Université de Montréal

Une famille de distributions symétriques et leptocurtiques représentée par la différence de deux variables aléatoires gamma

par

Maciej Augustyniak

Département de mathématiques et de statistique
Faculté des arts et des sciences

Mémoire présenté à la Faculté des études supérieures
en vue de l'obtention du grade de
Maître ès sciences (M.Sc.)
en Statistique

octobre 2008

Université de Montréal
Faculté des études supérieures

Ce mémoire intitulé

Une famille de distributions symétriques et leptocurtiques représentée par la différence de deux variables aléatoires gamma

présenté par

Maciej Augustyniak

a été évalué par un jury composé des personnes suivantes :

Mylène Bédard
(président-rapporteur)

Louis G. Doray
(directeur de recherche)

Alejandro Murua
(membre du jury)

Mémoire accepté le:
Octobre 2008

SOMMAIRE

Dans ce mémoire, nous introduisons une famille de distributions symétriques et leptocurtiques représentée par la différence de deux variables aléatoires gamma. Nous la définissons par sa fonction caractéristique puisque les fonctions de densité et de répartition ne possèdent pas de forme fermée. Nous consacrons une partie de ce mémoire à la fonction caractéristique et aux propriétés importantes qui y sont rattachées. De plus, nous présentons la classe d'estimateurs basés sur une méthode minimisant une distance quadratique introduite par Luong et Thompson (1987). Nous suggérons d'estimer les paramètres de la nouvelle famille en minimisant la distance quadratique entre les parties réelles des fonctions caractéristiques empiriques et théoriques. Nous montrons que l'estimateur qui en découle est convergent, robuste et asymptotiquement normal. Des tests d'ajustement basés sur la distance quadratique sont également introduits. Avant d'ajuster un échantillon à la famille de distributions symétriques, il est nécessaire de vérifier la compatibilité de l'échantillon avec le modèle. Pour ce faire, nous présentons deux tests de symétrie. Finalement, une étude de simulation est menée afin d'analyser les propriétés des estimateurs obtenus.

Mots clés : Fonction caractéristique, estimateur basé sur la distance quadratique, estimation de paramètres, leptocurtose, loi gamma, loi Laplace, tests d'ajustement, test de symétrie.

SUMMARY

In this thesis, we introduce a family of leptokurtic symmetric distributions represented by the difference of two gamma variates. We define it by its characteristic function since the density and distribution functions do not have closed forms. Thus, a review of properties of characteristic functions is provided. Moreover, we introduce the class of quadratic distance estimators presented by Luong and Thompson (1987). We suggest estimating the parameters of the new family by minimizing a quadratic distance between the real parts of the empirical and theoretical characteristic functions. The quadratic distance estimator obtained is shown to be consistent, robust and asymptotically normally distributed. Goodness-of-fit tests based on the quadratic distance are presented. Before fitting data to this symmetric family, it is necessary to verify that the data exhibit this characteristic. Thus, two tests of symmetry are also introduced. Simulation results are provided and properties of the estimators are studied.

Key Words : Characteristic function, quadratic distance estimator, parameter estimation, leptokurtosis, gamma distribution, Laplace distribution, goodness-of-fit, test of symmetry.

TABLE DES MATIÈRES

Sommaire iii

Summary iv

Remerciements 1

Introduction 2

Chapitre 1. La fonction caractéristique 4

1.1. Définitions et théorèmes 4

1.2. Exemples 8

1.2.1. La loi normale 9

1.2.2. La loi Laplace 11

1.2.3. La loi gamma 12

1.3. La fonction caractéristique empirique 15

Chapitre 2. Tests de symétrie 19

2.1. Médiane connue 19

2.1.1. Le test hybride 20

2.2. Médiane inconnue 22

2.2.1. Le test des triplets 22

Chapitre 3. Estimateurs basés sur une méthode minimisant une distance quadratique (DQ) 25

3.1. Définitions et propriétés asymptotiques 25

3.2. Tests d'ajustement 28
3.2.1. Hypothèse simple 29
3.2.2. Hypothèse composée 29
3.3. La DQ avec la fonction caractéristique 30
Chapitre 4. L'article 32
Conclusion 61
Bibliographie 63
Annexe A. Procédures codées avec Maple 11.0 ..... A-i

REMERCIEMENTS

Dans un premier temps, je tiens à remercier mon directeur de recherche, Louis Doray. Il a su susciter mon intérêt et ma curiosité intellectuelle en me présentant une idée dans laquelle j'ai pu utiliser mes connaissances antérieures et en approfondir de nouvelles afin de produire le présent travail de recherche.

Dans un second temps, je tiens à remercier tous les professeurs des universités Concordia et de Montréal qui ont contribué à mon éducation. Plus particulièrement, j'aimerais remercier les professeurs José Garrido, Ewa Duma, Pawel Góra, Charles Dugas et Christian Léger pour m'avoir encouragé à poursuivre des études supérieures et pour m'avoir donné de judicieux conseils reliés à mon éducation et à ma carrière professionnelle.

Je tiens également à remercier ma famille qui m'a toujours encouragé dans mes études et qui m'a offert un encadrement propice à ma réussite scolaire. Finalement, je tiens à remercier le Conseil de Recherches en Sciences Naturelles et en Génie du Canada (CRSNG) pour leur aide financière durant la durée complète de ma maîtrise.

INTRODUCTION

Il existe plusieurs distributions continues et symétriques en statistique telles les distributions normale, Laplace, Cauchy et logistique. La plus populaire d'entre elles est certainement la loi normale qui est présente dans presque tous les domaines de la statistique. Cependant, si utile soit-elle, elle ne peut servir de modèle statistique pour tous les phénomènes aléatoires observés dans la vie courante. Dans certains cas, des modèles non gaussiens sont plus appropriés. Par exemple, Mandelbrot (1963) mentionne que les prix dans les marchés financiers sont mieux modélisés à l'aide de distributions avec des ailes plus lourdes que la loi normale.

Ceci nous amène à présenter une famille de distributions symétriques représentée par la différence de deux variables aléatoires gamma. Puisque c'est une famille de distributions leptocurtiques, elle est composée de distributions avec des ailes plus lourdes que celle de la normale. L'idée est motivée par la loi Laplace qui peut être représentée par la différence de deux variables aléatoires exponentielles. Puisque la loi exponentielle est un cas particulier de la loi gamma, la distribution Laplace est un cas spécial de la nouvelle famille de distributions. De plus, nous montrerons que la distribution normale en est également un cas particulier.

Le but de ce mémoire est donc de présenter cette nouvelle famille de distributions, d'étaler ses propriétés et de proposer une méthode qui nous permettra d'ajuster un jeu de données à ce nouveau modèle. Le mémoire est divisé en 4 chapitres. Dans les 3 premiers chapitres, nous développons la théorie nécessaire afin d'introduire tout ce qui a trait à la nouvelle famille de distributions. Le chapitre 4 est une copie intégrale de l'article intitulé "A Leptokurtic Symmetric Family of Distributions Represented by the Difference of Two Gamma Variates" soumis à la revue scientifique *Communications in Statistics - Simulation and Computation*. Le premier auteur est également l'auteur du présent mémoire.

Dans le chapitre 1, nous définissons la fonction caractéristique et énonçons certains théorèmes qui y sont rattachés. De plus, nous présentons les lois normale, Laplace et gamma et trouvons leurs fonctions caractéristiques. Nous discutons également de certaines propriétés de ces distributions. Puisque les fonctions de densité et de répartition de la nouvelle famille de distributions ne possèdent pas de forme fermée, nous la définissons en terme de sa fonction caractéristique.

Dans le chapitre 2, nous présentons deux tests de symétrie. Le premier est le test hybride proposé par Moddares et Gastwirth (1998). Ce test est applicable quand la vraie médiane de l'échantillon est connue. Le deuxième est le test des triplets proposé par Randles, Fligner, Policello et Wolfe (1980). Ce test est adéquat quand la vraie médiane est inconnue. Avant d'ajuster un jeu de données à la nouvelle famille de distributions, il est essentiel de vérifier l'hypothèse de symétrie de notre échantillon sans quoi il ne sera pas compatible avec le modèle.

Dans le chapitre 3, nous présentons la classe d'estimateurs basés sur une méthode minimisant une distance quadratique entre des transformées des fonctions de répartition empirique et théorique introduite par Luong et Thompson (1987). De plus, nous construisons des statistiques pour des tests d'ajustement à partir de la distance quadratique et énonçons certaines propriétés de cette classe d'estimateurs. Nous utiliserons cette classe d'estimateurs afin d'estimer les paramètres de la nouvelle famille de distributions à partir d'un échantillon.

Dans le chapitre 4, nous présentons tout ce qui a trait à la nouvelle famille de distributions sous la forme d'un article. Donc, nous l'introduisons, étalons ses propriétés, proposons une méthode afin d'y ajuster un jeu de données, construisons des statistiques pour des tests d'ajustement et menons une étude de simulation afin d'analyser les propriétés des estimateurs obtenus.

Chapitre 1

LA FONCTION CARACTÉRISTIQUE

Dans ce chapitre, nous introduisons la fonction caractéristique et énonçons certains théorèmes qui y sont rattachés. De plus, nous présentons les lois normale, Laplace et gamma et trouvons leurs fonctions caractéristiques. Nous discutons également de certaines propriétés de ces distributions. Finalement, nous définissons la fonction caractéristique empirique et formulons certaines propriétés asymptotiques qui y sont reliées.

1.1. DÉFINITIONS ET THÉORÈMES

Dans cette section, nous introduisons la fonction caractéristique et énumérons quelques-unes de ses propriétés.

Définition 1.1.1. *La fonction caractéristique d'une variable aléatoire X est la fonction $\phi_X(t)$ où*

$$\phi_X(t) = \mathrm{E}[e^{itX}] = \mathrm{E}[\cos(tX) + i\sin(tX)], \quad i = \sqrt{-1} \quad et \quad t \in \mathbb{R}.$$

Si $F_X(x)$ est la fonction de répartition de X, alors la fonction caractéristique est donnée par la transformée de Fourier-Stieltjes :

$$\phi_X(t) = \mathrm{E}[e^{itX}] = \int_{-\infty}^{\infty} e^{itx} dF_X(x).$$

Si la fonction de densité $f_X(x)$ existe, nous obtenons

$$\phi_X(t) = \mathrm{E}[e^{itX}] = \int_{-\infty}^{\infty} e^{itx} f_X(x) dx.$$

Nous allons maintenant énoncer plusieurs théorèmes relatifs à la fonction caractéristique. Le Théorème 1.1.1 implique que la fonction caractéristique existe pour toutes les variables aléatoires tandis que le Théorème 1.1.2 démontre que chaque variable aléatoire possède une fonction caractéristique unique. Les Théorèmes 1.1.3 et 1.1.4 décrivent comment trouver la fonction caractéristique d'une somme de variables aléatoires indépendantes. De plus, le Théorème 1.1.5 illustre la relation existant entre la fonction caractéristique et les moments d'une variable aléatoire. Finalement, la Définition 1.1.2 introduit le concept d'une distribution infiniment divisible.

Théorème 1.1.1. *La fonction caractéristique $\phi_X(t)$ d'une variable aléatoire X satisfait :*

(a) $\phi_X(0) = 1$, $|\phi_X(t)| \leq 1$ *pour tout* $t \in \mathbb{R}$,

(b) $\phi_X(t)$ *est uniformément continue sur* $\mathbb{R}$,

(c) $\phi_X(t)$ *est définie non-négative, ce qui veut dire que* $\sum_{j,k} \phi_X(t_j - t_k) z_j \overline{z}_k \geq 0$ *pour tout réel* $t_1, t_2, \ldots, t_n$ *et pour tout complexe* $z_1, z_2, \ldots, z_n$.

DÉMONSTRATION. (a) Nous avons que $\phi_X(0) = \mathrm{E}(1) = 1$. Aussi, puisque $|e^{itx}| = \sqrt{\cos^2(tx) + \sin^2(tx)} = 1$, nous obtenons

$$|\phi_X(t)| \leq \int_{-\infty}^{\infty} |e^{itx}| dF_X(x) = \int_{-\infty}^{\infty} dF_X(x) = 1.$$

(b) Nous avons que

$$|\phi_X(t+h) - \phi_X(t)| = |\,\mathrm{E}(e^{i(t+h)X} - e^{itX})| \leq \mathrm{E}\,|e^{itX}(e^{ihX} - 1)| \leq \mathrm{E}(Y(h)),$$

où $Y(h) = |e^{ihX} - 1|$. Cependant, $|Y(h)| \leq 2$ et $Y(h) \to 0$ quand $h \to 0$ et donc, $\mathrm{E}(Y(h)) \to 0$ par le théorème de la convergence dominée.

(c) Nous avons que

$$\begin{aligned}
\sum_{j,k} \phi_X(t_j - t_k) z_j \overline{z}_k &= \sum_{j,k} \int_{-\infty}^{\infty} [z_j \exp(it_j x)][\overline{z}_k \exp(-it_k x)] dF_X(x) \\
&= \int_{-\infty}^{\infty} \sum_{j,k} [z_j \exp(it_j x)][\overline{z_k \exp(it_k x)}] dF_X(x) \\
&= \int_{-\infty}^{\infty} \left[\sum_j z_j \exp(it_j x)\right] \left[\sum_k \overline{z_k \exp(it_k x)}\right] dF_X(x) \\
&= \int_{-\infty}^{\infty} \left[\sum_j z_j \exp(it_j x)\right] \left[\overline{\sum_j z_j \exp(it_j x)}\right] dF_X(x) \\
&= \mathrm{E}\left(\left|\sum_j z_j \exp(it_j X)\right|^2\right) \geq 0.
\end{aligned}$$

□

Théorème 1.1.2. *Définissons ϕ_X et ϕ_Y comme les fonctions caractéristiques des variables aléatoires X et Y respectivement. Si $\phi_X = \phi_Y$, alors $X \stackrel{\mathcal{D}}{=} Y$ et vice-versa. $X \stackrel{\mathcal{D}}{=} Y$ veut dire égalité en distribution et donc, $P(X \leq x) = P(Y \leq x)$.*

DÉMONSTRATION. Voir Lukacs (1970). □

Théorème 1.1.3. *Si $a, b, c \in \mathbb{R}$ et si X et Y sont des variables aléatoires indépendantes, alors*

$$\phi_{aX+bY+c}(t) = e^{itc} \phi_X(at) \phi_Y(bt), \quad t \in \mathbb{R}.$$

DÉMONSTRATION. Nous avons que

$$\begin{aligned}
\phi_{aX+bY+c}(t) &= \mathrm{E}(e^{it(aX+bY+c)}) = \mathrm{E}(e^{itc} e^{i(at)X} e^{i(bt)Y}) \\
&= e^{itc}\, \mathrm{E}(e^{i(at)X})\, \mathrm{E}(e^{i(bt)Y}) = e^{itc} \phi_X(at) \phi_Y(bt).
\end{aligned}$$

□

Théorème 1.1.4. *Définissons $X_1, X_2, \ldots, X_n$ comme des variables aléatoires indépendantes et identiquement distribuées avec les fonctions caractéristiques correspondantes $\phi_{X_1}, \phi_{X_2}, \ldots, \phi_{X_n}$ et $S_n = X_1 + X_2 + \cdots + X_n$. La fonction caractéristique de S_n est*

$$\phi_{S_n}(t) = [\phi_{X_1}(t)]^n, \quad t \in \mathbb{R}.$$

DÉMONSTRATION. En utilisant le Théorème 1.1.3, nous obtenons

$$\phi_{S_n}(t) = \phi_{X_1}(t)\phi_{X_2}(t)\cdots\phi_{X_n}(t), \quad t \in \mathbb{R}.$$

Puisque $X_1, X_2, \ldots, X_n$ sont des variables aléatoires identiquement distribuées, le Théorème 1.1.2 implique que

$$\phi_{X_1}(t) = \phi_{X_2}(t) = \ldots = \phi_{X_n}(t).$$

Donc, nous obtenons

$$\phi_{S_n}(t) = \phi_{X_1}(t)\phi_{X_1}(t)\cdots\phi_{X_1}(t) = [\phi_{X_1}(t)]^n, \quad t \in \mathbb{R}.$$

□

Définition 1.1.2. *La distribution d'une variable aléatoire X est dite infiniment divisible si, pour toutes les valeurs de $n = 1, 2, 3, \ldots$, la fonction caractéristique $\phi_X(t)$ admet la représentation*

$$\phi_X(t) = [\phi(t)]^n,$$

où $\phi(t)$ est la fonction caractéristique d'une certaine variable aléatoire.

Théorème 1.1.5. *(a) Si $\phi_X^{(k)}(0) = \frac{d^k}{dt^k}\phi_X(t)\Big|_{t=0}$ existe, alors*

$$\begin{cases} \mathrm{E}\,|X^k| < \infty & \text{pour } k \text{ pair,} \\ \mathrm{E}\,|X^{k-1}| < \infty & \text{pour } k \text{ impair.} \end{cases}$$

(b) Si $\mathrm{E}\,|X^k| < \infty$, alors $\phi_X^{(k)}(0) = i^k\,\mathrm{E}(X^k)$.

DÉMONSTRATION. Voir Lukacs (1970). □

La fonction caractéristique peut parfois être difficile à calculer puisqu'il faut évaluer une intégrale sur un plan complexe. Il est souvent plus facile de trouver la fonction caractéristique à partir de la fonction génératrice des moments. Nous introduisons maintenant la fonction génératrice des moments.

Définition 1.1.3. *La fonction génératrice des moments d'une variable aléatoire X est la fonction $M_X(t)$, où*

$$M_X(t) = \mathrm{E}[e^{tX}],$$

pourvu que $M_X(t)$ existe pour $|t| < h$ et $h > 0$.

Il est à noter que l'existence de la fonction génératrice des moments n'est pas garantie pour toutes les variables aléatoires. Par exemple, elle n'existe pas pour la loi de Cauchy, car l'intégrale

$$\mathrm{E}(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f_X(x)dx = \int_{-\infty}^{\infty} \frac{e^{tx}}{\pi(1+x^2)}dx$$

est divergente ($f_X(x) = [\pi(1+x^2)]^{-1}$ est la fonction de densité de la loi de Cauchy). Cependant, si on adopte un t purement imaginaire, alors $M_X(t)$ existe toujours, car $M_X(t)$ devient une fonction caractéristique. Par conséquent, nous avons la relation $\phi_X(t) = M_X(it)$ à condition que $M_X(t)$ existe. Dans la section suivante, nous présentons les distributions normale, Laplace et gamma.

1.2. Exemples

Pour chacune des distributions, nous allons définir la fonction de densité, calculer la fonction caractéristique et trouver la moyenne (μ), la variance (σ^2), le coefficient d'asymétrie (β_1) et le coefficient d'aplatissement (β_2). Une distribution a une asymétrie négative si elle présente une queue vers la gauche, c'est-à-dire qu'il y a plus d'observations vers les valeurs négatives. Nous dirons qu'elle a une asymétrie positive si elle présente une queue vers la droite, c'est-à-dire qu'il y a plus d'observations vers les valeurs positives. L'aplatissement d'une distribution réfère à la concentration relative des observations autour du mode. Par exemple, nous verrons que le coefficient d'aplatissement de la loi normale est 3. Une distribution avec un aplatissement inférieur à 3 sera plus aplatie que la distribution normale. Nous dirons donc que c'est une distribution platycurtique. Dans le cas où l'aplatissement est supérieur à 3, nous dirons que c'est une distribution leptocurtique.

1.2.1. La loi normale

Définissons X comme une variable aléatoire normale avec moyenne μ et variance σ^2, donc $X \sim N(\mu, \sigma^2)$. La fonction de densité de X est

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty.$$

Aussi,

$$\begin{aligned} M_X(t) = E(e^{tX}) &= \int_{-\infty}^{\infty} \exp(tx) \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right) dx \\ &= \int_{-\infty}^{\infty} \exp(t(y+\mu)) \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-y^2}{2\sigma^2}\right) dy \\ &= \exp(\mu t) \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp(ty - y^2/2\sigma^2) dy. \end{aligned}$$

Puisque

$$ty - y^2/2\sigma^2 = -(y - \sigma^2 t)^2/2\sigma^2 + \sigma^2 t^2/2,$$

nous obtenons

$$M_X(t) = \exp(\mu t) \exp(\sigma^2 t^2/2) \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(y-\sigma^2 t)^2}{2\sigma^2}\right) dy.$$

L'intégrale précédente représente l'intégrale de la fonction de densité d'une $N(\sigma^2 t, \sigma^2)$ et est donc égale à 1. Nous obtenons

$$M_X(t) = \exp(\mu t) \exp(\sigma^2 t^2/2) = e^{\mu t + \sigma^2 t^2/2}, \quad -\infty < t < \infty.$$

Nous pouvons donc trouver la fonction caractéristique à l'aide de la relation,

$$\phi_X(t) = M_X(it) = e^{i\mu t - \sigma^2 t^2/2}, \quad -\infty < t < \infty.$$

Nous allons maintenant développer une formule pour les moments centrés d'une $N(\mu, \sigma^2)$.

Proposition 1.2.1. *Si $X \sim N(\mu, \sigma^2)$, alors les moments centrés impairs sont* 0 *et les moments centrés pairs sont obtenus à partir de la formule suivante :*

$$\mathrm{E}[(X-\mu)^{2k}] = \frac{(2k)!}{2^k k!}\sigma^{2k}, \quad k = 1, 2, 3, \ldots.$$

DÉMONSTRATION. Premièrement, nous trouvons la fonction caractéristique de $(X - \mu)$. Avec le Théorème 1.1.3, nous obtenons

$$\phi_{X-\mu}(t) = e^{-i\mu t}e^{i\mu t - \sigma^2 t^2/2} = e^{-\sigma^2 t^2/2}.$$

Par l'expansion en série de Taylor de la fonction exponentielle, nous avons

$$\begin{aligned}\phi_{X-\mu}(t) = e^{-\sigma^2t^2/2} &= \sum_{k=0}^{\infty} \frac{\left(-\sigma^2 t^2/2\right)^k}{k!} = \sum_{k=0}^{\infty} \frac{(-1)^k \sigma^{2k} t^{2k}}{2^k k!} \\ &= 1 - \frac{\sigma^2}{2}t^2 + \frac{\sigma^4}{8}t^4 - \frac{\sigma^6}{48}t^6 + \cdots.\end{aligned}$$

De plus, nous avons

$$\phi^{(j)}_{X-\mu}(0) = \begin{cases} \frac{(-1)^{j/2} j!}{2^{j/2}(j/2)!}\sigma^j & \text{pour } j \text{ pair,} \\ \\ 0 & \text{pour } j \text{ impair.} \end{cases}$$

En appliquant le Théorème 1.1.5 et en remplaçant j par $2k$, la démonstration est complète. □

À l'aide de la Proposition 1.2.1, nous pouvons déterminer la moyenne, la variance, l'asymétrie et l'aplatissement de $X \sim N(\mu, \sigma^2)$.

$$\begin{aligned}\mathrm{E}(X-\mu) = 0 \quad &\Rightarrow \quad \mathrm{E}(X) = \mu, \\ \mathrm{Var}(X) &= \mathrm{E}[(X-\mu)^2] = \sigma^2, \\ \beta_1 &= \frac{\mathrm{E}[(X-\mu)^3]}{[\mathrm{Var}(X)]^{3/2}} = 0, \\ \beta_2 &= \frac{\mathrm{E}[(X-\mu)^4]}{[\mathrm{Var}(X)]^2} = 3.\end{aligned}$$

Puisque $\beta_1 = 0$, la loi normale est une distribution symétrique. Par conséquent, la moyenne, la médiane et le mode de la distribution sont tous égaux à μ.

1.2.2. La loi Laplace

Définissons X comme une variable aléatoire Laplace classique avec paramètre de position μ et paramètre d'échelle $s > 0$, dénotée $X \sim \mathcal{CL}(\mu, s)$. La fonction de densité de X est

$$f_X(x) = \frac{1}{2s} e^{-\frac{|x-\mu|}{s}}, \quad -\infty < x < \infty.$$

De plus, la fonction de répartition de X est

$$F_X(x) = \begin{cases} \frac{1}{2} e^{-\frac{|x-\mu|}{s}} & \text{si } x \leq \mu, \\ 1 - \frac{1}{2} e^{-\frac{|x-\mu|}{s}} & \text{si } x > \mu. \end{cases}$$

Si $Y \sim \mathcal{CL}(0, 1)$, alors

$$\begin{aligned} M_Y(t) = E(e^{tY}) &= \int_{-\infty}^{\infty} e^{ty} \frac{1}{2} e^{-|y|} dy \\ &= \frac{1}{2} \int_{0}^{\infty} \left[e^{-(1-t)y} + e^{-(t+1)y} \right] dy \qquad (1.2.2.a) \\ &= \frac{1}{2} \left[\frac{1}{1-t} + \frac{1}{t+1} \right] = \frac{1}{1-t^2}. \end{aligned}$$

$M_Y(t)$ existe seulement pour $|t| < 1$, puisque l'intégrale (1.2.2.a) diverge pour $|t| \geq 1$. La fonction caractéristique de Y est

$$\phi_Y(t) = M_Y(it) = \frac{1}{1+t^2}, \quad -\infty < t < \infty.$$

Puisque $X = sY + \mu$, à l'aide du Théorème 1.1.3, nous obtenons

$$\phi_X(t) = e^{it\mu} \phi_Y(st) = \frac{e^{it\mu}}{1+s^2t^2}, \quad -\infty < t < \infty.$$

Proposition 1.2.2. *Si $X \sim \mathcal{CL}(\mu, s)$, alors les moments centrés impairs sont* 0 *et les moments centrés pairs sont obtenus à partir de la formule suivante :*

$$\mathrm{E}[(X-\mu)^{2k}] = (2k)! s^{2k}, \quad k = 1, 2, 3, \ldots.$$

DÉMONSTRATION. La formule précédente est un cas spécial de l'expression générale dérivée dans la Proposition 1 de l'article du chapitre 4. Il suffit d'appliquer le Théorème 1.1.3 pour trouver la fonction caractéristique de $(X - \mu)$,

$$\phi_{X-\mu}(t) = e^{-i\mu t}\frac{e^{it\mu}}{1+s^2t^2} = \frac{1}{1+s^2t^2}.$$

En posant $\lambda = 1$ et $\theta = s^2$ dans la Proposition 1 de l'article du chapitre 4, nous parvenons à la formule ci-dessus. $\square$

À l'aide de la Proposition 1.2.2, nous pouvons déterminer la moyenne, la variance, l'asymétrie et l'aplatissement de $X \sim \mathcal{CL}(\mu, s)$.

$$\begin{aligned}
\mathrm{E}(X-\mu) = 0 \quad &\Rightarrow \quad \mathrm{E}(X) = \mu, \\
\sigma^2 = \mathrm{Var}(X) &= \mathrm{E}[(X-\mu)^2] = 2s^2, \\
\beta_1 &= \frac{\mathrm{E}[(X-\mu)^3]}{[\mathrm{Var}(X)]^{3/2}} = 0, \\
\beta_2 &= \frac{\mathrm{E}[(X-\mu)^4]}{[\mathrm{Var}(X)]^2} = \frac{s^4 4!}{(2s^2)^2} = 6.
\end{aligned}$$

Puisque $\beta_1 = 0$, la loi Laplace est une distribution symétrique tout comme la loi normale. De plus, elle est moins aplatie que la normale puisque $\beta_2 = 6 > 3$. C'est donc une distribution leptocurtique.

1.2.3. La loi gamma

Définissons X comme une variable aléatoire gamma avec paramètre de forme $\alpha > 0$ et paramètre d'échelle $s > 0$, donc $X \sim \Gamma(\alpha, s)$. La fonction de densité de X est

$$f_X(x) = x^{\alpha-1}\frac{e^{-x/s}}{s^\alpha \Gamma(\alpha)}, \quad x > 0,$$

où

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx.$$

Si $Y \sim \Gamma(\alpha, 1)$, alors

$$
\begin{aligned}
M_Y(t) = E(e^{tY}) &= \int_0^\infty e^{ty} y^{\alpha-1} \frac{e^{-y}}{\Gamma(\alpha)} dy \\
&= \frac{1}{\Gamma(\alpha)} \int_0^\infty y^{\alpha-1} e^{-(1-t)y} dy \qquad (1.2.2.b) \\
&= \frac{1}{\Gamma(\alpha)(1-t)^\alpha} \int_0^\infty z^{\alpha-1} e^{-z} dz = \frac{1}{(1-t)^\alpha}.
\end{aligned}
$$

$M_Y(t)$ existe seulement pour $t < 1$, puisque l'intégrale (1.2.2.b) diverge pour $t \geq 1$. La fonction caractéristique de Y est

$$
\phi_Y(t) = M_Y(it) = \frac{1}{(1-it)^\alpha}, \quad -\infty < t < \infty.
$$

Puisque $X = sY$, à l'aide du Théorème 1.1.3, nous obtenons

$$
\phi_X(t) = \phi_Y(st) = \frac{1}{(1-ist)^\alpha}, \quad -\infty < t < \infty.
$$

Proposition 1.2.3. *Si $X \sim \Gamma(\alpha, s)$, alors les moments sont obtenus à partir de la formule suivante :*

$$
\mathrm{E}[X^k] = s^k \frac{\Gamma(\alpha+k)}{\Gamma(\alpha)} = s^k (\alpha+k-1)\cdots\alpha, \quad k = 1, 2, 3, \ldots.
$$

DÉMONSTRATION. Premièrement, si $Y \sim \Gamma(\alpha, 1)$, nous trouvons

$$
\mathrm{E}[Y^k] = \int_0^\infty y^{\alpha+k-1} \frac{e^{-y}}{\Gamma(\alpha)} dx = \frac{\Gamma(\alpha+k)}{\Gamma(\alpha)}.
$$

Puisque $X = sY$, nous obtenons

$$
\mathrm{E}[X^k] = \mathrm{E}[(sY)^k] = s^k\,\mathrm{E}[Y^k] = s^k \frac{\Gamma(\alpha+k)}{\Gamma(\alpha)}.
$$

Maintenant, en intégrant par parties, nous obtenons

$$\begin{aligned}\Gamma(\alpha) &= \int_0^\infty x^{\alpha-1}e^{-x}dx \\ &= \left[-x^{\alpha-1}e^{-x}\right]\Big|_0^\infty + \int_0^\infty (\alpha-1)x^{\alpha-2}e^{-x}dx \\ &= (\alpha-1)\Gamma(\alpha-1).\end{aligned}$$

Ceci nous amène à l'expression suivante :

$$\begin{aligned}\mathrm{E}[X^k] = s^k\frac{\Gamma(\alpha+k)}{\Gamma(\alpha)} &= s^k\frac{(\alpha+k-1)(\alpha+k-2)\cdots(\alpha+1)\alpha\Gamma(\alpha)}{\Gamma(\alpha)} \\ &= s^k(\alpha+k-1)(\alpha+k-2)\cdots\alpha, \quad k=1,2,3,\ldots.\end{aligned}$$

□

À l'aide de la Proposition 1.2.3, nous pouvons déterminer la moyenne, la variance, l'asymétrie et l'aplatissement de $X \sim \Gamma(\alpha, s)$.

$$\begin{aligned}\mu &= \mathrm{E}[X] = s\alpha, \\ \sigma^2 &= \mathrm{Var}(X) = \mathrm{E}[(X-\mu)^2] = \mathrm{E}[X^2] - \mu^2 = s^2(\alpha+1)\alpha - (s\alpha)^2 = s^2\alpha, \\ \beta_1 &= \frac{\mathrm{E}[(X-\mu)^3]}{[\mathrm{Var}(X)]^{3/2}} = \frac{\sum_{k=0}^3\binom{3}{k}(-1)^{k+1}\mu^{3-k}\,\mathrm{E}[X^k]}{\sigma^3} = \frac{2}{\sqrt{\alpha}} > 0, \\ \beta_2 &= \frac{\mathrm{E}[(X-\mu)^4]}{[\mathrm{Var}(X)]^2} = \frac{\sum_{k=0}^4\binom{4}{k}(-1)^k\mu^{4-k}\,\mathrm{E}[X^k]}{\sigma^4} = 3 + \frac{6}{\alpha} > 3.\end{aligned}$$

Puisque $\beta_1 > 0$, la loi gamma a une asymétrie positive et possède donc une queue plus prononcée vers la droite. De plus, elle est moins aplatie que la normale puisque $\beta_2 > 3$. C'est donc une distribution leptocurtique.

Deux lois connues sont des cas particuliers de la distribution gamma, soit les lois khi-deux et exponentielle. Si $X \sim \Gamma(\nu/2, 2)$, alors X est également une variable aléatoire khi-deux avec ν degrés de liberté, dénotée par χ^2_ν. De plus, si $X \sim \Gamma(1, s)$, alors X est identique à une variable aléatoire exponentielle avec paramètre d'échelle s, dénotée par $Exp(s)$. Prenons $Z = X_1 - X_2$, où X_1 et X_2 sont des variables aléatoires indépendantes $Exp(s)$. En appliquant le Théorème

1.1.3, la fonction caractéristique de Z est

$$\phi_Z(t) = \phi_{X_1}(t)\phi_{X_2}(-t) = \left(\frac{1}{1-ist}\right)\left(\frac{1}{1+ist}\right) = \frac{1}{1+s^2t^2}.$$

Notons que $\phi_Z(t)$ est également la fonction caractéristique d'une variable aléatoire $\mathcal{CL}(0,s)$. Ceci implique que la distribution Laplace est simplement la différence de deux distributions exponentielles. D'ailleurs, la loi Laplace est parfois appelée la loi doublement exponentielle.

1.3. La fonction caractéristique empirique

Dans cette section, nous introduisons la fonction caractéristique empirique et formulons quelques résultats qui y sont rattachés.

Définition 1.3.1. *Si $X_1, X_2, \ldots, X_n$ est un échantillon indépendant provenant d'une distribution avec la fonction de répartition $F_X(x)$ et la fonction caractéristique $\phi_X(t) = \int e^{itx} dF_X(x)$, alors la fonction caractéristique empirique est*

$$\phi_n(t) = \int_{-\infty}^{\infty} e^{itx} dF_n(x) = \frac{1}{n}\sum_{j=1}^{n} e^{itX_j} = \frac{1}{n}\sum_{j=1}^{n}[\cos(tX_j) + i\sin(tX_j)],$$

où $F_n(x)$ est la fonction de répartition empirique, $F_n(x) = \frac{N(x)}{n}$, où $N(x)$ est le nombre de $X_j \leq x$ pour $1 \leq j \leq n$.

Théorème 1.3.1. *Si $X_1, X_2, \ldots, X_n$ est un échantillon indépendant provenant d'une distribution avec la fonction caractéristique $\phi_X(t)$, alors*

(a) $\mathrm{E}[\phi_n(t)] = \phi_X(t)$,

(b) $\phi_n(t) \to \phi_X(t)$ *presque sûrement si $n \to \infty$ pour chaque t,*

(c) $\mathrm{Cov}(\phi_n(t), \phi_n(s)) = \frac{1}{n}[\phi_X(t+s) - \phi_X(t)\phi_X(s)]$,

(d) $\mathrm{E}[\mathrm{Re}\,\phi_n(t), \mathrm{Im}\,\phi_n(t)] = [\mathrm{Re}\,\phi_X(t), \mathrm{Im}\,\phi_X(t)]$,

(e) $\mathrm{Cov}(\mathrm{Re}\,\phi_n(t), \mathrm{Re}\,\phi_n(s)) = \frac{1}{2n}[\mathrm{Re}\,\phi_X(t+s) + \mathrm{Re}\,\phi_X(t-s) - 2\,\mathrm{Re}\,\phi_X(t)\,\mathrm{Re}\,\phi_X(s)]$,

(f) $\mathrm{Cov}(\mathrm{Im}\,\phi_n(t), \mathrm{Im}\,\phi_n(s)) = \frac{1}{2n}[\mathrm{Re}\,\phi_X(t-s) - \mathrm{Re}\,\phi_X(t+s) - 2\,\mathrm{Im}\,\phi_X(t)\,\mathrm{Im}\,\phi_X(s)]$,

(g) $\mathrm{Cov}(\mathrm{Re}\,\phi_n(t), \mathrm{Im}\,\phi_n(s)) = \frac{1}{2n}[\mathrm{Im}\,\phi_X(t+s) - \mathrm{Im}\,\phi_X(t-s) - 2\,\mathrm{Re}\,\phi_X(t)\mathrm{Im}\phi_X(s)]$.

Note : $\mathrm{Re}[\cdot]$ *et* $\mathrm{Im}[\cdot]$ *représentent les parties réelle et imaginaire d'un nombre complexe.*

Démonstration. (a) Nous avons que

$$\mathrm{E}[\phi_n(t)] = \mathrm{E}\left[\frac{1}{n}\sum_{j=1}^{n} e^{itX_j}\right] = \frac{1}{n}\sum_{j=1}^{n}\mathrm{E}[e^{itX_j}] = \frac{1}{n}\sum_{j=1}^{n}\phi_X(t) = \phi_X(t),$$

puisque $X_1, X_2, \ldots, X_n$ possèdent une fonction caractéristique commune $\phi_X(t)$.

(b) Puisque $X_1, X_2, \ldots, X_n$ sont indépendants et identiquement distribués, la loi forte des grands nombres implique que

$$\phi_n(t) \to \mathrm{E}[e^{itX_1}] = \phi_X(t) \text{ presque sûrement si } n \to \infty \text{ pour chaque } t.$$

(c) Nous avons que

$$\begin{aligned}
\mathrm{Cov}(\phi_n(t), \phi_n(s)) &= \mathrm{E}[\phi_n(t)\phi_n(s)] - \mathrm{E}[\phi_n(t)]\,\mathrm{E}[\phi_n(s)] \\
&= \mathrm{E}\left[\left(\frac{1}{n}\sum_{j=1}^{n} e^{itX_j}\right)\left(\frac{1}{n}\sum_{j=1}^{n} e^{isX_j}\right)\right] - \phi_X(t)\phi_X(s) \\
&= \frac{1}{n^2}\left[\sum_{j=1}^{n}\mathrm{E}[e^{i(t+s)X_j}] + n(n-1)\phi_X(t)\phi_X(s)\right] - \phi_X(t)\phi_X(s) \\
&= \frac{1}{n}[\phi_X(t+s) - \phi_X(t)\phi_X(s)]
\end{aligned}$$

(d) Nous avons que $\phi_n(t) = \mathrm{Re}\,\phi_n(t) + i\,\mathrm{Im}\,\phi_n(t)$. En (a), nous avons déterminé que

$\mathrm{E}[\phi_n(t)] = \phi_X(t)$ et donc,

$$\mathrm{E}[\mathrm{Re}\,\phi_n(t) + i\,\mathrm{Im}\,\phi_n(t)] = \mathrm{Re}\,\phi_X(t) + i\,\mathrm{Im}\,\phi_X(t).$$

L'expression précédente implique que

$$\mathrm{E}[\mathrm{Re}\,\phi_n(t)] + i\,\mathrm{E}[\mathrm{Im}\,\phi_n(t)] = \mathrm{Re}\,\phi_X(t) + i\,\mathrm{Im}\,\phi_X(t)$$

d'où nous obtenons le résultat voulu.

(e) $\mathrm{E}[\operatorname{Re}\phi_n(t)\operatorname{Re}\phi_n(s)]$

$$= \mathrm{E}\left[\left(\frac{1}{n}\sum_{j=1}^{n}\cos(tX_j)\right)\left(\frac{1}{n}\sum_{j=1}^{n}\cos(sX_j)\right)\right]$$

$$= \frac{1}{n^2}\left[\sum_{j=1}^{n}\mathrm{E}[\cos(tX_j)\cos(sX_j)] + n(n-1)\operatorname{Re}\phi_X(t)\operatorname{Re}\phi_X(s)\right]$$

$$= \frac{1}{n^2}\left[\sum_{j=1}^{n}\mathrm{E}[\frac{1}{2}\{\cos((t+s)X_j) + \cos((t-s)X_j)\}] + n(n-1)\operatorname{Re}\phi_X(t)\operatorname{Re}\phi_X(s)\right]$$

$$= \frac{1}{n^2}\left[\sum_{j=1}^{n}\frac{1}{2}[\operatorname{Re}\phi_X(t+s) + \operatorname{Re}\phi_X(t-s)] + n(n-1)\operatorname{Re}\phi_X(t)\operatorname{Re}\phi_X(s)\right]$$

$$= \frac{1}{2n}[\operatorname{Re}\phi_X(t+s) + \operatorname{Re}\phi_X(t-s)] + (1-\frac{1}{n})\operatorname{Re}\phi_X(t)\operatorname{Re}\phi_X(s)$$

Le résultat est obtenu directement en posant

$$\begin{aligned}\operatorname{Cov}(\operatorname{Re}\phi_n(t),\operatorname{Re}\phi_n(s)) &= \mathrm{E}[\operatorname{Re}\phi_n(t)\operatorname{Re}\phi_n(s)] - \mathrm{E}[\operatorname{Re}\phi_n(t)]\,\mathrm{E}[\operatorname{Re}\phi_n(s)]\\ &= \mathrm{E}[\operatorname{Re}\phi_n(t)\operatorname{Re}\phi_n(s)] - \operatorname{Re}\phi_X(t)\operatorname{Re}\phi_X(s)\end{aligned}$$

(f) Les étapes de la démonstration sont similaires à celles de (e). Il faut simplement faire usage d'une identité trigonométrique différente et noter que

$$\begin{aligned}\mathrm{E}[\sin(tX_j)\sin(sX_j)] &= \mathrm{E}[\frac{1}{2}\{\cos((t-s)X_j) - \cos((t+s)X_j)\}]\\ &= \frac{1}{2}[\operatorname{Re}\phi_X(t-s) - \operatorname{Re}\phi_X(t+s)].\end{aligned}$$

(g) Similairement à (e) et (f), nous avons que

$$\begin{aligned}\mathrm{E}[\cos(tX_j)\sin(sX_j)] &= \mathrm{E}[\frac{1}{2}\{\sin((t+s)X_j) - \sin((t-s)X_j)\}]\\ &= \frac{1}{2}[\operatorname{Im}\phi_X(t+s) - \operatorname{Im}\phi_X(t-s)].\end{aligned}$$

□

Le théorème précédent nous amène à formuler un résultat important apparaissant dans Feuerverger et Mureika (1977).

Théorème 1.3.2. *Considérons le processus stochastique complexe*

$$Y_n(t) = \sqrt{n}[\phi_n(t) - \phi_X(t)].$$

Nous avons que $\mathrm{E}[Y_n(t)] = 0$ *et* $\mathrm{Cov}(Y_n(t), Y_n(s)) = \phi_X(t+s) - \phi_X(t)\phi_X(s)$. *La structure de la covariance des parties réelle et imaginaire de* $Y_n(t)$ *est :*

$$\mathrm{Cov}(\mathrm{Re}\,Y_n(t), \mathrm{Re}\,Y_n(s)) = \frac{1}{2}[\mathrm{Re}\,\phi_X(t+s) + \mathrm{Re}\,\phi_X(t-s) - 2\,\mathrm{Re}\,\phi_X(t)\,\mathrm{Re}\,\phi_X(s)],$$

$$\mathrm{Cov}(\mathrm{Im}\,Y_n(t), \mathrm{Im}\,Y_n(s)) = \frac{1}{2}[\mathrm{Re}\,\phi_X(t-s) - \mathrm{Re}\,\phi_X(t+s) - 2\,\mathrm{Im}\,\phi_X(t)\,\mathrm{Im}\,\phi_X(s)],$$

$$\mathrm{Cov}(\mathrm{Re}\,Y_n(t), \mathrm{Im}\,Y_n(s)) = \frac{1}{2}[\mathrm{Im}\,\phi_X(t+s) - \mathrm{Im}\,\phi_X(t-s) - 2\,\mathrm{Re}\,\phi_X(t)\mathrm{Im}\phi_X(s)].$$

Puisque $\phi_n(t)$ *est une moyenne de processus bornés, le théorème limite centrale multidimensionel implique, pour un nombre fini de points* $t_1, t_2, \ldots, t_k$, *la convergence en distribution de* $Y_n(t_1), Y_n(t_2), \ldots, Y_n(t_k)$ *vers* $Y(t_1), Y(t_2), \ldots, Y(t_k)$, *où* $Y(t)$ *est un processus gaussien complexe satisfaisant* $Y(t) = \overline{Y(-t)}$ *avec moyenne* 0 *et structure de covariance identique à* $Y_n(t)$.

Démonstration. Avec le Théorème 1.3.1(a), $\mathrm{E}[\phi_n(t)] = \phi_X(t)$ et donc,

$$\mathrm{E}[Y_n(t)] = \sqrt{n}(\mathrm{E}[\phi_n(t)] - \phi_X(t)) = 0.$$

De plus, en utilisant le Théorème 1.3.1(c) nous avons

$$\begin{aligned}\mathrm{Cov}(Y_n(t), Y_n(s)) &= \mathrm{Cov}(\sqrt{n}[\phi_n(t) - \phi_X(t)], \sqrt{n}[\phi_n(s) - \phi_X(t)]) \\ &= n\,\mathrm{Cov}(\phi_n(t), \phi_n(s)) = \phi_X(t+s) - \phi_X(t)\phi_X(s).\end{aligned}$$

Similairement, les expressions pour la structure de la covariance des parties réelle et imaginaire de $Y_n(t)$ sont obtenues à partir des résultats (e), (f) et (g) du Théorème 1.3.1. Voir Feuerverger et Mureika (1977) pour une démonstration complète de la convergence faible de $Y_n(t)$ vers $Y(t)$. □

Définissons les vecteurs suivants pour un nombre fini de points $t_1, t_2, \ldots, t_k$:

$$\mathbf{Z}_n = [\mathrm{Re}\,\phi_n(t_1), \ldots, \mathrm{Re}\,\phi_n(t_k), \mathrm{Im}\,\phi_n(t_1), \ldots, \mathrm{Im}\,\phi_n(t_k)]' \tag{1.3.1}$$

$$\mathbf{Z}(\cdot) = [\mathrm{Re}\,\phi_X(t_1), \ldots, \mathrm{Re}\,\phi_X(t_k), \mathrm{Im}\,\phi_X(t_1), \ldots, \mathrm{Im}\,\phi_X(t_k)]'. \tag{1.3.2}$$

Le Théorème 1.3.2 implique que $\sqrt{n}[\mathbf{Z}_n - \mathbf{Z}(\cdot)]$ converge en distribution vers une loi normale multivariée centrée à 0 avec la structure de covariance spécifiée dans le Théorème 1.3.2.

Chapitre 2

TESTS DE SYMÉTRIE

Dans ce chapitre, nous présentons deux tests de symétrie. Le premier est le test hybride proposé par Moddares et Gastwirth (1998). Ce test est applicable quand la vraie médiane de l'échantillon est connue. Le deuxième est le test des triplets proposé par Randles, Fligner, Policello et Wolfe (1980). Ce test est adéquat quand la vraie médiane est inconnue.

2.1. MÉDIANE CONNUE

Définissons $x_1, \ldots, x_n$ comme n observations indépendantes d'une variable aléatoire continue X avec la fonction de répartition F, la fonction de densité f et le centre de symétrie connu μ_0. Nous considérons le test d'hypothèses

$$H_0 : F(\mu_0 - x) = 1 - F(\mu_0 + x)$$

contre

$$H_a : F(\mu_0 - x) \neq 1 - F(\mu_0 + x).$$

Nous sommes donc intéressés à vérifier si la fonction de densité f est symétrique par rapport à la médiane connue μ_0 ou non.

Plusieurs tests de symétrie sont présentés dans la littérature (voir Hollander et Wolfe, 1973; Lehman, 1975; Randles et Wolfe, 1979). McWilliams (1990) et Moddares et Gastwirth (1996) ont utilisé des tests basés sur des statistiques de séquences. Tajuddin (1994) et Thas, Rayner et Best (2005) ont exploité la statistique des rangs signés de Wilcoxon. Aussi, Cheng et Balakrishnan (2004) proposent un test du signe modifié.

Nous suggérons d'utiliser le test hybride présenté par Moddares et Gastwirth (1998) pour tester la symétrie autour d'une médiane connue. Nous favorisons ce test pour sa puissance élevée et sa simplicité. D'ailleurs, Thas, Rayner et Best (2005) ont accompli plusieurs simulations comparant la puissance de différents tests de symétrie ; elles ont révélé que le test hybride est plus puissant que la plupart des alternatives considérées.

2.1.1. Le test hybride

Puisque nous testons la symétrie autour de la médiane connue μ_0, nous pouvons supposer sans perte de généralité que $\mu_0 = 0$. Le test hybride comporte deux étapes. Dans l'étape I, le test du signe est utilisé au niveau $\alpha_1 < \alpha$. Si H_0 est acceptée dans l'étape I, alors le test de Wilcoxon tronqué pour 2 échantillons est appliqué dans l'étape II au niveau $\alpha_2 < \alpha$. Le test hybride est un test de niveau α, où $\alpha = \alpha_1 + (1 - \alpha_1)\alpha_2$. Les valeurs suggérées pour les niveaux des tests sont $\alpha_1 = 0.01$ et $\alpha_2 = 0.0404$ générant un niveau total de $\alpha = 0.05$.

Dans la première étape, le test du signe compte le nombre d'observations plus grandes que 0 et rejette H_0 si ce compte est trop loin de la valeur espérée sous l'hypothèse de symétrie. Donc, $S = \sum_{k=1}^{n} I_k$, où

$$I_k = \begin{cases} 1 & \text{si } x_k > 0, \\ 0 & \text{sinon.} \end{cases}$$

Sous H_0, la distribution exacte de S est binomiale avec paramètres n et $1/2$. La p-valeur du test est $2 \cdot P_{H_0}(S - n/2 \geq |s - n/2|)$, où s est la valeur observée de S. Nous rejetons donc H_0 si la p-valeur $\leq \alpha_1$. De plus, nous pouvons également utiliser la distribution asymptotique de S afin d'approximer la p-valeur. S est asymptotiquement normal avec $\mathrm{E}(S) = n/2$ et $\mathrm{Var}(S) = n/4$.

Dans la deuxième étape, nous appliquons le test de Wilcoxon tronqué pour 2 échantillons. Premièrement, nous devons séparer les l observations positives et les m observations négatives dans 2 échantillons différents. Notons que $n = l+m$. Les échantillons avec les observations positives et négatives sont dénommées, échantillon 1 et échantillon 2, respectivement. Ensuite, nous prenons la valeur absolue des observations de l'échantillon 2. Nous possédons ainsi deux échantillons d'observations positives. Nous combinons maintenant les deux échantillons et plaçons les observations en ordre croissant. Les statistiques $V_1, \ldots, V_n$ sont associées à l'ordre respectif des observations.

Plus particulièrement, définissons

$$V_k = \begin{cases} 1 & \text{si la } k^{\text{ième}} \text{ observation provient de l'échantillon 1,} \\ 0 & \text{si la } k^{\text{ième}} \text{ observation provient de l'échantillon 2.} \end{cases}$$

La statistique proposée par Moddares et Gastwirth (1998) est

$$W_p = \sum_{k=1+np}^{n} (k - np)V_k,$$

où p représente la proportion des observations autour de la médiane qui est tronquée (np est traité comme un entier). Si nous tenons compte de toutes les observations (c.-à-d. $p = 0$), nous obtenons la statistique proposée par Tajuddin (1994). Moddares et Gastwirth (1998) stipulent que W_p est asymptotiquement normale avec

$$\mathrm{E}(W_p) = \frac{1}{2}l(1-p)[n(1-p)+1]$$

et

$$\mathrm{Var}(W_p) = \frac{lm}{12(n-1)}(1-p)[n(1-p)+1][n(1-p)(3p+1)+3p-1].$$

La p-valeur du test est $2 \cdot P_{H0}(Z_w \geq |z_w|)$, où

$$Z_w = \frac{W_p - \mathrm{E}(W_p)}{[\mathrm{Var}(W_p)]^{\frac{1}{2}}}$$

et z_w est la valeur observée de Z_w en appliquant la correction pour la continuité. Nous rejetons donc H_0 si la p-valeur $\leq \alpha_2$.

Moddares et Gastwirth (1998) suggèrent d'utiliser une proportion tronquée $p = 0.70$ pour de petits échantillons ($n < 50$) et $p = 0.80$ pour de plus grands échantillons ($n \geq 50$). À travers plusieurs simulations, ils ont déterminé qu'il y a une augmentation substantielle de la puissance du test quand une proportion tronquée est utilisée. Il n'est pas intuitif que la puissance du test soit plus élevée quand nous nous débarrassons d'une fraction p de l'échantillon. Cependant, il y est suggéré que les observations dans les ailes de la distribution contiennent plus d'information que les observations proches de la médiane.

2.2. Médiane inconnue

Dans cette section, nous sommes intéressés à déterminer si une distribution est symétrique par rapport à un centre de symétrie inconnu. Ce problème a reçu moins d'attention dans la littérature que celui présenté dans la section 2.1.

Définissons $x_1, \ldots, x_n$ comme n observations indépendantes d'une variable aléatoire continue X avec la fonction de répartition F, la fonction de densité f et le centre de symétrie inconnu μ. Un test populaire est celui basé sur le coefficient d'asymétrie de l'échantillon

$$b_1 = \frac{m_3}{(m_2)^{3/2}},$$

où

$$m_k = \frac{1}{n}\sum_{j=1}^{n}(x_j - \overline{x})^k, \quad k = 1, 2, \ldots$$

et $\overline{x}$ représente la moyenne de l'échantillon. Gupta (1967) démontre que la statistique b_1 est asymptotiquement normale pourvu que les 6 premiers moments de la distribution de F existent. Il suggère donc d'utiliser la statistique b_1 dans le test de symétrie $H_0 : \beta_1 = 0$ contre $H_a : \beta_1 \neq 0$. β_1 est le coefficient d'asymétrie de la distribution de F et a déjà été défini dans la section 1.2.1. Randles, Fligner, Policello et Wolfe (1980) ont comparé ce test au test qu'ils ont introduit, le test des triplets. Ils ont découvert que le test des triplets est en général plus puissant que celui utilisant b_1. Dans la section suivante, nous présentons le test des triplets.

2.2.1. Le test des triplets

Randles, Fligner, Policello et Wolfe (1980) proposent la statistique

$$\hat{\eta} = \binom{n}{3}^{-1} \sum_{i<j<k} T(x_i, x_j, x_k) \tag{2.2.1}$$

où

$$T(x_i, x_j, x_k) = \frac{1}{3}[I(x_i + x_j - 2x_k) + I(x_i + x_k - 2x_j) + I(x_j + x_k - 2x_i)],$$

$$I(u) = \begin{cases} 1 & \text{si } u > 0, \\ 0 & \text{si } u = 0, \\ -1 & \text{si } u < 0 \end{cases}$$

et i, j et k sont trois entiers distincts satisfaisant $1 \leq i, j, k \leq n$. Si $T(x_i, x_j, x_k) = \frac{1}{3}$, nous disons que (x_i, x_j, x_k) forment un triplet vers la droite (TD), car l'observation du milieu est plus proche de la petite observation que de la grande. Si $T(x_i, x_j, x_k) = -\frac{1}{3}$, elles forment un triplet vers la gauche (TG). Nous pouvons donc redefinir $\hat{\eta}$ par

$$\hat{\eta} = \frac{\text{nombre de TD} - \text{nombre de TG}}{3\binom{n}{3}}.$$

Intuitivement, un échantillon d'une distribution symétrique devrait contenir autant de TD que de TG, impliquant que $\hat{\eta} = 0$. Dans les Propositions 2.2.1 et 2.2.2, nous trouvons l'espérance de $\hat{\eta}$ et démontrons que $\eta = 0$ pour une distribution symétrique.

Proposition 2.2.1. *Considérons $\hat{\eta}$ défini en* (2.2.1). *Nous avons que*

$$\eta = \mathrm{E}[\hat{\eta}] = P(X_1 + X_2 - 2X_3 > 0) - P(X_1 + X_2 - 2X_3 < 0).$$

DÉMONSTRATION. Nous avons que

$$\mathrm{E}[\hat{\eta}] = \binom{n}{3}^{-1} \sum_{i<j<k} \mathrm{E}[T(X_i, X_j, X_k)] = \mathrm{E}[T(X_1, X_2, X_3)],$$

puisqu'il y a $\binom{n}{3}$ combinaisons différentes des triplets (X_i, X_j, X_k) et que

$$\mathrm{E}[T(X_i, X_j, X_k)] = \mathrm{E}[T(X_1, X_2, X_3)], \quad 1 \leq i < j < k \leq n.$$

Maintenant,

$$\mathrm{E}[I(X_1 + X_2 - 2X_3)] = P(X_1 + X_2 - 2X_3 > 0) - P(X_1 + X_2 - 2X_3 < 0).$$

De plus,

$$\mathrm{E}[I(X_1 + X_2 - 2X_3)] = \mathrm{E}[I(X_1 + X_3 - 2X_2)] = \mathrm{E}[I(X_2 + X_3 - 2X_1)],$$

puisque X_1, X_2 et X_3 sont identiquement distribués. Nous obtenons $\mathrm{E}[T(X_1, X_2, X_3)]$

$$= \frac{1}{3}\left(\mathrm{E}[I(X_1 + X_2 - 2X_3)] + \mathrm{E}[I(X_1 + X_3 - 2X_2)] + \mathrm{E}[I(X_2 + X_3 - 2X_1)]\right)$$

$$= \mathrm{E}[I(X_1 + X_2 - 2X_3)]$$

$$= P(X_1 + X_2 - 2X_3 > 0) - P(X_1 + X_2 - 2X_3 < 0).$$

□

Proposition 2.2.2. *Si X_1, X_2 et X_3 proviennent de la même distribution symétrique F, alors*

$$\eta = P(X_1 + X_2 - 2X_3 > 0) - P(X_1 + X_2 - 2X_3 < 0) = 0.$$

Démonstration. Sous l'hypothèse de symétrie de F, nous avons la relation suivante

$$\begin{aligned} F(\mu + y) &= 1 - F(\mu - y) \\ P(X \leq \mu + y) &= 1 - P(X \leq \mu - y) \\ P(X - \mu \leq y) &= P(X > \mu - y) \\ &= P(\mu - X < y). \end{aligned}$$

Ceci implique que $(X - \mu)$ et $(\mu - X)$ sont identiquement distribués, donc $X - \mu \stackrel{\mathcal{D}}{=} \mu - X$. Nous avons que

$$\begin{aligned} P(X_1 + X_2 - 2X_3 > 0) &= P((X_1 - \mu) + (X_2 - \mu) - 2(X_3 - \mu) > 0) \\ &= P((\mu - X_1) + (\mu - X_2) - 2(\mu - X_3) > 0) \\ &= P(-X_1 - X_2 + 2X_3 > 0) \\ &= P(X_1 + X_2 - 2X_3 < 0) \end{aligned}$$

et donc $\eta = 0$. □

Le test $H_0 : \eta = 0$ contre $H_a : \eta \neq 0$ peut donc être interprété comme un test de symétrie. Pour la description explicite du test, voir Randles, Fligner, Policello et Wolfe (1980).

Chapitre 3

ESTIMATEURS BASÉS SUR UNE MÉTHODE MINIMISANT UNE DISTANCE QUADRATIQUE (DQ)

Dans ce chapitre, nous présentons la classe d'estimateurs basés sur une méthode minimisant une distance quadratique $d(F_n, F_{\boldsymbol{\theta}})$ entre des transformées des fonctions de répartition empirique (F_n) et théorique $(F_{\boldsymbol{\theta}})$ introduite par Luong et Thompson (1987). Des statistiques pour des tests d'ajustement peuvent être construites à partir de la mesure $d(F_n, F_{\boldsymbol{\theta}})$ et donc, l'estimation de paramètres et la validation de modèles sont traitées de manière analogue. Par exemple, la classe d'estimateurs basés sur la DQ inclut la méthode du khi-deux minimum et la méthode $K-L$ pour fonctions caractéristiques proposée par Feuerverger et McDunnough (1981a). Nous introduisons maintenant la classe d'estimateurs basés sur la DQ et dérivons certaines propriétés asymptotiques qui y sont rattachées.

3.1. Définitions et propriétés asymptotiques

Définition 3.1.1. *Définissons $x_1, \ldots, x_n$ comme n observations indépendantes d'une variable aléatoire X avec la fonction de répartition $F_{\boldsymbol{\theta}}$, où $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_m] \in \Theta \subseteq \mathbb{R}^m$. Dénotons aussi $\overline{\Theta}$ comme l'adhérence de Θ dans l'espace $\mathbb{R}^m$. La classe d'estimateurs basés sur la DQ est celle des estimateurs minimisant les distances de la forme*

$$d(F_n, F_{\boldsymbol{\theta}}) = [\mathbf{z}(F_n) - \mathbf{z}(F_{\boldsymbol{\theta}})]'\mathbf{Q}(F_{\boldsymbol{\theta}})[\mathbf{z}(F_n) - \mathbf{z}(F_{\boldsymbol{\theta}})],$$

où

(i) pour un ensemble fixe de fonctions $h_j, j = 1, \ldots, k$, la transformée de la fonction de répartition F est le vecteur $\mathbf{z}(F) = [z_1(F), z_2(F), ..., z_k(F)]'$, où

$$z_j(F) = \int_{-\infty}^{\infty} h_j(x) dF(x), \quad j = 1, \ldots, k,$$

si les h_j sont intégrables par rapport à F,

(ii) pour tout $\boldsymbol{\theta} \in \overline{\Theta}$, la matrice $\mathbf{Q}(F_{\boldsymbol{\theta}})$ est symétrique, continue et définie positive de dimension $k \times k$,

(iii) sous chaque $F_{\boldsymbol{\theta}}$, la matrice de variance-covariance de $[h_1(x), h_2(x), ..., h_k(x)]$ existe ; nous la dénotons par $\boldsymbol{\Sigma}(F_{\boldsymbol{\theta}}) = (\sigma_{ij})_{1 \leq i,j \leq k}$, où

$$\sigma_{ij} = \int_{-\infty}^{\infty} h_i(x) h_j(x) dF_{\boldsymbol{\theta}}(x) - z_i(F_{\boldsymbol{\theta}}) z_j(F_{\boldsymbol{\theta}}).$$

Notons que c'est également la matrice de variance-covariance de $\sqrt{n}\mathbf{z}(F_n)$ sous $F_{\boldsymbol{\theta}}$.

Dans ce qui suit, afin de simplifier la notation, nous utiliserons $\mathbf{z}(\boldsymbol{\theta}), \mathbf{z}_n, \mathbf{Q}(\boldsymbol{\theta}), \boldsymbol{\Sigma}(\boldsymbol{\theta})$ à la place de $\mathbf{z}(F_{\boldsymbol{\theta}}), \mathbf{z}(F_n), \mathbf{Q}(F_{\boldsymbol{\theta}}), \boldsymbol{\Sigma}(F_{\boldsymbol{\theta}})$ respectivement. L'estimateur basé sur la DQ, dénoté par $\hat{\boldsymbol{\theta}}_n$, est la valeur de $\boldsymbol{\theta}$ qui minimise la distance

$$[\mathbf{z}_n - \mathbf{z}(\boldsymbol{\theta})]' \mathbf{Q}(\boldsymbol{\theta}) [\mathbf{z}_n - \mathbf{z}(\boldsymbol{\theta})]. \tag{3.1.1}$$

Une condition suffisante pour la convergence de l'estimateur $\hat{\boldsymbol{\theta}}_n$ est établie dans le lemme suivant apparaissant dans Luong et Thompson (1987).

Lemme 3.1.1. *Comme auparavant, définissons $\mathbf{Q}(\boldsymbol{\theta})$ comme une matrice symétrique, continue en $\boldsymbol{\theta}$ et définie positive pour tout $\boldsymbol{\theta} \in \overline{\Theta}$. Aussi, $\hat{\boldsymbol{\theta}}_n$ est défini comme un estimateur minimisant une distance de la forme* (3.1.1). *Si $k > m$ et la fonction $\boldsymbol{\theta} \to \mathbf{z}(\boldsymbol{\theta})$ de Θ à $\mathbb{R}^k$ possède un inverse continu à $\boldsymbol{\theta}_0$ (c'est-à-dire que toutes les fois que $\mathbf{z}(\boldsymbol{\theta}_n) \to \mathbf{z}(\boldsymbol{\theta}_0)$, nous avons $\boldsymbol{\theta}_n \to \boldsymbol{\theta}_0$), alors $\hat{\boldsymbol{\theta}}_n \xrightarrow{\mathcal{P}} \boldsymbol{\theta}_0$, où $F_{\boldsymbol{\theta}_0}$ est la vraie fonction de répartition de X ($\xrightarrow{\mathcal{P}}$ dénote la convergence en probabilité).*

DÉMONSTRATION. Puisque $\hat{\boldsymbol{\theta}}_n$ est l'estimateur minimisant une distance de la forme (3.1.1), nous devons nécessairement avoir

$$0 \leq [\mathbf{z}_n - \mathbf{z}(\hat{\boldsymbol{\theta}}_n)]'\mathbf{Q}(\hat{\boldsymbol{\theta}}_n)[\mathbf{z}_n - \mathbf{z}(\hat{\boldsymbol{\theta}}_n)] \leq [\mathbf{z}_n - \mathbf{z}(\boldsymbol{\theta}_0)]'\mathbf{Q}(\boldsymbol{\theta}_0)[\mathbf{z}_n - \mathbf{z}(\boldsymbol{\theta}_0)]. \quad (3.1.2)$$

Puisque $\mathbf{z}_n \xrightarrow{\mathcal{P}} \mathbf{z}(\boldsymbol{\theta}_0)$ par la loi faible des grands nombres ($\mathbf{z}_n$ est la moyenne empirique de $\mathbf{z}(\boldsymbol{\theta}_0)$), l'expression à la droite de l'équation (3.1.2) converge vers 0 en probabilité et donc, le terme du milieu aussi. Ceci est seulement possible si $\mathbf{z}_n - \mathbf{z}(\hat{\boldsymbol{\theta}}_n) \xrightarrow{\mathcal{P}} 0$ impliquant que $\mathbf{z}(\hat{\boldsymbol{\theta}}_n) \xrightarrow{\mathcal{P}} \mathbf{z}(\boldsymbol{\theta}_0)$. Puisque nous avons fait l'hypothèse que la fonction $\boldsymbol{\theta} \to \mathbf{z}(\boldsymbol{\theta})$ de Θ à $\mathbb{R}^k$ possède un inverse continu à $\boldsymbol{\theta}_0$, nous obtenons $\hat{\boldsymbol{\theta}}_n \xrightarrow{\mathcal{P}} \boldsymbol{\theta}_0$. □

Les lemme et remarque suivants apparaissant dans Luong et Thompson (1987) nous permettent d'établir que les estimateurs basés sur la DQ sont asymptotiquement normaux.

Lemme 3.1.2. *Supposons que $k > m$, que $\hat{\boldsymbol{\theta}}_n \xrightarrow{\mathcal{P}} \boldsymbol{\theta}_0$ et que les conditions suivantes sont satisfaites :*

(i) la matrice $\mathbf{Q}$ est symétrique, continue et définie positive (et les dérivés partielles par rapport au vecteur $\boldsymbol{\theta}$ existent),

(ii) avec probabilité 1, l'estimateur $\hat{\boldsymbol{\theta}}_n$ satisfait le système d'équations m-dimensionnel

$$\sqrt{n}\frac{\partial}{\partial \boldsymbol{\theta}}\left[[\mathbf{z}_n - \mathbf{z}(\hat{\boldsymbol{\theta}}_n)]'\mathbf{Q}(\hat{\boldsymbol{\theta}}_n)[\mathbf{z}_n - \mathbf{z}(\hat{\boldsymbol{\theta}}_n)]\right] = 0,$$

(iii) chacun des $\partial z_i(\boldsymbol{\theta})/\partial\theta_j$ est continu à $\boldsymbol{\theta}_0$, $i = 1, \ldots, k$ et $j = 1, \ldots, m$,

(iv) la matrice de dimension $k \times m$, $\mathbf{S}(\boldsymbol{\theta}) = (s_{ij})$, où $s_{ij} = \partial z_i(\boldsymbol{\theta})/\partial\theta_j$, est de rang m quand $\boldsymbol{\theta} = \boldsymbol{\theta}_0$,

(v) $[\partial\mathbf{Q}(\boldsymbol{\theta})/\partial\theta_1, \ldots, \partial\mathbf{Q}(\boldsymbol{\theta})/\partial\theta_m]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_n}$ est borné en probabilité.

Alors, nous avons

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = (\mathbf{S}'\mathbf{Q}\mathbf{S})^{-1}\mathbf{S}'\mathbf{Q}\sqrt{n}[\mathbf{z}_n - \mathbf{z}] + o_p(1),$$

où $\mathbf{S} = \mathbf{S}(\boldsymbol{\theta}_0)$, $\mathbf{Q} = \mathbf{Q}(\boldsymbol{\theta}_0)$, $\mathbf{z} = \mathbf{z}(\boldsymbol{\theta}_0)$ et $o_p(1)$ représente une expression convergeant vers 0 en probabilité.

DÉMONSTRATION. Voir Luong et Thompson (1987). □

Remarque 3.1.1. *Luong et Thompson (1987) ont démontré que* $\sqrt{n}[\mathbf{z}_n - \mathbf{z}] \xrightarrow{\mathcal{D}} N(0, \boldsymbol{\Sigma})$, *où* $\boldsymbol{\Sigma}$ *est la matrice de variance-covariance de* $\sqrt{n}[\mathbf{z}_n - \mathbf{z}]$ *sous l'hypothèse* $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ *et* $\xrightarrow{\mathcal{D}}$ *veut dire convergence en distribution. Donc, en utilisant le Lemme 3.1.2, nous obtenons que*

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{\mathcal{D}} N(0, \boldsymbol{\Sigma}_1),$$

où $\boldsymbol{\Sigma}_1 = (\mathbf{S}'\mathbf{Q}\mathbf{S})^{-1}\mathbf{S}'\mathbf{Q}\boldsymbol{\Sigma}\mathbf{Q}\mathbf{S}(\mathbf{S}'\mathbf{Q}\mathbf{S})^{-1}$. *De plus, si* $\boldsymbol{\Sigma}$ *est inversible, alors le choix optimal de* $\mathbf{Q}$ *afin de minimiser la norme de la matrice* $\boldsymbol{\Sigma}_1$ *est* $\boldsymbol{\Sigma}^{-1}$. *Nous avons alors que*

$$\begin{aligned}
\boldsymbol{\Sigma}_1 &= (\mathbf{S}'\boldsymbol{\Sigma}^{-1}\mathbf{S})^{-1}\mathbf{S}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}\mathbf{S}(\mathbf{S}'\boldsymbol{\Sigma}^{-1}S)^{-1} \\
&= (\mathbf{S}'\boldsymbol{\Sigma}^{-1}\mathbf{S})^{-1}(\mathbf{S}'\boldsymbol{\Sigma}^{-1}\mathbf{S})(\mathbf{S}'\boldsymbol{\Sigma}^{-1}\mathbf{S})^{-1} \\
&= (\mathbf{S}'\boldsymbol{\Sigma}^{-1}\mathbf{S})^{-1}.
\end{aligned}$$

Dans la section suivante, nous présentons des statistiques pour des tests d'ajustement construites à partir de la distance quadratique $d(F_n, F_{\boldsymbol{\theta}})$.

3.2. TESTS D'AJUSTEMENT

Puisque nous avons bâti des statistiques reposant sur une distance minimum entre des parties empiriques et théoriques, il est naturel de les utiliser dans le cadre de la validation de modèles. Nous allons présenter des tests d'ajustement pour l'hypothèse simple $H_0 : F = F_{\boldsymbol{\theta}_0}$ et l'hypothèse composée $H_0 : F = F_{\boldsymbol{\theta}}$ pour un $\boldsymbol{\theta} \in \Theta$. Les statistiques pour ces tests suivent une loi asymptotique khi-deux. Avant de construire ces statistiques, nous énonçons un théorème et une remarque apparaissant dans Luong et Thompson (1987).

Théorème 3.2.1. *Supposons que le vecteur aléatoire* $\mathbf{Y}$ *de dimension* k *est* $N(0, \boldsymbol{\Sigma})$ *et* $\mathbf{Q}$ *est une matrice symétrique semi-définie positive de dimension* $k \times k$ *; alors la forme quadratique* $\mathbf{Y}'\mathbf{Q}\mathbf{Y}$ *suit une loi khi-deux avec* ν *degrés de liberté si* $\boldsymbol{\Sigma}\mathbf{Q}$ *est idempotente et* $\mathrm{trace}(\boldsymbol{\Sigma}\mathbf{Q}) = \nu$. *(Le même résultat est vrai asymptotiquement si* $\mathbf{Q}$ *est remplacé par un estimateur convergent* $\hat{\mathbf{Q}}$ *et* $\mathbf{Y} \xrightarrow{\mathcal{D}} N(0, \boldsymbol{\Sigma})$*).*

DÉMONSTRATION. Voir Rao (1973). □

3.2.1. Hypothèse simple

Afin de tester l'hypothèse simple $H_0 : F = F_{\boldsymbol{\theta}_0}$, Luong et Thompson (1987) proposent la statistique suivante basée sur la DQ :

$$nd(F_n, F_{\boldsymbol{\theta}_0}) = n[\mathbf{z}_n - \mathbf{z}]'\boldsymbol{\Sigma}^{-1}[\mathbf{z}_n - \mathbf{z}] = \mathbf{v}_n'\boldsymbol{\Sigma}^{-1}\mathbf{v}_n,$$

où $\mathbf{v}_n = \sqrt{n}[\mathbf{z}_n - \mathbf{z}]$ et $\boldsymbol{\Sigma}$ est la matrice de variance-covariance de $\mathbf{v}_n$. En utilisant la Remarque 3.1.1 et le Théorème 3.2.1, nous constatons que $\mathbf{v}_n'\boldsymbol{\Sigma}^{-1}\mathbf{v}_n \xrightarrow{\mathcal{D}} \chi^2_\nu$, où

$$\nu = \text{trace}(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}) = \text{trace}(\mathbf{I}_k) = k,$$

car $\mathbf{I}_k$, la matrice identité de dimension $k \times k$, est idempotente ($\mathbf{I}_k^2 = \mathbf{I}_k$).

3.2.2. Hypothèse composée

Afin de tester l'hypothèse composée $H_0 : F = F_{\boldsymbol{\theta}}$ pour un $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, Luong et Thompson (1987) proposent la statistique suivante basée sur la DQ :

$$nd(F_n, F_{\hat{\boldsymbol{\theta}}_n}) = n[\mathbf{z}_n - \mathbf{z}(\hat{\boldsymbol{\theta}}_n)]'\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\theta}}_n)[\mathbf{z}_n - \mathbf{z}(\hat{\boldsymbol{\theta}}_n)] = \mathbf{v}_n'(\hat{\boldsymbol{\theta}}_n)\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\theta}}_n)\mathbf{v}_n(\hat{\boldsymbol{\theta}}_n),$$

où $\mathbf{v}_n(\hat{\boldsymbol{\theta}}_n) = \sqrt{n}[\mathbf{z}_n - \mathbf{z}(\hat{\boldsymbol{\theta}}_n)]$ et $\hat{\boldsymbol{\theta}}_n$ est l'estimateur obtenu en minimisant la distance (3.1.1) avec $\mathbf{Q}(\boldsymbol{\theta}) = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})$. Afin de trouver la distribution asymptotique de $nd(F_n, F_{\hat{\boldsymbol{\theta}}_n})$, nous devons faire usage de la proposition suivante.

Proposition 3.2.1. *La distribution asymptotique de* $\mathbf{v}_n(\hat{\boldsymbol{\theta}}_n) \xrightarrow{\mathcal{D}} N(0, \boldsymbol{\Sigma}_2)$, *où*

$$\boldsymbol{\Sigma}_2 = \left[\mathbf{I}_k - \mathbf{S}(\mathbf{S}'\boldsymbol{\Sigma}^{-1}\mathbf{S})^{-1}\mathbf{S}'\boldsymbol{\Sigma}^{-1}\right]\boldsymbol{\Sigma}\left[\mathbf{I}_k - \boldsymbol{\Sigma}^{-1}\mathbf{S}(\mathbf{S}'\boldsymbol{\Sigma}^{-1}\mathbf{S})^{-1}\mathbf{S}'\right].$$

DÉMONSTRATION. Par l'expansion en série de Taylor, nous obtenons

$$\mathbf{v}_n(\hat{\boldsymbol{\theta}}_n) = \mathbf{v}_n - \mathbf{S}\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + o_p(1).$$

En utilisant le Lemme 3.1.2 avec $\mathbf{Q} = \boldsymbol{\Sigma}^{-1}$, nous avons

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = (\mathbf{S}'\boldsymbol{\Sigma}^{-1}\mathbf{S})^{-1}\mathbf{S}'\boldsymbol{\Sigma}^{-1}\mathbf{v}_n + o_p(1)$$

qui nous amène à

$$\begin{aligned}\mathbf{v}_n(\hat{\boldsymbol{\theta}}_n) &= \mathbf{v}_n - \mathbf{S}(\mathbf{S}'\boldsymbol{\Sigma}^{-1}\mathbf{S})^{-1}\mathbf{S}'\boldsymbol{\Sigma}^{-1}\mathbf{v}_n + o_p(1)\\ &= [\mathbf{I}_k - \mathbf{S}(\mathbf{S}'\boldsymbol{\Sigma}^{-1}\mathbf{S})^{-1}\mathbf{S}'\boldsymbol{\Sigma}^{-1}]\mathbf{v}_n + o_p(1).\end{aligned}$$

Puisque $\mathbf{v}_n \xrightarrow{\mathcal{D}} N(0, \boldsymbol{\Sigma})$ (Remarque 3.1.1), nous obtenons $\mathbf{v}_n(\hat{\boldsymbol{\theta}}_n) \xrightarrow{\mathcal{D}} N(0, \boldsymbol{\Sigma}_2)$, où

$$\begin{aligned}\boldsymbol{\Sigma}_2 &= \left[\mathbf{I}_k - \mathbf{S}(\mathbf{S}'\boldsymbol{\Sigma}^{-1}\mathbf{S})^{-1}\mathbf{S}'\boldsymbol{\Sigma}^{-1}\right] \boldsymbol{\Sigma} \left[\mathbf{I}_k - \mathbf{S}(\mathbf{S}'\boldsymbol{\Sigma}^{-1}\mathbf{S})^{-1}\mathbf{S}'\boldsymbol{\Sigma}^{-1}\right]'\\ &= \left[\mathbf{I}_k - \mathbf{S}(\mathbf{S}'\boldsymbol{\Sigma}^{-1}\mathbf{S})^{-1}\mathbf{S}'\boldsymbol{\Sigma}^{-1}\right] \boldsymbol{\Sigma} \left[\mathbf{I}_k - \boldsymbol{\Sigma}^{-1}\mathbf{S}(\mathbf{S}'\boldsymbol{\Sigma}^{-1}\mathbf{S})^{-1}\mathbf{S}'\right].\end{aligned}$$

□

En utilisant le Théorème 3.2.1, nous pouvons conclure que $nd(F_n, F_{\hat{\boldsymbol{\theta}}_n}) \xrightarrow{\mathcal{D}} \chi^2_{\nu}$, où $\nu = \mathrm{trace}(\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}^{-1})$. La matrice $\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}^{-1}$ est idempotente, car

$$\begin{aligned}\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}^{-1} &= \left[\mathbf{I}_k - \mathbf{S}(\mathbf{S}'\boldsymbol{\Sigma}^{-1}\mathbf{S})^{-1}\mathbf{S}'\boldsymbol{\Sigma}^{-1}\right] \boldsymbol{\Sigma} \left[\mathbf{I}_k - \boldsymbol{\Sigma}^{-1}\mathbf{S}(\mathbf{S}'\boldsymbol{\Sigma}^{-1}\mathbf{S})^{-1}\mathbf{S}'\right] \boldsymbol{\Sigma}^{-1}\\ &= \left[\mathbf{I}_k - \mathbf{S}(\mathbf{S}'\boldsymbol{\Sigma}^{-1}\mathbf{S})^{-1}\mathbf{S}'\boldsymbol{\Sigma}^{-1}\right] \left[\mathbf{I}_k - \mathbf{S}(\mathbf{S}'\boldsymbol{\Sigma}^{-1}\mathbf{S})^{-1}\mathbf{S}'\boldsymbol{\Sigma}^{-1}\right]\\ &= \left[\mathbf{I}_k - \mathbf{S}(\mathbf{S}'\boldsymbol{\Sigma}^{-1}\mathbf{S})^{-1}\mathbf{S}'\boldsymbol{\Sigma}^{-1}\right]^2\\ &= \mathbf{I}_k - 2\mathbf{S}(\mathbf{S}'\boldsymbol{\Sigma}^{-1}\mathbf{S})^{-1}\mathbf{S}'\boldsymbol{\Sigma}^{-1} + \mathbf{S}(\mathbf{S}'\boldsymbol{\Sigma}^{-1}\mathbf{S})^{-1}\mathbf{S}'\boldsymbol{\Sigma}^{-1}\mathbf{S}(\mathbf{S}'\boldsymbol{\Sigma}^{-1}\mathbf{S})^{-1}\mathbf{S}'\boldsymbol{\Sigma}^{-1}\\ &= \mathbf{I}_k - 2\mathbf{S}(\mathbf{S}'\boldsymbol{\Sigma}^{-1}\mathbf{S})^{-1}\mathbf{S}'\boldsymbol{\Sigma}^{-1} + \mathbf{S}(\mathbf{S}'\boldsymbol{\Sigma}^{-1}\mathbf{S})^{-1}\mathbf{S}'\boldsymbol{\Sigma}^{-1}\\ &= \mathbf{I}_k - \mathbf{S}(\mathbf{S}'\boldsymbol{\Sigma}^{-1}\mathbf{S})^{-1}\mathbf{S}'\boldsymbol{\Sigma}^{-1}.\end{aligned}$$

Lorsqu'une matrice $\boldsymbol{A}$ est idempotente, nous avons que $\mathrm{trace}(\boldsymbol{A}) = \mathrm{rang}(\boldsymbol{A})$. Donc, nous obtenons

$$\begin{aligned}\nu = \mathrm{trace}(\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}^{-1}) = \mathrm{rang}(\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}^{-1}) &= \mathrm{rang}(\mathbf{I}_k - \mathbf{S}(\mathbf{S}'\boldsymbol{\Sigma}^{-1}\mathbf{S})^{-1}\mathbf{S}'\boldsymbol{\Sigma}^{-1})\\ &= \mathrm{rang}(\mathbf{I}_k) - \mathrm{rang}(\mathbf{S}(\mathbf{S}'\boldsymbol{\Sigma}^{-1}\mathbf{S})^{-1}\mathbf{S}'\boldsymbol{\Sigma}^{-1}) = k - m\end{aligned}$$

3.3. La DQ avec la fonction caractéristique

Dans le chapitre 4, nous proposons d'estimer les paramètres de la famille représentée par la différence de deux variable aléatoires gamma en utilisant la DQ basée sur la transformée de Fourier-Stieltjes (c.-à-d. la fonction caractéristique). L'idée est de trouver les paramètres qui minimisent la distance entre les parties réelles et imaginaires des fonctions caractéristiques empiriques et théoriques. Nous voulons donc minimiser la DQ entre les vecteurs $\mathbf{Z}_n$ et $\mathbf{Z}(\boldsymbol{\theta})$ de dimension $2k$ définis dans les équations (1.3.1) et (1.3.2) respectivement. Plus précisément, nous sommes intéressés à trouver l'estimateur $\hat{\boldsymbol{\theta}}_n$ de $\boldsymbol{\theta}$ qui minimise la distance

$$d(\boldsymbol{\theta}) = [\mathbf{Z}_n - \mathbf{Z}(\boldsymbol{\theta})]'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})[\mathbf{Z}_n - \mathbf{Z}(\boldsymbol{\theta})],$$

où les éléments de $\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})$ sont trouvés à partir du Théorème 1.3.2. Il est facile de constater que l'estimateur $\hat{\boldsymbol{\theta}}_n$ fait partie de la classe des estimateurs définie dans la section 3.1 avec les fonctions $h_j(x)$, $j = 1, \ldots, 2k$, où

$$h_j(x) = \begin{cases} \cos(t_j x) & \text{si } j = 1, \ldots, k, \\ \sin(t_j x) & \text{si } j = k+1, \ldots, 2k. \end{cases}$$

Nous pouvons donc supposer que la méthode utilisant la DQ basée sur la transformée de Fourier-Stieltjes générera des estimateurs adéquats puisque les fonctions trigonométriques sont des fonctions bornées, continues et dérivables.

Chapitre 4

L'ARTICLE

Le présent chapitre est une copie intégrale de l'article intitulé "A Leptokurtic Symmetric Family of Distributions Represented by the Difference of Two Gamma Variates" soumis à la revue scientifique *Communications in Statistics - Simulation and Computation*. Le premier auteur est également l'auteur du présent mémoire.

A Leptokurtic Symmetric Family of Distributions Represented by the Difference of Two Gamma Variates

Maciej Augustyniak and Louis G. Doray

Département de mathématiques et de statistique, Université de Montréal, C.P. 6128, Succursale Centre-ville, Montréal, Québec, Canada H3C 3J7

Key Words and Phrases: Empirical characteristic function, quadratic distance, parameter estimation, leptokurtosis, gamma distribution, Laplace distribution, goodness-of-fit, normality test, test of symmetry.

ABSTRACT

We introduce a family of leptokurtic symmetric distributions represented by the difference of two gamma variates. Properties of this family are discussed. The Laplace, the sums of Laplace and the normal distributions all arise as special cases of this family. We propose a two-step method for fitting data to this family. First, we perform a test of symmetry, and second, we estimate the parameters by minimizing the quadratic distance between the real parts of the empirical and theoretical characteristic functions. The quadratic distance estimator obtained is consistent, robust and asymptotically normally distributed. We develop a statistical test for goodness-of-fit and introduce a test of normality of the data. A simulation study is provided to validate the theory.

1 INTRODUCTION

We introduce a family of leptokurtic symmetric distributions by presenting its characteristic function. Consider X and Y to be independent and identically distributed random variables from a gamma distribution with shape $\frac{1}{\lambda}$ and scale $\sqrt{\lambda\theta}$ (i.e. $X, Y \sim \Gamma(\frac{1}{\lambda}, \sqrt{\lambda\theta})$), where the parameters λ and θ are defined on the positive real line. The new family is represented by the random variable Z, where $Z = X - Y$ with characteristic function

$$\phi_Z(t) = \phi_X(t)\phi_{-Y}(t) = (\frac{1}{1 - it\sqrt{\lambda\theta}})^{\frac{1}{\lambda}}(\frac{1}{1 + it\sqrt{\lambda\theta}})^{\frac{1}{\lambda}} = (\frac{1}{1 + t^2\lambda\theta})^{\frac{1}{\lambda}}. \quad (1.1)$$

In the limit, as $\lambda \to 0$, we have $\phi_Z(t) \to e^{-t^2\theta}$, which is the characteristic function of a normal random variable centered at 0 and with variance 2θ. Hence, we define this family of symmetric distributions by its characteristic function $\phi(t)$, where

$$\phi(t) = \begin{cases} (\frac{1}{1+t^2\lambda\theta})^{\frac{1}{\lambda}} & \text{for } \lambda, \theta > 0, \\ e^{-t^2\theta} & \text{for } \lambda = 0, \theta > 0. \end{cases} \quad (1.2)$$

Subsequently, we use the designation that a random variable follows the double gamma difference (DGD) distribution family with parameters (λ, θ) if its characteristic function is given by (1.2). For $\lambda < 0$ or $\theta < 0$, $\phi(t)$ is not a characteristic function because $|\phi(t)|$ is not bounded by 1. Please refer to Lukacs (1970) for more details on the properties of characteristic functions. When $\lambda = 1$, the family becomes the difference of two independent exponentially distributed variates with mean $\sqrt{\theta}$. Kotz, Kozubowski and Podgórski (2001) proved that a Laplace random variable centered at the origin can be represented as the difference between two independent exponentially distributed variates. This is easily seen by letting $\lambda = 1$ in (1.1) and noting that the characteristic function of a classical Laplace random variable centered at 0 with scale paramter s is

$$\frac{1}{1 + t^2s^2}.$$

Hence, the classical Laplace distribution is a special case of the DGD family with parameters $(\lambda = 1, \theta = s^2)$. Moreover, when $\lambda = \frac{1}{n}$, $n \in \mathbb{N}$, the difference of two independent gamma variates can also be seen as the sum of n differences of two independent exponential variates which in turn is simply the sum of n independent Laplace variates. Hence, in the limit, when $n \to \infty$ (i.e. $\lambda \to 0$), our result is consistent with the Central Limit Theorem as a sum of n independent Laplace variates converges to the normal distribution.

We now list some properties and particularities of this family. Since its characteristic function is real and even; it is a family of symmetric distribution functions (i.e. symmetric distributions centered at 0, refer to Lukacs (1970) for more details about this relationship). Thus, the odd moments are 0 and the positive even moments can be calculated from the formula given in the following proposition.

Proposition 1. *Let Z be a $DGD(\lambda, \theta)$ random variable with characteristic function $\phi(t)$ as defined by* (1.2). *Then,*

$$\mathrm{E}[Z^{2k}] = \frac{\theta^k (2k)!}{k!} \prod_{j=0}^{k-1} (1 + j\lambda), \quad k = 1, 2, 3, \ldots .$$

Proof. From the generalized binomial theorem, we obtain the binomial series for $\phi(t)$, where

$$\begin{aligned}\phi(t) = (1 + t^2\lambda\theta)^{-\frac{1}{\lambda}} &= \sum_{k=0}^{\infty} \binom{-\frac{1}{\lambda}}{k} (t^2\lambda\theta)^k = \sum_{k=0}^{\infty} \binom{-\frac{1}{\lambda}}{k} \lambda^k \theta^k t^{2k} \\ &= 1 + \binom{-\frac{1}{\lambda}}{1} \lambda\theta t^2 + \binom{-\frac{1}{\lambda}}{2} \lambda^2\theta^2 t^4 + \binom{-\frac{1}{\lambda}}{3} \lambda^3\theta^3 t^6 + \cdots .\end{aligned}$$

From the relationship between moments of a random variable and the derivatives of its characteristic function, we have

$$\mathrm{E}[Z^{2k}] = i^{-2k} \left. \frac{d^{2k}}{dt^{2k}} \phi(t) \right|_{t=0} = (-1)^k \phi^{(2k)}(0).$$

$\phi^{(2k)}(0)$ corresponds to the $(k+1)^{\text{th}}$ term from the binomial series for $\phi(t)$ differentiated $2k$ times. Thus,

$$\mathrm{E}[Z^{2k}] = (-1)^k \phi^{(2k)}(0) = (-1)^k \binom{-\frac{1}{\lambda}}{k} (2k)! \lambda^k \theta^k.$$

Since the binomial coefficients admit the representation

$$\binom{-\frac{1}{\lambda}}{k} = \frac{1}{k!}\prod_{j=0}^{k-1}(-\frac{1}{\lambda} - j) = \frac{(-1)^k}{k!\lambda^k}\prod_{j=0}^{k-1}(1 + j\lambda),$$

we get the following expression:

$$\mathrm{E}[Z^{2k}] = (-1)^k(2k)!\lambda^k\theta^k\left[\frac{(-1)^k}{k!\lambda^k}\prod_{j=0}^{k-1}(1+j\lambda)\right] = \frac{\theta^k(2k)!}{k!}\prod_{j=0}^{k-1}(1+j\lambda).$$

We note here that the proof only applies for values of $\lambda > 0$. However, the same result holds for $\lambda = 0$. The expression for $\lambda = 0$ can be derived similarly by using the series expansion for $e^{-t^2\theta}$. □

From Proposition 1, we obtain the variance and the kurtosis of a $DGD(\lambda, \theta)$ random variable which are 2θ and $(3+3\lambda)$ respectively. Since $\lambda \geq 0$, the kurtosis is always greater or equal to 3. Thus, the family is leptokurtic because the kurtosis is always at least that of the normal distribution. Let $Z_1, Z_2, \ldots, Z_n$ be independent and identically distributed $DGD(n\lambda, \theta/n)$ variates and consider $Z = Z_1 + Z_2 + \ldots + Z_n$, then

$$\phi_Z(t) = \phi_{Z_1}(t)\phi_{Z_2}(t)\cdots\phi_{Z_n}(t) = [\phi_{Z_1}(t)]^n = (\frac{1}{1+t^2\lambda\theta})^{\frac{1}{\lambda}}. \tag{1.3}$$

Clearly, Z is a $DGD(\lambda, \theta)$ random variable. Also, if $a \in \mathbb{R}$, then the characteristic function of aZ is $\phi_{aZ}(t)$, where

$$\phi_{aZ}(t) = \phi_Z(at) = (\frac{1}{1+t^2a^2\lambda\theta})^{\frac{1}{\lambda}}.$$

This entails that aZ is a $DGD(\lambda, a^2\theta)$ random variable. Thus, the family is closed under scale and convolution operations but not under the translation operation as the center of symmetry is fixed at the origin. Moreover, from (1.3) we can recognize that its characteristic function is infinitely divisible. When the family reduces to Lapace or normal random variables, the density function can be expressed in a closed form. For example, when $\lambda = 1$, we have already established that we obtain the characteristic function of a classical Laplace random variable centered at 0 with corresponding density function denoted by $f(z; \lambda = 1, \theta = s^2)$, where

$$f(z; \lambda = 1, \theta = s^2) = \frac{1}{2s} e^{-|z|/s}.$$

Also, when $\lambda = \frac{1}{n}$, $n \in \mathbb{N}$, and $\theta = n$, we retrieve the characteristic function of the sum of n standard classical Laplace variates. Kotz, Kozubowski and Podgórski (2001) obtained the following formula for its density function denoted by $f(z; \lambda = 1/n, \theta = n)$, where

$$f(z; \lambda = 1/n, \theta = n) = \frac{e^{-|z|}}{(n-1)!2^n} \sum_{j=0}^{n-1} \frac{(n-1+j)!}{(n-1-j)!j!} \cdot \frac{|z|^{n-1-j}}{2^j}.$$

However, in the general case the density function is not tractable and does not have a closed form expression. This is the reason why we introduced the DGD family by presenting its characteristic function.

Since the family consists of symmetric leptokurtic distributions, this suggests that data exhibiting the properties of being symmetric around the origin and of having excess kurtosis can be fitted to this family. In the following section, we develop a two-step method for fitting data to the DGD family. The first step comprises model validation and the second, parameter estimation. In Section 3, goodness-of-fit tests for the simple and composite hypotheses are presented. The test statistics are shown to follow a chi-square distribution asymptotically. In addition, we explain how the parameter λ can be employed to test for distributional assumptions. More precisely, a test of normality of the data is presented. In Section 4, we provide simulation results for the methods developed.

2 FITTING TO THE DGD FAMILY

2.1 Introduction

We suggest a two-step method for fitting data to the DGD family. The first step consists of assessing the compatibility between the data and the family. Since it is a family of leptokurtic symmetric distributions, the data have to exhibit those characteristics. We verify them by performing a test of symmetry and by only considering data with a kurtosis greater than 3. Once we have confirmed that the family is well-suited for the data, we proceed with parameter estimation which is the second step. Parameter estimation is achieved through a minimum-distance method based on the characteristic function. We choose the parameters which minimize the distance between the real parts of the theoretical characteristic function and the empirical characteristic function. The estimators obtained are consistent, robust and asymptotically normal.

2.2 Testing Symmetry

2.2.1 Introduction

Let $x_1, \ldots, x_n$ be n independent observations from a continuous random variable X with distribution function F, density f and known center μ_0. We consider the problem of testing

$$H_0 : F(\mu_0 - x) = 1 - F(\mu_0 + x)$$

against

$$H_a : F(\mu_0 - x) \neq 1 - F(\mu_0 + x).$$

Thus, we are interested in testing whether the density f is symmetric about the known median μ_0 or skewed.

Many tests of symmetry have been described in the literature (see Lehman, 1975; Randles and Wolfe, 1979). McWilliams (1990) and Moddares and Gastwirth (1996) used tests based on a runs statistic. Tajuddin (1994) and Thas, Rayner and Best (2005) used tests based on the Wilcoxon signed rank statistic. Also, Cheng and Balakrishnan (2004) proposed a modified sign test for symmetry.

We suggest using the hybrid test proposed by Moddares and Gastwirth (1998) to test the hypothesis of symmetry around a known median. We favor this test due to its high power and simplicity. Thas, Rayner and Best (2005) performed extensive simulations comparing the power of different tests of symmetry, which revealed that the hybrid test is more powerful than most alternatives considered.

2.2.2 Hybrid test

The hybrid test is defined in two stages. Stage I consists of the sign test at level $\alpha_1 < \alpha$. If H_0 is accepted in stage I, then the percentile-modified two-sample Wilcoxon test is performed in stage II at level $\alpha_2 < \alpha$. The hybrid procedure is an α-level test, where $\alpha = \alpha_1 + (1 - \alpha_1)\alpha_2$. Suggested values for the levels of the tests are $\alpha_1 = 0.01$ and $\alpha_2 = 0.0404$ generating an overall level of $\alpha = 0.05$. Please refer to Moddares and Gastwirth (1998) for more details on the hybrid procedure.

The first step of our method involves validating the compatibility between the data and the DGD family. It consists of two elements: the kurtosis of the data has to be greater than 3 and the hybrid test must not reject the hypothesis of symmetry around μ_0. If the data qualify, then we can carry on with the second step, parameter estimation. For the DGD family, μ_0 is conveniently set to 0. However, in the particular case where μ_0 is known and $\mu_0 \neq 0$, then μ_0 must be subtracted from the data and the shifted data can be fitted. If μ_0 is unknown, our model must be extended by adding a third parameter for location. This will be discussed in the conclusion.

2.3 Parameter Estimation

2.3.1 Introduction

We will estimate the parameters through a minimum-distance method based on the characteristic function. There is an extensive literature involving the characteristic function in parameter estimation. For example, it is a widely used method with stable distributions. References include Paulson, Holcomb and Leitch (1975), Feuerverger and McDunnough (1981a), Csörgő (1987), Gürtler and Henze (2000) and Matsui and Takemura (2005a, 2005b). Moreover, Yu (2004) shows how techniques relying on the characteristic function are used in mixtures of normal distributions, in the variance gamma distribution, in stable ARMA processes, and in a diffusion model.

Traditionally, the maximum likelihood approach is widely favored due to its generality and asymptotic efficiency. However, the likelihood function is not always tractable or does not have a closed form expression, as is the case with stable laws. When this occurs, the characteristic function can have a closed form expression and may be more tractable. Yu (2004) justifies the use of the characteristic function by noting that it is the Fourier-Stietjes transform of the cumulative distribution function and hence, there is a one-to-one correspondence between them. As a consequence, the empirical characteristic function retains all the information in the sample. This suggests that estimation and inference via the empirical characteristic function should work as efficiently as the likelihood-based approaches. Feuerverger and McDunnough (1981a) showed that the asymptotic variance-covariance matrix of the parameters estimated using a minimum-distance method based on the characteristic function can be made arbitrarily close to the Cramér-Rao bound so that the method can attain arbitrarily high asymptotic efficiency. Moreover, the estimators obtained are consistent, robust and asymptotically normally distributed. Feuerverger and McDunnough (1981b) noted that the robustness properties for procedures associated with the empirical characteristic function are the result of a bounded influence curve for the estimators. For more details on the influence curve, see Hampel (1974). Next, we present the empirical characteristic function and rationalize using only its real part to estimate the parameters of the DGD family.

2.3.2 The Empirical Characteristic Function

Consider $Z_1, Z_2, \ldots, Z_n$ to be independent and identically distributed observations from the $DGD(\lambda, \theta)$. Let us define the empirical and theoretical characteristic functions at a specific point t_0 as $\phi_n(t_0)$ and $\phi(t_0)$ respectively, where

$$\phi_n(t_0) = \frac{1}{n}\sum_{j=1}^{n} e^{it_0 Z_j} = \frac{1}{n}\sum_{j=1}^{n}[\cos(t_0 Z_j) + i\sin(t_0 Z_j)]$$

and

$$\phi(t_0) = (\frac{1}{1 + t_0^2\lambda\theta})^{\frac{1}{\lambda}}.$$

Thus, $\phi(t_0)$ only has a real part and let us denote the real part of $\phi_n(t_0)$ as $\phi_n^{Re}(t_0)$, where

$$\phi_n^{Re}(t_0) = \frac{1}{n}\sum_{j=1}^{n}\cos(t_0 Z_j). \tag{2.1}$$

For any fixed t_0, $\phi_n(t_0)$ is an average of bounded independent and identically distributed random variables having mean $\phi(t_0)$ and finite variance. Therefore, it follows by the strong law of large numbers that $\phi_n(t_0)$ converges almost surely to $\phi(t_0)$. Furthermore, Feuerverger and Mureika (1977) proved, for fixed $T < \infty$, the convergence of

$$\sup_{|t|\leq T} |\phi_n(t) - \phi(t)| \to 0$$

almost surely as $n \to \infty$ and assert that $\phi_n^{Re}(t)$ will become uniformly close to $\phi(t)$ when the underlying distribution is symmetric. This implies that the imaginary part of $\phi_n(t)$, denoted by $\phi_n^{Im}(t)$, is approximately 0 for large n. Thus, any discrepancies observed between $\phi_n^{Im}(t)$ and 0 will be due to sampling error and consequently $\phi_n^{Im}(t)$ will not hold any information about the parameters λ and θ. Hence, since we are only fitting data that are symmetric around the origin, we will only consider the real parts of $\phi_n(t)$ and $\phi(t)$ to estimate the parameters as the imaginary parts will be uninformative.

2.3.3 Quadratic Distance

The method used is a form of nonlinear weighted least squares estimation. It is similar to the $k-L$ procedure introduced by Feuerverger and McDunnough (1981a) and it is a special distance within the class of quadratic distances introduced by Luong and Thompson (1987), where a unified theory for estimation and goodness-of-fit is developed. More precisely, the technique consists in choosing the parameters which minimize the quadratic distance between the real parts of the theoretical characteristic function and the empirical characteristic function.

Let us choose the points $t_1, t_2, \ldots, t_k > 0$ and let us define the vectors

$$\mathbf{Z}_n = [\phi_n^{Re}(t_1), \ldots, \phi_n^{Re}(t_k)]'$$
$$\mathbf{Z}(\lambda, \theta) = [\phi(t_1), \ldots, \phi(t_k)]'.$$

The quadratic distance estimator based on the characteristic function, denoted by $(\hat{\lambda}^*, \hat{\theta}^*)$, is defined as the value of (λ, θ) which minimizes the distance

$$d(\lambda, \theta) = [\mathbf{Z}_n - \mathbf{Z}(\lambda, \theta)]' \, \mathbf{Q}(\lambda, \theta) \, [\mathbf{Z}_n - \mathbf{Z}(\lambda, \theta)], \tag{2.2}$$

where $\mathbf{Q}(\lambda, \theta)$ is a positive definite matrix which may depend on (λ, θ). Luong and Thompson (1987) showed that an optimal choice of $\mathbf{Q}(\lambda, \theta)$ in the sense of minimizing the norm of the variance-covariance matrix of the estimated parameters is $\mathbf{Q}(\lambda, \theta) = \mathbf{\Sigma}^{-1}(\lambda, \theta)$, where $\mathbf{\Sigma}(\lambda, \theta)$ is the variance-covariance matrix of $\mathbf{Y}_n(\lambda, \theta) = \sqrt{n}[\mathbf{Z}_n - \mathbf{Z}(\lambda, \theta)]$.

Let $Y_n(t) = \sqrt{n}[\phi_n^{Re}(t) - \phi(t)]$, then $\mathbf{\Sigma}(\lambda, \theta) = (\sigma_{ij})$ is the $k \times k$ symmetric matrix with elements:

$$\sigma_{ij} = \mathrm{Cov}[Y_n(t_i), Y_n(t_j)] = \frac{1}{2}[\phi(t_i + t_j) + \phi(t_i - t_j)] - \phi(t_i)\phi(t_j).$$

This result follows because $E[\cos(tZ)] = \phi(t)$ and

$$E[\cos(tZ)\cos(sZ)] = E[\frac{1}{2}(\cos((t+s)Z) + \cos((t-s)Z))] = \frac{1}{2}[\phi(t+s) + \phi(t-s)].$$

Since minimization of $d(\lambda, \theta)$ involves the inverse of the matrix $\mathbf{\Sigma}(\lambda, \theta)$ which is dependent on the parameters, a simpler procedure would be to replace $\mathbf{\Sigma}(\lambda, \theta)$ by a consistent estimate $\hat{\mathbf{\Sigma}}$ and minimize

$$d'(\lambda, \theta) = [\mathbf{Z}_n - \mathbf{Z}(\lambda, \theta)]' \, \hat{\mathbf{\Sigma}}^{-1} \, [\mathbf{Z}_n - \mathbf{Z}(\lambda, \theta)]. \tag{2.3}$$

Let (λ_0, θ_0) be the true value of (λ, θ) and $\mathbf{\Sigma}(\lambda_0, \theta_0) = \mathbf{\Sigma}$, then, if $\hat{\mathbf{\Sigma}} \xrightarrow{\mathcal{P}} \mathbf{\Sigma}$ (i.e. $\hat{\mathbf{\Sigma}}$ is a consistent estimate of $\mathbf{\Sigma}$), Luong and Doray (2002, 2007) assert that minimization of (2.2) and (2.3) yields asymptotically equivalent estimators. For example, $\mathbf{\Sigma}_n^{Re}$ defined analogously to $\mathbf{\Sigma}$ in terms of $\phi_n^{Re}(t)$ is a consistent estimate of $\mathbf{\Sigma}$. More precisely, $\mathbf{\Sigma}_n^{Re} = (a_{ij})$ is the $k \times k$ matrix with elements:

$$a_{ij} = \frac{1}{2}[\phi_n^{Re}(t_i + t_j) + \phi_n^{Re}(t_i - t_j)] - \phi_n^{Re}(t_i)\phi_n^{Re}(t_j).$$

Luong and Doray (2002) suggested an iterative procedure to estimate $\mathbf{\Sigma}$. First, we obtain $(\tilde{\lambda}, \tilde{\theta})$ by choosing $\mathbf{Q}(\lambda, \theta) = \mathbf{I}$, the identity matrix. Despite the fact that $(\tilde{\lambda}, \tilde{\theta})$ is less efficient, it can be used to estimate $\mathbf{\Sigma}$, by letting $\hat{\mathbf{\Sigma}} = \mathbf{\Sigma}(\tilde{\lambda}, \tilde{\theta})$. We then can use $\hat{\mathbf{\Sigma}}$ to obtain the first iteration for $(\hat{\lambda}, \hat{\theta})$ and this procedure can be repeated with $\mathbf{\Sigma}$ reestimated at each step; $(\hat{\lambda}, \hat{\theta})$ is defined as the convergent vector value of the procedure. Next, we establish the asymptotic properties of the quadratic distance estimator.

2.3.4 Asymptotic Properties of the Quadratic Distance Estimator

From (2.1), we observe that $\phi_n^{Re}(t)$ is an average of bounded processes and it follows, by means of the multivariate Central Limit Theorem, that $\mathbf{Y}_n(\lambda_0, \theta_0) = \mathbf{Y}_n$ converges in distribution to a multivariate normal distribution with zero mean and covariance structure $\mathbf{\Sigma}$. Thus, we have

$$\mathbf{Y}_n = \sqrt{n}\,[\mathbf{Z}_n - \mathbf{Z}(\lambda_0, \theta_0)] \xrightarrow{\mathcal{D}} N(0, \mathbf{\Sigma}). \tag{2.4}$$

Let $(\hat{\lambda}^*, \hat{\theta}^*)$ be the estimator obtained by minimizing (2.2) with $\mathbf{Q}(\lambda, \theta) = \mathbf{\Sigma}^{-1}(\lambda, \theta)$. Under the conditions that $d(\lambda, \theta)$ attains its minimum at an interior point of $\mathbf{\Theta} = \{\lambda, \theta \in \mathbb{R}; \lambda \geq 0, \theta > 0\}$ and that $\mathbf{Z}(\lambda, \theta)$ and $\mathbf{Q}(\lambda, \theta)$ are differentiable, the estimator $(\hat{\lambda}^*, \hat{\theta}^*)$ may also be defined implicitly as a root of the 2-dimensional system of estimating equations

$$\frac{\partial}{\partial(\lambda, \theta)} \left\{ [\mathbf{Z}_n - \mathbf{Z}(\lambda, \theta)]' \, \mathbf{\Sigma}^{-1}(\lambda, \theta) \, [\mathbf{Z}_n - \mathbf{Z}(\lambda, \theta)] \right\} = 0.$$

Using lemmas (2.4.2) and (3.4.1) in Luong and Thompson (1987), we can conclude that:

(i) $(\hat{\lambda}^*, \hat{\theta}^*) \xrightarrow{\mathcal{P}} (\lambda_0, \theta_0)$, i.e. $(\hat{\lambda}^*, \hat{\theta}^*)$ is a consistent estimator of (λ_0, θ_0),

(ii) $(\hat{\lambda}^*, \hat{\theta}^*)$ satisfies $\dfrac{\partial \mathbf{Z}'(\hat{\lambda}^*, \hat{\theta}^*)}{\partial(\lambda, \theta)} \left\{ \mathbf{\Sigma}^{-1}(\hat{\lambda}^*, \hat{\theta}^*) \mathbf{Y}_n(\hat{\lambda}^*, \hat{\theta}^*) \right\} + o_p(1) = 0$,

(iii) $\sqrt{n}[(\hat{\lambda}^*, \hat{\theta}^*) - (\lambda_0, \theta_0)] = (\mathbf{S}'\mathbf{\Sigma}^{-1}\mathbf{S})^{-1}\mathbf{S}'\mathbf{\Sigma}^{-1}\mathbf{Y}_n + o_p(1)$,

(iv) $\mathbf{Y}_n(\hat{\lambda}^*, \hat{\theta}^*) = \mathbf{Y}_n - \{\mathbf{S} + o_p(1)\}\sqrt{n}[(\hat{\lambda}^*, \hat{\theta}^*) - (\lambda_0, \theta_0)]$,

(v) $\sqrt{n}[(\hat{\lambda}^*, \hat{\theta}^*) - (\lambda_0, \theta_0)] \xrightarrow{\mathcal{D}} N(0, (\mathbf{S}'\mathbf{\Sigma}^{-1}\mathbf{S})^{-1})$.

The symbol $o_p(1)$ denotes an expression converging to 0 in probability (i.e. $o_p(1) \xrightarrow{\mathcal{P}} 0$), $\mathbf{S}$ is a matrix of dimension $k \times 2$ defined as

$$\mathbf{S} = \begin{pmatrix} \frac{\partial Z_1(\lambda,\theta)}{\partial \lambda} & \frac{\partial Z_1(\lambda,\theta)}{\partial \theta} \\ \vdots & \vdots \\ \frac{\partial Z_k(\lambda,\theta)}{\partial \lambda} & \frac{\partial Z_k \lambda,\theta)}{\partial \theta} \end{pmatrix} = \begin{pmatrix} \frac{\partial \phi(t_1)}{\partial \lambda} & \frac{\partial \phi(t_1)}{\partial \theta} \\ \vdots & \vdots \\ \frac{\partial \phi(t_k)}{\partial \lambda} & \frac{\partial \phi(t_k)}{\partial \theta} \end{pmatrix},$$

where

$$\frac{\partial\phi(t)}{\partial\lambda} = \frac{(1+\lambda\theta t^2)\ln(1+\lambda\theta t^2) - \lambda\theta t^2}{\lambda^2(1+\lambda\theta t^2)^{1+\frac{1}{\lambda}}} \quad \text{and}$$

$$\frac{\partial\phi(t)}{\partial\theta} = -\frac{t^2}{(1+\lambda\theta t^2)^{1+\frac{1}{\lambda}}} \quad ,$$

all quantities being evaluated at (λ_0, θ_0). Thus, the estimator $(\hat{\lambda}^*, \hat{\theta}^*)$ is consistent and asymptotically normally distributed with variance-covariance matrix $(\mathbf{S}'\mathbf{\Sigma}^{-1}\mathbf{S})^{-1}$. The same results hold for $(\hat{\lambda}, \hat{\theta})$, the estimator obtained by minimizing (2.3).

The choice of points $t_1, \ldots, t_k$ affects $(\mathbf{S}'\mathbf{\Sigma}^{-1}\mathbf{S})^{-1}$ and thus we must choose them with care. Feuerverger and McDunnough (1981a) showed that by using a sufficiently extensive grid $\{t_i\}_{i=1}^k$, $(\mathbf{S}'\mathbf{\Sigma}^{-1}\mathbf{S})^{-1}$ can be made arbitrarily close to the Cramér-Rao bound. However, by choosing more points, the $k \times k$ matrix $\mathbf{\Sigma}$ can become near singular and computational problems may arise. Clear guidelines on the choices of finite points $t_1, \ldots, t_k$ are still not available in the literature. For our simulation study, we will consider sets of points having the general form:

$$\{t_i\}_{i=1}^k = \left\{\frac{Mi}{k}\right\}_{i=1}^k = \left\{\frac{M}{k}, \frac{2M}{k}, \ldots, M\right\}, \tag{2.5}$$

where M is an arbitrary number. More precisely, we will use values of $M = 0.01, 0.1, 1, 2$ and 3 and examine the effect on our estimation when $k = 5, 10, 20$ or 30. We will determine the choices of points for which the variances of the estimated parameters are a minimum.

Note that the number of entries in the matrix $\mathbf{\Sigma}$ is four times smaller than it could have been, have we considered the real and imaginary parts of $\phi_n(t)$. With the imaginary part, $\mathbf{Z}_n$ is a vector with $2k$ elements and the corresponding covariance matrix would be of dimension $2k \times 2k$. Hence, we may suggest that the quadratic distance method based on the characteristic function will be computationally more tractable when considering families of symmetric distribution functions. Next, goodness-of-fit tests for the simple and composite hypotheses are presented.

3 HYPOTHESIS TESTING

3.1 Goodness-of-fit

3.1.1 Introduction

Since we built statistics based on a minimum distance between empirical and theoretical parts, it is natural to use them for testing goodness-of-fit. Luong and Thompson (1987) developed a unified theory for estimation and goodness-of-fit when quadratic distances are employed. They showed that test statistics for goodness-of-fit follow a chi-square distribution asymptotically. Their results generalize the tests based on the characteristic function proposed by Koutrouvelis (1980) and Koutrouvelis and Kellermeir (1981). We now present the test statistics for the simple and composite hypotheses respectively. The following theorem appearing in Luong and Doray (2002) is needed; its proof can be found in Rao (1973).

Theorem 1. *Suppose that the random vector $\mathbf{Y}_n$ of dimension k is $N(0, \mathbf{\Sigma})$ and $\mathbf{Q}$ is any $k \times k$ symmetric positive semi-definite matrix; then the quadratic form $\mathbf{Y}_n'\mathbf{Q}\mathbf{Y}_n$ is chi-square distributed with ν degrees of freedom if $\mathbf{\Sigma Q}$ is idempotent and* $\operatorname{trace}(\mathbf{\Sigma Q}) = \nu$. *(The same result holds asymptotically if $\mathbf{Q}$ is replaced by a consistent estimate $\hat{\mathbf{Q}}$ and $\mathbf{Y}_n \xrightarrow{\mathcal{D}} N(0, \mathbf{\Sigma})$).*

3.1.2 Simple hypothesis

To test the simple hypothesis $H_0 : Z_1, \ldots, Z_n$ come from a specified DGD distribution with parameters (λ_0, θ_0), the following test statistic can be used:

$$\begin{aligned} nd(\lambda_0, \theta_0) &= n[\mathbf{Z}_n - \mathbf{Z}(\lambda_0, \theta_0)]' \, \mathbf{\Sigma}^{-1} \, [\mathbf{Z}_n - \mathbf{Z}(\lambda_0, \theta_0)] \\ &= \mathbf{Y}_n' \, \mathbf{\Sigma}^{-1} \, \mathbf{Y}_n. \end{aligned}$$

It follows from (2.4) and Theorem 1, that $nd(\lambda_0, \theta_0) \xrightarrow{\mathcal{D}} \chi^2_\nu$.

$$\nu = \text{trace}(\mathbf{\Sigma}\mathbf{\Sigma}^{-1}) = \text{trace}(\mathbf{I}_k) = k,$$

where $\mathbf{I}_k$ is the $k \times k$ identity matrix. Thus, the test statistic follows a limiting chi-square distribution with $\nu = k$ degrees of freedom under H_0. To test the hypothesis H_0 at significance level α, compute the value of the test statistic $nd(\lambda_0, \theta_0)$ from the sample. The null hypothesis H_0 should be rejected if $nd(\lambda_0, \theta_0) > \chi^2_{k,1-\alpha}$, where $\chi^2_{k,1-\alpha}$ is the $100(1-\alpha)^{\text{th}}$ quantile of a χ^2 distribution with k degrees of freedom.

3.1.3 Composite hypothesis

To test the composite hypothesis $H_0 : Z_1, \ldots, Z_n$ come from a DGD distribution where the values of the parameters are not specified, we first calculate the quadratic distance estimator $(\hat{\lambda}^*, \hat{\theta}^*)$ by minimizing (2.2) with $\mathbf{Q}(\lambda, \theta) = \mathbf{\Sigma}^{-1}(\lambda, \theta)$ or an equivalent expression where $\mathbf{\Sigma}(\lambda, \theta)$ is replaced by a consistent estimate. Luong and Thompson (1987) showed that the test statistic

$$\begin{aligned} nd(\hat{\lambda}^*, \hat{\theta}^*) &= n[\mathbf{Z}_n - \mathbf{Z}(\hat{\lambda}^*, \hat{\theta}^*)]' \, \mathbf{\Sigma}^{-1}(\hat{\lambda}^*, \hat{\theta}^*) \, [\mathbf{Z}_n - \mathbf{Z}(\hat{\lambda}^*, \hat{\theta}^*)] \\ &= \mathbf{Y}_n'(\hat{\lambda}^*, \hat{\theta}^*) \, \mathbf{\Sigma}^{-1}(\hat{\lambda}^*, \hat{\theta}^*) \, \mathbf{Y}_n(\hat{\lambda}^*, \hat{\theta}^*) \end{aligned}$$

follows an asymptotic chi-square distribution with $\nu = k-2$ degrees of freedom under H_0. Again, $\mathbf{\Sigma}(\hat{\lambda}^*, \hat{\theta}^*)$ can be replaced by a consistent estimate $\hat{\mathbf{\Sigma}}$.

Analogous to the case for the simple null hypothesis, a significance level α test can be performed to test H_0.

3.2 The parameter λ

In Section 2.3.4, we showed that the estimator $(\hat{\lambda}, \hat{\theta})$ is asymptotically normally distributed with variance-covariance matrix $(\mathbf{S}'\mathbf{\Sigma^{-1}S})^{-1}$. Thus, we can easily construct individual and joint $(1-\alpha)\%$ confidence intervals for the parameters λ and θ. Joint confidence intervals can be created via the Bonferroni procedure which provides a lower bound to the $(1-\alpha)\%$ family confidence interval. Since we do not know (λ_0, θ_0), $(\mathbf{S}'\mathbf{\Sigma^{-1}S})^{-1}$ must be replaced by a consistent estimate, $(\hat{\mathbf{S}}'\hat{\mathbf{\Sigma}}^{-1}\hat{\mathbf{S}})^{-1}$ where all quantities are evaluated at $(\hat{\lambda}, \hat{\theta})$.

However, of more practical interest is testing for the parameter λ. In Section 1, we saw that particular values of λ define specific distributions within the DGD distribution family. For example, when $\lambda = 0$ or $\lambda = 1$, we obtain the normal and the Laplace distributions respectively. This suggests using the parameter λ to test distributional assumptions. A test of normality of the data can be constructed by testing

$$H_0 : \lambda = 0 \quad \text{versus} \quad H_a : \lambda > 0. \tag{3.1}$$

In Section 1, we noted that the kurtosis of a $DGD(\lambda, \theta)$ random variable Z is $(3 + 3\lambda)$. Kurtosis is defined as the normalized fourth central moment and it is a measure of peakedness and of heaviness of the tails. Let β_2 be the kurtosis of Z, we have

$$\beta_2 = \frac{\mathrm{E}(Z^4)}{(\mathrm{Var}(Z))^2} = \frac{(\theta^2)(12)(1+\lambda)}{(2\theta)^2} = 3 + 3\lambda.$$

Thus, if we have a sample from Z, $\hat{\beta}_2 = (3 + 3\hat{\lambda})$ is a consistent estimate of the population kurtosis, β_2. Moreover, since β_2 is a simple linear function of λ, the hypotheses identified in (3.1) are equivalent to

$$H_0 : \beta_2 = 3 \quad \text{versus} \quad H_a : \beta_2 > 3.$$

This implies that in (3.1) we are testing the normal distribution against symmetric distributions with heavier tails. Thus, it would be interesting to compare the power of this test to a normality test based on the sample kurtosis. D'Agostino and Pearson (1973) describe such a test. Moreover, when the alternative is the Laplace distribution, the power of the test can be compared to the likelihood ratio test. Kotz, Kozubowski and Podgórski (2001) assert that the likelihood ratio test is the most powerful scale invariant test for testing the normal against the Laplace when the center of symmetry is known. In the following section, we provide a simulation study for estimating parameters and testing hypotheses with the methods presented previously.

4 SIMULATION STUDY

4.1 General estimation

We now provide simulation results for estimating the parameters λ and θ using the techniques developed in Section 2. While the expressions for quadratic distance estimators may seem complex, they are relatively simple to implement using a computer software with built-in statistical functions such as Maple, MATHEMATICA, SAS or S-plus. The quadratic distance estimator can be computed numerically using a nonlinear least squares method. All of our simulations were completed using Maple 11.0.

We first generated 3 samples from a $DGD(\lambda = 1, \theta = 1)$ random variable of sizes 100, 500 and 1000. For each sample, we estimated the parameters using the method of moments (MOM), ordinary least squares (OLS) (i.e. using (2.2) with $\mathbf{Q}(\lambda, \theta) = \mathbf{I}$, the identity matrix) and weighted least squares (WLS) (i.e. using (2.2) with an appropriate choice of $\mathbf{Q}(\lambda, \theta)$). OLS and WLS methods were implemented using 20 different sets of points, $\{t_i\}_{i=1}^{k}$, in order to determine which are the best choices. All the sets have the general form defined by (2.5). As stated previously in Section 2.3.4, values of $M = 0.01, 0.1, 1, 2$ and 3 and values of $k = 5, 10, 20$ and 30 were used to define $\{t_i\}_{i=1}^{k}$.

Tables 1, 2, and 3 summarize the pertinent results for sample sizes of 100, 500 and 1000 respectively. Each table provides estimated values of λ and θ using the MOM, OLS and WLS. The WLS estimates were obtained using the iterative procedure to estimate $\mathbf{\Sigma}$ presented in Section 2.3.3. The asymptotic standard deviations of the estimated parameters are given for the WLS estimates under the columns 'Stdev'. Results for values of $M = 0.01$ and 0.1 are not presented as the WLS method rarely found an improved estimate over the OLS method. Consequently, we do not recommend using those choices of M. The other choices of M all yielded good estimates but we suggest using $M = 3$ as this choice generated the lowest asymptotic standard deviations of the estimated parameters. Moreover, increasing the value of k (i.e. increasing the number of points in the sets) generally improved estimates. However, the improvement when using $k = 30$ over $k = 10$ or 20 is not substantial and perhaps not worthy of the additional computation time. Thus, we suggest using values of $k = 10$ or 20 for a fast and efficient estimation.

From the tables, we can observe that the asymptotic standard deviations of $\hat{\theta}$'s are roughly half of those for $\hat{\lambda}$'s. Thus, we can generally estimate θ with more precision than λ. For a sample size of 100, the estimation of λ may not be very accurate as the asymptotic standard deviations were in the neighborhood of 0.42. For sample sizes of 500 and 1000, those values diminished and were in neighborhoods of 0.18 and 0.13 respectively. We also performed WLS estimation with choices of $\mathbf{Q}(\lambda, \theta) = \mathbf{\Sigma}_n^{Re}$ and $\mathbf{\Sigma}$. With $\mathbf{\Sigma}_n^{Re}$, we obtained poor estimates and often they did not converge to a solution. The choice of $\mathbf{\Sigma}$ produced comparable estimates to the ones obtained in the tables under the WLS columns. However, the choice of $\mathbf{\Sigma}$ is not a viable selection in practice as the true parameters (λ_0, θ_0) are always unknown. Next, we present simulation results for goodness-of-fit.

Table 1: Estimated values of the parameters using a sample size of 100

		λ				θ			
M	k	MOM	OLS	WLS	Stdev	MOM	OLS	WLS	Stdev
1	5	0.5134	0.9107	0.9771	0.5140	0.8894	0.9180	0.9027	0.1977
2	5	0.5134	1.0748	0.9292	0.4164	0.8894	0.9474	0.8974	0.1928
3	5	0.5134	1.3867	1.3082	0.4755	0.8894	1.0779	0.9797	0.2445
1	10	0.5134	0.8908	1.0158	0.4955	0.8894	0.9148	0.9059	0.2001
2	10	0.5134	1.0895	0.9657	0.4157	0.8894	0.9528	0.9015	0.1950
3	10	0.5134	1.3543	1.1752	0.4248	0.8894	1.0687	0.9267	0.2106
1	20	0.5134	0.8784	0.9897	0.4623	0.8894	0.9128	0.9035	0.1976
2	20	0.5134	1.0968	1.0324	0.4220	0.8894	0.9553	0.9091	0.2000
3	20	0.5134	1.3274	1.1326	0.4139	0.8894	1.0561	0.9210	0.2070
1	30	0.5134	0.8738	0.8738	0.4181	0.8894	0.9120	0.9120	0.1926
2	30	0.5134	1.0993	1.0782	0.4268	0.8894	0.9562	0.9146	0.2035
3	30	0.5134	1.3179	1.1041	0.4070	0.8894	1.0516	0.9171	0.2046

Table 2: Estimated values of the parameters using a sample size of 500

		λ				θ			
M	k	MOM	OLS	WLS	Stdev	MOM	OLS	WLS	Stdev
1	5	0.7789	1.1542	1.1050	0.2392	1.1698	1.2122	1.1834	0.1199
2	5	0.7789	1.1285	1.1212	0.2002	1.1698	1.2023	1.1928	0.1213
3	5	0.7789	1.0414	1.1024	0.1917	1.1698	1.1585	1.2068	0.1314
1	10	0.7789	1.1463	1.0735	0.2231	1.1698	1.2102	1.1799	0.1182
2	10	0.7789	1.1300	1.0793	0.1917	1.1698	1.2033	1.1804	0.1175
3	10	0.7789	1.0303	1.0735	0.1789	1.1698	1.1507	1.1801	0.1169
1	20	0.7789	1.1413	1.0646	0.2120	1.1698	1.2090	1.1788	0.1176
2	20	0.7789	1.1298	1.0483	0.1863	1.1698	1.2032	1.1757	0.1160
3	20	0.7789	1.0294	1.0734	0.1777	1.1698	1.1502	1.1800	0.1168
1	30	0.7789	1.1395	1.0746	0.2083	1.1698	1.2085	1.1801	0.1179
2	30	0.7789	1.1297	1.0353	0.1836	1.1698	1.2032	1.1737	0.1154
3	30	0.7789	1.0294	1.0730	0.1770	1.1698	1.1501	1.1799	0.1167

Table 3: Estimated values of the parameters using a sample size of 1000

		λ				θ			
M	k	MOM	OLS	WLS	Stdev	MOM	OLS	WLS	Stdev
1	5	0.7178	0.9273	0.9404	0.1557	1.0091	1.0319	1.0208	0.0699
2	5	0.7178	1.3283	1.1599	0.1460	1.0091	1.1642	1.0477	0.0759
3	5	0.7178	1.1622	1.0770	0.0732	1.0091	1.0787	1.0465	0.0781
1	10	0.7178	0.9183	0.9975	0.1528	1.0091	1.0301	1.0263	0.0713
2	10	0.7178	1.3129	1.1217	0.1401	1.0091	1.1583	1.0394	0.0740
3	10	0.7178	1.1977	1.0389	0.1259	1.0091	1.0993	1.0279	0.0714
1	20	0.7178	0.9127	1.0696	0.1520	1.0091	1.0290	1.0338	0.0731
2	20	0.7178	1.3017	1.1027	0.1366	1.0091	1.1536	1.0369	0.0734
3	20	0.7178	1.2134	1.0454	0.1253	1.0091	1.1075	1.0288	0.0715
1	30	0.7178	0.9106	1.1073	0.1516	1.0091	1.0286	1.0380	0.0740
2	30	0.7178	1.2976	1.0949	0.1348	1.0091	1.1519	1.0358	0.0732
3	30	0.7178	1.2188	1.0588	0.1254	1.0091	1.1103	1.0308	0.0719

4.2 Goodness-of-fit testing

We now perform goodness-of-fit testing for the simple hypothesis as presented in Section 3.1.2. First, we wish to determine if the test has a correct size when using critical values from the chi-square distribution for sample sizes of $n = 100, 500$ and 1000. For each sample size n, we generated 5000 samples from a $DGD(\lambda = 1, \theta = 1)$ random variable and calculated the test statistics $nd(\lambda_0 = 1, \theta_0 = 1)$. We repeated the procedure for samples from a $DGD(\lambda = 2, \theta = 1)$. We were thus able to obtain simulated critical values for a level α test by taking the $100(1-\alpha)^{\text{th}}$ quantiles from the empirical distributions of the test statistics. All of the test statistics were obtained using the set of points defined by (2.5) with values of $M = 3$ and $k = 10$. We present our results in Tables 4 and 5.

Table 4: Actual sizes of the test when using $\chi^2_{10,1-\alpha}$ based on 5000 simulation runs

		Actual sizes of the test		
α	(λ_0, θ_0)	$n = 100$	$n = 500$	$n = 1000$
0.100	(1,1)	0.1250	0.1242	0.1168
	(2,1)	0.1420	0.1350	0.1194
0.050	(1,1)	0.0966	0.0834	0.0764
	(2,1)	0.1138	0.0898	0.0738
0.025	(1,1)	0.0798	0.0594	0.0544
	(2,1)	0.0930	0.0628	0.0448
0.010	(1,1)	0.0658	0.0406	0.0310
	(2,1)	0.0756	0.0412	0.0264

The results of Table 4 indicate that the goodness-of-fit test has an incorrect size that is severe enough to warrant a recommendation that the test should

Table 5: Critical values obtained for sample sizes of 100, 500 and 1000

		Critical values			
α	(λ_0, θ_0)	$n = 100$	$n = 500$	$n = 1000$	$\chi^2_{10,1-\alpha}$
0.100	(1,1)	17.9453	17.1660	16.8809	15.9872
	(2,1)	19.6814	17.5147	16.9050	15.9872
0.050	(1,1)	27.2592	21.5835	20.9238	18.3070
	(2,1)	27.6104	21.9962	20.0182	18.3070
0.025	(1,1)	37.0803	27.4926	24.4911	20.4832
	(2,1)	37.6163	25.9949	23.5219	20.4832
0.010	(1,1)	68.5516	39.5960	29.3234	23.2093
	(2,1)	58.1972	32.5478	28.3944	23.2093

not be used without appropriately sized critical values. From Table 5, we remark that the test statistic $nd(\lambda_0, \theta_0)$ converges very slowly to a chi-square random variable. Even for sample sizes of 1000, the approximation is not satisfactory. The real distribution of the test statistic will generally have a heavier right tail than the chi-square distribution, even for large sample sizes, and thus the test will always be oversized when using the critical value $\chi^2_{k,1-\alpha}$.

We will now assess the power of the goodness-of-fit test for the simple hypothesis $H_0 : (\lambda_0 = 1, \theta_0 = 1)$ against alternatives $H_a : (\lambda_a, \theta_a = 1)$, where $\lambda_a = 0, 0.5, 1, 1.5$ and 2. A level $\alpha = 0.05$ and sample sizes of $100, 500$ and 1000 were employed. We determined the power of the test by generating 5000 samples for each of the alternatives considered. Appropriately sized critical values (CV) were calculated by taking the average of the two critical values obtained in Table 5 for each sample size. Results are shown in Table 6.

The goodness-of-fit test performed poorly in rejecting the selected alternatives for a sample size of 100. When $n = 500$, the test did very well for

Table 6: Power of the test ($\alpha = 0.05$) based on 5000 simulation runs on selected alternatives

		Alternatives (λ_a, θ_a)				
n	CV	$(0, 1)$	$(0.5, 1)$	$(1, 1)$	$(1.5, 1)$	$(2, 1)$
100	27.4348	0.0626	0.0164	0.0518	0.1070	0.1852
500	21.7899	0.9986	0.2692	0.0516	0.3656	0.9394
1000	20.4710	1.0000	0.8000	0.0572	0.7524	1.0000

alternatives of λ_a one unity away of $\lambda_0 = 1$ but not so well when λ_a was half a unity away. For a large sample size of 1000, the test was powerful for all alternatives considered. For sample sizes of 500 and 1000, the recorded powers for alternatives $(0, 1)$ and $(2, 1)$ were close to or equal to 100%. This suggests that for a large enough sample size the test is well suited for discriminating between the fits of normal, Laplace and heavier tailed symmetric distributions. Moreover, by using adjusted critical values instead of $\chi^2_{k,1-\alpha}$, the tests had an adequate size. The discrepancies between the actual sizes and $\alpha = 0.05$ are due to the precision of the simulated critical values and to the large variability of the test statistic.

5 Conclusion

We introduced the double gamma difference family, which is a family of leptokurtic symmetric distributions. The Laplace, the sums of Laplace and the normal distributions all arise as special cases of this family. While there is in general no closed form expression for the density function, the characteristic function is simple to work with. Parameters can be estimated through a minimum quadratic distance method based on the characteristic function. The estimators obtained were shown to be consistent, robust and asymptotically normally distributed. Goodness-of-fit tests for the simple and composite hypotheses were presented and the test statistics shown to follow a chi-square distribution asymptotically. Moreoever, we suggested employing the parameter λ to test for distributional assumptions. Simulations revealed that large sample sizes are required to get a reasonable amount of precision for estimating the parameters. Also, the goodness-of-fit tests must be carried out with appropriate simulated critical values for the tests to have a correct size because the convergence to the chi-square distribution is slow.

The family can be extended by adding a third parameter for location μ. The characteristic function $\phi^*(t)$ would then both have real and imaginary parts, where

$$\phi^*(t) = \begin{cases} e^{it\mu} \cdot \left(\frac{1}{1+t^2\lambda\theta}\right)^{\frac{1}{\lambda}} & \text{for } \lambda, \theta > 0, \\ e^{it\mu - t^2\theta} & \text{for } \lambda = 0, \theta > 0. \end{cases}$$

Parameter estimation could still be achieved through a minimum distance method based on the characteristic function. However, both real and imaginary parts would have to be taken into account. For more details on the minimum distance method when the real and imaginary parts are involved, see Feuerverger and McDunnough (1981b). Before fitting data to this family, it is still necessary to verify symmetry. For testing symmetry around the unknown median μ we suggest using the triples test introduced by Randles, Fligner, Policello and Wolfe (1980).

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the financial support of the Natural Sciences and Engineering Research Council of Canada.

References

[1] CHENG W.H. AND BALAKRISHNAN N. (2004). A Modified Sign Test for Symmetry. *Comm. Statist. Simulation Comput.*, **33**, 703-709.

[2] CSÖRGŐ S. (1987). Testing for stability. In *Goodness-of-fit (Debrecen, 1984)*, volume 45 of *Colloq. Math. Soc. János Bolyai.* North-Holland, Amsterdam, pp. 101-132.

[3] D'AGOSTINO R. AND PEARSON E. S. (1973). Tests for departures from normality. Empirical results for the distribution of b_2 and $\sqrt{b_1}$. *Biometrika*, **60**, 613-622.

[4] FEUERVERGER A. AND MCDUNNOUGH P. (1981a). On the efficiency of empirical characteristic function procedures. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **43**, 20-27.

[5] FEUERVERGER A. AND MCDUNNOUGH P. (1981b). On some Fourier methods for inference. *J. Amer. Statist. Assoc.*, **76**, 379-387.

[6] FEUERVERGER A. AND MUREIKA R.A. (1977). The empirical characteristic function and its applications. *Ann. Statist.*, **1**, 88-97.

[7] GÜRTLER N. AND HENZE N. (2000). Goodness-of-fit tests for the Cauchy distribution based on the empirical characteristic function. *Ann. Inst. Statist. Math.*, **52**, 267-286.

[8] HAMPEL F.R. (1974). The influence curve and its role in robust estimation. *J. Amer. Statist. Assoc.*, **69**, 383-393.

[9] KOTZ S. AND KOZUBOWSKI T.J. AND PODGÓRSKI K. (2001). *The Laplace distribution and generalizations.* Birkhäuser, Boston.

[10] KOUTROUVELIS I.A. (1980). A goodness-of-fit test of simple hypotheses based on the empirical characteristic function. *Biometrika*, **67**, 238-240.

[11] KOUTROUVELIS I.A. AND KELLERMEIER J. (1981). A goodness-of-fit test based on the empirical characteristic function when parameters must be estimated. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **43**, 173-176.

[12] LEHMANN E. (1975). *Nonparametrics: Statistical methods based on ranks.* McGraw-Hill, New York.

[13] LUKACS E. (1970). *Characteristic Functions.* Hafner Publishing Co., New York. Second edition, revised and enlarged.

[14] LUONG A. AND THOMPSON M.E. (1987). Minimum-distance methods based on quadratic distances for transforms. *Canad. J. Statist.*, **15**, 239-251.

[15] LUONG A. AND DORAY L.G. (2002). General quadratic distance methods for discrete distributions definable recursively. *Insurance Math. Econom.*, **30**, 255-267.

[16] LUONG A. AND DORAY L.G. (2008). Inference for the positive stable laws based on a special quadratic distance. *Stat. Methodol.*, accepted.

[17] MATSUI M. AND TAKEMURA A. (2005a). Empirical characteristic function approach to goodness-of-fit tests for the Cauchy distribution with parameters estimated by MLE or EISE. *Ann. Inst. Statist. Math.*, **57**, 183-199.

[18] MATSUI M. AND TAKEMURA A. (2005b). Goodness-of-fit tests for symmetric stable distributions - empirical characteristic function approach. CIRJE F-Series CIRJE-F-384, CIRJE, Faculty of Economics, Univer-

sity of Tokyo. available at http://ideas.repec.org/p/tky/fseres/2005cf384.html.

[19] McWilliams T. (1990). A distribution-free test for symmetry based on a runs statistic. *J. Amer. Statist. Assoc.*, **85**, 1130-1133.

[20] Modarres R. and Gastwirth J. (1996). A modified runs test for symmetry. *Statist. Probab. Lett.*, **31**, 107-112.

[21] Modarres R. and Gastwirth J. (1998). Hybrid test for the hypothesis of symmetry. *J. Appl. Stat.*, **25**, 777-783.

[22] Paulson A.S. and Halcomb E.W. and Leitch R.A. (1975). The estimation of the parameters of the stable laws. *Biometrika*, **62**, 163-170.

[23] Randles R.H. and Wolfe D.A. (1979). *Introduction to the Theory of Nonparametric Statistics.* Wiley, New York.

[24] Randles R. H., Fligner M. A., Policello II G. E. and Wolfe D. A. (1980). An asymptotically distribution-free test for symmetry versus asymmetry. *J. Amer. Statist. Assoc.*, **75**, 168-172.

[25] Rao C.R. (1973). *Linear Statistical Inference and its Applications.* Wiley, New York.

[26] Tajuddin I. (1994). Distribution-free test for symmetry based on Wilcoxon two-sample test. *J. Appl. Stat.*, **21**, 409-416.

[27] Thas O. and Rayner J.C.W. and Best D.J. (2005). Tests for symmetry based on the one-sample wilcoxon signed rank statistic. *Comm. Statist. Simulation Comput.*, **34**, 957-973.

[28] Yu J. (2004). Empirical characteristic function estimation and its applications. *Econometric Rev.*, **23**, 93-123.

CONCLUSION

Dans ce mémoire, nous avons introduit une famille de distributions symétriques et leptocurtiques représentée par la différence de deux variables aléatoires gamma. Nous avons montré que les distributions Laplace et normale sont des cas particuliers de cette famille.

Nous avons également présenté une méthode permettant d'ajuster un jeu de données à cette famille. La méthode est composée de deux étapes. La première étape consistait à vérifier la compatibilité de l'échantillon avec la nouvelle famille par l'entremise d'un test de symétrie. Nous avons suggéré d'utiliser le test hybride dans le cas où la vraie médiane de l'échantillon est connue et le test des triplets dans le cas où elle est inconnue. La deuxième étape englobait l'estimation des paramètres de la nouvelle famille. Puisqu'il n'y avait pas de forme fermée pour les fonctions de densité et de répartition, nous avons proposé d'estimer les paramètres en minimisant la distance quadratique entre les parties réelles des fonctions caractéristiques empiriques et théoriques. Nous avons montré que l'estimateur qui en découle est convergent, robuste et asymptotiquement normal.

De plus, nous avons construit des statistiques pour des tests d'ajustement à partir de la distance quadratique afin de pouvoir valider si la nouvelle famille est un modèle adéquat pour un échantillon donnée. Nous avons montré que ces statistiques suivent une loi asymptotique khi-deux.

L'étude de simulation que nous avons menée a démontré que la taille de l'échantillon doit être grande afin d'obtenir un degré de précision raisonnable lors de l'estimation des paramètres. Aussi, nous avons réalisé que la convergence des statistiques des tests d'ajustement vers la distribution asymptotique est très lente. Nous recommandons donc de déterminer les valeurs critiques des tests par l'entremise de simulations afin que les niveaux des tests soient adéquats.

Nous notons finalement que l'estimateur basé sur la distance quadratique que nous avons présenté dans les chapitres 3 et 4 était calculé à partir d'un espace discret de points. Cependant, nous avons vu dans le chapitre 1 que la fonction caractéristique est définie pour tous les nombres réels. Donc, de manière analogue nous aurions pu trouver l'estimateur qui minimise l'intégrale

$$\int_{-\infty}^{\infty} [\phi_n^{Re}(t) - \phi(t)]^2 g(t) dt, \tag{4.0.1}$$

où la fonction $g(t)$ est une fonction de poids et les fonctions $\phi_n^{Re}(t)$ et $\phi(t)$ ont été définies dans la section 2.3.2 de l'article du chapitre 4. Yu (2004) démontre qu'en théorie si la fonction de poids $g(t)$ est choisie judicieusement, l'estimateur qui minimise l'intégrale précédente aura les mêmes propriétés asymptotiques que l'estimateur du maximum de vraisemblance. Cependant, cette fonction de poids n'a pas toujours de forme fermée ou est inconnue ce qui rend son application impossible dans certaines situations. En pratique, une fonction de poids exponentielle est souvent utilisée. Il serait intéressant de comparer les estimateurs provenant de la méthode qui minimise l'intégrale (4.0.1) et de celle que nous avons présentée dans le chapitre 4.

BIBLIOGRAPHIE

[1] Cheng W.H. et Balakrishnan N. (2004). A Modified Sign Test for Symmetry. *Comm. Statist. Simulation Comput.*, **33**, 703-709.

[2] Csörgő S. (1987). Testing for stability. In *Goodness-of-fit (Debrecen, 1984)*, volume 45 de *Colloq. Math. Soc. János Bolyai*. North-Holland, Amsterdam, pp. 101-132.

[3] D'Agostino R. et Pearson E. S. (1973). Tests for departures from normality. Empirical results for the distribution of b_2 and $\sqrt{b_1}$. *Biometrika*, **60**, 613-622.

[4] Feuerverger A. et McDunnough P. (1981a). On the efficiency of empirical characteristic function procedures. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **43**, 20-27.

[5] Feuerverger A. et McDunnough P. (1981b). On some Fourier methods for inference. *J. Amer. Statist. Assoc.*, **76**, 379-387.

[6] Feuerverger A. et Mureika R.A. (1977). The empirical characteristic function and its applications. *Ann. Statist.*, **1**, 88-97.

[7] Grimmett G. R. et Stirzaker D. R. (2001). *Probability and random processes.* Oxford University Press, New York, troisième édition.

[8] Groparu-Cojocaru I. (2007). *Quadratic distance methods applied to generalized normal Laplace distribution.* Mémoire de maîtrise, Département de mathématiques et de statistique, Université de Montréal, Montréal.

[9] Gupta M.K. (1967). An asymptotically nonparametric test of symmetry. *Ann. Math. Stat.*, **38**, 849-866.

[10] Gürtler N. et Henze N. (2000). Goodness-of-fit tests for the Cauchy distribution based on the empirical characteristic function. *Ann. Inst. Statist. Math.*, **52**, 267-286.

[11] HAMPEL F.R. (1974). The influence curve and its role in robust estimation. *J. Amer. Statist. Assoc.*, **69**, 383-393.

[12] HOLLANDER R. ET WOLFE D. (1973). *Nonparametric Statistical Methods.* Wiley, New York.

[13] KLUGMAN S. A., PANJER H. H. ET WILLMOT G. E. (2004). *Loss models.* Wiley Series in Probability and Statistics. Wiley-Interscience, Hoboken, NJ, deuxième édition. From data to decisions.

[14] KOTZ S., KOZUBOWSKI T.J. ET PODGÓRSKI K. (2001). *The Laplace distribution and generalizations.* Birkhäuser, Boston.

[15] KOUTROUVELIS I.A. (1980). A goodness-of-fit test of simple hypotheses based on the empirical characteristic function. *Biometrika*, **67**, 238-240.

[16] KOUTROUVELIS I.A. ET KELLERMEIER J. (1981). A goodness-of-fit test based on the empirical characteristic function when parameters must be estimated. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **43**, 173-176.

[17] LEHMANN E. (1975). *Nonparametrics : Statistical methods based on ranks.* McGraw-Hill, New York.

[18] LUKACS E. (1970). *Characteristic Functions.* Hafner Publishing Co., New York, deuxième édition. Second edition, revised and enlarged.

[19] LUONG A. ET THOMPSON M.E. (1987). Minimum-distance methods based on quadratic distances for transforms. *Canad. J. Statist.*, **15**, 239-251.

[20] LUONG A. ET DORAY L.G. (2002). General quadratic distance methods for discrete distributions definable recursively. *Insurance Math. Econom.*, **30**, 255-267.

[21] LUONG A. ET DORAY L.G. (2008). Inference for the positive stable laws based on a special quadratic distance. Stat. Methodol., accepted.

[22] MANDELBROT B. B. (1963). The variation of certain speculative prices. *The Journal of Business*, **36**, 394-419.

[23] MATSUI M. ET TAKEMURA A. (2005a). Empirical characteristic function approach to goodness-of-fit tests for the Cauchy distribution with parameters estimated by MLE or EISE. *Ann. Inst. Statist. Math.*, **57**, 183-199.

[24] MATSUI M. ET TAKEMURA A. (2005b). Goodness-of-fit tests for symmetric stable distributions - empirical characteristic function approach. CIRJE F-Series CIRJE-F-384, CIRJE, Faculty of Economics, University of Tokyo. disponible à `http://ideas.repec.org/p/tky/fseres/2005cf384.html`.

[25] MCWILLIAMS T. (1990). A distribution-free test for symmetry based on a runs statistic. *J. Amer. Statist. Assoc.*, **85**, 1130-1133.

[26] MODARRES R. ET GASTWIRTH J. (1996). A modified runs test for symmetry. *Statist. Probab. Lett.*, **31**, 107-112.

[27] MODARRES R. ET GASTWIRTH J. (1998). Hybrid test for the hypothesis of symmetry. *J. Appl. Stat.*, **25**, 777-783.

[28] PAULSON A.S., HALCOMB E.W. ET LEITCH R.A. (1975). The estimation of the parameters of the stable laws. *Biometrika*, **62**, 163-170.

[29] RANDLES R.H. ET WOLFE D.A. (1979). *Introduction to the Theory of Nonparametric Statistics.* Wiley, New York.

[30] RANDLES R. H., FLIGNER M. A., POLICELLO II G. E. ET WOLFE D. A. (1980). An asymptotically distribution-free test for symmetry versus asymmetry. *J. Amer. Statist. Assoc.*, **75**, 168-172.

[31] RAO C.R. (1973). *Linear Statistical Inference and its Applications.* Wiley, New York.

[32] TAJUDDIN I. (1994). Distribution-free test for symmetry based on Wilcoxon two-sample test. *J. Appl. Stat.*, **21**, 409-416.

[33] THAS O., RAYNER J.C.W. ET BEST D.J. (2005). Tests for symmetry based on the one-sample wilcoxon signed rank statistic. *Comm. Statist. Simulation Comput.*, **34**, 957-973.

[34] YU J. (2004). Empirical characteristic function estimation and its applications. *Econometric Rev.*, **23**, 93-123.

Annexe A

PROCÉDURES CODÉES AVEC MAPLE 11.0

Les procédures suivantes codées avec Maple 11.0 permettent de générer des observations provenant de la famille de distributions représentée par la différence de deux variables aléatoires gamma, de trouver l'estimateur qui minimise la distance quadratique entre les parties réelles et imaginaires des fonctions caractéristiques empiriques et théoriques et de procéder à des tests d'ajustement.

```
#Procédures permettant de générer des observations
#et de trouver l'estimateur de la DQ
restart;
with(Statistics):
with(Optimization):
with(LinearAlgebra):
interface(rtablesize=20):
Digits:=10;

GenerateSample:=proc(alpha::nonnegative,theta::nonnegative,n::nonnegint)
 local X,Y,Z,RV;
 global Sample1,returnSample;
 if (alpha=0) then
  RV:=RandomVariable(Normal(0,sqrt(2*theta))):
 else
  X:=RandomVariable(Gamma(sqrt(alpha*theta),1/alpha)):
  Y:=RandomVariable(Gamma(sqrt(alpha*theta),1/alpha)):
  Z:=X-Y:
  RV:=Z:
 end if;
 Sample1:=Sample(RV,n):
```

```
 returnSample:=['MOM',(Kurtosis(Sample1)-3)/3,Variance(Sample1)/2];
 return returnSample;
end proc;

OLSProc:=proc(n::nonnegint,m::nonnegint,maxt::nonnegative,opttolerance:=
   1e-15,optevallimit:=200)
 local Zn,QDI,Res2,Res3,Res4,ModifiedNewtonResTrim;
 global T,Y,ModifiedNewtonResOLS,starta,startb,returnOLS;
 T:=Vector([seq(maxt*i/m,i=1..m)],datatype=float):
 Y:=Vector([seq(evalf(1/n*add(cos(T[j]*Sample1[i]),i=1..n))-
   (a*b*T[j]^2+1)^(-1/a),j=1..m)]):
 QDI:=Transpose(Y).Y:
 starta:=(Kurtosis(Sample1)-3)/3;
 if (starta<=0) then starta:=0.5 end if;
 startb:=Variance(Sample1)/2;
 ModifiedNewtonResOLS:=NLPSolve(QDI,a=0..1000,b=0..1000,method =
   modifiednewton,initialpoint=[a=starta,b=startb],evaluationlimit=
     optevallimit,optimalitytolerance=opttolerance);
 ModifiedNewtonResTrim:=['OLS',parse(sprintf("%.10f",
   eval(a,ModifiedNewtonResOLS[2]))),parse(sprintf("%.10f",
     eval(b,ModifiedNewtonResOLS[2])))];
 returnOLS:=ModifiedNewtonResTrim;
 return returnOLS;
end proc;

WLSProc:=proc(n::nonnegint,m::nonnegint,maxt::nonnegative,opttolerance:=
  1e-15,optevallimit:=200,opttolerancelevel:=1e-12)
 local f,lastRes,lastResa,lastResb,QD,CovVar,ModifiedNewtonResTrim;
 global aa,bb,CovVarInv,ModifiedNewtonRes,count,returnWLS;
 aa:=eval(a,ModifiedNewtonResOLS[2]);
 bb:=eval(b,ModifiedNewtonResOLS[2]);
 if (aa>0) then f:=t->(1/(1+aa*bb*t^2))^(1/aa)
 else f:=t->limit((1/(1+aaa*bb*t^2))^(1/aaa),aaa=0)
 end if:
 lastResa:=aa+1:
 count:=0:
 CovVar:=convert(Matrix(m,m,(i,j)->evalf(1/2*(f(maxt*(i+j)/m)+
   f(maxt*(i-j)/m))-f(maxt*i/m)*f(maxt*j/m))),Matrix, datatype=float):
 CovVarInv:=CovVar^(-1):
```

```
QD:=Transpose(Y).CovVarInv.Y:
while abs(aa-lastResa)>opttolerancelevel or abs(bb-lastResb)>
  opttolerancelevel do
 count:=count+1;
 lastResa:=aa;
 lastResb:=bb;
 ModifiedNewtonRes:=NLPSolve(QD,a=0..1000,b=0..1000,method =
   modifiednewton,initialpoint=[a=aa+1e-30,b=bb],evaluationlimit=optevallimit,
     optimalitytolerance=opttolerance);
 aa:=eval(a,ModifiedNewtonRes[2]);
 bb:=eval(b,ModifiedNewtonRes[2]);
 if (aa>0) then f:=t->(1/(1+aa*bb*t^2))^(1/aa)
 else f:=t->limit((1/(1+aaa*bb*t^2))^(1/aaa),aaa=0)
 end if:
 CovVar:=convert(Matrix(m,m,(i,j)->evalf(1/2*(f(maxt*(i+j)/m)+
   f(maxt*(i-j)/m))-f(maxt*i/m)*f(maxt*j/m))),Matrix, datatype=float):
 CovVarInv:=CovVar^(-1):
 QD:=Transpose(Y).CovVarInv.Y
end do:
ModifiedNewtonResTrim:=['WLS',parse(sprintf("%.10f",aa)),
  parse(sprintf("%.10f",bb))];
returnWLS:=ModifiedNewtonResTrim,['ITERATIONS',count,'_'];
return returnWLS;
end proc;

COVProc:=proc(n::nonnegint,m::nonnegint,maxt::nonnegative)
 local phi;
 global SCovSInv,S,returnCOV,sdAlpha,sdTheta,varAlpha,varTheta;
 phi:=(a,b,t)->(1/(1+a*b*t^2))^(1/a):
 if (aa>0) then
  S:=convert(Matrix(m,2,(i,j)->evalf(D[j](phi)(aa,bb,T[i]))),Matrix,
    datatype=float):
 else
  S:=convert(Matrix(m,2,(i,j)->evalf(limit(D[j](phi)(aaa,bb,T[i]),aaa=0))),
    Matrix, datatype=float):
 end if;
 SCovSInv:=(Transpose(S).CovVarInv.S)^(-1)*1/n:
 varAlpha:=SCovSInv[1,1];
 varTheta:=SCovSInv[2,2];
```

```
 if (varAlpha>=0) then
  sdAlpha:=varAlpha^(0.5);
  sdTheta:=varTheta^(0.5);
  returnCOV:=['sigma',parse(sprintf("%.10f",sdAlpha)),parse(sprintf("%.10f",
    sdTheta))],['CONF_INT',[parse(sprintf("%.10f",aa-1.96*sdAlpha)),
      parse(sprintf("%.10f",aa+1.96*sdAlpha))],[parse(sprintf("%.10f",
        bb-1.96*sdTheta)),parse(sprintf("%.10f",bb+1.96*sdTheta))]];
 else
  sdTheta:=varTheta^(0.5);
  returnCOV:=['sigma',parse(sprintf("%.10f",varAlpha)),parse(sprintf("%.10f",
    sdTheta))],['CONF_INT','NA',[parse(sprintf("%.10f",bb-1.96*sdTheta)),
      parse(sprintf("%.10f",bb+1.96*sdTheta))]];
 end if;
 return returnCOV;
end proc;

WLS_nProc:=proc(n::nonnegint,m::nonnegint,maxt::nonnegative,
  opttolerance:=1e-15,optevallimit:=200)
 local aa,bb,PHI_n,QD_n,CovVar_n,ModifiedNewtonRes,ModifiedNewtonResTrim;
 global CovVarInv_n,returnWLS_n;
 aa:=eval(a,ModifiedNewtonResOLS[2]);
 bb:=eval(b,ModifiedNewtonResOLS[2]);
 PHI_n:=Vector([seq(evalf(1/n*add(cos((maxt*i/m)*Sample1[k]),k=1..n))
   ,i=0..2*m)],datatype=float):
 CovVar_n:=Matrix(m,m,(i,j)->evalf(1/2*(PHI_n[i+j+1]+PHI_n[abs(i-j)+1])-
   PHI_n[i+1]*PHI_n[j+1])):
 CovVarInv_n:=(CovVar_n)^(-1):
 QD_n:=Transpose(Y).(CovVarInv_n).Y:
 ModifiedNewtonRes:=NLPSolve(QD_n,a=0..1000,b=0..1000,method =
   modifiednewton,initialpoint=[a=aa+1e-30,b=bb],evaluationlimit=
     optevallimit,optimalitytolerance=opttolerance);
 ModifiedNewtonResTrim:=['WLS_n',parse(sprintf("%.10f",
   eval(a,ModifiedNewtonRes[2]))),parse(sprintf("%.10f",
     eval(b,ModifiedNewtonRes[2])))];
 returnWLS_n:=ModifiedNewtonResTrim;
 return returnWLS_n;
end proc;

WLS_0Proc:=proc(alpha::nonnegative,theta::nonnegative,n::nonnegint,
```

```
 m::nonnegint,maxt::nonnegative,opttolerance:=1e-15,optevallimit:=200)
local aa,bb,f,QD_0,CovVar_0,ModifiedNewtonRes,ModifiedNewtonResTrim;
global CovVarInv_0,returnWLS_0;
aa:=eval(a,ModifiedNewtonResOLS[2]);
bb:=eval(b,ModifiedNewtonResOLS[2]);
if (aa>0) then f:=t->(1/(1+aa*bb*t^2))^(1/aa)
else f:=t->limit((1/(1+aaa*bb*t^2))^(1/aaa),aaa=0)
end if:
CovVar_0:=convert(Matrix(m,m,(i,j)->evalf(1/2*(f(maxt*(i+j)/m)+
  f(maxt*(i-j)/m))-f(maxt*i/m)*f(maxt*j/m))),Matrix, datatype=float):
CovVarInv_0:=CovVar_0^(-1):
QD_0:=Transpose(Y).CovVarInv_0.Y:
ModifiedNewtonRes:=NLPSolve(QD_0,a=0..1000,b=0..1000,method =
  modifiednewton,initialpoint=[a=aa+1e-30,b=bb],evaluationlimit=
    optevallimit,optimalitytolerance=opttolerance);
ModifiedNewtonResTrim:=['WLS_0',parse(sprintf("%.10f",eval(a,
  ModifiedNewtonRes[2]))),parse(sprintf("%.10f",eval(b,
    ModifiedNewtonRes[2])))];
returnWLS_0:=ModifiedNewtonResTrim;
return returnWLS_0;
end proc;

QDEProc:=proc(alpha::nonnegative,theta::nonnegative,n::nonnegint,
 m::nonnegint,maxt::nonnegative,opttolerance:=1e-15,optevallimit:=200,
   opttolerancelevel:=1e-12)
Digits:=10;
GenerateSample(alpha,theta,n);
Digits:=50;
OLSProc(n,m,maxt,opttolerance,optevallimit);
WLSProc(n,m,maxt,opttolerance,optevallimit,opttolerancelevel);
COVProc(n,m,maxt);
WLS_nProc(n,m,maxt,opttolerance,optevallimit);
WLS_0Proc(alpha,theta,n,m,maxt,opttolerance,optevallimit);
return Matrix([['N'=n,'M'=m,'MAXT'=maxt],['_','ALPHA','THETA'],
  ['TRUE',alpha,theta],returnSample,returnOLS,returnWLS,returnCOV,
    returnWLS_n,returnWLS_0]);
end proc;

#QDEProc(ALPHA,THETA,N,M,MAXT,CONVERGENCE PRECISION,
```

```
  MAX ITERATIONS,CONVERGENCE LEVEL FOR PARAMETERS)
QDEProc(1,1,500,10,3,1e-25,1000,1e-12);

#Procédures permettant de procéder à des tests d'ajustement
restart;
with(Statistics):
with(LinearAlgebra):
interface(rtablesize=20):
Digits:=10;

GenerateRV:=proc(alpha1::nonnegative,theta1::nonnegative,n::nonnegint)
 local X,Y,Z;
 global RV;
 if (alpha1=0) then
  RV:=RandomVariable(Normal(0,sqrt(2*theta1))):
 else
  X:=RandomVariable(Gamma(sqrt(alpha1*theta1),1/alpha1)):
  Y:=RandomVariable(Gamma(sqrt(alpha1*theta1),1/alpha1)):
  Z:=X-Y:
  RV:=Z:
 end if;
end proc;

GOFSpeProc:=proc(alpha0::nonnegative,theta0::nonnegative,
  alpha1::nonnegative,theta1::nonnegative,n::nonnegint,
    m::nonnegint,maxt::nonnegative,nbtrials::integer,CV::nonnegative)
 local T,Sample1,Y,f,aa,bb,counter,CovVar;
 global CovVarInv,RejectCount,QD;
 GenerateRV(alpha1,theta1,n);
 Digits:=50;
 aa:=alpha0;
 bb:=theta0;
 if (aa>0) then f:=t->(1/(1+aa*bb*t^2))^(1/aa)
 else f:=t->limit((1/(1+aaa*bb*t^2))^(1/aaa),aaa=0)
 end if:
 CovVar:=convert(Matrix(m,m,(i,j)->evalf(1/2*(f(maxt*(i+j)/m)+f(maxt*(i-j)/m))-
   f(maxt*i/m)*f(maxt*j/m))),Matrix, datatype=float):
 CovVarInv:=CovVar^(-1):
 RejectCount:=0;
```

```
T:=Vector([seq(maxt*i/m,i=1..m)],datatype=float):
QD:=NULL;
Digits:=10;
Sample1:=Sample(RV,n):
Digits:=50;
Y:=Vector([seq(evalf(1/n*add(cos(T[j]*Sample1[i]),i=1..n))-f(T[j]),j=1..m)]):
QD:= QD, n*Transpose(Y).CovVarInv.Y;
if (is(QD>CV)=true) then RejectCount:=RejectCount+1 end if;
for counter from 2 to nbtrials do
 Digits:=10;
 Sample1:=Sample(RV,n):
 Digits:=50;
 Y:=Vector([seq(evalf(1/n*add(cos(T[j]*Sample1[i]),i=1..n))-f(T[j]),j=1..m)]):
 QD:= QD, n*Transpose(Y).CovVarInv.Y;
 if (is(QD[counter]>CV)=true) then RejectCount:=RejectCount+1 end if;
end do:
return Matrix([['_','alpha','theta'],['Null',alpha0,theta0],
  ['Alt',alpha1,theta1],['Rejected||Accepted',RejectCount,nbtrials-RejectCount],
    ['Q_Chi2||Q_Data', parse(sprintf("%.5f",Quantile('ChiSquare'(m),0.95))),
      parse(sprintf("%.5f",Quantile([QD],0.95,method=5)))]]);
end proc;

#GOFSpeProc(ALPHA NULL, THETA NULL, ALPHA ALT,THETA ALT,N,M,MAXT,TRIALS,CV)
GOFSpeProc(1,1,1,1,100,10,3,10,27.4348);
```