ est la fonction de répartition de la loi normale centrée réduite.

Lors de l'estimation de quantiles, puis d'expectiles extrêmes, nous avons à chaque fois scindé l'étude en deux parties, selon la vitesse de convergence de notre suite α_n vers 1. Nous allons effectuer une distinction similaire avec la vitesse de convergence de d_n vers $+\infty$. Pour cela, intéressons nous d'abord au cas où le nombre de covariables tend vers l'infini relativement lentement, c'est à dire tel que $d_n/n \rightarrow 0$ quand $n \rightarrow +\infty$. Ce cas là est le plus simple. En effet, l'assertion $d_n < n$ permet notamment de fournir, via la méthode des moments ou les M-estimateurs, une matrice de covariance empirique $\hat{\boldsymbol{\Sigma}}$ inversible. De plus, avec un échantillon d'apprentissage de taille n , la vitesse de convergence de ces méthodes est relativement rapide (en $\sqrt{n}$). Concernant $\hat{\xi}_{\mathbf{x}, d_n}$, il est aisé de montrer que

$$\sqrt{d_n} \left(\hat{\xi}_{\mathbf{x}, d_n} - \xi_{\mathbf{x}} \right) \xrightarrow{n \rightarrow +\infty} \mathcal{N} \left(0, \frac{\xi_{\mathbf{x}}^2}{2} \right).$$

Cette approche naïve est donc censée fournir une estimation correcte du quantile conditionnel. Cependant, nous verrons dans la Section [V.3](#) qu'elle est extrêmement volatile. De plus, ces méthodes ne sont plus applicables si le nombre de covariables d_n est supérieur à la taille de l'échantillon n , la matrice $\hat{\boldsymbol{\Sigma}}$ n'étant plus inversible. On propose donc dans la section suivante d'autres approximations. On verra d'ailleurs dans l'étude numérique que même si $d_n < n$, ces dernières semblent bien meilleures que l'approche précédente.

V.2. Estimation des paramètres en (très) grande dimension

Dans cette section, on considère le cas où $d_n/n \rightarrow c \in]0, +\infty]$ quand $n \rightarrow +\infty$. Le but est toujours de fournir des estimateurs de $\boldsymbol{\mu}$ et $\boldsymbol{\Sigma}$, et ainsi d'en déduire un estimateur de $q_{\alpha}(Y|\mathbf{X} = \mathbf{x})$. Des estimateurs robustes de $\boldsymbol{\mu}$ peuvent être obtenus dans la littérature. On peut notamment citer [\[Shahriari and Ahmadi, 2015\]](#), qui en propose plusieurs. On peut également donner comme exemple [\[Cellier et al., 1989\]](#) qui se concentre sur les distributions elliptiques, et applique une méthode appelée shrinkage. Nous reviendrons plus en détail sur cette méthode plus tard. En attendant, ces méthodes existantes font que l'estimation du vecteur de moyenne $\boldsymbol{\mu}$ n'est pas la principale difficulté de cette section. On considérera alors, pour l'estimation des matrices $\boldsymbol{\Sigma}$, puis $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1}$, que la distribution est centrée, l'opération de recentrage des données ayant été effectuée en amont.

L'estimation de la matrice Σ est également soumise à beaucoup de recherches récentes. Cependant, il est en général plus compliqué d'estimer la matrice Ω . Dans un cadre classique, i.e. en dimension raisonnable, une possibilité est d'estimer la matrice de covariance empirique

$$S_n = \frac{1}{n} \sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^\top, \quad (\text{V.3})$$

puis de prendre simplement $\hat{\Omega}_n = S_n^{-1}$ comme estimateur de Ω (appelée matrice de précision). Seulement, une telle approche se heurte à deux problèmes majeurs. Premièrement, si $d_n + 1 > n$, ce qui est le cas ici si $c > 1$, la matrice S est singulière avec probabilité 1 (voir Théorème 3.1.4 de [Muirhead, 2009]). Par ailleurs, $\mathbb{E}[\mathbf{Z}\mathbf{Z}^\top]$ peut ne pas exister, si la queue de distribution de ξ est trop lourde.

Pour contourner ce second problème, [Maronna, 1976] a proposé des M-estimateurs, qui vérifient le type de relation suivante :

$$\mathbf{V}_n = \frac{1}{n} \sum_{i=1}^n u(\mathbf{z}_i^\top \mathbf{V}_n^{-1} \mathbf{z}_i) \mathbf{z}_i \mathbf{z}_i^\top, \quad (\text{V.4})$$

où u doit vérifier certaines conditions. Un algorithme de type point-fixe permet alors d'obtenir rapidement $\mathbf{V}_n$. Le papier de [Maronna, 1976] donne notamment comme exemple la fonction $u(t) = (\nu + d_n + 1)/(\nu + t)$, $\nu > 0$. Cette fonction est dite "de Student" car une estimation paramétrique de Σ via un algorithme EM dans le cas d'une distribution Student à ν degrés de liberté mène exactement à la même relation de point fixe. Plus tard, [Tyler, 1987] a donné des résultats d'existence dans le cas de la fonction $u(t) = (d_n + 1)/t$. Bien que ne vérifiant pas les conditions de [Maronna, 1976], cette fonction est appréciée, car elle simplifie le calcul de certains moments, utiles pour obtenir les paramètres de shrinkage optimaux que nous verrons par la suite.

Une telle approche permet, on l'a dit, d'estimer une matrice Σ même pour des distributions à queue lourde. Cependant, comme pour l'estimateur précédent, la matrice $\hat{\Sigma}$ est singulière si $d_n + 1 > n$. On présente alors par la suite quelques méthodes abordées dans la littérature pour pallier ce problème.

V.2.a. Les méthodes de sparsité

Une première approche pour estimer $\Omega = \Sigma^{-1}$ est de considérer que cette dernière est parcimonieuse (ou sparse), c'est à dire qu'un grand nombre de ses éléments est nul. Dans ce cadre, on s'autorise l'hypothèse $d_n/n \rightarrow +\infty$ lorsque $n \rightarrow +\infty$. On notera également s_n le nombre d'éléments de Ω non-nuls (diagonale non-comprise).

Une première approche naïve serait d'estimer cette matrice par maximum de vraisemblance, sous l'hypothèse gaussienne en pénalisant en norme L_1 . Cette approche, vue comme une sorte de généralisation des méthodes LASSO (voir [Tibshirani, 1997]), est

appelée GLASSO, et est assez répandue. [Banerjee et al., 2008] ou [Friedman et al., 2008] sont de bons exemples. L'estimateur $\hat{\Omega}$ de Ω est alors :

$$\hat{\Omega} = \arg \min_{\Omega} -\ln(|\Omega|) + \text{tr}(\mathbf{S}_n \Omega) + \lambda_n \sum_{i,j} |\Omega_{ij}| \quad (\text{V.5})$$

où $\lambda_n > 0$ est un paramètre de régularisation. [Lam and Fan, 2009] proposent des résultats de convergence pour cette méthode. En se basant sur une idée similaire, [Cai et al., 2011] ont proposé une méthode d'estimation de Ω colonne par colonne. En effet, cet estimateur $\tilde{\Omega}$ appelé CLIME est défini de la manière suivante :

$$\begin{cases} \tilde{\Omega}_{\cdot j} = \arg \min_{\Omega} \|\Omega_{\cdot j}\|_1 \\ \text{s.c } \|\mathbf{S}_n \Omega_{\cdot j} - \mathbf{I}_{d_n, j}\|_{\infty} \leq \lambda_n, \forall j = 1, \dots, d_n + 1 \end{cases}, \quad (\text{V.6})$$

où $\Omega_{\cdot j}$ représente la j^e colonne de Ω , et $\mathbf{I}_{d_n, j}$ la j^e colonne de $\mathbf{I}_{d_n}$. Le papier de [Cai et al., 2011] donne également des taux de convergence similaires à ceux de la méthode GLASSO. En se basant sur ces deux méthodes, on peut alors construire deux estimateurs de quantiles conditionnels similaires à celui de la Définition 29

Définition 30 (Prédicteurs sparses). Soit $\mathbf{Z} = (\mathbf{X}, Y) \sim \mathcal{EC}_{d_n+1}(\mathbf{0}_{\mathbb{R}^{d_n+1}}, \Sigma, \xi)$. On introduit les estimateurs de $q_{\alpha}(Y|\mathbf{X} = \mathbf{x})$ suivants :

$$\begin{cases} \hat{q}_{\alpha}^{GLASSO}(Y|\mathbf{X} = \mathbf{x}) = \hat{\mu}_{Y|\mathbf{x}} + \sqrt{\hat{\Sigma}_{Y|\mathbf{X}}} \sqrt{\frac{\mathbf{x}^{\top} \hat{\Omega}_{\mathbf{X}} \mathbf{x}}{d_n}} \Phi^{-1}(\alpha), \\ \hat{q}_{\alpha}^{CLIME}(Y|\mathbf{X} = \mathbf{x}) = \tilde{\mu}_{Y|\mathbf{x}} + \sqrt{\tilde{\Sigma}_{Y|\mathbf{X}}} \sqrt{\frac{\mathbf{x}^{\top} \tilde{\Omega}_{\mathbf{X}} \mathbf{x}}{d_n}} \Phi^{-1}(\alpha), \end{cases} \quad (\text{V.7})$$

où $\hat{\Omega}$ et $\tilde{\Omega}$ sont respectivement introduits dans les Equations (V.2.a) et (V.2.a), Φ est la fonction de répartition de la loi normale centrée réduite, et

$$\begin{cases} \hat{\mu}_{Y|\mathbf{x}} = \hat{\Sigma}_{Y\mathbf{X}} \hat{\Omega}_{\mathbf{X}} \mathbf{x}, \\ \hat{\Sigma}_{Y|\mathbf{X}} = \hat{\Sigma}_Y - \hat{\Sigma}_{Y\mathbf{X}} \hat{\Omega}_{\mathbf{X}} \hat{\Sigma}_{\mathbf{X}Y}, \\ \tilde{\mu}_{Y|\mathbf{x}} = \tilde{\Sigma}_{Y\mathbf{X}} \tilde{\Omega}_{\mathbf{X}} \mathbf{x}, \\ \tilde{\Sigma}_{Y|\mathbf{X}} = \tilde{\Sigma}_Y - \tilde{\Sigma}_{Y\mathbf{X}} \tilde{\Omega}_{\mathbf{X}} \tilde{\Sigma}_{\mathbf{X}Y}, \end{cases}$$

avec $\hat{\Sigma} = \hat{\Omega}^{-1}$ et $\tilde{\Sigma} = \tilde{\Omega}^{-1}$.

Parlons maintenant du choix du paramètre λ_n pour ces deux méthodes. Pour cela, on s'intéresse à leurs propriétés asymptotiques, à commencer par la méthode GLASSO. Malheureusement, on se heurte à un premier problème : le modèle étant basé sur une vraisemblance gaussienne, le résultat de convergence que l'on va énoncer est valable si la distribution sous-jacente a une queue de distribution légère. Cependant, la méthode étant suffisamment robuste, on énonce tout de même le théorème suivant, issu de [Lam and Fan, 2009].

Théorème 52 (Taux de convergence de la méthode GLASSO). *Soit $\mathbf{Z} \sim \mathcal{EC}_{d_n+1}(\mathbf{0}_{\mathbb{R}^{d_n+1}}, \mathbf{\Sigma}, \xi)$, avec ξ à queue de distribution légère. Sous les hypothèses :*

(i) *il existe τ_1, τ_2 tels que*

$$0 < \tau_1 < \lambda_{\min}(\mathbf{\Sigma}) \leq \lambda_{\max}(\mathbf{\Sigma}) < \tau_2 < +\infty,$$

(ii) $\lambda_n = O\left(\left(1 + \frac{d_n+1}{s_n+1}\right) \sqrt{\frac{\ln(d_n+1)}{n}}\right)$ et $\min_{i,j} \frac{|\Omega_{ij}|}{\lambda_n} \rightarrow +\infty$ quand $n \rightarrow +\infty$,

(iii) $\ln\left(\frac{d_n+1}{n}\right) = O(\lambda_n^2)$ quand $n \rightarrow +\infty$,

alors $\hat{\mathbf{\Omega}}$ solution de (V.2.a.) vérifie

$$\|\hat{\mathbf{\Omega}} - \mathbf{\Omega}\|_F^2 = O_{\mathbb{P}}\left((d_n + 1 + s_n) \frac{\ln(d_n + 1)}{n}\right).$$

Ce résultat, bien qu'inutilisable dans notre cadre de distributions à queue lourde, nous donne un ordre de grandeur pour le choix du paramètre λ_n . On s'y attachera dans l'étude numérique de la Section V.3.

Passons maintenant à l'estimateur CLIME, pour lequel [Cai et al., 2011] propose des résultats de convergence, même pour certaines distributions à queue lourde. Le cadre est alors donné dans l'hypothèse ci-dessous, et adapté aux MEGC.

Hypothèse 5. *Soit $\mathbf{Z} \sim \mathcal{EC}_{d_n+1}(\mathbf{0}_{\mathbb{R}^{d_n+1}}, \mathbf{\Sigma}, \xi)$. On considère qu'il existe $k, c > 0$ tels que $d_n + 1 \leq cn^k$, et également $\delta > 0$ tel que*

$$\mathbb{E}[\xi^{4k+4+\delta}] < +\infty.$$

On suppose de plus que $\mathbf{\Omega} = \mathbf{\Sigma}^{-1}$ appartient à l'ensemble des matrices

$$\mathcal{U}(q, s_o(d_n + 1)) = \left\{ \mathbf{\Omega} > 0 : \|\mathbf{\Omega}\|_1 \leq M, \max_{1 \leq i \leq d_n+1} \sum_{j=1}^{d_n+1} |\Omega_{ij}|^q \leq s_o(d_n + 1) \right\}$$

pour $q \in [0, 1[$.

Cette hypothèse, qui élargit certes le cadre de l'estimateur GLASSO, reste tout de même assez restrictive, car nécessite l'existence de moments relativement élevés de la variable d'échelle ξ . Rajoutons enfin la notation :

$$\theta = \max_{i,j} \mathbb{E}[(\mathbf{Z}_i \mathbf{Z}_j - \Sigma_{ij})^2].$$

Ce paramètre θ , introduit dans [Cai et al., 2011], peut être réécrit comme suit si $\mathbf{Z}$ est un MEGC.

Lemme 53 (Expression de θ). *Soit $\mathbf{Z} \sim \mathcal{EC}_{d_n+1}(\mathbf{0}_{\mathbb{R}^{d_n+1}}, \mathbf{\Sigma}, \xi)$. On a alors :*

$$\theta = \max_{i,j} (2\mathbb{E}[\xi^4] - 2\mathbb{E}[\xi^2] + 1) \Sigma_{ij}^2 + \mathbb{E}[\xi^4] \Sigma_{ii} \Sigma_{jj}. \quad (\text{V.8})$$

Preuve : Intéressons nous à la loi du couple $(\mathbf{Z}_i, \mathbf{Z}_j)$. On a clairement

$$(\mathbf{Z}_i, \mathbf{Z}_j) \sim \mathcal{EC}_2 \left((0, 0), \begin{pmatrix} \Sigma_{ii} & \Sigma_{ij} \\ \Sigma_{ij} & \Sigma_{jj} \end{pmatrix}, \xi \right).$$

De même, au regard du Théorème 7 de [Frahm, 2004], ce vecteur admet la représentation stochastique suivante :

$$\begin{pmatrix} \mathbf{Z}_i \\ \mathbf{Z}_j \end{pmatrix} \sim \begin{pmatrix} \sqrt{\Sigma_{ii}}\xi\sqrt{\chi_2^2 U_1^{(1)}}\sqrt{\beta} \\ \frac{\Sigma_{ij}}{\sqrt{\Sigma_{ii}}}\xi\sqrt{\chi_2^2 U_1^{(1)}}\sqrt{\beta} + \sqrt{\Sigma_{jj} - \frac{\Sigma_{ij}^2}{\Sigma_{ii}}}\xi\sqrt{\chi_2^2}\sqrt{1 - \beta}U_2^{(1)} \end{pmatrix}$$

où β est une loi Bêta de paramètres $1/2$, $U_1^{(1)}$ et $U_2^{(1)}$ valent -1 ou 1 avec probabilité $1/2$. Enfin, ξ , χ_2^2 , β , $U_1^{(1)}$ et $U_2^{(1)}$ sont mutuellement indépendantes. En utilisant les moments d'ordre 1 et 2 des lois Bêta et du χ^2 , on obtient facilement les quantités $\mathbb{E}[\mathbf{Z}_i \mathbf{Z}_j]$ et $\mathbb{E}[\mathbf{Z}_i^2 \mathbf{Z}_j^2]$, d'où le résultat [53]. $\square$

Sous toutes les conditions que l'on vient d'exprimer, [Cai et al., 2011] donne le résultat de convergence suivant.

Théorème 54 (Taux de convergence de la méthode CLIME). *Sous l'Hypothèse [5], avec $\lambda_n = \sqrt{(5 + \tau)(\theta + 1)}M\sqrt{\frac{\ln(d_n + 1)}{n}}$, $\tau > 0$, alors $\tilde{\Omega}$ solution de [V.2.a.] vérifie :*

$$\|\tilde{\Omega} - \Omega\|_2 \leq 48\sqrt{(5 + \tau)(\theta + 1)}M^{2-2q}s_o(d_n) \left(\frac{\ln(d_n + 1)}{n} \right)^{\frac{1-q}{2}}$$

avec une probabilité tendant vers 1 quand $n \rightarrow +\infty$.

Ce résultat nous donne un critère pour choisir notre paramètre λ_n . Il faudra en effet le considérer proportionnel à $\sqrt{\frac{\ln(d_n + 1)}{n}}$. On propose dans l'étude numérique de la Section [V.3] une illustration. On utilisera d'ailleurs les deux estimateurs GLASSO et CLIME, et verra qu'ils sont plutôt performants en grande dimension.

Pour compléter notre étude, on s'intéresse dans la section suivante au cas où le nombre de covariables d_n et la taille de l'échantillon n sont du même ordre de grandeur.

V.2.b. Les méthodes de shrinkage

Les méthodes de shrinkage que l'on va présenter ont été introduites pour la première fois dans [Ledoit and Wolf, 2004]. Ces méthodes proposent de combiner l'estimateur de Σ avec une matrice diagonale, de manière à la rendre inversible. Elles ont l'avantage d'être simples à comprendre, ainsi qu'à mettre en place. Nous distinguons alors par la suite les cas de la matrice de covariance empirique $\mathbf{S}_n$ et celui du M-estimateur $\mathbf{V}_n$. Avant cela, nous établissons l'hypothèse suivante pour ce paragraphe.

Hypothèse 6. *On considère que :*

- (i) $d_n/n \rightarrow c$ quand $n \rightarrow +\infty$, avec $0 < c < +\infty$.
- (ii) $\text{tr}(\Sigma^i)/d_n \rightarrow \lambda_i$ quand $n \rightarrow +\infty$, avec $0 < \lambda_i < +\infty$, $\forall i = 1, \dots, 4$.

On se place alors dans le cadre où la dimension de la covariable d_n et la taille de l'échantillon n sont asymptotiquement proportionnels.

V.2.b.1. Le cas de $\mathbf{S}_n$

Il est clair, au regard de l'Equation (V.2.), que $\mathbf{S}_n$ est centré en $\mathbb{E}[\xi^2] \Sigma$ dans le cadre des Mélanges d'Echelle Gaussiens Consistants. Par souci d'identifiabilité, une hypothèse récurrente dans la littérature est de supposer que $\mathbb{E}[\xi^2] = 1$. On va alors se placer dans ce cadre là qui permet notamment d'avoir la relation $\mathbb{E}[\mathbf{Z}\mathbf{Z}^\top] = \Sigma$. Les exemples du Chapitre II peuvent être réécrits en normalisant les variables d'échelle ξ par $\sqrt{\mathbb{E}[\xi^2]}$. Définissons l'estimateur de shrinkage $\mathbf{S}_{n,\omega}$ basé sur la matrice $\mathbf{S}_n$, comme combinaison convexe de $\mathbf{S}_n$ et de la matrice identité.

Définition 31 (Matrice de shrinkage). *On définit la matrice $\mathbf{S}_{n,\omega}$ comme étant :*

$$\mathbf{S}_{n,\omega} = \omega \mathbf{S}_n + (1 - \omega) \mathbf{I}_{d_n+1},$$

où $\mathbf{S}_n$ est donnée dans l'Equation (V.2.) et $\mathbf{I}_{d_n+1}$ est la matrice identité.

De manière naturelle, on peut se poser la question du choix optimal de ω . Pour cela, il existe plusieurs critères, qui consistent pour la plupart à minimiser l'écart entre $\mathbf{S}_{n,\omega}$ et Σ au sens d'une fonction de perte $\mathcal{P}$.

$$\omega^* = \arg \min_{\omega \in [0,1]} \mathcal{P}(\mathbf{S}_{n,\omega}, \Sigma)$$

Sans être totalement exhaustif, voici un bref aperçu des fonctions $\mathcal{P}$ utilisées dans la littérature :

- La fonction de perte de Stein. Introduite notamment dans [Dey and Srinivasan, 1985], elle pénalise la trace ainsi que le logarithme du déterminant de la matrice $\mathbf{S}_{n,\omega} \Sigma^{-1}$. Elle a la forme suivante :

$$\mathcal{P}_S(\mathbf{S}_{n,\omega}, \Sigma) = \mathbb{E} [\text{tr}(\mathbf{S}_{n,\omega} \Sigma^{-1}) - \ln(|\mathbf{S}_{n,\omega} \Sigma^{-1}|)] - d_n.$$

- La norme de Frobenius. Cette approche, très simple à appréhender, prend la différence quadratique terme à terme des deux matrices, d'où la fonction :

$$\mathcal{P}_F(\mathbf{S}_{n,\omega}, \Sigma) = \mathbb{E} [\|\mathbf{S}_{n,\omega} - \Sigma\|_F^2] = \mathbb{E} \left[\text{tr} \left\{ (\mathbf{S}_{n,\omega} - \Sigma) (\mathbf{S}_{n,\omega} - \Sigma)^\top \right\} \right] \quad (\text{V.9})$$

Le paramètre ω^* minimisant cette fonction est appelé paramètre ORACLE dans la littérature. On peut notamment le trouver dans le papier initial de [Ledoit and Wolf, 2004] ou encore plus récemment dans [Chen et al., 2011].

— La fonction de perte quadratique :

$$\mathcal{P}_Q(\mathbf{S}_{n,\omega}, \mathbf{\Sigma}) = \mathbb{E} \left[\text{tr} \left(\mathbf{S}_{n,\omega} \mathbf{\Sigma}^{-1} - I \right)^2 \right] \quad (\text{V.10})$$

Cette fonction de coût est notamment utilisée dans de nombreux papiers tels que [Krishnamoorthy and Gupta, 1989], [Yang and Berger, 1994] ou [Wang et al., 2015].

Nous donnons alors quelques résultats concernant les paramètres de shrinkage optimaux, particulièrement dans notre cadre des distributions elliptiques consistantes. Le premier concerne le paramètre ORACLE.

Proposition 55 (Paramètre ORACLE optimal). *Soit $\mathbf{Z} \sim \mathcal{EC}_{d_n+1}(\mathbf{0}_{\mathbb{R}^{d_n+1}}, \mathbf{\Sigma}, \xi)$ avec $\mathbb{E}[\xi^2] = 1$. Le paramètre ORACLE ω_O de $\mathbf{Z}$ est donné par :*

$$\omega_O = \frac{\text{tr}(\mathbf{\Sigma}^2) - 2\text{tr}(\mathbf{\Sigma}) + d_n + 1}{\left(\frac{2\mathbb{E}[\xi^4] - 1}{n} + 1 \right) \text{tr}(\mathbf{\Sigma}^2) + \frac{\mathbb{E}[\xi^4]}{n} \text{tr}(\mathbf{\Sigma})^2 - 2\text{tr}(\mathbf{\Sigma}) + d_n + 1}. \quad (\text{V.11})$$

Preuve : Le développement de la fonction de perte (V.2.b.1.) implique que ω_0 est l'argmin de

$$\omega^2 \mathbb{E}[\text{tr}(\mathbf{S}_n^2)] + 2\omega(1-\omega) \mathbb{E}[\text{tr}(\mathbf{S}_n)] + (1-\omega)^2(d_n+1) - 2\omega \mathbb{E}[\text{tr}(\mathbf{S}_n \mathbf{\Sigma})] - 2(1-\omega) \text{tr}(\mathbf{\Sigma}).$$

On déduit de [Ollila, 2017] l'égalité $\mathbb{E}[\text{tr}(\mathbf{S}_n^2)] = \left(\frac{2\mathbb{E}[\xi^4] - 1}{n} + 1 \right) \text{tr}(\mathbf{\Sigma}^2) + \frac{\mathbb{E}[\xi^4]}{n} \text{tr}(\mathbf{\Sigma})^2$.

De plus, on remarque que $\mathbb{E}[\text{tr}(\mathbf{S}_n \mathbf{\Sigma})] = \mathbb{E}[\xi^2] \text{tr}(\mathbf{\Sigma}^2) = \text{tr}(\mathbf{\Sigma}^2)$. Enfin, on a $\mathbb{E}[\text{tr}(\mathbf{S}_n)] = \text{tr}(\mathbf{\Sigma})$. Une simple approche par gradient donne ensuite le résultat. $\square$

On s'aperçoit de manière assez logique que le paramètre ORACLE optimal dépend uniquement de la variable d'échelle ξ . En reprenant la Table II.2 on propose dans la Table V.1 les moments d'ordre 4 de ξ pour les exemples du Chapitre II. On peut en déduire directement les valeurs de ω_O pour chaque distribution (normalisée).

Passons maintenant à l'estimation d'un tel paramètre ω_O . L'Equation (55) nous montre que cette quantité dépend de $\mathbb{E}[\xi^4]$, $\text{tr}(\mathbf{\Sigma})$ et $\text{tr}(\mathbf{\Sigma}^2)$. Commençons donc par estimer le paramètre que nous appellerons de kurtosis $\mathbb{E}[\xi^4]$. Dans le cas d'une distribution elliptique (i.e. sans l'hypothèse de consistance), estimer cette quantité pose problème étant donné que la dimension d_n tend vers l'infini au moins aussi vite que n . Heureusement, avec l'hypothèse de consistance, le fait que la variable d'échelle ξ ne dépende pas de la dimension nous fait échapper à ce problème. On définit alors un estimateur basé uniquement sur une composante du vecteur, et donne ses propriétés asymptotiques.

Proposition 56 (Estimateur de $\mathbb{E}[\xi^4]$). *Soient $\mathbf{X}_1, \dots, \mathbf{X}_n$ des répétitions indépendantes et identiquement distribuées de même loi que $\mathbf{X} = (X^{(1)}, \dots, X^{(d_n)}) \sim \mathcal{EC}_{d_n}(\mathbf{0}_{\mathbb{R}^{d_n}}, \mathbf{\Sigma}_X, \xi)$*

Loi	ξ	$\mathbb{E} [\xi^4]$
Normale	1	1
Normale contaminée normalisée	$\frac{\sum_{k=1}^m p_k \delta_{\theta_k}}{\sqrt{\sum_{k=1}^m p_k \theta_k^2}}$	$\frac{\sum_{k=1}^m p_k \theta_k^4}{\left(\sum_{k=1}^m p_k \theta_k^2\right)^2}$
Student normalisée, $\nu > 4$	$\sqrt{\frac{\nu-2}{\chi_\nu^2}}$	$\frac{\nu-2}{\nu-4}$
Laplace normalisée, $\lambda > 0$	$\sqrt{\frac{1}{\lambda} \mathcal{E} \left(\frac{1}{\lambda} \right)}$	2
Slash normalisée, $a > 4$	$\sqrt{\frac{a-2}{a} \mathcal{P} (1, a)}$	$\frac{(a-2)^2}{a(a-4)}$

TABLE V.1 – Exemples de plusieurs Mélanges d'Echelle Gaussiens Consistants

avec $\mathbb{E} [\xi^2] = 1$. On définit l'estimateur de kurtosis :

$$\hat{\kappa} = \frac{n^{-1} \sum_{i=1}^n X_i^{(1)4}}{3 \left(n^{-1} \sum_{i=1}^n X_i^{(1)2} \right)^2}. \quad (\text{V.12})$$

Cet estimateur satisfait alors la normalité asymptotique suivante :

$$\sqrt{n} (\hat{\kappa} - \mathbb{E} [\xi^4]) \xrightarrow{n \rightarrow +\infty} \mathcal{N} \left(0, 12\mathbb{E} [\xi^4]^3 - \mathbb{E} [\xi^4]^2 - 20\mathbb{E} [\xi^4] \mathbb{E} [\xi^6] + \frac{35}{3}\mathbb{E} [\xi^8] \right). \quad (\text{V.13})$$

Preuve : On rappelle que $X^{(1)}$ égale en loi $\sigma \sqrt{\chi_1^2} \xi U^{(1)}$, avec $\sigma > 0$. L'hypothèse $\mathbb{E} [\xi^2] = 1$, ainsi que de simples calculs de moments de la loi χ_1^2 mènent à :

$$\sqrt{n} \begin{pmatrix} \frac{1}{n} \sum_{i=1}^n X_i^{(1)2} - \sigma^2 \\ \frac{1}{n} \sum_{i=1}^n X_i^{(1)4} - 3\mathbb{E} [\xi^4] \sigma^4 \end{pmatrix} \xrightarrow{n \rightarrow +\infty} \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} (3\mathbb{E} [\xi^4] - 1) \sigma^4 & (15\mathbb{E} [\xi^6] - 3\mathbb{E} [\xi^4]) \sigma^6 \\ (15\mathbb{E} [\xi^6] - 3\mathbb{E} [\xi^4]) \sigma^6 & (105\mathbb{E} [\xi^8] - 9\mathbb{E} [\xi^4]^2) \sigma^8 \end{pmatrix} \right).$$

La méthode delta complète la preuve. $\square$

Il paraît intéressant de regarder le comportement de cette loi limite sur l'exemple le plus courant, à savoir la distribution gaussienne.

Exemple : Dans le cas gaussien, en prenant $\xi = 1$ presque sûrement, on obtient la loi limite suivante :

$$\sqrt{n} (\hat{\kappa} - 1) \xrightarrow{n \rightarrow +\infty} \mathcal{N} \left(0, \frac{8}{3} \right).$$

On remarque qu'un tel estimateur fait apparaître des moments d'ordre 8, ce qui assez gênant si on considère des distributions à queue lourde. Si l'estimation de $\mathbb{E}[\xi^4]$ conduit à un résultat asymptotique satisfaisant, celles des traces s'avèrent plus délicates. S'il paraît illusoire d'obtenir les distributions asymptotiques, [Ollila, 2017] a proposé des estimateurs consistants pour ces quantités en se basant notamment sur [Zhang and Wiesel, 2016]. Il faut pour cela se placer dans le cadre de l'Hypothèse [6] i.e. ajouter quelques conditions sur les relations asymptotiques entre d_n et n , puis sur le spectre de Σ .

Proposition 57 (Estimateur consistant de ω_O). *Soient $\mathbf{Z}_1, \dots, \mathbf{Z}_n$ des répétitions indépendantes et identiquement distribuées de même loi que $\mathbf{Z} \sim \mathcal{EC}_{d_n+1}(\mathbf{0}_{\mathbb{R}^{d_n+1}}, \Sigma, \xi)$ avec $\mathbb{E}[\xi^2] = 1$. On introduit la matrice de signe empirique :*

$$\mathbf{S}_{sgn} = \frac{1}{n} \sum_{i=1}^n \frac{\mathbf{Z}_i \mathbf{Z}_i^\top}{\|\mathbf{Z}_i\|^2}.$$

Alors sous l'Hypothèse [6]

$$\hat{T} = (d_n + 1) \text{tr}(\mathbf{S}_{sgn}^2) - \frac{d_n + 1}{n} \quad (\text{V.14})$$

est un estimateur consistant de $T = (d_n + 1) \text{tr}(\Sigma^2) / \text{tr}(\Sigma)^2$. L'estimateur $\hat{\omega}_O$ ci-dessous est par conséquent un estimateur consistant de ω_O :

$$\hat{\omega}_O = \frac{\hat{T} \text{tr}(\mathbf{S}_n)^2 - 2 \text{tr}(\mathbf{S}_n)(d_n + 1) + (d_n + 1)^2}{\left(\frac{2\hat{\kappa}-1}{n} + 1\right) \hat{T} \text{tr}(\mathbf{S}_n)^2 + \frac{\hat{\kappa}(d_n+1)}{n} \text{tr}(\mathbf{S}_n)^2 - 2 \text{tr}(\mathbf{S}_n)(d_n + 1) + (d_n + 1)^2}.$$

Avec ce paramètre optimal, on définit l'estimateur suivant pour le quantile conditionnel, basé sur l'approximation gaussienne et l'estimation de la matrice de précision.

Définition 32 (Estimateur ORACLE de quantile conditionnel). *Sous les Hypothèses [2] et [6], on définit l'estimateur suivant de $q_\alpha(Y|\mathbf{X} = \mathbf{x})$:*

$$\hat{q}_\alpha^{ORACLE} = \hat{\mu}_{Y|\mathbf{x}}^O + \sqrt{\hat{\Sigma}_{Y|\mathbf{x}}^O} \sqrt{\frac{\mathbf{x}^\top \mathbf{S}_{\mathbf{X},n,\hat{\omega}_O}^{-1} \mathbf{x}}{d_n}} \Phi^{-1}(\alpha), \quad (\text{V.15})$$

avec

$$\begin{cases} \hat{\mu}_{Y|\mathbf{x}}^O = & \mathbf{S}_{Y\mathbf{X},n,\hat{\omega}_O} \mathbf{S}_{\mathbf{X},n,\hat{\omega}_O}^{-1} \mathbf{x} \\ \hat{\Sigma}_{Y|\mathbf{x}}^O = & \mathbf{S}_{Y,n,\hat{\omega}_O} - \mathbf{S}_{Y\mathbf{X},n,\hat{\omega}_O} \mathbf{S}_{\mathbf{X},n,\hat{\omega}_O}^{-1} \mathbf{S}_{\mathbf{X}Y,n,\hat{\omega}_O} \end{cases}.$$

On propose par la suite une étude numérique pour tester un tel estimateur. On s'aperçoit que la fonction de perte ([V.2.b.1.]) fait en sorte que l'estimation de Σ soit optimale. Cependant, il pourrait être intéressant de se baser sur la matrice de précision $\Omega = \Sigma^{-1}$. Pour

cela, on peut alors considérer la fonction de coût quadratique $\mathcal{P}_Q$ de l'Equation (V.2.b.1.). Cette fonction fait en effet intervenir la matrice Σ^{-1} , et promet donc une matrice empirique $\mathbf{S}_{n,\omega}^{-1}$ plus fidèle à Ω . La proposition suivante donne le paramètre ω_Q optimal dans ce cas de figure.

Proposition 58 (Paramètre optimal au sens quadratique). *Soit $\mathbf{Z} \sim \mathcal{EC}_{d_n+1}(\mathbf{0}_{\mathbb{R}^{d_n+1}}, \Sigma, \xi)$ avec $\mathbb{E}[\xi^2] = 1$. Le paramètre ω_Q de $\mathbf{Z}$ relatif à $\mathcal{P}_Q$ est le suivant :*

$$\omega_Q = \frac{n \left(\text{tr}(\Sigma^{-1}) - d_n - 1 \right)^2}{(d_n + 1)(n(d_n + 1) + 1) - 2n(d_n + 1)\text{tr}(\Sigma^{-1}) + n\text{tr}(\Sigma^{-1})^2}. \quad (\text{V.16})$$

Preuve : Le développement de la fonction $\mathcal{P}_Q$ de (V.2.b.1.) donne :

$$\begin{aligned} \omega_Q = \arg \min_{\omega} \omega^2 \mathbb{E} \left[\text{tr}(\mathbf{S}_n \Sigma^{-1})^2 \right] + 2\omega(1-\omega) \mathbb{E} \left[\text{tr}(\mathbf{S}_n \Sigma^{-1}) \right] \text{tr}(\Sigma^{-1}) + (1-\omega)^2 \text{tr}(\Sigma^{-1})^2 \\ - 2\omega(d_n + 1) \mathbb{E} \left[\text{tr}(\mathbf{S}_n \Sigma^{-1}) \right] - 2(d_n + 1)(1-\omega) \text{tr}(\Sigma^{-1}) + (d_n + 1)^2. \end{aligned}$$

En utilisant les résultats de [Nagar and Gupta, 2011], on obtient $\mathbb{E} \left[\text{tr}(\mathbf{S}_n \Sigma^{-1})^2 \right] = \frac{(d_n+1)(n(d_n+1)+1)}{n}$, ainsi que $\mathbb{E} \left[\text{tr}(\mathbf{S}_n \Sigma^{-1}) \right] = d_n + 1$. ω_Q annulant le gradient de la précédente expression, on obtient facilement le résultat (58). $\square$

On s'aperçoit qu'ici, le paramètre de shrinkage dépend uniquement de la trace de Ω . Il nous faut alors un estimateur pour cette quantité, qui nous mènera directement à un estimateur de ω_Q . Une idée serait de prendre la trace de la matrice $\hat{\Omega}$ ou $\tilde{\Omega}$ obtenues respectivement par les méthodes GLASSO et CLIME. Cependant, une telle approche donne des résultats très décevants. On ne développera donc pas plus cette option.

V.2.b.2. Le cas de $\mathbf{V}_n$

L'approche avec la matrice de covariance empirique $\mathbf{S}_n$, comme on a pu le constater, est efficace dans le cas des distributions à queue légère ou modérément lourde. Pour éviter cette restriction, on traite alors le cas des matrices empiriques $\mathbf{V}_n$ de [Maronna, 1976] ou [Tyler, 1987]. Comme mentionné dans ces papiers, les matrices obtenues via la relation de point fixe (V.2.) sont uniques à un coefficient près. Par souci d'identifiabilité, il faut alors faire une hypothèse supplémentaire sur Σ . On propose alors de prendre de manière récurrente $\text{tr}(\Sigma) = d_n + 1$, comme fréquemment dans la littérature.

Cependant, les M-estimateurs $\mathbf{V}_n$ souffrent du même problème que $\mathbf{S}_n$, à savoir une singularité presque sûre si $d_n + 1 > n$. Certains auteurs comme [Chen et al., 2011] ont alors proposé une approche de type shrinkage en modifiant l'Equation (V.2.) comme suit :

$$\mathbf{V}_{n,\tau} = \tau \frac{1}{n} \sum_{i=1}^n u(\mathbf{Z}_i^\top \mathbf{V}_{n,\tau}^{-1} \mathbf{Z}_i) \mathbf{Z}_i \mathbf{Z}_i^\top + (1-\tau) \mathbf{I}_{d_n+1}.$$

Pour l'estimateur de [Tyler, 1987], i.e. avec $u(t) = (d_n + 1)/t$, l'existence, l'unicité (sous la condition $\text{tr}(\mathbf{V}_{n,\tau}) = d_n + 1$) ainsi que la semi-définie positivité sont montrées dans [Chen et al., 2011], pour tout $\tau \in]0, 1[$. La question que l'on se pose naturellement concerne alors le paramètre optimal τ^* . On introduit pour cela la notation suivante :

$$\tilde{\Sigma}(\tau) = \tau \frac{d_n + 1}{n} \sum_{i=1}^n \frac{\mathbf{Z}_i \mathbf{Z}_i^\top}{\mathbf{Z}_i^\top \Sigma^{-1} \mathbf{Z}_i} + (1 - \tau) \mathbf{I}_{d_n+1}.$$

Comme pour la matrice $\mathbf{S}_{n,\omega}$, le paramètre de shrinkage optimal τ^* est défini par le critère suivant :

$$\tau^* = \arg \min_{\tau \in [0,1]} \mathcal{P}(\tilde{\Sigma}(\tau), \Sigma).$$

Les fonctions de perte $\mathcal{P}$ utilisées sont alors les mêmes que précédemment. Le résultat suivant, issu de [Chen et al., 2011], donne le paramètre optimal dans le cas où $\mathcal{P}$ est la distance de Frobenius.

Proposition 59 (Paramètre Tyler optimal). *Soit $\mathbf{Z} \sim \mathcal{EC}_{d_n}(\mathbf{0}_{\mathbb{R}^{d_n+1}}, \Sigma, \xi)$ avec $\text{tr}(\Sigma) = d_n + 1$. Le paramètre ORACLE τ_O de $\mathbf{Z}$ est donné par :*

$$\tau_O = \frac{n \left(1 + \frac{2}{d_n+1}\right) \text{tr}(\Sigma^2) - n(d_n + 3)}{\left(n + 1 + \frac{2(n-1)}{d_n+1}\right) \text{tr}(\Sigma^2) + d_n + 1 - n(d_n + 3)}.$$

On constate que ce paramètre dépend uniquement de $\text{tr}(\Sigma^2)$. En reprenant notre estimateur $\hat{T}$ de l'Equation (57), on peut alors construire un estimateur consistant de τ_O .

Proposition 60 (Estimateur consistant de τ_O). *Soient $\mathbf{Z}_1, \dots, \mathbf{Z}_n$ des répétitions indépendantes et identiquement distribuées de même loi que $\mathbf{Z} \sim \mathcal{EC}_{d_n+1}(\mathbf{0}_{\mathbb{R}^{d_n}}, \Sigma, \xi)$ avec $\mathbb{E}[\xi^2] = 1$. Sous l'Hypothèse 6, l'estimateur $\hat{\tau}_O$ ci-dessous est alors un estimateur consistant de τ_O :*

$$\hat{\tau}_O = \frac{n(d_n + 3) (\hat{T} - 1)}{((n + 1)(d_n + 1) + 2(n - 1)) \hat{T} + (d_n + 1)^2 - n(d_n + 3)}.$$

Cette approche a l'avantage de ne pas faire intervenir l'estimateur de kurtosis $\hat{\kappa}$, qui nécessitait des moments d'ordre 8. On en déduit directement l'estimateur suivant du quantile conditionnel.

Définition 33 (Estimateur Tyler de quantile conditionnel). *Sous les Hypothèses 2 et 6, on définit l'estimateur suivant de $q_\alpha(Y|\mathbf{X} = \mathbf{x})$:*

$$\hat{q}_\alpha^{TYLER} = \hat{\mu}_{Y|\mathbf{x}}^\tau + \sqrt{\hat{\Sigma}_{Y|\mathbf{x}}^\tau} \sqrt{\frac{\mathbf{x}^\top \mathbf{V}_{n,\hat{\tau}_O}^{-1} \mathbf{x}}{d_n}} \Phi^{-1}(\alpha). \quad (\text{V.17})$$

avec

$$\begin{cases} \hat{\mu}_{Y|X}^\tau = & \mathbf{V}_{YX,n,\hat{\tau}_O} \mathbf{V}_{X,n,\hat{\tau}_O}^{-1} \mathbf{x} \\ \hat{\Sigma}_{Y|X}^\tau = & \mathbf{V}_{Y,n,\hat{\tau}_O} - \mathbf{V}_{YX,n,\hat{\tau}_O} \mathbf{V}_{X,n,\hat{\tau}_O}^{-1} \mathbf{V}_{XY,n,\hat{\tau}_O} \end{cases}.$$

On a défini dans cette section des estimateurs de shrinkage, qui consistent à perturber l'estimateur de la matrice de covariance. Il est donc évident qu'une telle méthode induit un biais sur l'estimation. Cependant, la variance d'estimation est en général bien plus faible. La Section [V.3.](#) met en évidence ce phénomène, en appliquant les estimateurs que l'on vient d'introduire, afin de comparer leurs performances.

V.3. Etude numérique

Par analogie avec les chapitres précédents, on propose une étude numérique sur une loi de Student. Pour cela, on va diviser notre étude en trois parties. La première traitera le cas $n < d_n$, la seconde prendra des valeurs semblables de n et d_n , quand le dernier prendra un nombre de covariables d_n bien plus grand que n .

V.3.a. En petite dimension

Avant de se consacrer aux cas $d_n > n$, il paraît intéressant de tester la qualité de l'approximation gaussienne dans les cas où la dimension de la covariable d_n est plus raisonnable. Pour cela, on considère par exemple que l'on dispose de $n = 200$ répétitions $\mathbf{Z}_1 = (\mathbf{X}_1, Y_1), \dots, \mathbf{Z}_n = (\mathbf{X}_n, Y_n)$ d'une loi de Student de dimension $d_n + 1 = 51$ à $\nu = 5$ degrés de liberté. Par commodité, on prendra $\boldsymbol{\mu} = \mathbf{0}_{\mathbb{R}^{d_n+1}}$ et $\boldsymbol{\Sigma} = \mathbf{I}_{d_n+1}$. On simule alors 100 jeux de données. Pour chaque échantillon, le but sera d'estimer le quantile à 95% de $Y|\mathbf{X} = \mathbf{x}$, où $\mathbf{x}$ est simulé. De manière à avoir un aperçu global, on applique l'estimateur [\(29\)](#) à chaque jeu de données, en utilisant respectivement les matrices $\mathbf{S}_n$ et $\mathbf{V}_n$, puis les estimateurs de shrinkage ORACLE et Tyler, pour finir avec les estimateurs sparses [\(30\)](#).

On s'intéresse en particulier au choix des paramètres λ_n pour les méthodes GLASSO et CLIME. Pour la première, le Théorème [52](#) nous pousse à choisir λ_n dominé par $\left(1 + \frac{d_n+1}{s_n+1}\right) \sqrt{\frac{\ln(d_n+1)}{n}} = 7.29$. On prendra alors $\lambda_n = 4.5$.

Concernant la méthode CLIME, on prend λ_n comme mentionné dans la Théorème [54](#) avec $\tau = 1$. De plus, dans le cas Student, et avec la matrice identité $\boldsymbol{\Sigma}$, on a $\theta = \mathbb{E}[\xi^4] = \frac{\nu^2}{(\nu-2)(\nu-4)} = \frac{25}{3}$ si $\nu = 5$. De même, la matrice de précision $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1}$ étant également la matrice identité, on prend $M = 1$ et on obtient un coefficient λ_n de 0.99. Dans cette configuration, l'Hypothèse [5](#) est vérifiée si il existe $\delta > 0$ tel que $\nu = 5 > 4k + 4 + \delta$, où k est tel que $d_n + 1 \leq cn^k$. d_n étant bien plus petit que n ici, on peut s'attendre à une

bonne performance pour cet estimateur, conformément au Théorème 54.

La formule théorique (7) des quantiles conditionnels dans le cas Student nous permet de comparer nos résultats. La Figure V.1 montre les boxplots des erreurs relatives de prédiction sur les 100 jeux de données.

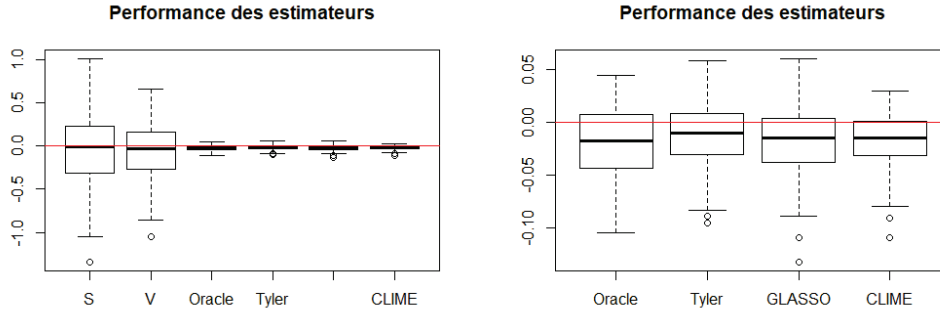


FIGURE V.1 – Boxplots de 100 répétitions des erreurs relatives $\frac{\hat{q}_\alpha(Y|\mathbf{X}=\mathbf{x})}{q_\alpha(Y|\mathbf{X}=\mathbf{x})} - 1$, pour les estimateurs basiques (29) avec la matrice $\mathbf{S}_n$ puis $\mathbf{V}_n$, ORACLE (32), Tyler (33), GLASSO et CLIME (30). A droite, un zoom sur les prédicteurs ORACLE, Tyler, GLASSO et CLIME.

On s'aperçoit que même si dans ce cas de figure, les matrices $\mathbf{S}_n$ et $\mathbf{V}_n$ sont inversibles et donc potentiellement exploitables, elles fournissent une prédiction certes centrée, mais extrêmement variable des quantiles. Même ici, il est donc plus judicieux d'opter pour les méthodes de grande dimension, légèrement biaisées mais bien moins volatiles. En particulier, l'estimateur Tyler semble être le plus efficace, les méthodes sparses paraissant légèrement plus biaisées, tout comme l'estimateur ORACLE. Les estimateurs basiques (29) sont donc généralement à bannir. Dans les exemples qui suivent, on considérera d'ailleurs des cas où $d_n > n$, sur lesquels ils ne pourront même pas être appliqués.

V.3.b. En "moyenne" dimension

On utilise ici le terme moyenne dimension pour faire référence au cas où la taille de l'échantillon n et la celle de la covariable d_n sont asymptotiquement proportionnelles, comme mentionné dans l'Hypothèse 6. On prendra dans cet exemple $n = 100$ et $d_n = 110$. Avec un tel choix, les matrices $\mathbf{S}_n$ et $\mathbf{V}_n$ ne sont plus inversibles. Dans ce cas, on a introduit des estimateurs de shrinkage que l'on va tester sur la même distribution de Student que précédemment. On applique alors le même protocole que celui décrit dans le Paragraphe V.3.a., avec les estimateurs des Définitions 32 et 33. La Figure V.3.b. donne un aperçu des différentes performances.

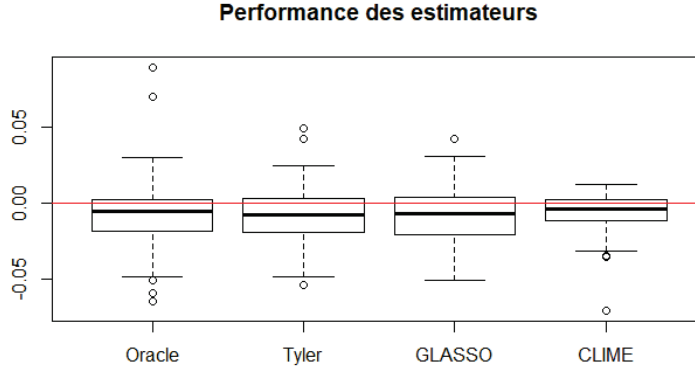


FIGURE V.2 – Boxplots de 100 répétitions des erreurs relatives $\frac{\hat{q}_\alpha(Y|\mathbf{X}=\mathbf{x})}{q_\alpha(Y|\mathbf{X}=\mathbf{x})} - 1$, pour les estimateurs ORACLE (32), Tyler (33), GLASSO et CLIME (30).

Parmi les estimateurs de shrinkage, l'estimateur de type Tyler semble être légèrement moins dispersé que son concurrent. Cela s'explique en partie par le fait que son paramètre optimal τ_O ne nécessite que l'estimation d'un paramètre, lorsque le paramètre ω_O en nécessite 3, notamment celle du kurtosis $\mathbb{E}[\xi^4]$ qui le rend plus variable. De plus, la distribution n'admet pas ici de moment d'ordre 8, ce qui compromet le résultat asymptotique (56) sur l'estimateur $\hat{\kappa}$ de l'Equation (56).

A titre comparatif, on teste également nos estimateurs GLASSO et CLIME. Pour le premier, on avait choisi un paramètre de pénalisation 4.5 pour $n = 200$ et $d_n = 50$. Conformément au Théorème 52, on prend λ_n proportionnel à $\left(1 + \frac{d_n+1}{s_n+1}\right) \sqrt{\frac{\ln(d_n+1)}{n}}$, soit environ 15 ici. Pour le prédicteur CLIME, le choix est plus délicat. En effet, avec d_n du même ordre de grandeur que n , nous n'entrons clairement plus dans le cadre de l'Hypothèse 5 (on rappelle qu'elle implique qu'il existe $\delta > 0$ tel que $5 > 4k + 4 + \delta$, avec k proche de 1 dans cet exemple). Le choix du λ_n du Théorème 54 nous expose donc à une très mauvaise performance. On décide donc de l'abaisser à 0.5.

On s'aperçoit que les estimateurs de sparsité semblent également assez efficaces, même pour $d_n \sim n$.

V.3.c. En grande dimension

Intéressons nous maintenant au cas où $d_n \gg n$. Pour cela, on va considérer la même loi de Student que précédemment, avec $n = 100$ et $d_n = 300$, puis effectuer le même procédé numérique sur nos estimateurs GLASSO ($\lambda_n = 45$) et CLIME ($\lambda_n = 0.5$). Le choix des paramètres λ_n découle du même raisonnement que celui du paragraphe précédent. On ajoute également nos estimateurs de shrinkage pour comparer leurs performances

respectives. La Figure [V.3.c.](#) compare nos quatre approches.

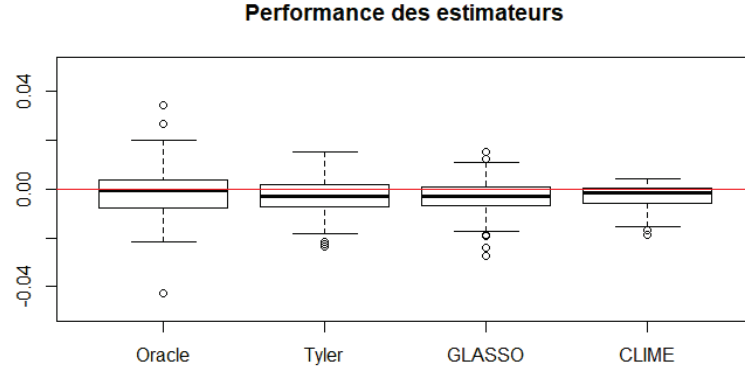


FIGURE V.3 – Boxplots de 100 répétitions des erreurs relatives $\frac{\hat{q}_\alpha(Y|\mathbf{X}=\mathbf{x})}{q_\alpha(Y|\mathbf{X}=\mathbf{x})} - 1$, pour les estimateurs ORACLE ([32](#)), Tyler ([33](#)), GLASSO et CLIME ([30](#)).

Le prédicteur ORACLE, bien que plus centré, semble être plus volatile que le Tyler, pour les raisons que l'on a déjà énoncées. De manière générale, les autres prédicteurs ont l'air relativement efficaces. Notamment, les estimateurs Tyler et GLASSO semblent se valoir en terme de biais, avec toutefois un léger avantage au GLASSO pour la variance d'estimation. Mais le plus efficace ici est le prédicteur CLIME, moins biaisé et volatile que ses deux concurrents.

CONCLUSION ET PERSPECTIVES

On propose dans ce chapitre final un bilan des sujets abordés dans ce manuscrit, ainsi que les potentielles pistes.

Dans cette thèse, nous nous sommes intéressés particulièrement aux quantile et expectile d'une variable aléatoire Y conditionnellement à un vecteur de covariables $\mathbf{X} = \mathbf{x}$ dans les Chapitres III et IV. De plus, on a considéré que le vecteur $\mathbf{Z} = (\mathbf{X}, Y)$ appartenait à une sous classe des distributions elliptiques dite consistante. Dans ce cadre, on a pu donner les paramètres optimaux des modèles de régression quantile puis expectile. Il apparaît alors que ces derniers peuvent s'avérer mauvais, en particulier pour des niveaux de quantile ou d'expectile extrêmes. On a donc proposé d'autres prédictors de quantile puis d'expectile extrêmes consistants et asymptotiquement normaux, sous des hypothèses de queue lourde. En se basant sur les prédictors de quantile extrêmes, on a pu déduire des prédictors d'autres mesures de risque telles que les L_p -quantiles ou les mesures d'Haezendonck-Goovaerts. Leurs propriétés asymptotiques sont également déduites. Cependant, ces prédictors extrêmes sont moins performants lorsque la taille de la covariable d est grande. Dans le Chapitre V qui fait office de perspective, on a alors montré que si cette dernière tendait vers $+\infty$, la distribution conditionnelle tendait vers une loi normale. On propose donc quelques approximations basées sur ce résultat, ainsi que sur des méthodes d'estimation en grande dimension. Tous nos prédictors sont illustrés par plusieurs études numériques tout au long du manuscrit.

On peut noter qu'il ne se dégage pas de méthode vraiment efficace pour estimer des quantiles non-extrêmes. Les modèles de régression présentés ne sont en effet pas recommandés. En prenant du recul, il paraît évident qu'une telle approche ne peut fonctionner. En effet, dans le cadre des distributions à queue lourde, on a vu via le Lemme 37 que l'indice de queue de distribution changeait pour la loi conditionnelle $Y|\mathbf{X} = \mathbf{x}$. Approcher le quantile par une combinaison affine de $\mathbf{x}$ paraît donc assez inapproprié, du moins dans ce cas de figure. Le passage au modèle linéaire généralisé, avec une fonction de lien proportionnelle à $t^{1/\eta}$, où η est donné dans l'Equation (15), fonctionnerait certainement bien mieux. Cependant, il paraît très peu évident que l'on obtienne des résultats théoriques comme ceux obtenus dans les Sections III.2 et IV.2. On pourrait également estimer directement la densité conditionnelle à partir des observations, puis en déduire les quantiles.

Dans le cas des mesures de risque extrêmes, nous travaillons sous l'Hypothèse 2 durant la majorité de l'étude. Il paraît évidemment intéressant de considérer d'autres cas de figures, comme par exemple [El Methni et al., 2018] ou [Mao and Hu, 2012] ont pu le faire récemment. Si le domaine d'attraction de Weibull, ne considérant que des points terminaux finis, ne présente pas spécialement d'intérêt dans le cadre des quantiles extrêmes, celui de Gumbel pourrait être extrêmement intéressant, car il englobe plusieurs distributions que l'on a donné en exemple dans la Section II telles que les lois normale, normale contaminée

et Laplace.

De manière plus générale, on pourrait également être tenté d'élargir le cadre des distributions elliptiques. Dans un premier temps, on pourrait supprimer l'hypothèse de consistance, ce qui compliquerait certes l'étude, mais paraît réalisable. Ensuite, nous pourrions considérer une famille de distributions généralisant les elliptiques, appelée les skew elliptiques. Ces distributions, que l'on peut trouver notamment dans [Genton, 2004] ou [Branco and Dey, 2001], ajoutent de l'asymétrie aux distributions elliptiques classiques, et sont appréciées en finance (voir [Framstad, 2011]).

Dans toute la thèse, on considère des mesures de risque sur $Y|\mathbf{X} = \mathbf{x}$, où Y est univariée. On pourrait très bien imaginer estimer des mesures de risque sur un vecteur $\mathbf{Y}|\mathbf{X} = \mathbf{x}$. En effet, ces dernières années, la littérature est riche de mesures de risques généralisées au cas multivarié. Sans être exhaustif, [Cousin and Di Bernardino, 2013], [Maume-Deschamps et al., 2017a] et [Cousin and Di Bernardino, 2014] ont proposé respectivement des quantiles, expectiles et TVaR multivariés. Plus précisément, le papier de [Landsman et al., 2016] s'est intéressé aux TVaR multivariées dans le cas des elliptiques. On pourrait donc tout à fait s'y consacrer dans un cadre conditionnel.

Enfin, il existe une multitude de mesures de risques auxquelles nous pourrions nous intéresser. Nous nous sommes cantonnés aux L_p -quantiles et mesures d'Haezendonck-Goovaerts, mais la littérature fournit régulièrement de nouvelles mesures. Par exemple, [Belles-Sampera et al., 2014] ont proposé la Glue Value at Risk. Par ailleurs, on trouve dans [Breckling and Chambers, 1988] une famille de mesures incluant les L_p -quantiles, nommée M -quantiles. D'autres mesures plus répandues telles que l'Expected Shortfall, estimée dans [Daouia et al., 2017a], pourraient également être considérées.

Pour conclure, une piste serait également, à l'image de l'estimateur de covariance de [Cressie, 1988], de fournir des estimateurs consistants dans un cadre spatial de nos paramètres extrémaux η et $\ell(\mathbf{x})$.

BIBLIOGRAPHIE

- [Abramowitz et al., 1966] Abramowitz, M., Stegun, I. A., et al. (1966). Handbook of mathematical functions. Applied mathematics series, 55(62) :39.
- [Artzner et al., 1999] Artzner, P., Delbaen, F., Eber, J., and Heath, D. (1999). Coherent measures of risk. Mathematical Finance, 9(3) :203–228.
- [Banerjee et al., 2008] Banerjee, O., Ghaoui, L., and D’Aspremont, A. (2008). Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data. Journal of Machine learning research, 9(Mar) :485–516.
- [Bargès et al., 2009] Bargès, M., Cossette, H., and Marceau, E. (2009). TVaR-based capital allocation with copulas. Insurance : Mathematics and Economics, 45(3) :348–361.
- [Belles-Sampera et al., 2014] Belles-Sampera, J., Guillén, M., and Santolino, M. (2014). Beyond value-at-risk : Gluevar distortion risk measures. Risk Analysis, 34(1) :121–134.
- [Bellini et al., 2014] Bellini, F., Klar, B., Muller, A., and Gianin, E. R. (2014). Generalized quantiles as risk measures. Insurance : Mathematics and Economics, 54 :41–48.
- [Bellini and Rosazza Gianin, 2008] Bellini, F. and Rosazza Gianin, E. (2008). On Haezendonck risk measures. Journal of Banking & Finance, 32 :986–994.
- [Bingham et al., 1989] Bingham, N. H., Goldie, C. M., and Teugels, J. L. (1989). Regular variation, volume 27. Cambridge university press.
- [Branco and Dey, 2001] Branco, M. D. and Dey, D. K. (2001). A general class of multivariate skew-elliptical distributions. Journal of Multivariate Analysis, 79(1) :99–113.
- [Breckling and Chambers, 1988] Breckling, J. and Chambers, R. (1988). M-quantiles. Biometrika, 75(4).
- [Cai and Weng, 2016] Cai, J. and Weng, C. (2016). Optimal reinsurance with expectile. Scandinavian Actuarial Journal, 2016(7) :624–645.
- [Cai et al., 2011] Cai, T., Liu, W., and Luo, X. (2011). A constrained ℓ_1 minimization approach to sparse precision matrix estimation. Journal of the American Statistical Association, 106(494) :594–607.
- [Cambanis et al., 1981] Cambanis, S., Huang, S., and Simons, G. (1981). On the theory of elliptically contoured distributions. Journal of Multivariate Analysis, 11 :368–385.
- [Cellier et al., 1989] Cellier, D., Fourdrinier, D., and Robert, C. (1989). Robust shrinkage estimators of the location parameter for elliptically symmetric distributions. Journal of Multivariate Analysis, 29(1) :39–52.
- [Chambers and Tzavidis, 2006] Chambers, R. and Tzavidis, N. (2006). M-quantile models for small area estimation. Biometrika, 93(2) :255–268.

- [Chen et al., 2011] Chen, Y., Wiesel, A., and Hero, A. O. (2011). Robust Shrinkage Estimation of High-Dimensional Covariance Matrices. IEEE Transactions on Signal Processing, 59(9) :4097–4107.
- [Chen, 1996] Chen, Z. (1996). Conditional Lp-quantiles and their application to the testing of symmetry in non-parametric regression. Statistics & Probability Letters, 29(2) :107 – 115.
- [Chiancone et al., 2017] Chiancone, A., Forbes, F., and Girard, S. (2017). Student Sliced Inverse Regression. Computational Statistics and Data Analysis, 113 :441–456.
- [Cousin and Di Bernardino, 2013] Cousin, A. and Di Bernardino, E. (2013). On multivariate extensions of Value-at-Risk. Journal of Multivariate Analysis, 119 :32–46.
- [Cousin and Di Bernardino, 2014] Cousin, A. and Di Bernardino, E. (2014). On multivariate extensions of conditional-tail-expectation. Insurance : Mathematics and Economics, 55 :272–282.
- [Cressie, 1988] Cressie, N. (1988). Spatial prediction and ordinary kriging. Mathematical Geology, 20(4) :405–421.
- [Daouia et al., 2013] Daouia, A., Gardes, L., and Girard, S. (2013). On kernel smoothing for extremal quantile regression. Bernoulli, 19(5B) :2557–2589.
- [Daouia et al., 2017a] Daouia, A., Girard, S., and Stupfler, G. (2017a). Estimation of tail risk based on Extreme Expectiles. Journal of the Royal Statistical Society : Series B.
- [Daouia et al., 2017b] Daouia, A., Girard, S., and Stupfler, G. (2017b). Extreme M-quantiles as risk measures : from L1 to Lp optimization. Bernoulli.
- [de Haan and Ferreira, 2006] de Haan, L. and Ferreira, A. (2006). Extreme value theory : an introduction. Springer Science & Business Media.
- [de Haan and Resnick, 1998] de Haan, L. and Resnick, S. (1998). On asymptotic normality of the hill estimator. Communications in Statistics, 14(4) :849–866.
- [de Haan and Rootzén, 1993] de Haan, L. and Rootzén, H. (1993). On the estimation of high quantiles. Journal of Statistical Planning and Inference, 35(1) :1–13.
- [de Valk, 2016] de Valk, C. (2016). Approximation of high quantiles from intermediate quantiles. Extremes, 19 :661–686.
- [Dekkers and De Haan, 1989] Dekkers, A. L. and De Haan, L. (1989). On the estimation of the extreme-value index and large quantile estimation. The Annals of Statistics, 17(4) :1795–1832.
- [Dey and Srinivasan, 1985] Dey, D. K. and Srinivasan, C. (1985). Estimation of a covariance matrix under Stein’s loss. The Annals of Statistics, 13(4) :1581–1591.
- [Di Bernardino and Prieur, 2017] Di Bernardino, E. and Prieur, C. (2017). Estimation of the multivariate conditional-tail-expectation for extreme risk levels : illustrations on environmental data-sets.

- [Djurčić and Torgašev, 2001] Djurčić, D. and Torgašev, A. (2001). Strong asymptotic equivalence and inversion of functions in the class kc . Journal of Mathematical Analysis and Applications, 255 :383–390.
- [El Methni et al., 2014] El Methni, J., Gardes, L., and Girard, S. (2014). Non-parametric Estimation of Extreme Risk Measures from Conditional Heavy-tailed Distributions. Scandinavian Journal of Statistics, 41(4) :988–1012.
- [El Methni et al., 2018] El Methni, J., Gardes, L., and Girard, S. (2018). Kernel estimation of extreme regression risk measures. Electronic Journal of Statistics, 12 :359–398.
- [El Methni et al., 2012] El Methni, J., Gardes, L., Girard, S., and Guillou, A. (2012). Estimation of extreme quantiles from heavy and light tailed distributions. Journal of Statistical Planning and Inference, 142(10) :2735–2747.
- [Eltoft et al., 2006] Eltoft, T., Kim, T., and Lee, T.-W. (2006). On the multivariate Laplace distribution. IEEE Signal Processing Letters, 13(5) :300–303.
- [Fougères and Soulier, 2010] Fougères, A. L. and Soulier, P. (2010). Limit conditional distributions for bivariate vectors with polar representation. Stochastic Models, 26 :54–77.
- [Frahm, 2004] Frahm, G. (2004). Generalized Elliptical Distributions : Theory and Applications. PhD thesis, Universität zu Köln.
- [Framstad, 2011] Framstad, N. C. (2011). Portfolio separation properties of the skew-elliptical distributions, with generalizations. Statistics & Probability Letters, 81(12) :1862–1866.
- [Friedman et al., 2008] Friedman, J., Hastie, T., and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. Biostatistics, 9(3) :432–441.
- [Frontini and Sormani, 2003] Frontini, M. and Sormani, E. (2003). Some variant of newton’s method with third-order convergence. Applied Mathematics and Computation, 140.
- [Gardes and Girard, 2005] Gardes, L. and Girard, S. (2005). Estimating extreme quantiles of weibull tail distributions. Communications in Statistics-Theory and Methods, 34 :1065–1080.
- [Genton, 2004] Genton, M. G. (2004). Skew-elliptical distributions and their applications : a journey beyond normality. CRC Press.
- [Gneiting, 2011] Gneiting, T. (2011). Making and evaluating point forecasts. Journal of the American Statistical Association, 106(494) :746–762.
- [Goovaerts et al., 2004] Goovaerts, M., Kaas, R., Dhaene, J., and Tang, Q. (2004). Some new classes of consistent risk measures. Insurance : Mathematics and Economics, 34 :505–516.
- [Grant and Gneiting, 2012] Grant, K. and Gneiting, T. (2012). Consistent scoring functions for quantiles. From Probability to Statistics and Back : High-Dimensional Models and Processes, 9 :163–173.

- [Gupta and Song, 1997] Gupta, A. and Song, D. (1997). L_p -norm spherical distribution. Journal of Statistical Planning and Inference, 60(2) :241–260.
- [Haezendonck and Goovaerts, 1982] Haezendonck, J. and Goovaerts, M. (1982). A new premium calculation principle based on Orlicz norms. Insurance : Mathematics and Economics, 1 :41–53.
- [Hashorva, 2006a] Hashorva, E. (2006a). Gaussian approximation of conditional elliptical random vectors. Stochastic Models, 22 :441–457.
- [Hashorva, 2006b] Hashorva, E. (2006b). On the regular variation of elliptical random vectors. Statistics & Probability Letters, 76 :1427–1434.
- [Hashorva, 2007] Hashorva, E. (2007). Extremes of conditioned elliptical random vectors. Journal of Multivariate Analysis, 98(8) :1583–1591.
- [Hashorva and Jaworski, 2012] Hashorva, E. and Jaworski, P. (2012). Gaussian approximation of conditional elliptical copulas. Journal of Multivariate Analysis, 111 :397–407.
- [Hashorva and Pakes, 2010] Hashorva, E. and Pakes, A. G. (2010). Tail asymptotics under beta random scaling. Journal of Mathematical Analysis and Applications, 372 :496–514.
- [Hill, 1975] Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. The Annals of Statistics, 3(5) :1163–1174.
- [Hua and Joe, 2011] Hua, L. and Joe, H. (2011). Second order regular variation and conditional tail expectation of multiple risks. Insurance : Mathematics and Economics, 49 :537–546.
- [Hult and Lindskog, 2002] Hult, H. and Lindskog, F. (2002). Multivariate extremes, aggregation and dependence in elliptical distributions. Advances in Applied Probability, 34(3) :587–608.
- [Hunter and Lange, 2000] Hunter, D. and Lange, K. (2000). Quantile regression via an mm algorithm. Journal of Computational and Graphical Statistics, 9(1) :60–77.
- [Hunter and Lange, 2004] Hunter, D. and Lange, K. (2004). A Tutorial on MM Algorithms. The American Statistician, 58(1) :30–37.
- [Jagger and Elsner, 2009] Jagger, T. H. and Elsner, J. B. (2009). Modeling tropical cyclone intensity with quantile regression. International Journal of Climatology, 29(10) :1351–1361.
- [Jessen and Mikosch, 2006] Jessen, H. A. and Mikosch, T. (2006). Regularly varying functions. Publications de l’Institut Mathématique, 80(94) :171–192.
- [Johnson, 1987] Johnson, M. (1987). Multivariate Statistical Simulation. Wiley & Sons.
- [Jones, 1994] Jones, M. (1994). Expectiles and M-quantiles are quantiles. Statistics & Probability Letters, 20 :149–153.
- [Jorion, 1996] Jorion, P. (1996). Risk2 : Measuring the risk in value at risk. Financial analysts journal, 52(6) :47–56.

- [Kallenberg, 2006] Kallenberg, O. (2006). Foundations of modern probability. Springer Science & Business Media.
- [Kano, 1994] Kano, Y. (1994). Consistency property of elliptical probability density functions. Journal of Multivariate Analysis, 51 :139–147.
- [Kelker, 1970] Kelker, D. (1970). Distribution theory of spherical distributions and a location-scale parameter generalization. Sankhya : The Indian Journal of Statistics, Series A, 32(4) :419–430.
- [Koenker and Bassett, 1978] Koenker, R. and Bassett, G. J. (1978). Regression quantiles. Econometrica, 46(1) :33–50.
- [Kratz and Resnick, 1996] Kratz, M. and Resnick, S. I. (1996). The qq-estimator and heavy tails. Stochastic Models, 12(4) :699–724.
- [Krishnamoorthy and Gupta, 1989] Krishnamoorthy, K. and Gupta, A. (1989). Improved minimax estimation of a normal precision matrix. The Canadian Journal of Statistics, 17(1) :91–102.
- [Lam and Fan, 2009] Lam, C. and Fan, J. (2009). Sparsistency and rates of convergence in large covariance matrix estimation. The Annals of Statistics, 37(6B) :4254–4278.
- [Landsman et al., 2016] Landsman, Z., Makov, U., and Shushi, T. (2016). Multivariate tail conditional expectation for elliptical distributions. Insurance : Mathematics and Economics, 70 :216–223.
- [Landsman and Valdez, 2003] Landsman, Z. and Valdez, E. (2003). Tail conditional expectations for elliptical distributions. North American Actuarial Journal, 7(4) :55–71.
- [Ledoit and Wolf, 2004] Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. Journal of Multivariate Analysis, 88 :365–411.
- [Li and Zhang, 2010] Li, F. and Zhang, N. R. (2010). Bayesian Variable Selection in Structured High-Dimensional Covariate Spaces With Applications in Genomics. Journal of the American Statistical Association, 105(491) :1202–1214.
- [Li, 1991] Li, K.-C. (1991). Sliced inverse regression for dimension reduction. Journal of the American Statistical Association, 86(414) :316–327.
- [Ligas and Kulczycki, 2010] Ligas, M. and Kulczycki, M. (2010). Simple spatial prediction by least squares prediction, simple kriging, and conditional expectation of normal vector. Geodasy and Cartography, 59(2) :69–81.
- [Linsmeier and Pearson, 2000] Linsmeier, T. J. and Pearson, N. D. (2000). Value at risk. Financial Analysts Journal, 56(2) :47–67.
- [Mao and Hu, 2012] Mao, T. and Hu, T. (2012). Second-order properties of the Haezendonck-Govaerts risk measure for extreme risks. Insurance : Mathematics and Economics, 51 :333–343.
- [Maronna, 1976] Maronna, R. A. (1976). Robust M-Estimators of Multivariate Location and Scatter. The Annals of Statistics, 4(1) :51–67.

- [Maume-Deschamps et al., 2017a] Maume-Deschamps, V., Rulliere, D., and Said, K. (2017a). Multivariate extensions of expectiles risk measures. Dependence Modeling, 5(1) :20–44.
- [Maume-Deschamps et al., 2017b] Maume-Deschamps, V., Rullière, D., and Usseglio-Carleve, A. (2017b). Quantile predictions for elliptical random fields. Journal of Multivariate Analysis, 159 :1 – 17.
- [Maume-Deschamps et al., 2018] Maume-Deschamps, V., Rullière, D., and Usseglio-Carleve, A. (2018). Spatial expectile predictions for elliptical random fields. Methodology and Computing in Applied Probability, 20(2) :643–671.
- [McNeil et al., 2015] McNeil, A. J., Frey, R., and Embrechts, P. (2015). Quantitative risk management : Concepts, techniques and tools. Princeton university press.
- [Muirhead, 2009] Muirhead, R. J. (2009). Aspects of multivariate statistical theory, volume 197. John Wiley & Sons.
- [Nadarajah and Kotz, 2004] Nadarajah, S. and Kotz, S. (2004). Multivariate T-Distributions and Their Applications. Cambridge University Press.
- [Nagar and Gupta, 2011] Nagar, D. K. and Gupta, A. K. (2011). Expectations of functions of complex Wishart matrix. Acta applicandae mathematicae, 113(3) :265–288.
- [Newey and Powell, 1987] Newey, W. and Powell, J. (1987). Asymmetric least squares estimation and testing. Econometrica, 55(4) :819–847.
- [Ollila, 2017] Ollila, E. (2017). Optimal high-dimensional shrinkage covariance estimation for elliptical distributions. In Signal Processing Conference (EUSIPCO), 2017 25th European, pages 1639–1643. IEEE.
- [Opitz, 2016] Opitz, T. (2016). Modeling asymptotically independent spatial extremes based on Laplace random fields. Spatial Statistics, 16 :1–18.
- [Owen and Rabinovitch, 1983] Owen, J. and Rabinovitch, R. (1983). On the class of elliptical distributions and their applications to the theory of portfolio choice. The Journal of Finance, 38(3) :745–752.
- [Parzen, 1962] Parzen, E. (1962). On estimation of a probability density function and mode. Annals of Mathematical Statistics, 33(3) :1065–1076.
- [Pickands, 1975] Pickands, J. (1975). Statistical inference using extreme order statistics. The Annals of Statistics, 3(1) :119–131.
- [Resnick and Stărică, 1997] Resnick, S. and Stărică, C. (1997). Smoothing the hill estimator. Advances in Applied Probability, 29(1) :271–293.
- [Rosenblatt, 1956] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. The Annals of Mathematical Statistics, 27(3) :832–837.
- [Saerens, 2000] Saerens, M. (2000). Building cost functions minimizing to some summary statistics. IEEE Transactions on neural networks, 11(6) :1263–1271.

- [Schultze and Steinebach, 1996] Schultze, J. and Steinebach, J. (1996). On least squares estimates of an exponential tail coefficient. Statistics & Decisions, 14(3) :353–372.
- [Shahriari and Ahmadi, 2015] Shahriari, H. and Ahmadi, O. (2015). Robust estimation of the mean vector for high-dimensional data set using robust clustering. Journal of Applied Statistics, 42(6) :1183–1205.
- [Sobotka et al., 2013] Sobotka, F., Kauermann, G., Waltrup, L. S., and Kneib, T. (2013). On confidence intervals for semiparametric expectile regression. Statistics and Computing, 23(2) :135–148.
- [Tang and Yang, 2012] Tang, Q. and Yang, F. (2012). On the Haezendonck-Gooovaerts risk measure for extreme risks. Insurance : Mathematics and Economics, 50 :217–227.
- [Taylor, 2008] Taylor, J. W. (2008). Estimating Value at Risk and Expected Shortfall Using Expectiles. Journal of Financial Econometrics, 6(2) :231–252.
- [Thibaud and Opitz, 2015] Thibaud, E. and Opitz, T. (2015). Efficient inference and simulation for elliptical pareto processes. Biometrika, 102(4) :855–870.
- [Thomson, 1979] Thomson, W. (1979). Eliciting production possibilities from a well-informed manager. Journal of Economic Theory, 20(3) :360–380.
- [Tibshirani, 1997] Tibshirani, R. (1997). The lasso method for variable selection in the Cox model. Statistics in medicine, 16(4) :385–395.
- [Tukey, 1960] Tukey, J. (1960). A survey of sampling from contaminated distributions. In Contributions to Probability and Statistics : Essays in Honor of Harold Hotelling, chapter 39, pages 448–485. Stanford University Press, California.
- [Tyler, 1987] Tyler, D. (1987). A Distribution-free M-estimator of Multivariate Scatter. The Annals of Statistics, 15(1) :234–251.
- [Usseglio-Carleve, 2017] Usseglio-Carleve, A. (2017). Estimation of conditional extreme risk measures from heavy-tailed elliptical random vectors. working paper.
- [Wang et al., 2015] Wang, C., Pan, G., Tong, T., and Zhu, L. (2015). Shrinkage estimation of large dimensional precision matrix using random matrix theory. Statistica Sinica, 25 :993–1008.
- [Wang et al., 2012] Wang, H. J., Li, D., and He, X. (2012). Estimation of high conditional quantiles for heavy-tailed distributions. Journal of the American Statistical Association, 107(500) :1453–1464.
- [Wang and Genton, 2006] Wang, J. and Genton, M. (2006). The multivariate skew-slash distribution. Journal of Statistical Planning and Inference, 136(1) :209–220.
- [Weissman, 1978] Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations. Journal of the American Statistical Association, 73(364) :812–815.
- [Xiao and Valdez, 2015] Xiao, Y. and Valdez, E. (2015). A Black-Litterman asset allocation under Elliptical distributions. Quantitative Finance, 15(3) :509–519.

- [Yang and Berger, 1994] Yang, R. and Berger, J. O. (1994). Estimation of a covariance matrix using the reference prior. The Annals of Statistics, 22(3) :1195–1211.
- [Yang et al., 2018] Yang, Y., Zhang, T., and Zou, H. (2018). Flexible Expectile Regression in Reproducing Kernel Hilbert Spaces. Technometrics, 60(1) :26–35.
- [Zhang and Wiesel, 2016] Zhang, T. and Wiesel, A. (2016). Automatic diagonal loading for tyler’s robust covariance estimator. In Statistical Signal Processing Workshop (SSP), 2016 IEEE, pages 1–5. IEEE.
- [Zheng, 2011] Zheng, S. (2011). Gradient descent algorithms for quantile regression with smooth approximation. International Journal of Machine Learning and Cybernetics, 2(3) :191–207.
- [Ziegel, 2016] Ziegel, J. (2016). Coherence and elicibility. Mathematical Finance, 26(4) :901–918.

FONCTIONS ET LOIS USUELLES

Définition 34 (Fonction Gamma). *La fonction Gamma Γ est définie sur $\mathbb{R}$ privé des entiers négatifs de la manière suivante :*

$$\Gamma(t) = \int_0^{+\infty} u^{t-1} e^{-u} du.$$

Définition 35 (Fonction Bêta). *La fonction Bêta B est définie sur $\mathbb{R}_+^* \times \mathbb{R}_+^*$ de la manière suivante :*

$$B(s, t) = \frac{\Gamma(s) \Gamma(t)}{\Gamma(s+t)}.$$

Définition 36 (Fonction Digamma). *La fonction Digamma est définie sur $\mathbb{R}$ privé des entiers négatifs comme la dérivée logarithmique de la fonction gamma :*

$$\Psi(t) = \frac{\Gamma'(t)}{\Gamma(t)}.$$

Définition 37 (Fonction de Bessel de second type modifiée). *La fonction de Bessel de second type d'ordre $m \in \mathbb{R}$ modifiée est définie sur $\mathbb{R}_+^*$ comme :*

$$K_m(t) = \frac{\pi}{2} \frac{I_{-m}(t) - I_m(t)}{\sin(m\pi)},$$

avec $I_m(t) = \sum_{k=0}^{+\infty} \frac{1}{k! \Gamma(k+m+1)} \left(\frac{t}{2}\right)^{2k+m}.$

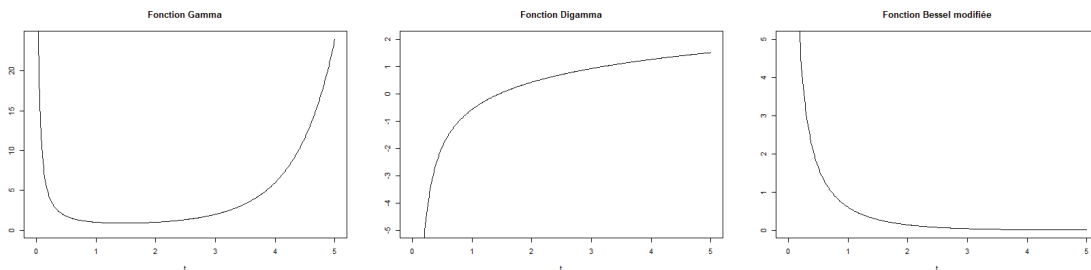


FIGURE A.1 – Fonctions Gamma, Digamma et Bessel modifiée ($m = 1$) sur $]0, 5]$

Définition 38 (Loi Bêta). La loi Bêta de paramètres $a, b > 0$, notée $\beta(a, b)$, admet pour densité de probabilité

$$\frac{t^{a-1}(1-t)^{b-1}}{B(a, b)}, \quad t \in [0, 1],$$

où B est la fonction Bêta donnée dans la Définition 35.

Définition 39 (Loi de Pareto). La loi de Pareto de paramètre $a > 0$, notée $\mathcal{P}(a)$, admet pour densité de probabilité

$$at^{-1-a}, \quad t \geq 1.$$

Notons que la loi de Pareto de paramètre a peut être vue comme l'inverse d'une loi Bêta de paramètres a et $b = 1$.

Définition 40 (Loi de Burr). La loi de Burr de paramètres $a, b > 0$ admet pour densité de probabilité

$$ab \frac{t^{a-1}}{(1+t^a)^{b+1}}, \quad t > 0.$$

Définition 41 (Loi de Fisher). La loi de Fisher de paramètres $\nu_1, \nu_2 > 0$ admet pour densité de probabilité

$$\frac{\left(\frac{\nu_1 t}{\nu_1 t + \nu_2}\right)^{\nu_1/2} \left(1 - \frac{\nu_1 t}{\nu_1 t + \nu_2}\right)^{\nu_2/2}}{t B\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)}, \quad t \geq 0,$$

où B est la fonction Bêta donnée dans la Définition 35.

Définition 42 (Loi Gamma). La loi Gamma de paramètres $k, \theta > 0$ admet pour densité de probabilité

$$t^{k-1} \frac{\theta^k e^{-\theta t}}{\Gamma(k)}, \quad t > 0,$$

où Γ est la fonction Gamma donnée dans la Définition 34.

Définition 43 (Loi exponentielle). La loi exponentielle de paramètre $\lambda > 0$, notée $\mathcal{E}(\lambda)$, admet pour densité de probabilité

$$\lambda e^{-\lambda t}, \quad t \geq 0.$$

On ajoute qu'une loi exponentielle de paramètre $\lambda > 0$ est une loi Gamma de paramètres $k = 1$ et $\theta = \lambda$.

Définition 44 (Loi du χ^2). La loi du χ^2 à $m \in \mathbb{N}^*$ degrés de liberté, notée χ_m^2 , admet pour densité de probabilité

$$\frac{1}{2^{m/2} \Gamma\left(\frac{m}{2}\right)} t^{m/2-1} e^{-t/2}, \quad t \geq 0.$$

CONDITIONS DU NOYAU K

On considère ici des répétitions $x_1, \dots, x_n$ indépendantes réelles d'une loi de densité $f : \mathbb{R} \rightarrow \mathbb{R}_+$, et on s'intéresse à l'estimateur à noyau de f pour un certain $t \in \mathbb{R}$:

$$\hat{f}_n(t) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{t - x_i}{h_n}\right),$$

où $(h_n)_{n \in \mathbb{N}}$ est une suite vérifiant $h_n \rightarrow 0$ et $nh_n \rightarrow +\infty$ quand $n \rightarrow +\infty$. Le résultat suivant donne un résultat de convergence de l'estimateur $\hat{f}_n$ sous certaines conditions concernant K .

Théorème 61 (Convergence de l'estimateur à noyau). *Si :*

- (i) $\int_{\mathbb{R}} K(u) du = 1$, $K(u) = K(-u)$, $\forall u \in \mathbb{R}$ et $\int_{\mathbb{R}} u^2 K(u) du \neq 0$,
- (ii) K est à support compact sur $] -1, 1[$ et bornée,
- (iii) Au voisinage de t , f est deux fois dérivable, et $f^{(m)}$ est continue et bornée, $\forall m = 0, 1, 2$,

alors

$$\begin{cases} \mathbb{E} [\hat{f}_n(t)] - f(t) &= O(h_n^2) \\ \text{Var} [\hat{f}_n(t)] &= O\left(\frac{1}{nh_n}\right) \end{cases}.$$

Si on suppose de plus que :

- (iv) $nh_n^5 \rightarrow 0$ quand $n \rightarrow +\infty$,

alors

$$\sqrt{nh_n} \left(\hat{f}_n(t) - f(t) \right) \xrightarrow{n \rightarrow +\infty} \mathcal{N} \left(0, f(t) \int K^2(u) du \right).$$

La Table [B.1](#), issue de [\[Parzen, 1962\]](#), donne quelques exemples de noyaux.

Noyau $K(t)$	$\int K^2(u)du$
$\frac{1}{2}\mathbb{1}_{\{ t \leq 1\}}$	$\frac{1}{2}$
$(1 - t)\mathbb{1}_{\{ t \leq 1\}}$	$\frac{2}{3}$
$\left(\frac{4}{3} - 8t^2 + 8 t ^3\right)\mathbb{1}_{\{ t <\frac{1}{2}\}} + \frac{8}{3}(1 - t)^3\mathbb{1}_{\{\frac{1}{2}\leq t \leq 1\}}$	0.96
$\frac{1}{2\pi}e^{-t^2/2}$	$\frac{1}{2\sqrt{\pi}}$
$\frac{1}{2}e^{- t }$	$\frac{1}{2}$
$\frac{1}{\pi(1+t^2)}$	$\frac{1}{\pi}$
$\frac{1}{2\pi}\left(2\frac{\sin(t/2)}{t}\right)^2$	$\frac{1}{3\pi}$

TABLE B.1 – Exemples de noyaux utilisables, et leur terme de variance associé.