Université de Montréal

Estimation de paramètres en exploitant les aspects calculatoires et numériques

par

Romain KADJE KENMOGNE

Département de mathématiques et de statistique

Faculté des arts et des sciences

Thèse présentée à la Faculté des études supérieures
en vue de l'obtention du grade de
Philosophiæ Doctor (Ph.D.)
en statistique

31 août 2017

SOMMAIRE

Dans cette thèse, nous nous intéressons principalement à l'estimation de paramètres. Elle est constituée de trois articles dans lesquels nous abordons des problèmes d'estimation dans des modèles précis.

Dans le premier article, nous considérons la famille paramétrique de copules de Farlie-Gumbel-Morgenstern. Les observations proviennent d'une loi qui fait intervenir à la fois la copule et les marges, au travers de la décomposition du théorème de Sklar. Les marges sont inconnues. Dans le cadre de l'estimation d'une fonction du paramètre de la copule, la pseudo-vraisemblance est souvent utilisée en lieu et place de la vraisemblance. C'est une approche qui va généralement produire des estimateurs non efficaces (variance non-minimale), spécialement pour de petites tailles d'échantillons. Dans la pseudo-vraisemblance, les marges sont remplacées par des estimateurs qui dépendent des rangs. Nous proposons d'utiliser la vraisemblance des rangs, qui est la vraisemblance basée sur la distribution de la statistique des rangs. Cette approche peut être complexe en pratique car on doit travailler sur le groupe des permutations et on doit calculer des intégrales multiples. Néanmoins, il est possible d'effectuer les calculs pour la famille de copules de Farlie-Gumbel-Morgenstern. Nous comparons les estimateurs obtenus de ces approches à l'aide de la méthode bayésienne. Les résultats des études numériques sont présentés.

Certains estimateurs rencontrés dans la littérature sont des rapports de variables aléatoires de lois normales. Dans le second article, nous nous intéressons à la distribution du rapport de deux variables aléatoires de lois normales. Plusieurs auteurs ont abordé ce sujet. La nouvelle paramétrisation que nous introduisons nous permet d'arriver assez facilement à de nouveaux résultats. Tout d'abord, nous montrons que l'expression de la densité du rapport s'écrit comme un mélange de densités appartenant à une nouvelle famille que nous créons. Nous proposons quelques propriétés et nous donnons l'expression analytique de la fonction caractéristique pour la nouvelle famille. Cette nouvelle famille est en fait une généralisation de la famille des densités des lois de Student de degrés de liberté impairs. Nous arrivons à des résultats de convergence similaires à la convergence observée des lois de Student lorsque le degré de liberté tend vers l'infini. Des résultats semblables de convergence ainsi que quelques propriétés sont développées pour la loi du rapport. Nous utilisons une approche bayésienne pour estimer le rapport de deux moyennes de variables aléatoires de lois normales. Des illustrations graphiques et des résultats des simulations sont présentés.

Dans le troisième sujet, nous étendons la famille de distributions issue de la différence de deux variables aléatoires indépendantes de loi gamma, en considérant le cas où ces variables aléatoires sont positivement corrélées. Les densités de cette famille sont trouvées. La forme complexe de ces densités rend fastidieuses les méthodes d'estimation basées sur la vraisemblance. La méthode d'estimation basée sur la fonction caractéristique est utilisée, et l'approche continue est comparée à l'approche discrète. Par ailleurs, nous aboutissons à deux algorithmes simples permettant de générer facilement des données de deux variables aléatoires positivement corrélées et identiquement distribuées de loi gamma. Le meilleur estimateur équivariant pour le paramètre d'échelle est obtenu dans le cas de l'indépendance.

Mots clés : convergence de variables aléatoires, copule, différence de variables de loi gamma, estimateur équivariant, estimation des paramètres, fonction caractéristique, méthode bayésienne, pseudo-vraisemblance, rapport de variables gaussiennes, vraisemblance des rangs.

SUMMARY

In this thesis, we are mainly interested in the estimation of parameters. It consists of three papers in which we discuss estimation problems in specific models.

In the first paper, we consider the Farlie-Gumbel-Morgenstern parametric family of copulas. The observations come from a distribution that involves both the copula and the margins, through the decomposition of the Sklar's theorem. Margins are unknown. In the estimation of a function of the parameter of the copula, pseudo-likelihood is often used instead of the likelihood. It is an approach that will generally produce non-efficient estimators, especially for small sample sizes. In pseudo-likelihood, margins are replaced by estimators that depend on the ranks. We propose to use the rank-likelihood, which is the likelihood based on the distribution of the rank statistics. This approach can be complex in practice because we work on the permutation group and we calculate multiple integrals. However, it is possible to do the calculations for the Farlie-Gumbel-Morgenstern family of copulas. We compare the estimators obtained from these approaches using the Bayesian method. The results of numerical studies are presented.

Some estimators found in the literature are ratios of normal random variables. In the second paper, we are interested in the distribution of the ratio of two normal random variables. Several authors have addressed this subject. The new parametrization that we introduced allows us to obtain new results faster and effortlessly. Firstly, we show that the expression of the density of the ratio can be written as a mixture of densities belonging to a new family that we create. We find some properties, and we give the analytic expression of the characteristic function for the new family. This new family is in fact a generalization of the family of densities of Student distributions having odd degrees of freedom. We obtain convergence results generalizing the one for Student distributions when the degree of freedom tends to infinity. Similar results of convergence as well as some properties are developed for the distribution of the ratio. We use a Bayesian approach to estimate the ratio of the means of two normal random variables. Graphical illustrations and simulation results are presented.

In the third paper, we extend the family of distributions, resulting from the difference of two independent random variables having a gamma distribution, by considering the case where these random variables are positively correlated. The densities in this family are found. Their complex forms make tedious the estimation methods based on the likelihood. The estimation method based on the characteristic function is used, and the continuous approach is compared to the discrete approach. Moreover, we developed two simple algorithms for generating data from two positively correlated and identically distributed random variables having a gamma distribution. The best equivariant estimator for the scale parameter is obtained in the case of independence.

Keywords: Bayesian method, characteristic function, convergence of random variables, copula, difference of gamma variates, equivariant estimator, parameter estimation, pseudo-likelihood, rank-likelihood, ratio of normal variables.

TABLE DES MATIÈRES

LISTE DES TABLEAUX

TABLE DES FIGURES

LISTE DES SIGLES ET DES ABRÉVIATIONS

cdf	Fonction de répartition
C.V.	Méthode ECF continue avec variable de contrôle
DGD	Différence de deux variables de lois Gamma, de l'anglais *Double Gamma Difference*
ECF	Fonction caractéristique empirique
FGM	Farlie-Gumbel-Morgenstern
i.i.d.	Indépendant identiquement distribué
MGF	Fonction génératrice des moments
MOM	Méthode des momemts
OLS	Moindres carrés ordinaires, de l'anglais *Ordinary Least Squares*
STAND	Approche standard de la méthode ECF continue
WLS	Moindres carrés pondérés, de l'anglais *Weighted Least Squares*

REMERCIEMENTS

La réalisation de cette thèse est rendue possible grâce au soutien de nombreuses personnes et institutions. Je voudrais dans ces quelques lignes leur témoigner toute ma reconnaissance.

Je tiens à remercier très chaleureusement mon directeur de thèse, le Prof. François PERRON, qui n'a ménagé aucun effort pour me soutenir à chaque étape de ma thèse. Il a été présent lorsqu'il fallait préparer les examens de synthèse en probabilité et en statistique mathématique. Les séances de rencontres hebdomadaires et régulières m'ont aidé à maintenir une évolution constante tout au long de la thèse.

Je voudrais remercier les membres de Jury de cette thèse pour leurs remarques et pour leurs questions. Il s'agit des Professeurs Louis DORAY, Maciej AUGUSTYNIAK, François PERRON, Eugène SYRIANI de l'Université de Montréal et du Professeur Sévérien NKURUNZIZA de l'Université de Windsor. Je remercie l'Institut des Sciences Mathématiques du Québec, le Centre de Recherche Mathématiques, la fondation J. Armand Bombardier, le département de Mathématiques et statistique de l'Université de Montréal, et la faculté des études supérieures et postdoctorales pour leur support financier.

Je désire grandement remercier la Famille MOMEYA qui m'a encouragé à m'inscrire au doctorat à l'Université de Montréal alors que j'effectuais encore mon Master II à l'Université de Douala. Je tiens également à lui exprimer ma reconnaissance pour son accueil chaleureux et pour l'agréable moment passé avec elle.

Je remercie mes enseignants de l'Université de Montréal ainsi que le personnel administratif pour leur disponibilité. Je remercie mes enseignants de l'Université de Douala pour m'avoir encouragé et soutenu dans cette aventure. J'exprime toute ma gratitude à mes encadreurs de Master II : Prof. FONO de l'Université de Douala et Prof. NGNIATEDEMA de Morgan State University, pour leur disponibilité, leur soutien et leurs conseils.

Je remercie en outre mes camarades des classes avec qui j'ai partagé ces années. Je remercie mes amis, les membres de la communauté Basil Moreau de l'Oratoire Saint Joseph, de la chorale Marie Reine des Apôtres et de l'association Pouahom Canada pour avoir contribué à embellir cette aventure. J'associe à ces remerciements les familles KOUAM, KENGNE, MBIELEU, KENMOGNE et MBOZO'O pour leur encouragement.

Je dis merci à Francine pour son soutien et son affection. Enfin, toute ma reconnaissance à ma famille pour son soutien multiforme et inconditionnel au cours de ces années.

INTRODUCTION

Considérons la famille paramétrique de copules de Farlie-Gumbel-Morgenstern, discutée successivement par Morgenstern (1956), Farlie (1960) et Gumbel (1960). Les observations sont générées d'une loi qui fait intervenir la copule C_θ et les marges F et G, au travers de la décomposition du théorème de Sklar :

$$H(x, y) = C_\theta(F(x), G(y)), \quad x,\, y \in \mathbb{R}, \quad \theta \in \Theta.$$

Dans la discussion "Copulas : Tales and facts" initiée par Mikosch (2006), Segers (2006) présente un sujet ouvert concernant l'estimation efficace des paramètres d'une copule lorsqu'aucune hypothèse paramétrique n'est faite sur les marges. La question que l'on se pose est : comment faire l'estimation de paramètres de façon efficace dans ce contexte ? Dans le premier chapitre de cette thèse, nous estimons le polynôme de Legendre de degré deux, donné par $P_2(\theta) = (3\theta^2 - 1)/2$. Les marges F et G sont des paramètres de nuisance. Nous discuterons de trois approches. La première étant lorsque les marges sont connues, la deuxième et la troisième ayant des marges inconnues et basées respectivement sur la pseudo-vraisemblance et la vraisemblance des rangs. Le but de ce chapitre est de comparer les estimateurs issus de ces trois approches, à l'aide du risque de Bayes avec une fonction de perte quadratique.

Les calculs effectués pour trouver l'estimateur à partir de la vraisemblance des rangs peuvent être très longs et compliqués. Le choix de la famille de copules Farlie-Gumbel-Morgenstern est motivé par le fait qu'on peut obtenir des expressions analytiques de chacun des estimateurs.

La première approche, où les marges F et G sont connues, n'est pas réaliste. Notre but est d'utiliser l'estimateur issu de cette approche comme un estimateur de référence et de le comparer avec les autres. En effet, la différence entre cet estimateur et un des autres estimateurs est considérée comme la perte due au fait que les marges soient inconnues. Plusieurs auteurs ont utilisé cette approche, voir par exemple Kim et coll. (2007) et Hofert et coll. (2012).

Les marges sont en général inconnues. Une façon de procéder est par exemple de remplacer les marges F et G par des fonctions de répartition empirique F_n et G_n respectivement, et de faire comme si la fonction obtenue (la pseudo-vraisemblance) était la vraisemblance, voir Genest et coll. (1995) et Grønneberg et Hjort (2014). Nous utiliserons la pseudo-vraisemblance avec une méthodologie bayésienne. Cette procédure n'a pas de justification théorique et produit généralement des estimateurs peu performants, particulièrement pour des échantillons de petite taille. De plus, l'estimateur que nous obtiendrons n'est pas un estimateur de Bayes. Il dépend tout de même de la statistique des rangs. Finalement, il est naturel de travailler avec la statistique des rangs. Nous proposons dans la troisième approche de baser l'estimation sur la vraisemblance des rangs. La loi des rangs ne dépend pas des marges. Nous travaillerons avec le groupe de permutations et le calcul de la distribution des rangs implique des intégrales multiples d'ordre $2n$. Ce qui constitue une limite en pratique. Cependant, pour la famille de copules Farlie-Gumbel-Morgenstern, il est possible d'effectuer les calculs. Il est attendu que l'estimateur obtenu soit meilleur que celui issu de la pseudo-vraisemblance. Nous utilisons également une approche bayésienne, avec de petites tailles d'échantillon. En effet, la méthode basée sur la pseudo-vraisemblance est une méthode basée sur la statistique des rangs et par conséquent elle ne saurait être plus performante que celle basée sur la vraisemblance des rangs.

Dans la Section 1.2, nous trouvons la distribution des rangs en utilisant le groupe de permutations. Cette distribution est développée pour les copules de la famille Farlie-Gumbel-Morgenstern dans la Section 1.3. Dans la Section 1.4, nous trouvons les expressions analytiques des différents estimateurs. Précisons que nous utilisons deux distributions a priori, une uniforme et l'autre dite de Jeffrey. Nous comparons de façon générale les estimateurs à la Section 1.5, à l'aide du risque de Bayes pour une perte quadratique. Finalement, à la Section 1.6, nous présentons les résultats des études de simulation réalisées.

Le rapport, X_1/X_2 de deux variables aléatoires de lois normales, a été discuté pour la première fois par Geary (1930), et ensuite par Fieller (1932), Marsaglia (1965, 2006), Hinkley (1969), Korhonen et Narula (1989), Pham-Gia et coll. (2006) etc. En effet, plusieurs estimateurs rencontrés dans la littérature sont le rapport de variables aléatoires de lois normales, voir par exemple Gill et Keating (2008). La densité de ce rapport a été utilisée pour résoudre plusieurs problèmes remarquables dans des domaines tels que l'économie, l'industrie, le commerce, l'éducation et la médecine, voir Öksoy et Aroian (1994), Soldan et coll. (2010), Barone (2012) et Białek (2015). Il est donc important d'avoir des outils théoriques afin de mieux exploiter les estimateurs ayant cette forme.

Dans le deuxième chapitre de cette thèse, nous étudions en profondeur les propriétés de ce rapport et nous l'utilisons dans le cadre de l'estimation du rapport de deux moyennes des variables aléatoires de lois normales. Plus précisément, dans la Section 2.2, nous introduisons une nouvelle famille de fonctions de densité. Cette famille vient généraliser la famille des densités de Student à degrés de liberté impairs. Le résultat bien connu de convergence des distributions de Student vers une distribution normale lorsque le degré de liberté tend vers l'infini, est aussi établi pour cette nouvelle famille. Quelques propriétés sont données et nous calculons les fonctions caractéristiques associées à ces densités.

La Section 2.3 est consacrée au développement des caractéristiques et propriétés du rapport étudié. Nous commençons par trouver deux expressions de la densité du rapport. L'une de ces expressions s'écrit comme un mélange des densités de la nouvelle famille définie à la Section 2.2, avec des poids qui sont les probabilités d'une distribution de Poisson. Par le passé, les auteurs trouvaient la densité de ce rapport en utilisant les mêmes paramètres que ceux des variables X_1 et X_2. Cela menait à des problèmes d'identifiabilité. Pour éviter ce problème, nous utilisons une nouvelle paramétrisation qui facilite par ailleurs les calculs.

Nous trouvons quelques propriétés. Il apparaît que la distribution du rapport de deux variables aléatoires de lois normales peut être très proche de la distribution d'une loi normale. En effet, Hinkley (1969) avait essayé de montrer la convergence en distribution du rapport lorsque l'un des paramètres tend vers l'infini. Il avait obtenu la convergence de la distribution vers une fonction qui n'était pas une distribution. Hayya et coll. (1975) et Marsaglia (2006) ont, à travers des simulations et des études empiriques, trouvé quelques conditions sous lesquelles la densité du rapport est proche de celle d'une loi normale. Récemment, Díaz-Francés et Rubio (2013) ont montré sous des conditions sévères que la différence entre la distribution du rapport et celle d'une loi normale pouvait être suffisamment petite dans un intervalle précis. Grâce à la nouvelle paramétrisation que nous utilisons, nous montrons la convergence en distribution du rapport vers une variable aléatoire de loi normale lorsque l'un des paramètres tend vers l'infini. Nous donnons quelques illustrations graphiques de la densité du rapport, des densités de la nouvelle famille, ainsi que de l'approximation de la loi de X_1/X_2 par celle d'une variable aléatoire normale.

Dans la Section 2.4, nous estimons le rapport des moyennes de deux variables aléatoires de lois normales. L'estimateur usuel utilisé dans la littérature est le rapport des moyennes échantillonnales, voir Pham-Gia et coll. (2006). Cet estimateur est le rapport de deux variables de lois normales. Il n'admet donc pas de moments et il est impossible de juger de sa qualité à l'aide des critères usuels tels que l'erreur de moindres carrés ou de moindres valeurs absolues. Nous définissons un critère approprié et dans une approche bayésienne, nous proposons d'estimer le rapport des moyennes par la médiane de sa distribution a posteriori. Des études de simulations sont effectuées afin de comparer les deux estimateurs.

La différence, $Y = X_1 - X_2$, de deux variables aléatoires i.i.d. de loi gamma de paramètres α et β, a été utilisée par Augustyniak et Doray (2012). Sa distribution est notée DGD$(\alpha, \beta, 0)$, pour *double gamma difference distribution*. Tel que mentionné par Klar (2015), la distribution DGD est un cas particulier d'une autre distribution appelée *variance gamma distribution* et étudiée par plusieurs auteurs tels que Madan et Seneta (1990) et Seneta (2004).

Au chapitre 3, nous considérons le cas où les deux variables aléatoires X_1 et X_2 sont indépendantes ou positivement corrélées et identiquement distribuées de loi gamma. À partir de la fonction génératrice des moments du vecteur (X_1, X_2) définie dans l'article de Chen et coll. (2014), nous trouvons la fonction caractéristique de Y. La nouvelle famille de distributions est notée $\text{DGD}(\alpha, \beta, \rho)$, avec ρ^2 le coefficient de corrélation linéaire entre X_1 et X_2. Elle conserve les mêmes propriétés que dans le cas de l'indépendance des variables X_1 et X_2 étudié par Augustyniak et Doray (2012). Par exemple, sa fonction de densité est symétrique, ses moments existent et sa fonction caractéristique est réelle. Nous nous intéressons dans ce chapitre à l'estimation des paramètres. Nous trouvons l'expression analytique de la densité de Y. Sa forme est complexe, ce qui rend compliquées toutes les méthodes d'estimation basées sur la vraisemblance. Une alternative est d'utiliser la méthode d'estimation basée sur la fonction caractéristique empirique. Nous utilisons deux approches : l'une dite continue et l'autre discrète. L'approche continue consiste à minimiser la fonction

$$g(\theta|\mathbf{y}) = \int_{\mathbb{R}} |\varphi_Y(t|\theta) - \psi_n(t|\mathbf{y})|^2 w(t|\theta) \mathrm{d}t,$$

où θ est le vecteur de paramètres, $\mathbf{y}$ est un vecteur constitué de n observations provenant de Y, $w(t|\theta)$ est la fonction de poids, φ_Y et ψ_n sont respectivement les fonctions caractéristiques théorique et empirique. La difficulté de cette méthode est le choix de la fonction de poids. En pratique, la fonction exponentielle est utilisée. Bien qu'elle soit facile à manipuler, l'estimateur résultant n'est pas toujours performant. Un choix optimal est proposé dans la littérature par Feuerverger et McDunnough (1981). Cependant, cela ne nous aidera pas, car il fait appel à la densité qui a, rappelons-le, une forme complexe. Nous utiliserons la densité d'une loi normale comme fonction de poids. La seconde façon est l'approche discrète, qui consiste à choisir q points , $t_1, \ldots, t_q$, et à minimiser

$$\left[\frac{1}{n}\sum_{j=1}^{n} Z(y_j|\theta)\right]' W_n \left[\frac{1}{n}\sum_{j=1}^{n} Z(y_j|\theta)\right],$$

où,

$$Z(y_j|\theta) = (Re[C(t_1|y_j,\theta)],\ldots,Re[C(t_q|y_j,\theta)],Im[C(t_1|y_j,\theta)],\ldots,Im[C(t_q|y_j,\theta)]),$$

$$C(t|y_j,\theta) = \exp(ity_j) - \varphi_Y(t|\theta),$$

$Re[\cdot]$ et $Im[\cdot]$ sont les parties réelle et imaginaire d'un nombre complexe, W_n est une matrice semi-définie qui converge presque sûrement vers une matrice définie positive W_0. Augustyniak et Doray (2012) ont utilisé cette approche pour l'estimation des paramètres de la famille DGD avec indépendance entre les variables X_1 et X_2. Notre objectif est tout d'abord de faire l'estimation dans le cas où le coefficient de corrélation linéaire ρ^2 est non nul. Ensuite, nous voulons comparer les approches continue et discrète.

Dans la Section 3.2, nous commençons par présenter le meilleur estimateur équivariant, pour une fonction de perte quadratique, du paramètre β lorsque α est connu et les variables X_1 et X_2 sont indépendantes. À la Section 3.3, nous trouvons la densité de Y et nous donnons deux procédures simples permettant de générer des observations de deux variables aléatoires positivement corrélées et identiquement distribuées de loi gamma. Dans la Section 3.4, nous expliquons comment trouver les différents estimateurs. Nous utilisons une variable de contrôle afin de réduire la variance des estimateurs dans l'approche continue. Les résultats obtenus dans les études de simulations réalisées sont présentés à la Section 3.5.

Références

Augustyniak, M. et Doray, L. G. (2012). Inference for a leptokurtic symmetric family of distributions represented by the difference of two gamma variates. *J. Stat. Comput. Simul.*, **82**, 1621–1634.

Barone, P. (2012). A diffusion equation for the density of the ratio of two jointly distributed gaussian variables and the exponential analysis problem. *SIAM Journal on Scientific Computing*, **34**, A1053–A1078.

Białek, J. (2015). Construction of confidence intervals for the Laspeyres price index. *Journal of Statistical Computation and Simulation*, **85**, 2962–2973.

Chen, L.-S., Tzeng, I.-S. et Lin, C.-T. (2014). Bivariate generalized gamma distributions of Kibble's type. *Statistics*, **48**, 933–949.

Díaz-Francés, E. et Rubio, F. J. (2013). On the existence of a normal approximation to the distribution of the ratio of two independent normal random variables. *Statistical Papers*, **54**, 309–323.

Farlie, D. J. G. (1960). The performance of some correlation coefficients for a general bivariate distribution. *Biometrika*, **47**, 307–323.

Feuerverger, A. et McDunnough, P. (1981). On some Fourier methods for inference. *Journal of the American Statistical Association*, **76**, 379–387.

Fieller, E. C. (1932). The distribution of the index in a normal bivariate population. *Biometrika*, **24**, 428–440.

Geary, R. C. (1930). The frequency distribution of the quotient of two normal variates. *Journal of the Royal Statistical Society*, **93**, 442–446.

Genest, C., Ghoudi, K. et Rivest, L.-P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, **82**, 543–552.

Gill, R. S. et Keating, J. P. (2008). Monitoring gasoline blending rates for compliance. *Environmetrics*, **19**, 429–437.

Grønneberg, S. et Hjort, N. L. (2014). The copula information criteria. *Scand. J. Stat.*, **41**, 436–459.

Gumbel, E. J. (1960). Bivariate exponential distributions. *J. Amer. Statist. Assoc.*, **55**, 698–707.

Hayya, J., Armstrong, D. et Gressis, N. (1975). A note on the ratio of two normally distributed variables. *Management Science*, **21**, 1338–1341.

Hinkley, D. V. (1969). On the ratio of two correlated normal random variables. *Biometrika*, **56**, 635–639.

Hofert, M., Mächler, M. et McNeil, A. J. (2012). Likelihood inference for Archimedean copulas in high dimensions under known margins. *J. Multivariate Anal.*, **110**, 133–150.

Kim, G., Silvapulle, M. J. et Silvapulle, P. (2007). Comparison of semiparametric and parametric methods for estimating copulas. *Comput. Statist. Data Anal.*, **51**, 2836–2850.

Klar, B. (2015). A note on gamma difference distributions. *J. Stat. Comput. Simul.*, **85**, 3708–3715.

Korhonen, P. J. et Narula, S. C. (1989). The probability distribution of the ratio of the absolute values of two normal variables. *J. Statist. Comput. Simulation*, **33**, 173–182.

Madan, D. B. et Seneta, E. (1990). The variance gamma (vg) model for share market returns. *Journal of Business*, 511–524.

Marsaglia, G. (1965). Ratios of normal variables and ratios of sums of uniform variables. *J. Amer. Statist. Assoc.*, **60**, 193–204.

Marsaglia, G. (2006). Ratios of normal variables. *Journal of Statistical Software*, **16**, 1–10.

Mikosch, T. (2006). Copulas : tales and facts. *Extremes*, **9**, 3–20.

Morgenstern, D. (1956). Einfache Beispiele zweidimensionaler Verteilungen. *Mitteilungsbl. Math. Statist.*, **8**, 234–235.

Öksoy, D. et Aroian, L. A. (1994). The quotient of two correlated normal variables with applications. *Comm. Statist. Simulation Comput.*, **23**, 223–241.

Pham-Gia, T., Turkkan, N. et Marchand, E. (2006). Density of the ratio of two normal random variables and applications. *Comm. Statist. Theory Methods*, **35**, 1569–1591.

Segers, J. (2006). Efficient estimation of copula parameter. Discussion of : "Copulas : tales and facts" [Extremes **9** (2006), no. 1, 3–20 ;] by T. Mikosch. *Extremes*, **9**, 51–53.

Seneta, E. (2004). Fitting the variance-gamma model to financial data. *J. Appl. Probab.*, **41A**, 177–187. Stochastic methods and their applications.

Soldan, D. L., Osborne, W. P. et Gruenbacher, D. (2010). Modeling the economic cost of inadequate teaching and mentoring. In *Frontiers in Education Conference (FIE), 2010 IEEE*, F3J–1. IEEE.

Chapitre 1

VRAISEMBLANCE DES RANGS VERSUS PSEUDO-VRAISEMBLANCE

Cet article sera soumis à la revue *Canadian Journal of Statistics*.

Les principales contributions de *Romain Kadje Kenmogne* à cet article sont présentées.

– Conduite de la revue de la littérature.

– Contribution dans la production des Sections 4, 5 et 6.

– Conception, écriture et validation des programmes R.

– Conduite des simulations.

– Rédaction partielle de l'article.

Rank-likelihood versus pseudo-likelihood

Romain Kadje Kenmogne and François Perron

Département de mathématiques et de statistique,

Université de Montréal,

C.P. 6128, Succursale Centre-ville,

Montréal, Canada H3C 3J7

Abstract

Let $H(x, y) = C_\theta(F(x), G(y))$ be the cumulative distribution function of a continuous bivariate random vector. Here C_θ is the copula. We want to estimate $P_2(\theta)$, a function of the parameter θ, in a Bayesian setup. The functions F and G are unknown cumulative distribution functions. The pseudo-likelihood is a plug-in approach where F and G are replaced by the empirical distribution functions F_n and G_n, which are based on rank statistics. The rank-likelihood is the likelihood based on the distribution of the rank statistics. In this paper we compare the Bayes estimator based on the ranks versus the one obtained from the pseudo-likelihood. The comparisons are given for the Farlie-Gumbel-Morgenstern family of copulas.

Keywords: Bayes estimator; copulas; group; permutation; pseudo-Bayes estimator; pseudo-likelihood; pseudo-observations; rank; rank-likelihood.

1.1. Introduction

Let $\{(X_i, Y_i),\ i = 1, \ldots, n\}$ be n independent bivariate random vectors having the same cumulative distribution function H with

$$H(x, y) = C_\theta(F(x), G(y)), \quad x, y \in \mathbb{R}, \quad \theta \in \Theta.$$

This representation is based on Sklar's theorem. The function C_θ is a copula, F and G are the margins. We assume that F and G are unknown, $\Theta = [-1, 1]$ and

$$C_\theta(u, v) = uv + \theta uv(1 - u)(1 - v), \quad u, v \in [0, 1]. \tag{1.1}$$

Here $\{C_\theta : \theta \in [-1, 1]\}$ is called the Farlie-Gumbel-Morgenstern (FGM) family of copulas. This family contains as members all copulas with quadratic sections in both u and v, and was discussed by Morgenstern (1956), Farlie (1960) and Gumbel (1960). The copula C_θ, is the cumulative distribution function of the independent and identically distributed random vectors $(U_i, V_i) = (F(X_i), G(Y_i)),\ i = 1, \ldots, n$. The copulas of this family are absolutely continuous and the corresponding density functions are given by

$$c_\theta(u, v) = \frac{\partial^2 C_\theta(u, v)}{\partial u \partial v} = 1 + \theta(2u - 1)(2v - 1), \quad \text{with} \quad \theta \in [-1, 1]. \tag{1.2}$$

In Discussion of "Copulas: Tales and facts" introduced by Mikosch (2006), Segers (2006) presents an open problem concerning the efficient estimation of the parameter of a copula when no parametric assumptions are made regarding the margins. How to draw inference on the unknown parameter in the most efficient way?

We want to estimate the Legendre's polynomial of degree two, $P_2(\theta) = (3\theta^2 - 1)/2$. The margins F and G are nuisance parameters. We shall discuss three approaches. The first one is based on known margins, the second and last one, obtained when margins are unknown, are based respectively on pseudo-likelihood and rank-likelihood. We shall compare the Bayes risk with the quadratic loss for each of the three estimators, this is the aim of the paper. The choice of the Farlie-Gumbel-Morgenstern family is motivated by the fact that we can obtain a closed-form expression involving a polynomial for each of the estimators.

The first estimator is the Bayes estimator given that we know the margins F and G. This is not a realistic approach, but we want to compare the Bayes risk in the best possible case (when we know F and G) with the one obtained in a realistic case when we do not know F and G. When F and G are known, the likelihood based on $(u_i, v_i) = (F(x_i), G(y_i))$, $i = 1, \ldots, n$, is given by

$$L(\theta) = \prod_{i=1}^{n} c_\theta(u_i, v_i).$$

The maximizer $\hat{\theta}_{MLE}$ of L which respect to θ, is the maximum likelihood estimator. Kim et al. (2007) used this approach to compare semi-parametric and parametric methods of estimating copulas, and they called $\hat{\theta}_{MLE}$ the benchmark estimator. The difference between the benchmark estimator and another estimator is the loss when margins are unknown. Hofert et al. (2012) have also considered a parametric estimation method based on the likelihood assuming known margins in inference for Archimedean copulas in high dimensions.

In practice, margins F and G are usually unknown. In the second approach, margins F and G are unknown and estimated by the modified empirical distribution functions F_n and G_n given by

$$F_n(x) = \frac{1}{n+1} \sum_{i=1}^{n} \mathbb{1}(X_i \leq x) \quad \text{and} \quad G_n(y) = \frac{1}{n+1} \sum_{i=1}^{n} \mathbb{1}(Y_i \leq y). \tag{1.3}$$

The random variables $F_n(X_i)$ and $G_n(Y_i)$, $i = 1, \ldots, n$, are called (in the literature) *pseudo-observations* and they take the values $1/(n+1), 2/(n+1), \ldots, n/(n+1)$ (instead of $1/n, 2/n, \ldots, 1$ as in Deheuvels (1979)) to avoid boundary problems. They have been used in several problems including estimation of copulas, see Genest et al. (1995), Panchenko (2005), Scaillet et al. (2007), Grønneberg and Hjort (2014). As described by Segers (2006), the procedure for estimation of the parameter θ in the literature works in two steps (see for example Genest et al. (1995) and the references therein) :

1. Estimate F and G by F_n and G_n, respectively.
2. Pretend that the likelihood is given by the pseudo-likelihood defined by

$$L_{pseudo}(\theta) = \prod_{i=1}^{n} c_\theta(\widehat{u_i}, \widehat{v_i}), \quad (\widehat{u_i}, \widehat{v_i}) = (F_n(x_i), G_n(y_i)),\ i = 1, \ldots, n.$$

In our paper, we shall proceed in the following way. We perform step 1. In step 2, we use the standard Bayesian methodology on L_{pseudo} assuming that L_{pseudo} is the likelihood. The obtained estimator will be called pseudo-Bayes estimator. This is an ad hoc method with no theoretical justification, especially when the sample size is small. In general, the pseudo-likelihood estimator is not efficient (the variance is not minimal) and the pseudo-Bayes estimator is not a Bayes estimator. This approach was used by Grønneberg and Hjort (2014) in the context of model selection.

The pseudo-Bayes estimator will depend on rank statistics. Therefore, it becomes natural to work with rank statistics. We use an approach, the third one, where margins are unknown and inference is based on the ranks $\{(\text{rank}(X_i), \text{rank}(Y_i)),\ i = 1, \ldots, n\}$. The idea of rank-likelihood has been used in a different way by Hoff (2007) for analysis of mixed continuous and discrete data. He has estimated the parameters of the Gaussian copula via a Markov chain Monte Carlo algorithm based on Gibbs sampling. The law of these ranks depends on the distribution function H through the copula C_θ only, in other words, it does not depend on the margins. In general, the calculation of the distribution of the rank statistics will involve multiple integration of the order $2n$. This is a severe limitation in practice. However, for the Farlie-Gumbel-Morgenstern family of copulas, we can evaluate the multiple integration. With this approach, we are going to improve the efficiency of estimation, compared to the one based on pseudo-observations, in terms of the Bayes risk. It can be used on any problem with non-parametric estimation of margins. In the literature, most results are asymptotic. We work with fixed sample sizes and particularly with small sample sizes.

It is expected, using the Bayes risk criteria, that the best estimator is given when margins are known, the second best estimator is the Bayes estimator obtained from the rank-likelihood and the worst estimator is the pseudo-Bayes estimator. Actually, the pseudo-likelihood approach produces an estimator based on the rank statistics, so it cannot give a better Bayes risk than the Bayes estimator based on the rank statistics.

The rest of this paper is organized as follows. In Section 1.2, we derive the distribution of ranks by using the group of permutations on the set $\{1, \ldots, n\}$. In Section 1.3, we apply the results of Section 1.2 in the case of the FGM Copula. In Section 1.4, we compute the estimators using a Bayesian procedure. In Section 1.5, we compare estimators in general using the Bayes risk. In Section 1.6, we provide simulation results for the three approaches.

1.2. The rank-likelihood

The random vectors $\{(X_i, Y_i), i = 1, \ldots, n\}$ are independent and have the same distribution H. The aim of this section is to find the probability distribution of the ranks, which will be used in the third approach described above. For that, we will use the group of permutations on the set $M_n = \{1, \ldots, n\}$, named S_n and characterized in the following way. The elements $\sigma \in S_n$ are the bijections $\sigma \colon S_n \to S_n$ and the group law (denoted $\circ$) is the composition of functions ($\sigma \circ \tau(i) = \sigma(\tau(i))$, $i = 1, \ldots, n$, with $\sigma, \tau \in S_n$). The identity element $e \in S_n$ satisfies $e \circ \sigma = \sigma = \sigma \circ e$ for all $\sigma \in S_n$ and is given by the identity function on M_n. The inverse of σ, $\sigma \in S_n$, is denoted by σ^{-1} and satisfies $\sigma \circ \sigma^{-1} = e = \sigma^{-1} \circ \sigma$, that is $\sigma^{-1}(\sigma(i)) = i$, $i = 1, \ldots, n$. We shall use the following notation

$$\sigma = (\sigma(1), \ldots, \sigma(n)), \quad \sigma \in S_n.$$

For instance, a particular permutation of the set $M_4 = \{1, 2, 3, 4\}$ can be written as $\sigma = (1, 4, 2, 3)$. This means that σ satisfies $\sigma(1) = 1$, $\sigma(2) = 4$, $\sigma(3) = 2$, and $\sigma(4) = 3$. The identity element is $e = (1, \ldots, 4)$ and $\sigma^{-1} = (1, 3, 4, 2)$.

Let $\mathcal{D}_n$ be the set of vectors in $\mathbb{R}^n$ having no ties. In other words,

$$\mathcal{D}_n = \{x \in \mathbb{R}^n \colon x_{(1)} < \cdots < x_{(n)}\}.$$

Let

$$r_x = (r_x(1), \ldots, r_x(n)), \quad \text{with} \quad r_x(i) = \text{rank}(x_i) \quad \text{and} \quad x \in \mathcal{D}_n,$$

where $\text{rank}(x_i)$ is the rank of x_i among $x_1, \ldots, x_n$. Notice that if $x \in \mathcal{D}_n$, then $r_x \in S_n$. Consider the statistics $R(X,Y)$ and $S(X,Y)$, where

$$R(x,y) = \begin{cases} (r_x, r_y) & \text{if} \quad x, y \in \mathcal{D}_n, \\ (e,e) & \text{if} \quad (x,y) \in \mathbb{R}^{2n} \setminus \mathcal{D}_n^2 \end{cases} \qquad \text{and} \qquad S(x,y) = r_y \circ r_x^{-1} =: s_{xy} \in S_n.$$

Notice that if $x, y \in \mathcal{D}_n$, then $\{\big(r_x(i), r_y(i)\big) \colon i = 1, \ldots, n\} = \{\big(i, s_{xy}(i)\big) \colon i = 1, \ldots, n\}$. For example, if $r_x = (1,4,2,3)$ and $r_y = (2,1,4,3)$, then $r_x^{-1} = (1,3,4,2)$, $s_{xy} = (2,4,3,1)$ and

$$\begin{aligned} \{\big(r_x(i), r_y(i)\big) \colon i = 1, \ldots, 4\} &= \{(1,2),(4,1),(2,4),(3,3)\} \\ &= \{(1,2),(2,4),(3,3),(4,1)\} \\ &= \{\big(i, s_{xy}(i)\big) \colon i = 1, \ldots, 4\}. \end{aligned}$$

The following proposition, which is the main result of this section, gives the probability of the event $\{S(X,Y) = s\}$, $s \in S_n$. The statistic $S(X,Y)$ represents the ranks of Y presented in order of the X-values.

Proposition 1.1. *Let c be the density related to the copula C. We have,*

$$\begin{aligned} P(S(X,Y) = s) &= n! P(R(U,V) = (e,s)) \\ &= \frac{1}{n!} E\left(\prod_{i=1}^{n} c\left(\sum_{j_1=1}^{i} W_{1,j_1}, \sum_{j_2=1}^{s(i)} W_{2,j_2} \right) \right), \quad s \in S_n, \end{aligned}$$

where $(W_{i,1}, \ldots, W_{i,n+1})$, $i = 1,2$ are i.i.d. random vectors, $(W_{i,1}, \ldots, W_{i,n+1}) \sim$ Dirichlet$(1, \ldots, 1)$.

Proof. Let $s \in S_n$ and $E_s = \{(u,v) \in [0,1]^{2n} \cap \mathcal{D}_n^2 \colon R(u,v) = (e,s)\}$. We have

$$\begin{aligned} P(S(X,Y) = s) &= P(R(X,Y) \in \{(t, s \circ t), t \in S_n\}) \\ &= n! P(R(X,Y) = (e,s)) \\ &= n! P(R(U,V) = (e,s)) \end{aligned}$$

$$
\begin{aligned}
&= n! \int_{E_s} \prod_{i=1}^{n} c(u_i, v_i)\, du_i dv_i \\
&= \frac{1}{n!} \int_{(0,1)^{2n}} \prod_{i=1}^{n} c(u_{(i)}, v_{(s(i))})\, du_i dv_i \\
&= \frac{1}{n!} E\Big[\prod_{i=1}^{n} c(U_{(i)}, V_{(s(i))})\Big], \quad U_1, \ldots, U_n, V_1, \ldots, V_n \text{ iid } \mathcal{U}(0,1) \\
&= \frac{1}{n!} E\Big(\prod_{i=1}^{n} c(\sum_{j_1=1}^{i} W_{1,j_1}, \sum_{j_2=1}^{s(i)} W_{2,j_2})\Big), \quad s \in S_n.
\end{aligned}
$$

$(W_{1,1}, \ldots, W_{1,n+1})$ being the spacings on $[0,1]$ based on $U_1, \ldots, U_n$ and $(W_{2,1}, \ldots, W_{2,n+1})$ being the spacings on $[0,1]$ based on $V_1, \ldots, V_n$. The random vectors $W_i = (W_{i,1}, \ldots, W_{i,n})$, $i = 1, 2$ are independent and uniformly distributed on the standard n-simplex

$$\Delta_n = \{(z_1, \ldots, z_n)^T \in [0,1]^n \colon \sum_{i=1}^{n} z_i \leq 1\}.$$

□

The pseudo-observations are the random variables given by $(F_n(X_i), G_n(Y_i)), i = 1, \cdots, n$, with F_n and G_n given in Equation (1.3). The approach with the pseudo-likelihood uses the likelihood of the copula and does as if the pseudo-observations represent a sample from the copula. Pseudo-observations are a function of the ranks. Our approach is to build an estimator based on ranks by using the rank-likelihood, defined in Proposition 1.1. We show, in the following section, how to calculate the rank-likelihood in the case where the model is given by a copula of the Farlie-Gumble-Morgenstein family.

1.3. Rank-likelihood: FGM family of copulas

In using the notations of Section 1.2, the rank-likelihood is defined as in Proposition 1.1 by

$$L_{rank}(\theta) = P_\theta(S(X,Y) = s) = n! P_\theta(R(U,V) = (e,s)).$$

Let us compute $Q(s) = P(R(U,V) = (e,s))$ for the parametric family of copulas called, the Farlie-Gumble-Morgenstein family of copulas and defined in Equation (1.1). The density is given by

$$c_\theta(u,v) = \frac{\partial^2 C_\theta(u,v)}{\partial u \partial v} = 1 + \theta(2u-1)(2v-1), \quad \text{with} \quad \theta \in [-1,1].$$

The probabilities $Q(s)$ are given by polynomials in θ. In the following we shall derive the coefficients of the polynomials.

Let s be a permutation on the set $M_n = \{1, 2, \ldots, n\}$ and let s^{-1} be the inverse of the permutation s. Let $0 < u_1 < u_2 < \cdots < u_n < 1$ and $0 < v_{s^{-1}(1)} < v_{s^{-1}(2)} < \cdots < v_{s^{-1}(n)} < 1$. Consider the spacings $t_1, \cdots, t_{n+1}$ and $w_1, \cdots, w_{n+1}$ given by

$$t_1 = u_1,\ w_1 = v_{s^{-1}(1)},\ t_{n+1} = 1 - u_n,\ w_{n+1} = 1 - v_{s^{-1}(n)}$$

and

$$t_i = u_i - u_{i-1},\ w_i = v_{s^{-1}(i)} - v_{s^{-1}(i-1)},\ \text{for } 2 \le i \le n.$$

We have

$$u_i = \sum_{j=1}^{i} t_j, \quad (1-u_i) = \sum_{j=i+1}^{n+1} t_j, \quad v_i = \sum_{j=1}^{s(i)} w_j \quad \text{and} \quad (1-v_i) = \sum_{j=s(i)+1}^{n+1} w_j$$

which gives

$$(2u_i - 1) = u_i - (1-u_i) = \sum_{j=1}^{i} t_j + \sum_{j=i+1}^{n+1} (-1) t_j = \sum_{j=1}^{n+1} (-1)^{\mathbb{1}(j>i)} t_j$$

and

$$(2v_i - 1) = v_i - (1-v_i) = \sum_{j=1}^{s(i)} w_j + \sum_{j=s(i)+1}^{n+1} (-1) w_j = \sum_{j=1}^{n+1} (-1)^{\mathbb{1}(j>s(i))} w_j.$$

Thus,

$$\begin{aligned}
\prod_{i=1}^{n} \{1 + \theta(2u_i-1)(2v_i-1)\} &= \prod_{i=1}^{n} \left\{ 1 + \theta \left(\sum_{j=1}^{n+1} (-1)^{\mathbb{1}(j>i)} t_j \right) \left(\sum_{j=1}^{n+1} (-1)^{\mathbb{1}(j>s(i))} w_j \right) \right\} \\
&= 1 + \sum_{\ell=1}^{n} \theta^\ell \sum_{1 \le i_1 < i_2 < \cdots < i_\ell \le n} b_{\ell,n}(i_1, \ldots, i_\ell; t) b_{\ell,n}(s(i_1), \ldots, s(i_\ell); w),
\end{aligned}$$

with

$$b_{\ell,n}(i_1,\ldots,i_\ell;t)=\sum_{j_1=1}^{n+1}\cdots\sum_{j_\ell=1}^{n+1}(-1)^{\mathbb{1}(j_1>i_1)+\cdots+\mathbb{1}(j_\ell>i_\ell)}t_{j_1}\cdots t_{j_\ell}.$$

Let $S=\{t\in(0,1)^n\colon t_1+\cdots+t_n\le 1\}$, we have

$$\begin{aligned}Q(s)=P(R(U,V)=(e,s))&=\int_S\int_S\prod_{i=1}^{n}\{1+\theta(2u_i-1)(2v_i-1)\}dtdw\\&=\frac{1}{[n!]^2}+\sum_{\ell=1}^{n}\frac{1}{[(n+\ell)!]^2}\theta^\ell\sum_{1\le i_1<i_2<\cdots<i_\ell\le n}c_{\ell,n}(i_1,\ldots,i_\ell)c_{\ell,n}(s(i_1),\ldots,s(i_\ell)),\end{aligned}$$

with

$$\begin{aligned}&c_{\ell,n}(i_1,\ldots,i_\ell)\\&\quad=\sum_{j_1=1}^{n+1}\cdots\sum_{j_\ell=1}^{n+1}(-1)^{\mathbb{1}(j_1>i_1)+\cdots+\mathbb{1}(j_\ell>i_\ell)}\left\{\sum_{k=1}^{\ell}\mathbb{1}(j_k=1)\right\}!\times\cdots\times\left\{\sum_{k=1}^{\ell}\mathbb{1}(j_k=n+1)\right\}!.\end{aligned}$$

Note that $c_{n,n}(1,2,\ldots,n)=0$ for all $n\ge 1$. Indeed, $Q(s)$ is a polynomial function in θ that can be written

$$Q(s)=\sum_{j=0}^{n}a_{j,n}(s)\theta^j=\sum_{j=0}^{n-1}a_{j,n}(s)\theta^i+\frac{1}{[(2n)!]^2}c_{n,n}^2(1,2,\ldots,n)\theta^n,\quad\text{for all }\ \theta\in[-1,1].$$

When s varies, $n!Q(s)$ gives probabilities. Thus, $\sum_{s\in S_n}n!Q(s)=1$ and

$$\frac{1}{n!}=\sum_{s\in S_n}Q(s)=\sum_{j=0}^{n-1}\sum_{s\in S_n}a_{j,n}(s)\theta^j+n!\Big(\frac{c_{n,n}(1,2,\ldots,n)}{(2n)!}\Big)^2\theta^n.$$

This proves that $c_{n,n}^2(1,2,\ldots,n)=0$.

We will find the estimators for the three approaches in the following section.

1.4. Calculation of estimators according to the Bayesian approach

The Bayesian approach to estimation starts with a prior distribution on the parameter of interest, θ here. There exists two types of priors: informative and non-informative. A non-informative prior provides little information relative to the experiment and expresses vague or general information about the parameter. Informative prior distributions, on the other hand, summarize the evidence about the parameters concerned from many sources and often have a considerable impact on the results. It expresses specific and definite information about the parameter.

We shall estimate Legendre's polynomial given by $P_2(\theta) = (3\theta^2 - 1)/2$ with a uniform prior on parameter θ. The calculation is also done for the Jeffreys prior.

1.4.1. Uniform prior

The parameter θ is supposed to be uniformly distributed on $[-1, 1]$, so θ is a random variable with the density function $\pi(\theta) = \frac{1}{2}\mathbb{1}_{[-1,1]}(\theta)$.

1.4.1.1. *First approach: known margins*

When margins are known, the observations $(u_i, v_i) = (F(x_i), F(y_i)),\ i = 1 \ldots n$, come directly from the copula C_θ and we can use the standard Bayesian approach. The likelihood is given by

$$L(\theta) = \prod_{i=1}^{n} c_\theta(u_i, v_i) = \prod_{i=1}^{n}(1 + \theta(2u_i - 1)(2v_i - 1)) = \sum_{i=0}^{n} A_i \theta^i,$$

where

$$A_i = \begin{cases} 1 & \text{if } i = 0, \\ \sum_{1 \leq k_1 < \cdots < k_i \leq n} \prod_{t \in \{k_1, \ldots, k_i\}} (2u_t - 1)(2v_t - 1) & \text{otherwise,} \end{cases}$$

and under quadratic loss function, the Bayes estimator of $P_2(\theta)$ is given by the posterior mean, that is

$$\begin{aligned} \widehat{P_2(\theta)} &= E[P_2(\theta) | (U_i, V_i) = (u_i, v_i),\ i = 1, \ldots, n] \\ &= \int_{-1}^{1} \prod_{i=1}^{n} c_\theta(u_i, v_i) P_2(\theta) \pi(\theta) d\theta \bigg/ \int_{-1}^{1} \prod_{i=1}^{n} c_\theta(u_i, v_i) \pi(\theta) d\theta \\ &= \sum_{k=0}^{n} \frac{A_k}{4} \int_{-1}^{1} \theta^k (3\theta^2 - 1) d\theta \bigg/ \sum_{k=0}^{n} \frac{A_k}{2} \int_{-1}^{1} \theta^k d\theta \\ &= \sum_{k=0,\ k \text{ even}}^{n} \frac{k}{(k+1)(k+3)} A_k \bigg/ \sum_{k=0,\ k \text{ even}}^{n} \frac{1}{k+1} A_k. \end{aligned}$$

1.4.1.2. *Second approach: unknown margins and pseudo-observations*

When margins are unknown, a basic approach is to replace the likelihood function with a substitution (the pseudo-likelihood), while specifying a prior for the parameters in the pseudo-likelihood. In this approach, $(F(X_i), G(Y_i))$, $i = 1, \ldots, n$, are estimated by their modified empirical counterpart given by

$$\left(\frac{r_X(i)}{n+1}, \frac{r_Y(i)}{n+1}\right), \quad i = 1, \ldots, n.$$

Since

$$\{\left(r_x(i), r_y(i)\right) : i = 1, \ldots, n\} = \{\left(i, s_{xy}(i)\right) : i = 1, \ldots, n\},$$

the likelihood is based on

$$\left(\frac{i}{n+1}, \frac{s_{xy}(i)}{n+1}\right), \quad i = 1, \ldots, n,$$

where s_{xy} is defined in Section 1.2 and will be noted s below. The substitution (called pseudo-likelihood) is given by

$$L_{pseudo}(\theta) = \prod_{i=1}^{n} c_\theta\left(\frac{i}{n+1}, \frac{s(i)}{n+1}\right) = \prod_{i=1}^{n}\left(1 + \theta\left(2\frac{i}{n+1} - 1\right)\left(2\frac{s(i)}{n+1} - 1\right)\right) = \sum_{i=0}^{n} D_i\theta^i,$$

where

$$D_i = \begin{cases} 1 & \text{if } i = 0, \\ \sum_{1\le k_1 < \cdots < k_i \le n} \prod_{t\in\{k_1,\ldots,k_i\}} \left(\frac{2t}{n+1} - 1\right)\left(\frac{2s(t)}{n+1} - 1\right) & \text{otherwise.} \end{cases}$$

The use of the pseudo-likelihood function in semi-parametric problems or for elimination of nuisance parameters is widely shared in the literature; see Ventura et al. (2009), Ventura et al. (2010) and the references therein.

The same calculations as in the first case lead to an estimator which is not really a Bayes' estimator because of the use of pseudo-likelihood. It is a pseudo-Bayes' estimator and is given by

$$\widehat{P_2(\theta)} = \sum_{k=0,\ k \text{ even}}^{n} \frac{k}{(k+1)(k+3)} D_k \Big/ \sum_{k=0,\ k \text{ even}}^{n} \frac{1}{k+1} D_k.$$

1.4.1.3. *Third approach: unknown margins and the likelihood of the rank statistics*

The rank-likelihood is given by

$$L_{rank}(\theta) = P(S = s) = n!Q(s) = \sum_{k=0}^{n} B_k \theta^k,$$

where

$$B_k = \begin{cases} \frac{1}{n!} & \text{if } k = 0, \\ \frac{n!}{((n+k)!)^2} \sum_{1 \le i_1 < \ldots < i_k \le n} c_{k,n}(i_1, \ldots, i_k) \times c_{k,n}(s(i_1), \ldots, s(i_k)) & \text{otherwise,} \end{cases}$$

with the function $c_{k,n}$ defined in Section 1.2. The Bayes estimator obtained with this approach using the same calculations as before is given by

$$\widehat{P_2(\theta)} = \sum_{k=0,\, k \text{ even}}^{n} \frac{k}{(k+1)(k+3)} B_k \Bigg/ \sum_{k=0,\, k \text{ even}}^{n} \frac{1}{k+1} B_k.$$

1.4.2. The Jeffreys' prior

One of the rules for selecting priors suggested in the literature is Jeffreys' rule. The Jeffreys prior, considered as a default prior, is a non-informative prior distribution on the parameter space that is proportional to the square root of the determinant of the Fisher information. In our study, which is a one-dimensional parameter case,

$$\pi(\theta) \propto \sqrt{I(\theta)},$$

where $I(\theta)$ is the Fisher information defined by

$$I(\theta) = E\left(\frac{\partial}{\partial\theta} \log c_\theta(U,V)\right)^2, \quad (U,V) \sim C_\theta,$$

and calculated as follows

$$\begin{aligned} I(\theta) &= \int_0^1 \int_0^1 \frac{(2u-1)^2(2v-1)^2}{1+\theta(2u-1)(2v-1)} du dv \\ &= \frac{1}{4} \int_{-1}^1 \int_{-1}^1 \frac{x^2 y^2}{1+\theta xy} dx dy \end{aligned}$$

$$= \frac{1}{4}\int_{-1}^{1} y^2 \left(\int_{-1}^{1} \frac{x^2}{1+\theta xy} dx\right) dy.$$

If $\theta = 0$, then

$$I(\theta) = \frac{1}{4}\int_{-1}^{1}\int_{-1}^{1} x^2 y^2 dx dy = \left(\frac{1}{2}\int_{-1}^{1} x^2 dx\right)^2 = \frac{1}{9}.$$

If $\theta \neq 0$ and $y \neq 0$, then since $0 \leq 1 + \theta xy$, we have

$$\begin{aligned}
\int_{-1}^{1} \frac{x^2}{1+\theta xy} dx &= \int_{-1}^{1} \frac{1}{(\theta y)^2} \frac{(\theta xy)^2}{1+\theta xy} dx \\
&= \int_{-1}^{1} \frac{1}{(\theta y)^2} \frac{(1+\theta xy - 1)^2}{1+\theta xy} dx \\
&= \frac{1}{(\theta y)^2}\left\{\int_{-1}^{1} (1+\theta xy) dx - 2\int_{-1}^{1} dx + \int_{-1}^{1} \frac{1}{1+\theta xy} dx\right\} \\
&= \frac{1}{(\theta y)^3} \log\left(\frac{1+\theta y}{1-\theta y}\right) - \frac{2}{(\theta y)^2}.
\end{aligned}$$

For $\theta \neq 0$, the function

$$y^2\left(\int_{-1}^{1} \frac{x^2}{1+\theta xy} dx\right) = \begin{cases} 0 & \text{if } y = 0, \\ \frac{1}{\theta^3 y} \log\left(\frac{1+\theta y}{1-\theta y}\right) - \frac{2}{\theta^2} & \text{otherwise,} \end{cases}$$

is continuous on $[-1, 1]$. Finally,

$$\begin{aligned}
I(\theta) &= \frac{1}{4}\int_{-1}^{1} \left\{\frac{1}{\theta^3 y} \log\left(\frac{1+\theta y}{1-\theta y}\right) - \frac{2}{\theta^2}\right\} dy \\
&= \frac{1}{2\theta^3}\int_{0}^{1} \frac{1}{y} \log\left(\frac{1+\theta y}{1-\theta y}\right) dy - \frac{1}{\theta^2} \\
&= \frac{1}{2\theta^3}\left\{\int_{0}^{1} \frac{\log(1+\theta y)}{y} dy - \int_{0}^{1} \frac{\log(1-\theta y)}{y} dy\right\} - \frac{1}{\theta^2} \\
&= \frac{1}{2\theta^3}\{\mathrm{LI}_2(\theta) - \mathrm{LI}_2(-\theta)\} - \frac{1}{\theta^2},
\end{aligned}$$

where

$$\mathrm{LI}_2(\theta) = -\int_{0}^{1} \frac{\log(1-\theta y)}{y} dy, \quad \theta \in [-1, 1],$$

is the dilogarithm function. See Zagier (2007) for some properties of this function. The Fisher information is given by

$$I(\theta) = \begin{cases} \frac{1}{9} & \text{if } \theta = 0, \\ \frac{1}{\theta^2}\left[\frac{\mathrm{LI}_2(\theta) - \mathrm{LI}_2(-\theta)}{2\theta} - 1\right] & \text{if } \theta \in [-1,0) \cup (0,1]. \end{cases}$$

One can use series representation,

$$\mathrm{LI}_2(\theta) = \sum_{n=1}^{\infty} \frac{\theta^n}{n^2}, \quad |\theta| < 1,$$

to obtain

$$I(\theta) = \sum_{n=0}^{\infty} \frac{\theta^{2n}}{(2n+3)^2}, \quad 0 < |\theta| < 1.$$

The function I is symmetric and increasing on $[0,1)$,

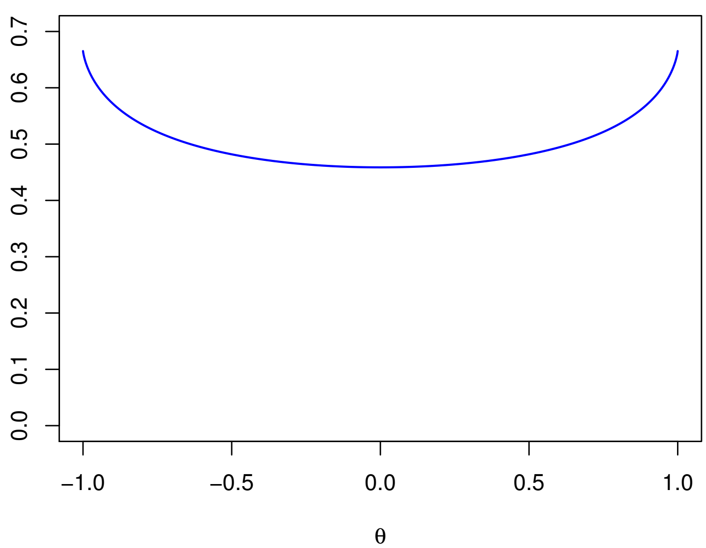
$$I(0) = \frac{1}{9} \quad \text{and} \quad \lim_{\theta \longrightarrow 1} I(\theta) = \frac{\pi^2}{8} - 1.$$

Thus, the function $\sqrt{I(\theta)}$ is integrable on $(-1,1)$. The prior density of θ is given by

$$\pi(\theta) = \frac{1}{c}\sqrt{I(\theta)}, \quad \text{with } c = \int_{-1}^{1} \sqrt{I(\theta)}d\theta,$$

and its curve is given in Figure 1.1.

Figure 1.1. The Jeffreys' prior density function.



It is more difficult in this case to calculate the estimators. Nevertheless, it can be done numerically. The Bayes estimator in the first approach is given by

$$\begin{aligned}\widehat{P_2(\theta)} &= E[P_2(\theta)|(U_i,V_i)=(u_i,v_i),\ i=1,\dots,n] \\ &= \int_{-1}^{1}\prod_{i=1}^{n} c_\theta(u_i,v_i)P_2(\theta)\pi(\theta)d\theta \Big/ \int_{-1}^{1}\prod_{i=1}^{n} c_\theta(u_i,v_i)\pi(\theta)d\theta \\ &= \sum_{k=0}^{n}\left(\int_{-1}^{1}\theta^k P_2(\theta)\pi(\theta)d\theta\right)A_k \Big/ \sum_{k=0}^{n}\left(\int_{-1}^{1}\theta^k \pi(\theta)d\theta\right)A_k.\end{aligned}$$

The same calculation is done for other approaches and we have

$$\widehat{P_2(\theta)} = \sum_{k=0}^{n} a_{k,\pi}E_k \Big/ \sum_{k=0}^{n} b_{k,\pi}E_k,$$

where

$$a_{k,\pi} = \int_{-1}^{1}\theta^k P_2(\theta)\pi(\theta)d\theta \quad \text{and} \quad b_{k,\pi} = \int_{-1}^{1}\theta^k\pi(\theta)d\theta,$$

with

$$E_i = \begin{cases} A_i & \text{in approach 1,} \\ B_i & \text{in approach 3,} \\ D_i & \text{in approach 2.} \end{cases}$$

1.5. Comparison of estimators

We consider a quadratic loss function. In order to compare the estimators, we use the frequentist risk and the Bayes risk. The frequentist risk is defined by

$$R(\theta,\hat{\theta}) = E_\theta[(\theta-\hat{\theta})^2].$$

The Bayes risk is given by

$$r(\pi,\hat{\theta}) = \int_\theta R(\theta,\hat{\theta})d\pi(\theta).$$

It is easy to see that in general the estimator with known margins is the best, the second best estimator is the rank-likelihood estimator and the worst estimator being the pseudo-likelihood estimator when the Bayes risk is used. In fact, suppose that the estimators of a function $g(\theta)$ are given by the posterior mean according to these three approaches. Let us call them $\widehat{g}$, $\widehat{g_0}$ and $\widehat{g_1}$. We have

$$\begin{aligned} r(\pi, \widehat{g_0}) &= E\Big[E_\theta[(g(\theta) - \widehat{g_0})^2]\Big] \\ &= E\Big[E_\theta[(g(\theta) - \widehat{g})^2 + (\widehat{g} - \widehat{g_0})^2 + 2(g(\theta) - \widehat{g})(\widehat{g} - \widehat{g_0})]\Big] \\ &= E\Big[E_\theta[(g(\theta) - \widehat{g})^2]\Big] + E\Big[E_\theta[(\widehat{g} - \widehat{g_0})^2]\Big] \\ &\geq r(\pi, \widehat{g}), \end{aligned}$$

since

$$\begin{aligned} E\Big[E_\theta[(g(\theta) - \widehat{g})(\widehat{g} - \widehat{g_0})]\Big] &= E\Big[E[(g(\theta) - \widehat{g})(\widehat{g} - \widehat{g_0})|(U_i, V_i),\ i = 1, \ldots, n]\Big] \\ &= E\Big[(E[g(\theta)|(U_i, V_i),\ i = 1, \ldots, n] - \widehat{g})(\widehat{g} - \widehat{g_0})\Big] \\ &= 0. \end{aligned}$$

In the same way, since

$$\begin{aligned} E\Big[E_\theta[(g(\theta) - \widehat{g_0})(\widehat{g_0} - \widehat{g_1})]\Big] &= E\Big[E[(g(\theta) - \widehat{g_0})(\widehat{g_0} - \widehat{g_1})|S]\Big] \\ &= E\Big[(E[g(\theta)|S] - \widehat{g_0})(\widehat{g_0} - \widehat{g_1})\Big] \\ &= 0, \end{aligned}$$

we have

$$r(\pi, \widehat{g_1}) \geq r(\pi, \widehat{g_0}).$$

Thus,

$$r(\pi, \widehat{g}) \leq r(\pi, \widehat{g_0}) \leq r(\pi, \widehat{g_1}).$$

Hence, according to the Bayes risk, the approach with known margins is the best and the approach with ranks is better than the one with pseudo-observations.

Why are we focusing on the estimation of Legendre's polynomials?

The rank-likelihood and the pseudo-likelihood are both polynomials with respect to the parameter θ. The three first coefficients of the pseudo-likelihood are given by

$$\begin{aligned} D_1 &= 1, \\ D_2 &= \sum_{t=1}^{n} \left(\frac{2t}{n+1} - 1\right)\left(\frac{2s(t)}{n+1} - 1\right) = \frac{1}{(n+1)^2} \sum_{t=1}^{n} h_n(k) \times h_n(s(k)) \quad \text{and} \\ D_3 &= \sum_{1 \le k_1 < k_2 \le n} \prod_{t \in \{k_1, k_2\}} \left(\frac{2t}{n+1} - 1\right)\left(\frac{2s(t)}{n+1} - 1\right), \end{aligned}$$

where h_n is a function on $\{1, 2, \ldots, n+1\}$ defined by $h_n(i) = 2i - (n+1)$. The first three coefficients of the rank-likelihood are given by

$$\begin{aligned} B_1 &= \frac{1}{n!}, \\ B_2 &= \frac{n!}{((n+1)!)^2} \sum_{i=1}^{n} c_{1,n}(i) \times c_{1,n}(s(i)) = \frac{1}{n!(n+1)^2} \sum_{t=1}^{n} h_n(k) \times h_n(s(k)) \quad \text{and} \\ B_3 &= \frac{n!}{((n+2)!)^2} \sum_{1 \le i_1 < i_2 \le n} c_{2,n}(i_1, i_2) \times c_{2,n}(s(i_1), s(i_2)). \end{aligned}$$

Note that the first two coefficients are proportional according to the following formulas

$$D_i = n! B_i, \quad i = 1 \text{ and } 2,$$

and they are respectively dominant among other coefficients, which are different. The estimator of $P_2(\theta)$ is given by its posterior mean, which involves the integration of the product of $P_2(\theta)$ and the likelihood. The first two terms of the likelihood are multiple of Legendre's polynomials $P_0(\theta) = 1$ and $P_1(\theta) = \theta$, which are all members of a family of orthogonal polynomials. The contribution of the two first coefficients of the likelihood will be cancelled. Thus, it will allow us to see the behaviour of the other coefficients when estimating a function of the parameter θ.

1.6. Simulation studies and results

In order to compare the performance of the three estimators obtained in Section 1.4 by the three approaches, we carried out simulations using artificial data. We did it for small sample sizes $n = 4, 5, \ldots, 9$. All our simulations were completed using R 3.2.1. The frequentist risk, which is the mean squared error, of every estimator is computed numerically by generating $m = 1000$ data sets of samples from the copula C_θ. The algorithm to compute the estimator is given by:

1) Fix a value for θ, $\theta \in [-1, 1]$, and draw observations (u_i, v_i), $i = 1, \ldots, n$, from the copula C_θ.

2) Compute $s = r_v \circ r_u^{-1}$, and the estimator $\widehat{P_2(\theta)}$.

3) Repeat steps 1) and 2) m times.

4) Approximate the frequentist risk function with

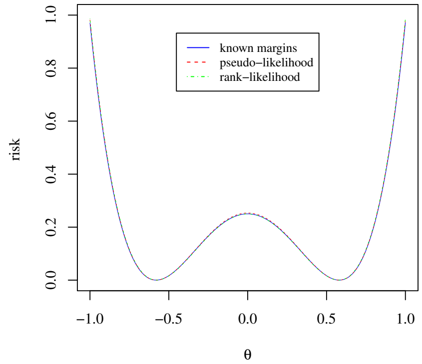
$$\text{risk}(\theta) = \frac{1}{m} \sum_{i=1}^{m} (\widehat{P_2(\theta)} - P_2(\theta))^2.$$

With Jeffreys' prior, we need to compute the coefficients $a_{k,\pi}$ and $b_{k,\pi}$ for $k = 1, \ldots, n$. These coefficients are given in the table 1. I. We find that $a_{0,\pi} = 0.0401$, $b_{0,\pi} = 1$ and $a_{k,\pi} = 0$, $b_{k,\pi} = 0$ for $k \leq n$ odd.

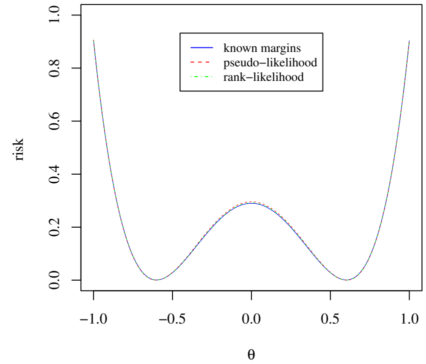
Table 1. I. Coefficients $a_{k,\pi}$ and $b_{k,\pi}$.

n	k	2	4	6	8
4	$a_{k,\pi}$	0.1567	0.1341		
	$b_{k,\pi}$	0.3601	0.2245		
5	$a_{k,\pi}$	0.1567	0.1342		
	$b_{k,\pi}$	0.3601	0.2245		
6	$a_{k,\pi}$	0.1567	0.1342	0.1128	
	$b_{k,\pi}$	0.3601	0.2245	0.1643	
7	$a_{k,\pi}$	0.1567	0.1342	0.1128	
	$b_{k,\pi}$	0.3601	0.2245	0.1643	
8	$a_{k,\pi}$	0.1567	0.1342	0.1128	0.0965
	$b_{k,\pi}$	0.3601	0.2245	0.1643	0.1300

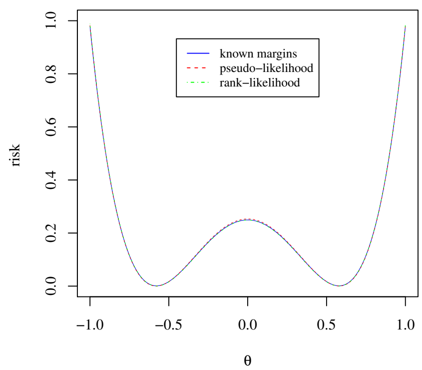
The frequentist risk functions of the three different estimators are plotted on the same graph in Figure 1.2 and 1.3 for each value of n and for uniform and Jeffreys' priors. The



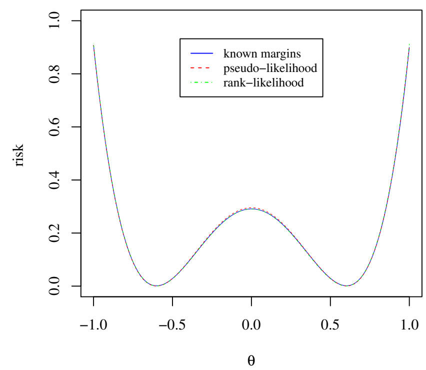
(a) $n = 4$, uniform prior



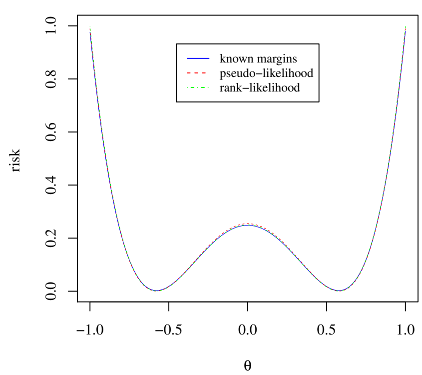
(b) $n = 4$, Jeffreys' prior



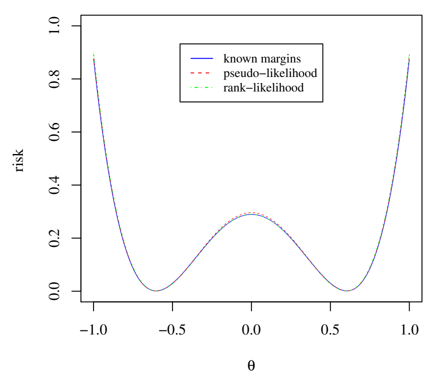
(c) $n = 5$, uniform prior



(d) $n = 5$, Jeffreys' prior

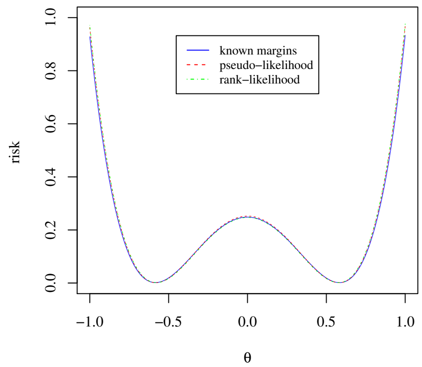


(e) $n = 6$, uniform prior

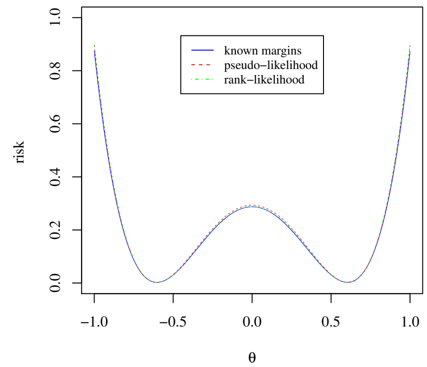


(f) $n = 6$, Jeffreys' prior

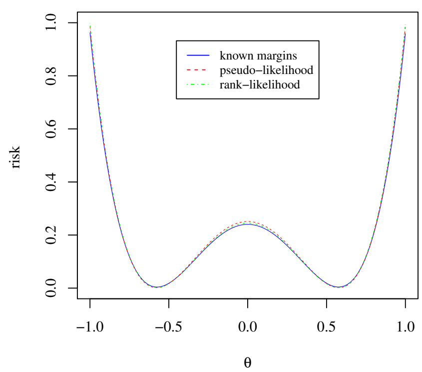
Figure 1.2. Frequentist risk function for n equal to 4 and 6



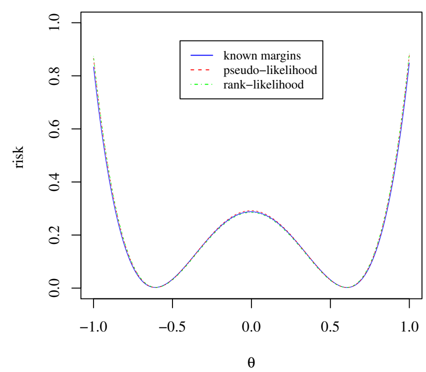
(a) $n = 7$, uniform prior



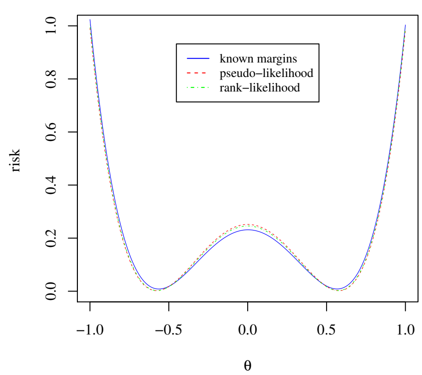
(b) $n = 7$, Jeffreys' prior



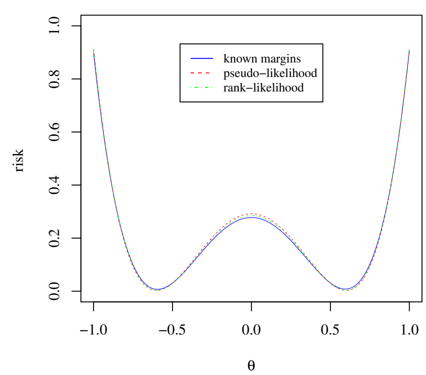
(c) $n = 8$, uniform prior



(d) $n = 8$, Jeffreys' prior



(e) $n = 9$, uniform prior



(f) $n = 9$, Jeffreys' prior

Figure 1.3. Frequentist risk function for n from 7 to 9

Table 1. II. Bayes risk of different estimators in the case of uniform prior.

n	Bayes risk			(c-a)/(b-a)
	known margins (a)	pseudo-likelihood (b)	rank-likelihood (c)	
4	0.1961979	0.1985544	0.198244	0.8683131
5	0.1978087	0.1984503	0.1980236	0.3348736
6	0.1982249	0.2002605	0.2002188	0.979549
7	0.1911326	0.1968416	0.1967608	0.9858371
8	0.1947729	0.199602	0.1998248	1.04614
9	0.2025235	0.2004639	0.2006711	0.8993965

Table 1. III. Bayes risk of different estimators in the case of Jeffreys' prior.

n	Bayes risk			(c-a)/(b-a)
	known margins (a)	pseudo-likelihood (b)	rank-likelihood (c)	
4	0.2108882	0.2124969	0.2121554	0.7877355
5	0.2114897	0.2126018	0.2123315	0.756995
6	0.2068153	0.2107277	0.2104528	0.9297533
7	0.2067585	0.2121606	0.2120593	0.9812447
8	0.2019722	0.2076137	0.207111	0.9108886
9	0.2118729	0.2140115	0.2140251	1.006375

three estimators have frequentist risks almost equal. When the parameter θ is near 0, a slight difference is perceptible. Actually, in this area, the estimator from the first approach is the best and the second best is the estimator based on the rank-likelihood. Tables 1. II and 1. III give the Bayes risk, respectively for the uniform prior and the Jeffreys prior. The Bayes risk is the mean of frequentist risk, and is obtained through a numerical integration on the interval $[-1, 1]$. The rectangle rule is used with subintervals of the same length 0.01. We showed theoretically that according to the Bayes risk that the approach which known margins is the best and the second best is the approach based on the rank-likelihood. We can see it in the tables 1. II and 1. III. Notice that because the risks are too close and because of the random effects simulations, this order is not respected for some values of n.

1.7. Conclusion

We considered three approaches in this work. The first one with known margins and the others with unknown margins. The second approach is based on the pseudo-likelihood and the last approach is based on the rank-likelihood. With a Bayesian procedure we used frequentist risk functions and Bayes risks to compare estimators obtained from the different approaches. It is known from the Bayesian theory that the rank-likelihood estimator from the third approach has a smaller Bayes risk than the pseudo-likelihood estimator from the second approach and the estimator from the first approach is the best regarding the Bayes risk. We derived the rank-likelihood in bivariate case.

In order to see how the rank-likelihood estimator performs compared to others, we used the Farlie-Gumble-Mogenstein family of copulas and the estimation of the Legendre's polynomial of degree two. We used this particular family of copulas since it was easy to make an analytical calculation of the estimator. According to the Bayes risk, the estimator with known margins is the best and the second best is the estimator based on the rank-likelihood.

For the rank-likelihood estimator, the calculations are long and complicated when the sample size, n, increases. For the pseudo-likelihood estimator, the calculations are simple regardless of the value of n. The numerical results are very similar, even for small values of n. We suggest to use the rank-likelihood when the sample size n is very small.

Acknowledgment

The author Romain KADJE KENMOGNE wishes to thank Institut des sciences Mathématiques, Centre de Recherches Mathématiques, Faculté des Études Supérieures et Postdoctorales of the University of Montreal, and Fondation J. Armand Bombardier for their financial support.

The author François Perron wishes to thank NSERC for their financial support.

References

Paul Deheuvels. La fonction de dépendance empirique et ses propriétés. Un test non paramétrique d'indépendance. *Acad. Roy. Belg. Bull. Cl. Sci. (5)*, 65(6):274–292, 1979.

D. J. G. Farlie. The performance of some correlation coefficients for a general bivariate distribution. *Biometrika*, 47:307–323, 1960.

C. Genest, K. Ghoudi, and L.-P. Rivest. A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82(3):543–552, 1995.

Steffen Grønneberg and Nils Lid Hjort. The copula information criteria. *Scand. J. Stat.*, 41(2): 436–459, 2014.

E. J. Gumbel. Bivariate exponential distributions. *J. Amer. Statist. Assoc.*, 55:698–707, 1960.

Marius Hofert, Martin Mächler, and Alexander J. McNeil. Likelihood inference for Archimedean copulas in high dimensions under known margins. *J. Multivariate Anal.*, 110:133–150, 2012.

Peter D. Hoff. Extending the rank likelihood for semiparametric copula estimation. *Ann. Appl. Stat.*, 1(1):265–283, 2007.

Gunky Kim, Mervyn J. Silvapulle, and Paramsothy Silvapulle. Comparison of semiparametric and parametric methods for estimating copulas. *Comput. Statist. Data Anal.*, 51(6):2836–2850, 2007.

Thomas Mikosch. Copulas: tales and facts. *Extremes*, 9(1):3–20, 2006.

Dietrich Morgenstern. Einfache Beispiele zweidimensionaler Verteilungen. *Mitteilungsbl. Math. Statist.*, 8:234–235, 1956.

Valentyn Panchenko. Goodness-of-fit test for copulas. *Phys. A*, 355(1):176–182, 2005.

Olivier Scaillet, Arthur Charpentier, and Jean-David Fermanian. The estimation of copulas: theory and practice. 2007.

Johan Segers. Efficient estimation of copula parameter. Discussion of: "Copulas: tales and facts" [Extremes **9** (2006), no. 1, 3–20;] by T. Mikosch. *Extremes*, 9(1):51–53, 2006.

L. Ventura, S. Cabras, and W. Racugno. Default prior distributions from quasi- and quasi-profile likelihoods. *J. Statist. Plann. Inference*, 140(11):2937–2942, 2010.

Laura Ventura, Stefano Cabras, and Walter Racugno. Prior distributions from pseudo-likelihoods in the presence of nuisance parameters. *J. Amer. Statist. Assoc.*, 104(486):768–774, 2009.

Don Zagier. The dilogarithm function. In *Frontiers in number theory, physics, and geometry II*, pages 3–65. Springer, 2007.

Chapitre 2

SUR LE RAPPORT DE DEUX VARIABLES ALÉATOIRES DE LOIS NORMALES

Cet article sera soumis à la revue *Scandinavian journal of Statistics*.

Les principales contributions de *Romain Kadje Kenmogne* à cet article sont présentées.

– Conduite de la revue de la littérature.

– Contribution dans l'élaboration du théorème 3.1, et des lemmes 2.2 et 3.5.

– Démonstration des lemmes 2.4, 3.7, 4.1 et 4.2.

– Conception, écriture et validation des programmes R.

– Conduite des simulations.

– Réalisation des illustrations graphiques.

– Rédaction partielle de l'article.

On the ratio of two normal variables

Romain Kadje Kenmogne and François Perron

Département de mathématiques et de statistique,

Université de Montréal,

C.P. 6128, Succursale Centre-ville,

Montréal, Canada H3C 3J7

Abstract

Let (X_1, X_2) be a random binormal vector. Let $Y = X_1/X_2$ be the ratio. In this paper, we show that the density of Y is a mixture of densities belonging to a family extending the Student distributions having odd degrees of freedom. We show that this family is closed under certain type of transformations. We show that the densities of this family converge to a normal density, and we obtain their characteristic functions. We find convergence results and we propose normal approximations for the distribution of Y. We give more precision on these approximations by developing upper bounds in the sup norm. Finally, we provide graphical illustrations and we explain how to estimate the ratio of the means in a Bayesian setting.

Keywords: Bayesian estimation, bivariate normal, characteristic function, convergence results, mixture density, normal approximation, ratio of means, Student distribution.

2.1. Introduction

The density of the ratio of two random variables having a joint bivariate Gaussian distribution has been derived by several authors, firstly by Geary (1930) and subsequently by

Fieller (1932), Marsaglia (1965, 2006), Hinkley (1969), Korhonen and Narula (1989), Pham-Gia et al. (2006) and so on. One of the motivations, in linear regression analysis of bivariate data, is that it is sometimes of interest to estimate the ratio of two population parameters. For example, if we consider the analysis of the simple linear model $y_i = \alpha + \beta x_i + \epsilon_i$ $(i = 1, \ldots, n)$, where $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ are independently normally distributed with mean zero and variance σ^2, then $-\alpha/\beta$ is the intercept of the regression line with the $x-$axis. Its maximum likelihood estimator is the ratio of two correlated normally distributed random variables. Another example can be found in Gill and Keating (2008), where the MLE of the blending coefficient parameter γ is also the ratio of two correlated normally distributed random variables. This density has many applications and has been used to solve many outstanding problems in all fields of science, economics, industry, commerce, education, and medicine, see Öksoy and Aroian (1994), Soldan et al. (2010), Barone (2012) and Białek (2015).

This paper is organized as follows. In Section 2.2, we define a new family of densities that contains all Student density functions with odd degrees of freedom. Some properties of this family, including convergence results and calculation of the characteristic functions are given.

In Section 2.3, we derive two expressions for the density of the ratio $Y = X_1/X_2$, where (X_1, X_2) are two normally distributed random variables. One of these expressions is a mixture of densities of the new family introduced in Section 2.2. We give some properties of this ratio and some convergence results. It was already suggested that the density of this ratio could be approximated by a normal distribution. Conditions for a reasonable normal approximation of the distribution of Y have been presented in scientific literature only through the study of specific cases, simulations and empirical results, see Hayya et al. (1975) and Marsaglia (2006). Hinkley (1969, p. 636 eq. (4)) tried to show the convergence in distribution of the ratio Y, as one of its parameters tends to infinity, to a normal distribution. He was not fully satisfied with his results. Recently, Díaz-Francés and Rubio (2013) pushed the idea further ahead. They showed that under severe restrictions, the difference in absolute value between the cumulative distribution function of Y and that of a normal random variable can be made sufficiently small.

There is an interest of estimating the ratio of means. The standard estimator often proposed in the literature is the ratio of two sample means. When data are from a normal distribution, this estimator is in fact a ratio of two normal random variables. It does not have moments, see Galeone and Pollastri (2012). Some authors found confidence intervals for this ratio, see for example Wang et al. (2015). In Section 2.4, we propose a new approach of estimating the ratio of means in a Bayesian setting and we compare our estimator with the one usually used.

2.2. On a family of distributions involved in the mixture representation for the distribution of the ratio of two normal random variables

In Theorem 2.1 of Section 2.3, one of the expression giving the distribution of the ratio of two normal random variables will involve a mixture distribution with Poisson weights and mixture density components $g_k(\cdot \mid \nu, \tau, \theta)$, $k = 0, 1, \ldots$. In this section we give a definition of these mixture components and develop some of the properties.

Lemma 2.1. *Let $\nu \in \mathbb{R}$, $\tau > 0$ and $\theta \in [0, \pi)$. For all integer $k \geq 0$, the function $g_k(\cdot|\nu, \tau, \theta) \colon \mathbb{R} \mapsto [0, \infty)$ given by*

$$g_k(y|\nu, \tau, \theta) = \frac{1}{\tau} \frac{1}{B\left(\frac{1}{2}, k + \frac{1}{2}\right)} \frac{\left[\cos(\theta) + \sin(\theta)\left(\frac{y-\nu}{\tau}\right)\right]^{2k}}{\left[1 + \left(\frac{y-\nu}{\tau}\right)^2\right]^{k+1}},$$

is a density, where $B(\cdot, \cdot)$ is the beta function.

Proof. Without any loss of generality, we suppose that $\nu = 0$ and $\tau = 1$. For $\theta \in [0, \pi)$, let $a = \sin(\theta)$ and $b = \cos(\theta)$. The function $g_k(\cdot|0, 1, \theta)$ is positive and continuous over $\mathbb{R}$. Let us show that its integral over $\mathbb{R}$ is equal to 1.

$$\begin{aligned}\int_{\mathbb{R}} g_k(t|0,1,\theta)dt &= \int_{\mathbb{R}} \frac{1}{\sqrt{\pi}} \frac{\Gamma(k+1)}{\Gamma(k+\frac{1}{2})} \frac{1}{(1+t^2)^{k+1}} (at+b)^{2k} dt \\ &= E[(aT_k+b)^{2k}],\end{aligned}$$

where the random variable $\sqrt{2k+1} \cdot T_k$ follows the Student distribution with $2k+1$ degrees of freedom.

Notice that

1) according to the Legendre's duplication formula (see Abramowitz and Stegun, 1964), we have

$$\Gamma\left(k+\frac{1}{2}\right) = \frac{(2k)!\Gamma\left(\frac{1}{2}\right)}{4^k k!}, \quad k \geq 0;$$

2) we have

$$E[T_k^j] = \begin{cases} \binom{k}{j/2} \Big/ \binom{2k}{j} & \text{if} \quad j \text{ is even}, \\ 0 & \text{if} \quad j \text{ is odd}, \end{cases} \qquad 0 \leq j \leq 2k.$$

Therefore,

$$\begin{aligned}\int_{\mathbb{R}} g_k(t|\nu,\tau,\theta))dt &= E[(aT_k+b)^{2k}] \\ &= \sum_{j=0}^{2k} \binom{2k}{j} a^j b^{2k-j} E[T_k^j] \\ &= \sum_{\ell=0}^{k} \binom{2k}{2\ell} a^{2\ell} b^{2(k-\ell)} \left\{ \binom{k}{\ell} \Big/ \binom{2k}{2\ell} \right\} \\ &= \sum_{\ell=0}^{k} \binom{k}{\ell} a^{2\ell} b^{2(k-\ell)} \\ &= (a^2+b^2)^k \\ &= [\cos^2(\theta)+\sin^2(\theta)]^k \\ &= 1.\end{aligned}$$

□

Since $g_k(\cdot \mid 0, \sqrt{2k+1}, 0)$ is the density of a Student distribution with $2k+1$ degrees of freedom, we obtain a generalization for the Student densities with odd degrees of freedom. When $k > 0$, the density $g_k(\cdot \mid \nu, \tau, \theta)$ reaches the zero value for all $\nu \in \mathbb{R}, \tau > 0, \theta \in (0, \pi)$. For $k = 0$, $g_k(\cdot|\nu, \tau, \theta)$ is the Cauchy density function with parameters ν and τ. The family of densities $g_k(\cdot|\nu, \tau, \theta)$, $k \geq 0$, is a location scale family and it is closed under the multiplicative inverse transformation as we shall see in the next lemma.

Lemma 2.2. *Let T be a random variable having the density function $g_k(\cdot|\nu, \tau, \theta)$, denoted by $T \sim g_k(\cdot|\nu, \tau, \theta)$. We have*

1. $aT + b \sim g_k(\cdot|a\nu + b, a\tau, \theta)$, $a > 0$, $b \in \mathbb{R}$;
2. $-T \sim \begin{cases} g_k(\cdot|-\nu, \tau, 0) & \text{if} \quad \theta = 0, \\ g_k(\cdot|-\nu, \tau, \pi - \theta) & \text{if} \quad \theta \in (0, \pi); \end{cases}$
3. $T^{-1} \sim g_k\left(\cdot \middle| \frac{\nu}{\nu^2+\tau^2}, \frac{\tau}{\nu^2+\tau^2}, \theta'\right)$ *with*

$$\theta' = \begin{cases} \arccos\left(\frac{\nu\cos(\theta)+\tau\sin(\theta)}{\sqrt{\nu^2+\tau^2}}\right) & \text{if} \quad \tau\cos(\theta) > \nu\sin(\theta), \\ 0 & \text{if} \quad \tau\cos(\theta) = \nu\sin(\theta), \\ \arccos\left(-\left[\frac{\nu\cos(\theta)+\tau\sin(\theta)}{\sqrt{\nu^2+\tau^2}}\right]\right) & \text{if} \quad \tau\cos(\theta) < \nu\sin(\theta). \end{cases}$$

PROOF. The proofs of point 1 and point 2 are straightforward. In point 3, let $X = T^{-1}$ and let h be the density of X. Let $\nu' = \nu/(\nu^2+\tau^2)$, $\tau' = \tau/(\nu^2+\tau^2)$. We obtain that

$$\begin{aligned} h(x) &= \frac{1}{\tau' B\left(\frac{1}{2}, k+\frac{1}{2}\right)} \frac{\left[\left(\frac{\nu\cos(\theta)+\tau\sin(\theta)}{\sqrt{\nu^2+\tau^2}}\right) + \left(\frac{\tau\cos(\theta)-\nu\sin(\theta)}{\sqrt{\nu^2+\tau^2}}\right)\left(\frac{x-\nu'}{\tau'}\right)\right]^{2k}}{\left[1+\left(\frac{x-\nu'}{\tau'}\right)^2\right]^{k+1}} \qquad (2.1) \\ &= \frac{1}{\tau' B\left(\frac{1}{2}, k+\frac{1}{2}\right)} \frac{\left[\cos(\theta') + \sin(\theta')\left(\frac{x-\nu'}{\tau'}\right)\right]^{2k}}{\left[1+\left(\frac{x-\nu'}{\tau'}\right)^2\right]^{k+1}} \end{aligned}$$

with $\sin(\theta') \geq 0$. □

It is well known that the Student densities converge to a normal density when the degrees of freedom tend to infinity. The same thing happens with our new family as we shall see in the next lemma.

Lemma 2.3. *Let T_k be a random variable having the density $g_k(\cdot|\nu,\tau,\theta)$.*

1. *For $\theta \neq \pi/2$ we obtain that*

$$\sqrt{2k+1}(T_k - \{\nu + \tau\tan(\theta)\}) \xrightarrow[k\to\infty]{\mathcal{L}} N(0, \tau^2\sec^4(\theta)).$$

2. *For $\tau\sin(\theta) + \nu\cos(\theta) \neq 0$ we obtain that*

$$\sqrt{2k+1}\left(\frac{1}{T_k} - \frac{1}{\nu^2+\tau^2}\left\{\nu + \tau\left[\frac{\tau\cos(\theta) - \nu\sin(\theta)}{\nu\cos(\theta)+\tau\sin(\theta)}\right]\right\}\right) \xrightarrow[k\to\infty]{\mathcal{L}} N\left(0, \frac{\tau^2}{(\nu\cos(\theta)+\tau\sin(\theta))^4}\right).$$

Proof. For the proof of part 1, without loss of generality, assume that $\nu = 0$ and $\tau = 1$. We have

$$\frac{\Gamma(x)}{\sqrt{\frac{2\pi}{x}}\left(\frac{x}{e}\right)^x} \xrightarrow[x\to\infty]{} 1.$$

Thus,

$$\begin{aligned}
\lim_{k\to\infty}\frac{\Gamma(k+1)}{\sqrt{\pi}\Gamma(k+\frac{1}{2})\sqrt{k+1/2}} &= \lim_{k\to\infty}\frac{\sqrt{\frac{2\pi}{k+1}}\left(\frac{k+1}{e}\right)^{k+1}}{\sqrt{\pi}\sqrt{\frac{2\pi}{k+1/2}}\left(\frac{k+1/2}{e}\right)^{k+1/2}\sqrt{k+1/2}} \\
&= \lim_{k\to\infty}\frac{1}{\sqrt{\pi}}\sqrt{\frac{k+1/2}{k+1}}\left(1+\frac{1/2}{k+1/2}\right)^{k+1/2}\left(\frac{k+1}{(k+1/2)\cdot e}\right)^{1/2} \\
&= \frac{1}{\sqrt{\pi}}.
\end{aligned}$$

$$\frac{(\cos(\theta)+t\sin(\theta))^2}{1+t^2} = 1 - \frac{(\sin(\theta)-t\cos(\theta))^2}{1+t^2}.$$

Let $T_k = \frac{c}{\sqrt{k+1/2}}U_k + \tan(\theta)$, the density of U_k is given by

$$\begin{aligned}
g_{U_k}(u) &= \frac{c}{\sqrt{k+1/2}}g_k\left(\frac{c}{\sqrt{k+1/2}}u + \tan(\theta)\right) \\
&= \frac{\Gamma(k+1)}{\sqrt{\pi}\Gamma\left(k+\frac{1}{2}\right)\sqrt{k+1/2}}\frac{c}{1+\left(\frac{c}{\sqrt{k+1/2}}u+\tan(\theta)\right)^2} \\
&\quad\times\left(1 - \frac{\left(\sin(\theta)-\cos(\theta)\left(\frac{c}{\sqrt{k+1/2}}u+\tan(\theta)\right)\right)^2}{1+\left(\frac{c}{\sqrt{k+1/2}}u+\tan(\theta)\right)^2}\right)^k
\end{aligned}$$

$$= \frac{\Gamma(k+1)}{\sqrt{\pi}\Gamma\left(k+\frac{1}{2}\right)\sqrt{k+1/2}} \frac{c}{1+\left(\frac{c}{\sqrt{k+1/2}}u+\tan(\theta)\right)^2}$$

$$\times \left(1 - \frac{\cos^2(\theta)c^2u^2}{(k+1/2)(1+\tan^2(\theta))} + o\left(\frac{1}{k}\right)\right)^k$$

$$= \frac{\Gamma(k+1)}{\sqrt{\pi}\Gamma\left(k+\frac{1}{2}\right)\sqrt{k+1/2}} \frac{c}{1+\left(\frac{c}{\sqrt{k+1/2}}u+\tan(\theta)\right)^2} \left(1 - \frac{\cos^4(\theta)c^2u^2}{k+1/2} + o\left(\frac{1}{k}\right)\right)^k$$

$$\underset{k\to\infty}{\longrightarrow} \frac{1}{\sqrt{\pi}} \frac{c}{1+\tan^2(\theta)} \exp(-\cos^4(\theta)c^2u^2) = \frac{1}{\sqrt{2\pi}} \exp(-u^2/2) \quad \text{when} \quad c = \sec^2(\theta)/\sqrt{2}.$$

The result comes from the Lehmann-Scheffé theorem. The proof of part 2 is a direct consequence of the one of part 1 and Lemma 2.2 with Equation (2.1). □

We shall derive the characteristic functions using the one of the standard Student distribution with odd degrees of freedom, see Fisher and Healy (1956).

Lemma 2.4. *Let T_k be a random variable having the density $g_k(\cdot|0,1,\theta)$. The characteristic function, φ_{T_k}, of the random variable T_k is given by*

$$\varphi_{T_k}(t) = \exp(-|t|) \sum_{\ell=0}^{k} \left\{ \sum_{j=0}^{2k} i^j \binom{2k}{j} (\sin(\theta))^j (\cos(\theta))^{2k-j} (-1)^{j\mathbb{1}(t<0)} \sum_{r=\ell}^{(j+l)\wedge k} (-1)^{\ell-r} \binom{j}{r-\ell} c_{k,r} \right\} \frac{|t|^\ell}{\ell!} \tag{2.2}$$

with

$$c_{k,r} = \frac{\pi}{4^k B(\frac{1}{2}, k+\frac{1}{2})} \binom{2k-r}{k} 2^r.$$

Notice that if $T_k \sim g_k(\cdot|0,1,\theta)$, then for $\tau > 0$ and $\nu \in \mathbb{R}$, $Y_k = \tau T_k + \nu \sim g_k(\cdot|\nu,\tau,\theta)$ and its characteristic function is given by $\varphi_{Y_k}(t) = e^{i\nu t}\varphi_{T_k}(\tau t)$. Therefore, there is no loss of generality in working on the density $g_k(\cdot|0,1,\theta)$.

Proof. We have

$$\varphi_{T_k}(t) = E[e^{itT_k}] = E[(\cos(\theta) + U_k \cdot \sin(\theta))^{2k} \exp(itU_k)]$$

$$= \sum_{j=0}^{2k} \binom{2k}{j} (\sin(\theta))^j (\cos(\theta))^{2k-j} E[U_k^j \exp(itU_k)],$$

where i is the imaginary number and $\sqrt{2k+1}U_k$ is a Student random variable with $2k+1$ degrees of freedom. Since the jth moment of U_k exists for $j = 1, \dots, 2k$, the jth derivative exists and is given by $\varphi_{U_k}^{(j)}(t) = i^j E[U_k^j e^{itU_k}]$, then $E[U_k^j e^{itU_k}] = (-i)^j \varphi_{U_k}^{(j)}(t)$. Thus,

$$\varphi_{T_k}(t) = \sum_{j=0}^{2k} \binom{2k}{j} (\sin(\theta))^j (\cos(\theta))^{2k-j} (-i)^j \varphi_{U_k}^{(j)}(t).$$

The characteristic function, $\varphi_X(\cdot|v)$, of a random variable X following the standard Student distribution with odd degrees of freedom v is given by

$$\varphi_X(t \mid v) = \exp(-|t\sqrt{v}|) \sum_{r=0}^{(v-1)/2} \frac{(v-r-1)!\left(\frac{v-1}{2}\right)!}{(v-1)!\left(\frac{v-1}{2}-r\right)!} \frac{2^r|t\sqrt{v}|^r}{r!} \qquad \text{(see Fisher and Healy, 1956)}$$

$$= \frac{\left[\left(\frac{v-1}{2}\right)!\right]^2}{(v-1)!} \exp(-|t\sqrt{v}|) \sum_{r=0}^{(v-1)/2} \left[\binom{v-r-1}{\frac{v-1}{2}} \frac{2^r|t\sqrt{v}|^r}{r!}\right].$$

In our case, $v = 2k+1$ and $X \stackrel{d}{=} \sqrt{2k+1}U_k$. Thus,

$$\varphi_{U_k}(t) = E\left[\exp\left(it\frac{X}{\sqrt{2k+1}}\right)\right] = \varphi_X\left(\frac{t}{\sqrt{2k+1}}\middle| 2k+1\right)$$

$$= \frac{(k!)^2}{(2k)!} \exp(-|t|) \sum_{r=0}^{k} \left[\binom{2k-r}{k} \frac{2^r|t|^r}{r!}\right]$$

$$= \frac{\pi}{4^k B(\frac{1}{2}, k+\frac{1}{2})} \exp(-|t|) \sum_{r=0}^{k} \left[\binom{2k-r}{k} \frac{2^r|t|^r}{r!}\right].$$

We want to calculate the jth derivative of this function, for $j = 1, \dots, 2k$.

For $t > 0$,

$$\varphi_{U_k}(t) = \frac{\pi}{4^k B(\frac{1}{2}, k+\frac{1}{2})} \exp(-t) \sum_{r=0}^{k} \left[\binom{2k-r}{k} \frac{2^r t^r}{r!}\right]$$

$$= \sum_{r=0}^{k} c_{k,r} \exp(-t) \frac{t^r}{r!},$$

where

$$c_{k,r} = \frac{\pi}{4^k B(\frac{1}{2}, k+\frac{1}{2})} \binom{2k-r}{k} 2^r.$$

The jth derivative of φ_{T_k} in this case is given by

$$\begin{aligned}
\varphi_{U_k}^{(j)}(t) &= \sum_{r=0}^{k} c_{k,r} \sum_{\ell=0}^{j\wedge r} \binom{j}{\ell} (-1)^{j-\ell} \exp(-t) \frac{t^{r-\ell}}{(r-\ell)!} \\
&= \exp(-t) \sum_{r=0}^{k} c_{k,r} \sum_{\ell=0\vee(r-j)}^{r} (-1)^{j+\ell-r} \binom{j}{r-\ell} \frac{t^\ell}{\ell!} \\
&= \exp(-t) \sum_{\ell=0}^{k} \left\{ \sum_{r=\ell}^{(j+l)\wedge k} (-1)^{j+\ell-r} \binom{j}{r-\ell} c_{k,r} \right\} \frac{t^\ell}{\ell!}. \qquad (2.3)
\end{aligned}$$

For $t < 0$

$$\varphi_{U_k}(t) = \sum_{r=0}^{k} c_{k,r} \exp(t) \frac{(-t)^r}{r!},$$

$$\begin{aligned}
\varphi_{U_k}^{(j)}(t) &= \sum_{r=0}^{k} c_{k,r} \sum_{\ell=0}^{j\wedge r} \binom{j}{\ell} (-1)^{r} \exp(t) \frac{t^{r-\ell}}{(r-\ell)!} \\
&= \exp(t) \sum_{r=0}^{k} (-1)^r c_{k,r} \sum_{\ell=0\vee(r-j)}^{r} \binom{j}{r-\ell} \frac{t^\ell}{\ell!} \\
&= \exp(t) \sum_{\ell=0}^{k} \left\{ \sum_{r=\ell}^{(j+l)\wedge k} (-1)^{r} \binom{j}{r-\ell} c_{k,r} \right\} \frac{t^\ell}{\ell!}. \qquad (2.4)
\end{aligned}$$

Finally, we have for $t < 0$

$$\begin{aligned}
\varphi_T(t) &= \sum_{j=0}^{2k} \binom{2k}{j} (\sin(\theta))^j (\cos(\theta))^{2k-j} (-i)^j \exp(t) \sum_{\ell=0}^{k} \left\{ \sum_{r=\ell}^{(j+l)\wedge k} (-1)^{r} \binom{j}{r-\ell} c_{k,r} \right\} \frac{t^\ell}{\ell!} \\
&= \exp(t) \sum_{\ell=0}^{k} \left\{ \sum_{j=0}^{2k} \binom{2k}{j} (\sin(\theta))^j (\cos(\theta))^{2k-j} (-i)^j \sum_{r=\ell}^{(j+l)\wedge k} (-1)^{r} \binom{j}{r-\ell} c_{k,r} \right\} \frac{t^\ell}{\ell!},
\end{aligned}$$

and for $t > 0$

$$\varphi_T(t) = \exp(-t) \sum_{\ell=0}^{k} \left\{ \sum_{j=0}^{2k} \binom{2k}{j} (\sin(\theta))^j (\cos(\theta))^{2k-j} (-i)^j \sum_{r=\ell}^{(j+l)\wedge k} (-1)^{j+\ell-r} \binom{j}{r-\ell} c_{k,r} \right\} \frac{t^\ell}{\ell!}.$$

Thus, for $t \neq 0$, the characteristic function of T at the point t is given by

$$\varphi_T(t) = \exp(-|t|) \sum_{\ell=0}^{k} \left\{ \sum_{j=0}^{2k} i^j \binom{2k}{j} (\sin(\theta))^j (\cos(\theta))^{2k-j} (-1)^{j \mathbb{1}(t<0)} \sum_{r=\ell}^{(j+l)\wedge k} (-1)^{\ell-r} \binom{j}{r-\ell} c_{k,r} \right\} \frac{|t|^\ell}{\ell!}.$$

It remains to show that φ_T, given in Equation (2.2), satisfies the equation $\varphi_T(0) = 1$. We used the continuity property of the characteristic function of U_k at 0. Since

$$i^j E[U_k^j] = \lim_{t \downarrow 0} \varphi_{U_k}^{(j)}(t) = \lim_{t \uparrow 0} \varphi_{U_k}^{(j)}(t),$$

we have

$$\sum_{r=0}^{j \wedge k} (-1)^r \binom{j}{r} c_{k,r} = \sum_{r=0}^{j \wedge k} (-1)^{j-r} \binom{j}{r} c_{k,r} = \begin{cases} 0 & \text{if } j \text{ is odd,} \\ i^j \frac{\binom{k}{j/2}}{\binom{2k}{j}} & \text{otherwise.} \end{cases} \tag{2.5}$$

The value of the function φ_T at 0 is given by

$$\begin{aligned} \varphi_T(0) &= \sum_{j=0}^{2k} i^j \binom{2k}{j} (\sin(\theta))^j (\cos(\theta))^{2k-j} \sum_{r=0}^{j \wedge k} (-1)^{-r} \binom{j}{r} c_{k,r} \\ &= \sum_{j=0,\, j \text{ even}}^{2k} i^{2j} \binom{k}{j/2} (\sin(\theta))^j (\cos(\theta))^{2k-j} \quad \text{(by using Equation (2.5))} \\ &= \sum_{m=0}^{k} \binom{k}{m} [\sin^2(\theta)]^m [\cos^2(\theta)]^{k-m} \\ &= (\cos^2(\theta) + \sin^2(\theta))^k \\ &= 1. \end{aligned}$$

□

2.3. Ratio distribution of two normal random variables

In this section, we find two expressions for the density of the normal ratio. The first expression is similar to the expressions found in Fieller (1932), Marsaglia (1965) and Hinkley (1969), and the second one is related to that of Pham-Gia et al. (2006). In the literature, the density of the ratio is parametrized with the same parameters as the one of the law of $X = (X_1, X_2)'$, which leads to some problems of identifiability. For a non-zero real number β, X_1/X_2 and $(\beta X_1)/(\beta X_2)$ have the same distribution, but the laws of (X_1, X_2) and $(\beta X_1, \beta X_2)$ are different. In this paper, we deal with the problem by introducing new notations. In using these notations, we can obtain the results faster and effortlessly. In particular, the mixture property appears clearly. This is an advantage compared to Pham-Gia et al. (2006). Let ϕ be the density function of the standard normal distribution and Φ be its cumulative distribution function.

Theorem 2.1. *Let $X \sim N(\mu, \Sigma)$ be a two-dimensional normal random vector with*

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix}, \Sigma^{-1} = \frac{1}{\sigma_2^2\tau^2}\begin{pmatrix} 1 & -\nu \\ -\nu & \sigma_1^2/\sigma_2^2 \end{pmatrix}.$$

Let

$$\nu = \rho\sigma_1/\sigma_2, \tau = \sqrt{1-\rho^2}\sigma_1/\sigma_2, \lambda = \sqrt{\mu'\Sigma^{-1}\mu} \text{ and } \theta = \begin{cases} \arccos(\mu_2/\lambda\sigma_2) & \text{if } \mu_1 > \nu\mu_2, \\ 0 & \text{if } \mu_1 = \nu\mu_2, \\ \arccos(-\mu_2/\lambda\sigma_2) & \text{if } \mu_1 < \nu\mu_2. \end{cases}$$

Let $\delta(\cdot \mid \theta) \colon \mathbb{R} \mapsto [-1, 1]$ be the function given by

$$\delta(z \mid \theta) = \frac{\cos(\theta) + \sin(\theta)z}{\sqrt{1+z^2}}, \quad \theta \in [0, 2\pi).$$

Let $Y = X_1/X_2$. Let $f_Y(\cdot \mid \nu, \tau, \lambda, \theta)$ be the density of Y. We obtain that

1. *If $\lambda = 0$, then the distribution of $(Y - \nu)/\tau$ is a standard Cauchy distribution.*
2. *If $\lambda > 0$ and $z = (y - \nu)/\tau$, then*

$$\begin{aligned} f_Y(y \mid \nu, \tau, \lambda, \theta) &= \frac{1}{\sqrt{2\pi}\tau}\frac{1}{1+z^2}\exp\Big(-\frac{\lambda^2}{2}[1-\delta^2(z \mid \theta)]\Big) \\ &\quad \times\Big\{2\phi(\lambda\sqrt{\delta^2(z \mid \theta)}) + \lambda\sqrt{\delta^2(z \mid \theta)}[2\Phi(\lambda\sqrt{\delta^2(z \mid \theta)}) - 1]\Big\} \\ &= \sum_{k=0}^{\infty} g_k(y \mid \nu, \tau, \theta)\exp\left(-\frac{\lambda^2}{2}\right)\left(\frac{\lambda^2}{2}\right)^k \Big/ k!. \end{aligned}$$

3. *If Z_1, Z_2 are independent standard normal random variables, $\nu \in \mathbb{R}$, $\tau, \lambda > 0$, $\theta \in [0, \pi)$ and*

$$\begin{aligned} X_1 &= \tau[Z_1 + \lambda\sin(\theta)] + \nu[Z_2 + \lambda\cos(\theta)], \\ X_2 &= Z_2 + \lambda\cos(\theta), \end{aligned}$$

then X is a bivariate normal vector with mean μ and covariance matrix Σ where

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \lambda \begin{pmatrix} \tau \sin(\theta) + \nu \cos(\theta) \\ \cos(\theta) \end{pmatrix} \text{ and } \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} \nu^2 + \tau^2 & \nu \\ \nu & 1 \end{pmatrix}.$$

Moreover,

$$\begin{aligned} \rho\sigma_1/\sigma_2 &= \nu, \\ \sqrt{1-\rho^2}\sigma_1/\sigma_2 &= \tau, \\ \mu'\Sigma^{-1}\mu &= \lambda^2, \\ \arccos(\mu_2/\lambda\sigma_2) &= \theta. \end{aligned}$$

Proof. The proof of part 1 is similar to the one of part 2 and the proof of part 3 is straightforward. We are going to develop the density of Z with $Z = (Y - \nu)/\tau$. Consider the change of variables

$$\begin{aligned} Z &= \left(\frac{X_1}{X_2} - \nu\right) \Big/ \tau, \\ U &= \sqrt{1+Z^2} X_2/\sigma_2, \end{aligned}$$

or, equivalently,

$$\begin{aligned} X_1 &= (\tau Z + \nu) \times \sigma_2 U \Big/ \sqrt{1+Z^2}, \\ X_2 &= \sigma_2 U \Big/ \sqrt{1+Z^2}. \end{aligned}$$

The Jacobian, in absolute value, is given by $\frac{\tau\sigma_2^2}{(1+Z^2)} \mid U \mid$. We have,

$$\begin{aligned} (x-\mu)'\Sigma^{-1}(x-\mu) &= x\Sigma^{-1}x - 2x\Sigma^{-1}\mu + \mu'\Sigma^{-1}\mu \\ &= u^2 - 2\Big\{\frac{\tau\mu_2 + (\mu_1 - \nu\mu_2)z}{\sigma_2\lambda\tau\sqrt{1+z^2}}\Big\}\lambda u + \lambda^2 \\ &= u^2 - 2\delta(z \mid \vartheta)\lambda u + \lambda^2 \\ &= \Big(u - \lambda\delta(z \mid \vartheta)\Big)^2 + \lambda^2\Big(1 - \delta^2(z \mid \vartheta)\Big). \end{aligned}$$

Remark 2.1. *Notice that at this point* $\cos(\vartheta) = \mu_2/\lambda\sigma_2$ *and* $\sin(\vartheta) = (\mu_1 - \nu\mu_2)/\sigma_2\lambda\tau$. *It is necessary to take* $\vartheta \in [0, 2\pi)$.

The density $f_X(\cdot \mid \mu, \Sigma)$ is given by

$$f_X(x \mid \mu, \Sigma) = \frac{1}{2\pi} \mid \Sigma \mid^{-1/2} \exp\left(-\frac{1}{2}(x-\mu)'(\Sigma^{-1}(x-\mu)\right).$$

The density $f_{ZU}(\cdot \mid \mu, \Sigma)$ of (Z, U) is then given by

$$\begin{aligned} f_{ZU}(z, u \mid \mu, \Sigma) &= f_X((\tau z + \nu) \times \sigma_2 u \big/ \sqrt{1+z^2}, \sigma_2 u \big/ \sqrt{1+z^2} \mid \mu, \Sigma) \times \frac{\tau\sigma_2^2}{(1+Z^2)} \mid u \mid \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{1+z^2} \exp\left(-\frac{\lambda^2}{2}\Big(1 - \delta^2(z \mid \vartheta)\Big)\right) \\ &\quad \times \mid u \mid \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\Big(u - \lambda\delta(z \mid \vartheta)\Big)^2\right) \end{aligned}$$

and

$$\begin{aligned} f_Z(z \mid \mu, \Sigma) &= \frac{1}{\sqrt{2\pi}} \frac{1}{1+z^2} \exp\left(-\frac{\lambda^2}{2}\Big(1 - \delta^2(z \mid \vartheta)\Big)\right) \\ &\quad \times \int_{-\infty}^{\infty} \mid u \mid \phi(u - \lambda\delta(z \mid \vartheta)) du \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{1+z^2} \exp\left(-\frac{\lambda^2}{2}\Big(1 - \delta^2(z \mid \theta)\Big)\right) \\ &\quad \times \int_{-\infty}^{\infty} \mid u \mid \phi(u - \lambda\delta(z \mid \theta)) du. \end{aligned}$$

Remark 2.2. *When we integrate* $f_{ZU}(z, u \mid \mu, \Sigma)$ *over* u *then the result remains the same changing* $\delta(z \mid \vartheta)$ *for* $-\delta(z \mid \vartheta)$. *Therefore, we define* $\theta \in [0, \pi)$ *such that* $(\cos(\theta), \sin(\theta)) = (\cos(\vartheta), \sin(\vartheta))$ *if* $\vartheta \in [0, \pi)$ *and* $(\cos(\theta), \sin(\theta)) = (-\cos(\vartheta), -\sin(\vartheta))$ *if* $\vartheta \in [\pi, 2\pi)$.

On the one hand, a simple calculation shows that

$$\int_{-\infty}^{+\infty} \mid u \mid \phi(u - \lambda\delta(z \mid \theta)) du = 2\phi(\lambda\delta(z \mid \theta)) - \lambda\delta(z \mid \theta)[1 - 2\Phi(\lambda\delta(z \mid \theta))] \qquad (2.6)$$

and we obtain part 2 of Theorem 2.1. On the other hand, Equation (2.6) represents $E[\sqrt{V}]$ where V is a noncentral chi-squared distribution of degree 1 with noncentral parameter $\lambda^2\delta^2(z \mid \theta)$. We use the mixture representation of the distribution of V, that is K has a Poisson$(\lambda^2\delta^2(z \mid \theta)/2)$ distribution and, conditionally on $K = k$, V has a chi-squared distribution with $2k+1$ degrees of freedom. Therefore,

$$\begin{aligned}
E[\sqrt{V}] &= \sum_{k=0}^{\infty} E[\sqrt{V} \mid K = k] \times \exp\left(-\frac{\lambda^2}{2}\delta^2(z \mid \theta)\right)\left(\frac{\lambda^2}{2}\delta^2(z \mid \theta)\right)^k \Big/ k! \\
&= \sum_{k=0}^{\infty} \frac{\sqrt{2}\Gamma(k+1)}{\Gamma(k+1/2)} \times \exp\left(-\frac{\lambda^2}{2}\delta^2(z \mid \theta)\right)\left(\frac{\lambda^2}{2}\delta^2(z \mid \theta)\right)^k \Big/ k! \\
&= \sqrt{2\pi}(1+z^2)\exp\left(\frac{\lambda^2}{2}[1-\lambda^2\delta^2(z \mid \theta)]\right) \times \sum_{k=0}^{\infty} g_k(z \mid 0,1,\theta)\exp\left(-\frac{\lambda^2}{2}\right)\left(\frac{\lambda^2}{2}\right)^k \Big/ k!
\end{aligned}$$

giving the second expression in part 2 of Theorem 2.1. □

Remark 2.3. *For the remaining part of this paper a distribution associated with the density given in part 2 of Theorem 2.1 will be denoted $H(\nu, \tau, \lambda, \theta)$ and the density will be denoted $h(\cdot \mid \nu, \tau, \lambda, \theta)$. The distribution will be called the Hinkley distribution because Hinkley was amongst the first to study the properties of this distribution.*

Lemma 2.5. *Let T be a random variable having the density function $h(\cdot|\nu, \tau, \lambda, \theta)$, denoted by $T \sim H(\cdot|\nu, \tau, \lambda, \theta)$. We have*

1. $aT + b \sim H(\cdot|a\nu + b, a\tau, \lambda, \theta)$, $a > 0$, $b \in \mathbb{R}$;
2. $-T \sim \begin{cases} H(\cdot|-\nu, \tau, \lambda, 0) & \text{if } \theta = 0, \\ H(\cdot|-\nu, \tau, \lambda, \pi - \theta) & \text{if } \theta \in (0, \pi); \end{cases}$
3. $T^{-1} \sim H\left(\cdot \middle| \frac{\nu}{\nu^2+\tau^2}, \frac{\tau}{\nu^2+\tau^2}, \lambda, \theta'\right)$ *with*

$$\theta' = \begin{cases} \arccos\left(\frac{\nu\cos(\theta)+\tau\sin(\theta)}{\sqrt{\nu^2+\tau^2}}\right) & \text{if } \tau\cos(\theta) > \nu\sin(\theta), \\ 0 & \text{if } \tau\cos(\theta) = \nu\sin(\theta), \\ \arccos\left(-\left[\frac{\nu\cos(\theta)+\tau\sin(\theta)}{\sqrt{\nu^2+\tau^2}}\right]\right) & \text{if } \tau\cos(\theta) < \nu\sin(\theta). \end{cases}$$

The proof of this lemma is a direct application of Lemma 2.2 and the mixture representation of $h(\cdot|\nu, \tau, \lambda, \theta)$. We can also use the mixture representation of $h(\cdot|\nu, \tau, \lambda, \theta)$ for finding the characteristic function.

In Figure 2.1, the black curve is the mixture density function of Y and the other curves are those of the new densities. For the parameter θ, we just focus on values in the interval $[0, \pi/2)$

because, for any $\theta \in [0, \pi/2)$, the symmetry with respect to the y axis of any of the considered densities will give a density in the same family with the parameter $\theta' = \pi - \theta \in [\pi/2, \pi)$, all other parameters staying the same. In fact, it follows from Lemma 2.2 and Lemma 2.5 that $g_k(-t|\nu, \tau, \theta) = g_k(t|\nu, \tau, \pi - \theta)$, $k \geq 0$, and $h(-t|\nu, \tau, \lambda, \theta) = h(t|\nu, \tau, \lambda, \pi - \theta)$ for all t in $\mathbb{R}$. We also see that the density of Y can be bimodal. Marsaglia (2006) gave conditions on the parameters $(\nu, \tau, \lambda, \theta)$ such that densities $h(\cdot|\nu, \tau, \lambda, \theta)$ become bimodal or unimodal.

Hinkley (1969, p. 636 eq. (4)) tried to show the convergence in distribution of the ratio Y, as one of its parameters tends to infinity, to a normal distribution without achieving results quite satisfactory. In fact, he showed that the cumulative distribution function of Y converges to a function, which is not a cumulative distribution function. Marsaglia (2006) has mentioned that many of the ratios of normal variates encountered in practice can themselves be taken as normally distributed. He gave some conditions for which the distribution of the ratio is close to the one of a normal random variable. We go further by showing the convergence in law of the ratio to a normal distribution when the parameter λ tends to infinity.

Lemma 2.6. *Let T_λ be a random variable having the density $h(\cdot|\nu, \tau, \lambda, \theta)$.*

1. *For $\theta \neq \pi/2$ we obtain that*

$$\lambda(T_\lambda - \{\nu + \tau \tan(\theta)\}) \xrightarrow[\lambda\to\infty]{\mathcal{L}} N(0, \tau^2 \sec^4(\theta)).$$

2. *For $\tau \sin(\theta) + \nu \cos(\theta) \neq 0$ we obtain that*

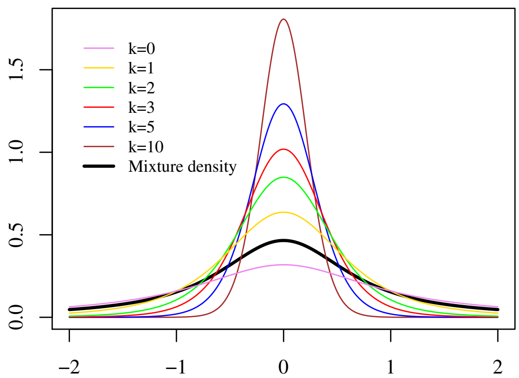
$$\lambda\left(\frac{1}{T_\lambda} - \frac{1}{\nu^2 + \tau^2}\left\{\nu + \tau\left[\frac{\tau\cos(\theta) - \nu\sin(\theta)}{\nu\cos(\theta) + \tau\sin(\theta)}\right]\right\}\right) \xrightarrow[\lambda\to\infty]{\mathcal{L}} N\left(0, \frac{\tau^2}{(\nu\cos(\theta) + \tau\sin(\theta))^4}\right).$$

3. *If $(X_{1i}, X_{2i})' \sim \mathcal{N}(\mu, \Sigma)$, $i = 1, \ldots, n$, are independent and $Y_1 = (X_{11}/X_{21}) \sim H(\nu, \mu, \lambda, \theta)$, then $Y_n = (\sum_{i=1}^n X_{1i}/\sum_{j=1}^n X_{2j}) \sim H(\nu, \mu, \sqrt{n}\lambda, \theta)$.*

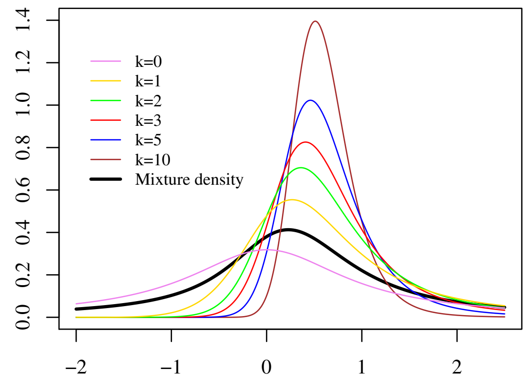
The proof of this lemma is a direct application of the Lehmann-Scheffé theorem using the first representation of $h(\cdot|\nu, \tau, \lambda, \theta)$ in part 2, and part 3 of Theorem 2.1.

In the last part of this section, we want to be more informative on the quality of the approximation by a normal distribution. Thus, in the following result, we study the quality

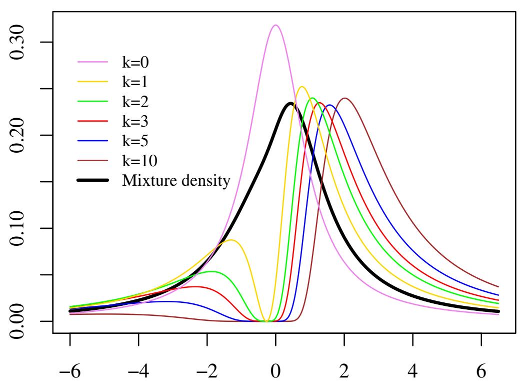
Figure 2.1. Density of the ratio Y and densities $g_k(\cdot|\nu, \tau, \theta)$ for $k = 0, 1, 2, 3, 5, 10$, $\nu = 0$, $\tau = 1$ and for some values of θ and λ.



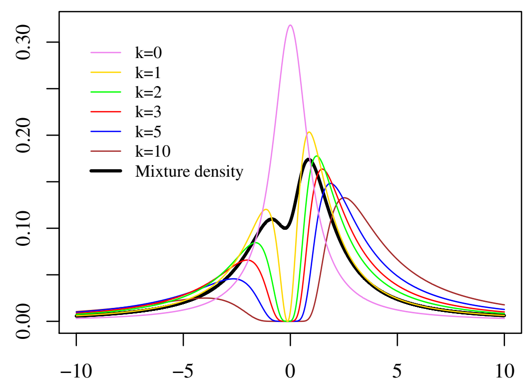
(a) $\theta = 0$ and $\lambda = 1$.



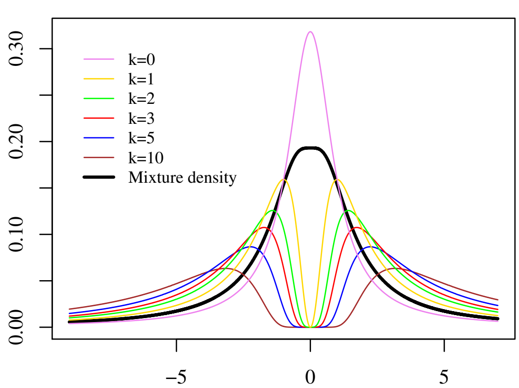
(b) $\theta = \pi/6$ and $\lambda = 1$.



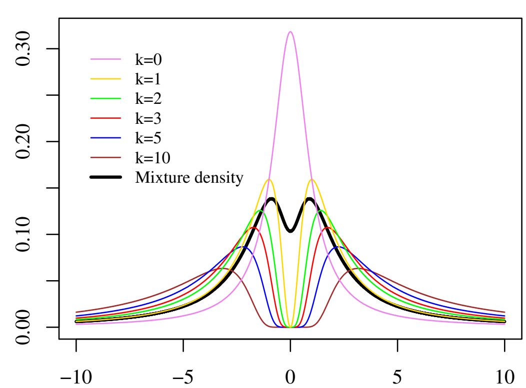
(c) $\theta = 5\pi/12$ and $\lambda = 1$.



(d) $\theta = 11\pi/24$ and $\lambda = 1.5$.



(e) $\theta = \pi/2$ and $\lambda = 1$.



(f) $\theta = \pi/2$ and $\lambda = 1.5$.

of the approximation of the cumulative distribution function of the ratio Y by the one of a normal random variable.

Lemma 2.7. *Let $Z \sim \mathcal{N}(0,1)$ and $Y \sim H(\nu, \tau, \lambda, \theta)$ with $\theta \in [0,\pi) \setminus \{\pi/2\}$. Let $M(\cdot : \theta) : (0,\infty) \longrightarrow (0,\infty)$ be a function satisfying*

1. $M(\lambda : \theta) \underset{\lambda \to +\infty}{\longrightarrow} +\infty$
2. $M(\lambda : \theta) \leq \lambda |\cos(\theta)|$, $\lambda > 0$
3. $M(\lambda : \theta)/\lambda \underset{\lambda \to +\infty}{\longrightarrow} 0.$

Let

$$Y' = \lambda \cos^2(\theta) \left[\left(\frac{Y - \nu}{\tau} \right) - \tan(\theta) \right].$$

We have,

$$\begin{aligned} |F_{Y'}(x) - \Phi(x)| &\leq 2P\Big(|Z| > M(\lambda : \theta)\Big) + \phi\left(|x| \left\{1 - \frac{M(\lambda : \theta)}{\lambda |\cos(\theta)|}\right\}\right) \frac{|x| M(\lambda : \theta)}{\lambda |\cos(\theta)|} \\ &\leq \frac{2e^{-M^2(\lambda,\theta)/2}}{M(\lambda : \theta)\sqrt{2\pi}} + \frac{e^{-1/2}}{\sqrt{2\pi}} \left(\frac{\frac{M(\lambda:\theta)}{\lambda|\cos(\theta)|}}{1 - \frac{M(\lambda:\theta)}{\lambda|\cos(\theta)|}} \right) \underset{\lambda \to +\infty}{\longrightarrow} 0 \quad \forall x \in \mathbb{R}. \end{aligned}$$

Where, $F_{Y'}$ and Φ are respectively the cumulative distribution function of Y' and of Z, while ϕ is the density function of Z.

In particular, we can take $M(\lambda : \theta) = |\cos(\theta)| \log(\lambda + 1)$.

Proof. For $\theta \in [0,\pi/2) \cup (\pi/2,\pi)$, let $M(\lambda : \theta) \leq \lambda |\cos(\theta)|$ and let $\operatorname{sgn}(\theta)$ be the sign of θ, where $M(\lambda : \theta)$ is a positive function such that $M(\lambda : \theta) \longrightarrow +\infty$ and $M(\lambda : \theta)/\lambda \longrightarrow 0$ as $\lambda \longrightarrow +\infty$.

$$\begin{aligned} |F_{Y'}(x) - \Phi(x)| &= \left| P\left(\frac{\lambda \cos^2(\theta)}{\tau} (Y - [\nu + \tan(\theta) \cdot \tau]) \leq x \right) - P(Z \leq x) \right| \\ &= \left| P\left(\frac{\lambda \cos^2(\theta)}{\tau} \left(\tau \frac{Z_1 + \lambda \sin(\theta)}{Z_2 + \lambda \cos(\theta)} + \nu - (\nu - \tan(\theta) \cdot \tau) \right) \leq x \right) - P(Z \leq x) \right|. \end{aligned}$$

Where Z_1 and Z_2 are two independent standard normal random variables.

$$\begin{aligned} |F_{Y'}(x) - \Phi(x)| &= \left| P\left(\lambda \cos^2(\theta) \left(\frac{Z_1 + \lambda \sin(\theta)}{Z_2 + \lambda \cos(\theta)} - \tan(\theta) \right) \leq x \right) - P(Z \leq x) \right| \\ &= \left| P\left(\cos^2(\theta) \frac{Z_1 - Z_2 \tan(\theta)}{\frac{Z_2}{\lambda} + \cos(\theta)} \leq x \right) - P(Z \leq x) \right| \\ &= \left| P\left(\cos^2(\theta) \frac{Z_1 \cdot \operatorname{sgn}(\cos(\theta)) - Z_2 \tan(\theta) \cdot \operatorname{sgn}(\cos(\theta))}{\frac{Z_2}{\lambda} \cdot \operatorname{sgn}(\cos(\theta)) + \cos(\theta) \cdot \operatorname{sgn}(\cos(\theta))} \leq x \right) - P(Z \leq x) \right| \end{aligned}$$

$$
\begin{aligned}
=&\left|P\left(\cos^2(\theta)\frac{Z_1 - Z_2\tan(\theta)}{\frac{Z_2}{\lambda} + |\cos(\theta)|} \leq x\right) - P(Z \leq x)\right| \\
=&\Bigg|P\left(Z_1 - Z_2\tan(\theta) \leq \frac{x}{\cos^2(\theta)}\left(\frac{Z_2}{\lambda} + |\cos(\theta)|\right), |Z_2| \leq M(\lambda:\theta)\right) \\
&+ \quad P\left(\cos^2(\theta)\frac{Z_1 - Z_2\tan(\theta)}{\frac{Z_2}{\lambda} + |\cos(\theta)|} \leq x, |Z_2| > M(\lambda:\theta)\right) - P(Z \leq x)\Bigg| \\
\leq&\left|P\left(Z_1 - Z_2\tan(\theta) \leq \frac{x}{\cos^2(\theta)}\left(\frac{Z_2}{\lambda} + |\cos(\theta)|\right), |Z_2| \leq M(\lambda:\theta)\right) - P(Z \leq x)\right| \\
&+ \quad P\Big(|Z_2| > M(\lambda:\theta)\Big) \\
=&|A - B| + P\Big(|Z_2| > M(\lambda:\theta)\Big),
\end{aligned}
$$

where

$$
A = P\left(Z_1 - Z_2\tan(\theta) \leq \frac{x}{\cos^2(\theta)}\left(\frac{Z_2}{\lambda} + |\cos(\theta)|\right), |Z_2| \leq M(\lambda:\theta)\right) \quad \text{and} \quad B = P(Z \leq x).
$$

Let,

$$
A_+ = P\left(Z_1 - Z_2\tan(\theta) \leq \frac{1}{\cos^2(\theta)}\left(\frac{|x|M(\lambda:\theta)}{\lambda} + x|\cos(\theta)|\right), |Z_2| \leq M(\lambda:\theta)\right)
$$

and

$$
A_- = P\left(Z_1 - Z_2\tan(\theta) \leq \frac{1}{\cos^2(\theta)}\left(\frac{-|x|M(\lambda:\theta)}{\lambda} + x|\cos(\theta)|\right), |Z_2| \leq M(\lambda:\theta)\right).
$$

We have,

$$
\begin{aligned}
|A_+ - B| =&\Bigg|P\left(Z_1 - Z_2\tan(\theta) \leq \frac{1}{\cos^2(\theta)}\left(\frac{|x|M(\lambda:\theta)}{\lambda} + x|\cos(\theta)|\right)\right) - P(Z \leq x) \\
&-P\left(Z_1 - Z_2\tan(\theta) \leq \frac{1}{\cos^2(\theta)}\left(\frac{|x|M(\lambda:\theta)}{\lambda} + x|\cos(\theta)|\right), |Z_2| > M(\lambda:\theta)\right)\Bigg| \\
\leq&\left|P\left(Z_1 - Z_2\tan(\theta) \leq \frac{1}{\cos^2(\theta)}\left(\frac{|x|M(\lambda:\theta)}{\lambda} + x|\cos(\theta)|\right)\right) - P(Z \leq x)\right| \\
&+ P\Big(|Z_2| > M(\lambda:\theta)\Big) \\
=&P\left(x \leq Z \leq x + \frac{|x|M(\lambda:\theta)}{\lambda|\cos(\theta)|}\right) + P\Big(|Z_2| > M(\lambda:\theta)\Big)
\end{aligned}
$$

$$\text{since,}\quad P\Big(Z_1 - Z_2\tan(\theta) \le t\Big) = P\Big(Z \le t|\cos(\theta)|\Big).$$

Similarly,

$$|A_- - B| \le P\left(x - \frac{|x|M(\lambda:\theta)}{\lambda|\cos(\theta)|} \le Z \le x\right) + P\Big(|Z_2| > M(\lambda:\theta)\Big).$$

Thus,

$$\begin{aligned}
|A - B| \le& \max\{|A_+ - B|, |A_- - B|\} \\
\le& P\Big(|Z_2| > M(\lambda:\theta)\Big) + \max\left\{P\left(x - \frac{|x|M(\lambda:\theta)}{\lambda|\cos(\theta)|} \le Z \le x\right), P\left(x \le Z \le x + \frac{|x|M(\lambda:\theta)}{\lambda|\cos(\theta)|}\right)\right\} \\
\le& P\Big(|Z_2| > M(\lambda:\theta)\Big) + P\left(|x| - \frac{|x|M(\lambda:\theta)}{\lambda|\cos(\theta)|} \le Z \le |x|\right) \\
\le& P\Big(|Z_2| > M(\lambda:\theta)\Big) + \phi\left(|x|\left\{1 - \frac{M(\lambda:\theta)}{\lambda|\cos(\theta)|}\right\}\right)\frac{|x|M(\lambda:\theta)}{\lambda|\cos(\theta)|}.
\end{aligned}$$

Hence,

$$\begin{aligned}
|F_{Y'}(x) - \Phi(x)| \le& 2P\Big(|Z_2| > M(\lambda:\theta)\Big) + \phi\left(|x|\left\{1 - \frac{M(\lambda:\theta)}{\lambda|\cos(\theta)|}\right\}\right)\frac{|x|M(\lambda:\theta)}{\lambda|\cos(\theta)|} \\
=& 2P\Big(|Z_2| > M(\lambda:\theta)\Big) + \frac{1}{\sqrt{2\pi}}e^{-x^2\left(1-\frac{M(\lambda:\theta)}{\lambda|\cos(\theta)|}\right)^2/2}\frac{|x|M(\lambda:\theta)}{\lambda|\cos(\theta)|} \\
\le& 2P\Big(|Z_2| > M(\lambda:\theta)\Big) + \frac{e^{-1/2}}{\sqrt{2\pi}}\left(\frac{\frac{M(\lambda:\theta)}{\lambda|\cos(\theta)|}}{1 - \frac{M(\lambda:\theta)}{\lambda|\cos(\theta)|}}\right).
\end{aligned}$$

Since

$$\int_x^{+\infty} \frac{1}{\sqrt{2\pi}}e^{-t^2/2}dt \le \int_x^{+\infty} \frac{t}{x}\frac{1}{\sqrt{2\pi}}e^{-t^2/2}dt = \frac{e^{-x^2/2}}{x\sqrt{2\pi}},$$

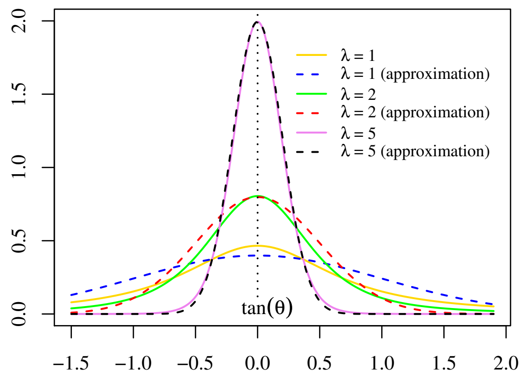
we have

$$|F_{Y'}(x) - \Phi(x)| \le \frac{2e^{-M^2(\lambda,\theta)/2}}{M(\lambda:\theta)\sqrt{2\pi}} + \frac{e^{-1/2}}{\sqrt{2\pi}}\left(\frac{\frac{M(\lambda:\theta)}{\lambda|\cos(\theta)|}}{1 - \frac{M(\lambda:\theta)}{\lambda|\cos(\theta)|}}\right).$$

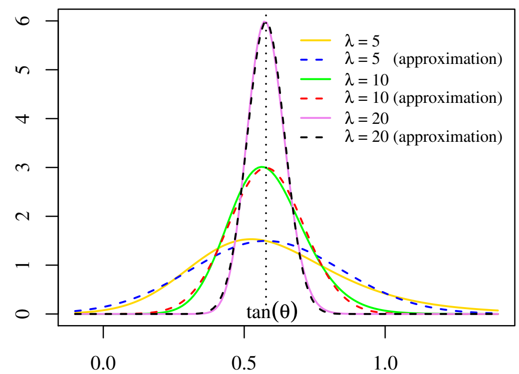
□

In Figure 2.2, we provide different curves of the density functions of Y (solid curves) and of normal approximation of Y (dotted curves), for some set of parameters. When θ is small, we have a good approximation with small values of λ. When θ is close to $\pi/2$, we need large values of λ to have good approximations.

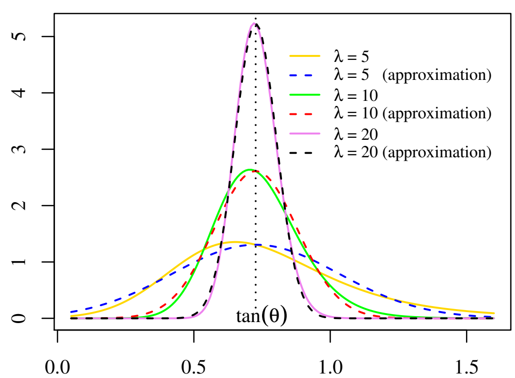
Figure 2.2. Densities of the ratio Y and its normal approximation for $\nu = 0$, $\tau = 1$ and for some values of λ and θ.



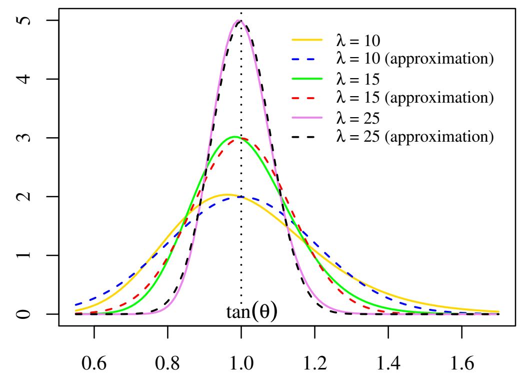
(a) $\theta = 0$.



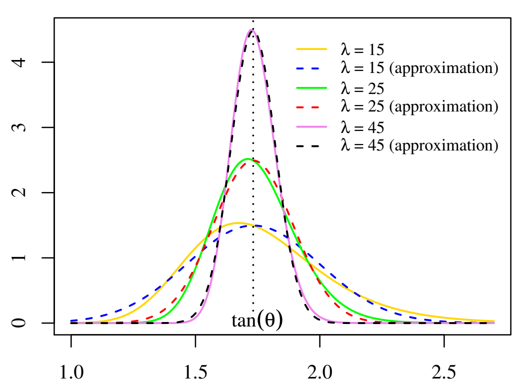
(b) $\theta = \pi/6$.



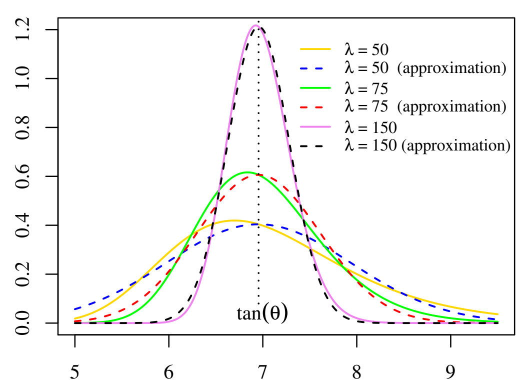
(c) $\theta = \pi/5$.



(d) $\theta = \pi/4$.



(e) $\theta = \pi/3$.



(f) $\theta = 5\pi/11$.

2.4. Estimation of the ratio of means using a Bayesian approach

Suppose that observations are from $X = (X_1, X_2)' \sim \mathcal{N}(\mu, \Sigma)$, where $\mu = (\mu_1, \mu_2)'$ and Σ is assumed known. We want to estimate the ratio of means $\alpha = \mu_1/\mu_2$ based on a sample of size n. Pham-Gia et al. (2006) used the standard approach by taking the ratio of the two sample means as an estimator. The standard estimator is given by

$$\widehat{\alpha}_{\text{std}} = \frac{\overline{X}_1}{\overline{X}_2}.$$

This estimator is Hinkley distributed. Thus, its expectation does not exist. We use a Bayesian approach. Let

$$\mu \sim \mathcal{N}(\vartheta, \Lambda), \quad \text{where} \quad \vartheta = (\vartheta_1, \vartheta_2)',$$

be the prior distribution. Our estimator is the median of the posterior distribution of μ_1/μ_2. This ratio of means is Hinkley distributed since the posterior distribution of the random vector μ is a normal distribution as we shall see in the following lemma.

Lemma 2.8. *If $(X_{1i}, X_{2i}) \sim \mathcal{N}(\mu, \Sigma)$ for $i = 1, \ldots, n$ are independent random vectors and $(Z_1, Z_2) \sim \mathcal{N}(\mu, (1/n)\Sigma)$, then we have*

1. *$\mathcal{L}(Z_1/Z_2|\mu, (1/n)\Sigma) = \mathcal{L}(\overline{X}_1/\overline{X}_2|\mu, \Sigma) = H(\nu, \tau, \sqrt{n}\lambda, \theta)$, where parameters ν, τ, λ and θ are defined in Section 2.3. $\mathcal{L}(Z_1/Z_2|\mu, (1/n)\Sigma)$ means the law of the ratio Z_1/Z_2 given μ and Σ.*
2. *In addition if $\mu \sim \mathcal{N}(\vartheta, \Lambda)$, then*

$$\begin{aligned}\mathcal{L}(\mu|X_{1\cdot} = x_{1\cdot}, X_{2\cdot} = x_{2\cdot}, \Sigma) &= \mathcal{L}(\mu|\overline{X}_1 = \overline{x}_1, \overline{X}_2 = \overline{x}_2, \Sigma) \\ &= N\Big(\vartheta + (\Lambda^{-1} + n\Sigma^{-1})^{-1} n\Sigma^{-1}(\overline{x} - \vartheta),\ (\Lambda^{-1} + n\Sigma^{-1})^{-1}\Big).\end{aligned}$$

Where $\overline{X}_j = 1/n \sum_{i=1}^{n} X_{ji}$ and $X_{j\cdot} = (X_{j1}, X_{j2}, \ldots, X_{jn})$ for $j = 1, 2$.

Proof. 1. It is easy to see that $Z_1/Z_2 \sim \text{H}(\nu, \tau, \sqrt{n}\lambda, \theta)$ and $\overline{X}_1/\overline{X}_2 \sim \text{H}(\nu, \tau, \sqrt{n}\lambda, \theta)$ since $\sqrt{n}\lambda = \sqrt{\mu' n\Sigma^{-1}\mu}$ (see the point 3 of Lemma 2.6).

2. We have

$$
\begin{aligned}
&f_{(X_{1\cdot},X_{2\cdot}|\mu)}(x_{1\cdot},x_{2\cdot}|\mu) \\
&\quad = \frac{|\Sigma|^{-n/2}}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(x_{\cdot i}-\mu)'\Sigma^{-1}(x_{\cdot i}-\mu)\right\} \\
&\quad = \frac{|\Sigma|^{-n/2}}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(x_{\cdot i}-\overline{x})'\Sigma^{-1}(x_{\cdot i}-\overline{x})\right\} \exp\left\{-\frac{n}{2}(\overline{x}-\mu)'\Sigma^{-1}(\overline{x}-\mu)\right\},
\end{aligned}
$$

where $x_{\cdot i} = (x_{1i}, x_{2i})$. Since $\overline{X} = (\overline{X}_1, \overline{X}_2)'$ is a sufficient statistic for μ, we have $\mathcal{L}(\mu|X_{1\cdot} = x_{1\cdot}, X_{2\cdot} = x_{2\cdot}, \Sigma) = \mathcal{L}(\mu|\overline{X}_1 = \overline{x}_1, \overline{X}_2 = \overline{x}_2, \Sigma)$. The density of $(X_{1\cdot}, X_{2\cdot}, \mu_1, \mu_2)$ is given by

$$
\begin{aligned}
&f_{(X_{1\cdot},X_{2\cdot}|\mu)}(x_{1\cdot},x_{2\cdot}|\mu)\cdot\pi(\mu) \\
&\quad = \frac{|\Sigma|^{-n/2}|\Lambda|^{-1/2}}{(2\pi)^{(n+1)/2}} \exp\left\{-\frac{1}{2}(\mu-\vartheta)'\Lambda^{-1}(\mu-\vartheta) - \frac{n}{2}(\overline{x}-\mu)'\Sigma^{-1}(\overline{x}-\mu)\right\} \\
&\qquad \times \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(x_{\cdot i}-\overline{x})'\Sigma^{-1}(x_{\cdot i}-\overline{x})\right\} \\
&\quad = \frac{|\Sigma|^{-n/2}|\Lambda|^{-1/2}}{(2\pi)^{(n+1)/2}} \exp\left\{-\frac{1}{2}((\overline{x}-\vartheta)', (\mu-\vartheta)')\Sigma_5^{-1}\begin{pmatrix}\overline{x}-\vartheta \\ \mu-\vartheta\end{pmatrix}\right\} \\
&\qquad \times \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(x_{\cdot i}-\overline{x})'\Sigma^{-1}(x_{\cdot i}-\overline{x})\right\},
\end{aligned}
$$

with

$$
\Sigma_5^{-1} = \begin{pmatrix} n\Sigma^{-1} & -n\Sigma^{-1} \\ -n\Sigma^{-1} & n\Sigma^{-1}+\Lambda^{-1}\end{pmatrix} \qquad \text{and} \qquad \Sigma_5 = \begin{pmatrix}\frac{1}{n}\Sigma+\Lambda & \Lambda \\ \Lambda & \Lambda\end{pmatrix}.
$$

Let $\pi_1(\cdot|x_{1\cdot}, x_{2\cdot})$ be the posterior density of μ. We have

$$
\begin{aligned}
\pi_1(\mu|x_{1\cdot},x_{2\cdot}) \quad &= \quad \frac{f_{(X_{1\cdot},X_{2\cdot}|\mu)}(x_{1\cdot},x_{2\cdot}|\mu)\cdot\pi(\mu)}{\int_{\mathbb{R}^2} f_{(X_{1\cdot},X_{2\cdot}|\mu)}(x_{1\cdot},x_{2\cdot}|\mu)\cdot\pi(\mu)d\mu} \\
&\propto \exp\left\{-\frac{1}{2}((\overline{x}-\vartheta)', (\mu-\vartheta)')\Sigma_5^{-1}\begin{pmatrix}\overline{x}-\vartheta \\ \mu-\vartheta\end{pmatrix}\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\mu-\vartheta-(\Lambda^{-1}+n\Sigma^{-1})^{-1}n\Sigma^{-1}(\overline{x}-\vartheta)\right]'(\Lambda^{-1}+n\Sigma^{-1})\right. \\
&\qquad \left.\times\left[\mu-\vartheta-(\Lambda^{-1}+n\Sigma^{-1})^{-1}n\Sigma^{-1}(\overline{x}-\vartheta)\right]\right\}.
\end{aligned}
$$

Therefore,

$$\begin{aligned}\mathcal{L}(\mu|X_{1\cdot}=x_{1\cdot},X_{2\cdot}=x_{2\cdot},\Sigma) &= \mathcal{L}(\mu|\overline{X}_1=\overline{x}_1,\overline{X}_2=\overline{x}_2,\Sigma)\\ &= N\Big(\vartheta+(\Lambda^{-1}+n\Sigma^{-1})^{-1}n\Sigma^{-1}(\overline{x}-\vartheta),\ (\Lambda^{-1}+n\Sigma^{-1})^{-1}\Big).\end{aligned}$$

□

It follows from Lemma 2.8 that making inferences based on a sample of size n from the normal distribution with mean μ and covariance matrix Σ is equivalent to making inferences based on a sample of size 1 from the normal distribution with mean μ and covariance matrix $(1/n)\Sigma$. Thus, without any loss of generality, we shall give a method for estimating the ratio of means based on a sample of size 1. In practice, the single observation can be $y=\overline{x}_1/\overline{x}_2$.

We have one observation y from Y. In fact, for simulation purposes and in order to compare the true parameter with its estimates, we choose the true value of μ and we generate $y=x_1/x_2$, where (x_1,x_2) is from $\mathcal{N}(\mu,\Sigma)$. Our estimator, the median of the posterior distribution of μ_1/μ_2, is not easy to evaluate. Nevertheless, we will rely on an approximation based on a sample median of $m=1000$ observations generated from the posterior distribution of μ_1/μ_2 for the numerical calculations. One initial observation of the parameter, $\mu^{(0)}=(\mu_1^{(0)},\mu_2^{(0)})$, is generated from the normal distribution with mean ϑ and covariance matrix Λ. The first observation is generated using $\mu^{(0)}$ and the two steps of the algorithm below. Other observations are generated recursively using the Gibbs sampler.

Algorithm - The Gibbs sampler -

Given $\left(x_{2k},\mu_1^{(k)},\mu_2^{(k)}\right)$, generate;

1) $X_{2\,k+1}\sim f_{X_2|(Y,\mu_1,\mu_2)}\left(x_2\middle|y,\mu_1^{(k)},\mu_2^{(k)}\right)$;

2) $\left(\mu_1^{(k+1)},\mu_2^{(k+1)}\right)\sim f_{(\mu_1,\mu_2)|(X_2,Y)}\left(\mu_1,\mu_2|x_{2\,k+1},y\right)$.

The numerical value of our estimator is

$$\widehat{\alpha}_{\texttt{Bayes}}=\texttt{Median}\left\{\frac{\mu_1^{(k)}}{\mu_2^{(k)}},\ k=1,\ldots,1000\right\}.$$

The following result gives the conditional densities in the above algorithm.

Lemma 2.9. *Let $Y = X_1/X_2$.*

1. *The conditional distribution of (μ_1, μ_2) given $(X_2, Y) = (x_2, y)$ is $\mathcal{N}(\vartheta + (\Sigma^{-1} + \Lambda^{-1})^{-1}\Sigma^{-1}((yx_2, x_2)' - \vartheta); (\Sigma^{-1} + \Lambda^{-1})^{-1})$.*
2. *The density of the random variable X_2 given $(Y, \mu_1, \mu_2) = (y, \mu_1, \mu_2)$ is given by*

$$f_{X_2|(Y,\mu_1,\mu_2)}(x_2|y, \mu_1, \mu_2)$$
$$= \frac{(1+z^2)|x_2|}{\sigma_2^2\sqrt{2\pi}} \frac{\exp\left\{-\frac{1}{2}\left(x_2\sqrt{1+z^2}/\sigma_2 - \lambda\delta(z|\theta)\right)^2\right\}}{2\phi(\lambda\sqrt{\delta^2(z|\theta)}) + \lambda\sqrt{\delta^2(z|\theta)}\left[2\Phi(\lambda\sqrt{\delta^2(z|\theta)}) - 1\right]},$$

with $z = (y - \nu)/\tau$.

PROOF. 1. It is well known that

$$(X_1, X_2, \mu_1, \mu_2)' \sim \mathcal{N}\left((\vartheta_1, \vartheta_2, \vartheta_1, \vartheta_2)'; \Sigma_4\right), \tag{2.7}$$

with

$$\Sigma_4 = \begin{pmatrix} \Sigma + \Lambda & \Lambda \\ \Lambda & \Lambda \end{pmatrix}.$$

We have $|\Sigma_4| = |\Lambda| \cdot |\Sigma + \Lambda - \Lambda\Lambda^{-1}\Lambda| = |\Sigma| \cdot |\Lambda|$ and $\Sigma_4^{-1} = \begin{pmatrix} \Sigma^{-1} & -\Sigma^{-1} \\ -\Sigma^{-1} & \Sigma^{-1} + \Lambda^{-1} \end{pmatrix}$.

The density of the random vector (X_1, X_2, μ_1, μ_2) is given by

$$f_{(X_1,X_2,\mu_1,\mu_2)}(x_1, x_2, \mu_1, \mu_2) = \frac{|\Sigma_4|^{-1/2}}{(2\pi)^2} \exp\left\{-\frac{1}{2}((x-\vartheta)', (\mu-\vartheta)')\Sigma_4^{-1}\begin{pmatrix} x - \vartheta \\ \mu - \vartheta \end{pmatrix}\right\},$$

and the one of (Y, X_2, μ_1, μ_2) is given by

$$f_{(Y,X_2,\mu_1,\mu_2)}(y, x_2, \mu_1, \mu_2)$$
$$= |x_2| f_{(X_1,X_2,\mu_1,\mu_2)}(yx_2, x_2, \mu_1, \mu_2)$$
$$= \frac{|x_2| \cdot |\Sigma_4|^{-1/2}}{(2\pi)^2} \exp\left\{-\frac{1}{2}((x_2c(y) - \vartheta)', (\mu-\vartheta)')\Sigma_4^{-1}\begin{pmatrix} x_2c(y) - \vartheta \\ \mu - \vartheta \end{pmatrix}\right\}.$$

It comes from (2.7) that

$$(X_1, X_2)' \sim \mathcal{N}(\vartheta; \Sigma + \Lambda),$$

then the density of (X_1, X_2) is

$$f_{(X_1,X_2)}(x_1, x_2) = \frac{|\Sigma + \Lambda|^{-1/2}}{2\pi} \exp\left\{-\frac{1}{2}(x - \vartheta)'(\Sigma + \Lambda)^{-1}(x - \vartheta)\right\}$$

and the one of (Y, X_2) is

$$f_{(Y,X_2)}(y, x_2) = \frac{|x_2| \cdot |\Sigma + \Lambda|^{-1/2}}{2\pi} \exp\left\{-\frac{1}{2}(x_2c(y) - \vartheta)'(\Sigma + \Lambda)^{-1}(x_2c(y) - \vartheta)\right\}$$

where $c(y) = (y, 1)'$. The conditional density of (μ_1, μ_2) given $(Y, X_2) = (y, x_2)$ is

$$\begin{aligned} & f_{(\mu_1,\mu_2|Y,X_2)}(\mu_1, \mu_2|y, x_2) \\ &= \frac{1}{2\pi}\left(\frac{|\Sigma\Lambda|}{|\Sigma + \Lambda|}\right)^{-1/2} \frac{\exp\left\{-\frac{1}{2}((x_2c(y) - \vartheta)', (\mu - \vartheta)')\Sigma_4^{-1}\begin{pmatrix} x_2c(y) - \vartheta \\ \mu - \vartheta \end{pmatrix}\right\}}{\exp\left\{-\frac{1}{2}(x_2c(y) - \vartheta)'(\Sigma + \Lambda)^{-1}(x_2c(y) - \vartheta)\right\}}. \end{aligned}$$

Since

$$\Sigma^{-1} - (\Sigma + \Lambda)^{-1} = \Sigma^{-1}(\Lambda^{-1} + \Sigma^{-1})^{-1}\Sigma^{-1},$$

$$\begin{aligned} & ((x_2c(y) - \vartheta)', (\mu - \vartheta)')\Sigma_4^{-1}\begin{pmatrix} x_2c(y) - \vartheta \\ \mu - \vartheta \end{pmatrix} - (x_2c(y) - \vartheta)'(\Sigma + \Lambda)^{-1}(x_2c(y) - \vartheta) \\ &= \Big(\mu - \vartheta - (\Sigma^{-1} + \Lambda^{-1})^{-1}\Sigma^{-1}(x_2c(y) - \vartheta)\Big)'(\Sigma^{-1} + \Lambda^{-1})\Big(\mu - \vartheta - (\Sigma^{-1} + \Lambda^{-1})^{-1}\Sigma^{-1}(x_2c(y) - \vartheta)\Big). \end{aligned}$$

We have

$$|(\Sigma^{-1} + \Lambda^{-1})^{-1}| = \frac{1}{|\Sigma^{-1} + \Lambda^{-1}|} = \frac{|\Lambda\Sigma|}{|\Lambda + \Sigma|},$$

then the posterior distribution of (μ_1, μ_2) is the normal distribution with mean

$$\mu^* = \vartheta + (\Sigma^{-1} + \Lambda^{-1})^{-1}\Sigma^{-1}((yx_2, x_2)' - \vartheta)$$

and the covariance matrix

$$\Sigma^* = (\Sigma^{-1} + \Lambda^{-1})^{-1};$$

that is

$$(\mu_1, \mu_2)|Y, X_2 = (y, x_2) \sim \mathcal{N}(\mu^*, \Sigma^*).$$

2.

$$\begin{aligned}
&f_{X_2|(Y,\mu_1,\mu_2)}(x_2|y, \mu_1, \mu_2) \\
&\quad = \frac{f_{(X_2,Y,\mu_1,\mu_2)}(x_2, y, \mu_1, \mu_2)}{f_{(Y,\mu_1,\mu_2)}(y, \mu_1, \mu_2)} \\
&\quad = \frac{f_{(X_2,Y)|(\mu_1,\mu_2)}(x_2, y|\mu_1, \mu_2)}{f_{Y|(\mu_1,\mu_2)}(y|\mu_1, \mu_2)} \\
&\quad = \frac{\frac{|x_2|\cdot|\Sigma|^{-1/2}}{2\pi} \exp\left\{-\frac{1}{2}(x_2c(y) - \mu)'\Sigma^{-1}(x_2c(y) - \mu)\right\}}{\frac{1}{\sqrt{2\pi\tau}}\frac{1}{1+z^2} \exp\left(- \frac{\lambda^2}{2}[1 - \delta^2(z \mid \theta)]\right)\left\{2\phi(\lambda\sqrt{\delta^2(z \mid \theta)}) + \lambda\sqrt{\delta^2(z \mid \theta)}[2\Phi(\lambda\sqrt{\delta^2(z \mid \theta)}) - 1]\right\}} \\
&\quad = \frac{(1 + z^2)|x_2|}{\sigma_2^2\sqrt{2\pi}} \frac{\exp\left\{-\frac{1}{2}\left(x_2\sqrt{1 + z^2}/\sigma_2 - \lambda\delta(z|\theta)\right)^2\right\}}{2\phi(\lambda\sqrt{\delta^2(z|\theta)}) + \lambda\sqrt{\delta^2(z|\theta)}\left[2\Phi(\lambda\sqrt{\delta^2(z|\theta)}) - 1\right]}.
\end{aligned}$$

□

It follows from the first part of Lemma 2.9 that the conditional distribution of μ_1/μ_2 given $(Y, X_2) = (y, x_2)$ is Hinkley distributed, then its expectation does not exist. We use the median of the posterior distribution of μ_1/μ_2 as our estimator.

The standard estimator is a ratio of two random normal variables. Thus, the associated least absolute error and the mean squared error are both equal to infinity. The rule of comparison that we use is a modified least absolute error criteria. The best estimator, $\widehat{\alpha}$, of $\alpha = \mu_1/\mu_2$ has the smallest value of the following risk

$$R(\alpha, \widehat{\alpha}) = E[|\widehat{\alpha} - \alpha| - |Y|].$$

2.4.1. Simulation results

Let $\widehat{\alpha}_{std}$ be the standard estimator and $\widehat{\alpha}_{Bayes}$ our estimator. When $\tau = 1$, $\nu = 0$, $\vartheta = (0, 0)$ and $\Lambda = \begin{pmatrix} 3600 & 300 \\ 300 & 2500 \end{pmatrix}$, the comparison of the two estimators of $\alpha = \mu_1/\mu_2 = \tan(\theta)$ is given in the following table.

Table 2. I. Estimates of risk for estimation of $\mu_1/\mu_2 = \tan(\theta)$.

	$\alpha = \sqrt{3}/3$ $(\theta = \pi/6)$		$\alpha = 1$ $(\theta = \pi/4)$		$\alpha = \sqrt{3}$ $(\theta = \pi/3)$	
λ	$\widehat{R}(\alpha, \widehat{\alpha}_{\text{std}})$	$\widehat{R}(\alpha, \widehat{\alpha}_{\text{Bayes}})$	$\widehat{R}(\alpha, \widehat{\alpha}_{\text{std}})$	$\widehat{R}(\alpha, \widehat{\alpha}_{\text{Bayes}})$	$\widehat{R}(\alpha, \widehat{\alpha}_{\text{std}})$	$\widehat{R}(\alpha, \widehat{\alpha}_{\text{Bayes}})$
0.25	0.1208	−7.4477	0.3141	−1.0247	0.7651	−4.1820
1	−0.0048	−1.7523	0.0523	−1.5769	0.3263	−3.4729
2	−0.1686	−4.4729	−0.3305	−1.1335	−0.3444	−2.9000
5	−0.3941	−0.5078	−0.7396	−0.7417	−1.2344	−1.5706
10	−0.4810	−0.4810	−0.8590	−0.8591	−1.4656	−1.4673
20	−0.5277	−0.5276	−0.9256	−0.9256	−1.5867	−1.5869
100	−0.5671	−0.5671	−0.9844	−0.9844	−1.7005	−1.7005

A better estimator has a smaller risk. The simulation studies we have done indicate that our estimator performs better, especially when λ is small. When the sample size becomes large or equivalently the value of λ is large, the two estimators have similar risks. All our simulations were completed using R 3.3.1.

2.5. Conclusion

In this work, we found two expressions of the density function of the ratio of two normal random variables. The parametrization that we proposed to handle the problem of identifiability allowed us to show that the ratio Y converges in law to a normal random variable as the parameter λ tends to infinity while the other parameters remain fixed. We gave explicit bounds on the quality of approximation, in absolute value, between the cumulative distribution function of Y and the one of the standard normal distribution. We also proposed an estimator of the ratio of means in a Bayesian setting. According to simulation results, this estimator seems to perform better than the standard one. For future research, one could show formally that our estimator performs better than the standard estimator.

Acknowledgment

The author Romain KADJE KENMOGNE wishes to thank Institut des sciences Mathématiques, Centre de Recherches Mathématiques, Faculté des Études Supérieures et Postdoctorales of the University of Montreal, and Fondation J. Armand Bombardier for their financial support.

The second author wishes to thank NSERC for their financial support.

References

Milton Abramowitz and Irene A. Stegun. *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, volume 55 of *National Bureau of Standards Applied Mathematics Series*. For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C., 1964.

Piero Barone. A diffusion equation for the density of the ratio of two jointly distributed gaussian variables and the exponential analysis problem. *SIAM Journal on Scientific Computing*, 34(2): A1053–A1078, 2012.

Jacek Białek. Construction of confidence intervals for the Laspeyres price index. *Journal of Statistical Computation and Simulation*, 85(14):2962–2973, 2015.

Eloísa Díaz-Francés and Francisco J Rubio. On the existence of a normal approximation to the distribution of the ratio of two independent normal random variables. *Statistical Papers*, 54(2): 309–323, 2013.

E. C. Fieller. The distribution of the index in a normal bivariate population. *Biometrika*, 24(3/4): 428–440, 1932.

R. A. Fisher and M. J. R. Healy. New tables of Behrens' test of significance. *J. Roy. Statist. Soc. Ser. B.*, 18:212–216, 1956.

Carlotta Galeone and Angiola Pollastri. Confidence intervals for the ratio of two means using the distribution of the quotient of two normals. *Statistics in Transition*, 13(3):451–472, 2012.

R. C. Geary. The frequency distribution of the quotient of two normal variates. *Journal of the Royal Statistical Society*, 93(3):442–446, 1930.

Ryan S. Gill and Jerome P. Keating. Monitoring gasoline blending rates for compliance. *Environmetrics*, 19(4):429–437, 2008.

Jack Hayya, Donald Armstrong, and Nicolas Gressis. A note on the ratio of two normally distributed variables. *Management Science*, 21(11):1338–1341, 1975.

D. V. Hinkley. On the ratio of two correlated normal random variables. *Biometrika*, 56(3):635–639, 1969.

Pekka J. Korhonen and Subhash C. Narula. The probability distribution of the ratio of the absolute values of two normal variables. *J. Statist. Comput. Simulation*, 33(3):173–182, 1989.

George Marsaglia. Ratios of normal variables and ratios of sums of uniform variables. *J. Amer. Statist. Assoc.*, 60:193–204, 1965.

George Marsaglia. Ratios of normal variables. *Journal of Statistical Software*, 16(4):1–10, 2006.

Dolun Öksoy and Leo A. Aroian. The quotient of two correlated normal variables with applications. *Comm. Statist. Simulation Comput.*, 23(1):223–241, 1994.

T. Pham-Gia, N. Turkkan, and E. Marchand. Density of the ratio of two normal random variables and applications. *Comm. Statist. Theory Methods*, 35(7-9):1569–1591, 2006.

David L Soldan, William P Osborne, and Don Gruenbacher. Modeling the economic cost of inadequate teaching and mentoring. In *Frontiers in Education Conference (FIE), 2010 IEEE*, pages F3J–1. IEEE, 2010.

Yanqing Wang, Suojin Wang, and Raymond J Carroll. The direct integral method for confidence intervals for the ratio of two location parameters. *Biometrics*, 71(3):704–713, 2015.

Chapitre 3

INFÉRENCE POUR LA DIFFÉRENCE DE DEUX VARIABLES ALÉATOIRES DE LOI GAMMA

Cet article sera soumis à la revue *Journal of Statistical Computation and Simulation.*

Les principales contributions de *Romain Kadje Kenmogne* à cet article sont présentées.

– Conduite de la revue de la littérature.

– Contribution dans tous les résultats de l'article.

– Conception, écriture et validation des programmes R.

– Conduite des simulations.

– Rédaction de l'article.

Inference for the Difference of Two Gamma Random Variables

Romain Kadje Kenmogne and François Perron

Département de mathématiques et de statistique,

Université de Montréal,

C.P. 6128, Succursale Centre-ville,

Montréal, Canada H3C 3J7

Abstract

In this paper, we extend a family of leptokurtic symmetric distributions of the difference of two independent gamma random variables. This family was used by Augustyniak and Doray (2012). We consider the difference of two positively correlated or independent gamma random variables and explain how to estimate the parameters using the continuous empirical characteristic function estimation method. A simulation study gives an idea on the performance of this approach compared to the discrete one previously used by Augustyniak and Doray (2012). We also describe two ways of generating data from the difference of two positively correlated gamma random variables. For the independence case, we derive an equivariant estimator of the scale parameter when the shape parameter is known.

Keywords: equivariant estimator, double gamma difference, gamma distribution, leptokurtic distribution, symmetric distribution, empirical characteristic function, quadratic distance, parameter estimation, data generation.

3.1. Introduction

The difference of two independent and identically distributed gamma random variables was used by Augustyniak and Doray (2012). Let $Y = X_1 - X_2$, X_i $(i = 1, 2)$ being independent and identically distributed random variables following a gamma distribution with parameters α and β. The density function is given by

$$f(x) = \frac{x^{\alpha-1}e^{-\frac{x}{\beta}}}{\Gamma(\alpha)\beta^{\alpha}}, \quad \alpha > 0,\ \beta > 0 \text{ and for all } x \in [0, +\infty).$$

Here, α and β are respectively the shape parameter and the scale parameter. The characteristic function of Y is given by

$$\begin{aligned}\varphi_Y(t) &= \varphi_{X_1-X_2}(t) = \varphi_{X_1}(t)\varphi_{X_2}(-t)\\ &= (1 - i\beta t)^{-\alpha}(1 + i\beta t)^{-\alpha}\\ &= \left(1 + \beta^2 t^2\right)^{-\alpha}.\end{aligned}$$

Here i is the unit imaginary number ($i^2 = -1$). The distribution of Y is called double gamma difference distribution. As mentioned by Klar (2015), the double gamma distribution (DGD) is a particular case of the variance gamma distribution studied by many authors, see Madan and Seneta (1990) and Seneta (2004).

We consider the case of the difference of two independent or positively correlated and identically distributed gamma random variables. Let (X_1, X_2) be a random vector following a bivariate gamma distribution such that its moment generating function (MGF) is defined as in Chen et al. (2014) by

$$M_{X_1,X_2}(t_1, t_2|\alpha_1, \alpha_2, \lambda_1, \lambda_2, \rho, \gamma) = \left(1 - \frac{t_1}{\lambda_1}\right)^{-\alpha_1}\left(1 - \frac{t_2}{\lambda_2}\right)^{-\alpha_2}\left(1 - \frac{\rho^2 t_1 t_2}{(\lambda_1 - t_1)(\lambda_2 - t_2)}\right)^{-\gamma},$$

where $t_1 < \lambda_1$, $t_2 < \lambda_2$, $\rho^2 t_1 t_2 < (\lambda_1 - t_1)(\lambda_2 - t_2)$, $\rho \in (-1, 1)$ and $\alpha_1, \alpha_2, \lambda_1, \lambda_2, \gamma > 0$. Since we want X_1 and X_2 to be identically distributed, we take $\lambda_1 = \lambda_2 = \lambda$ and $\alpha_1 = \alpha_2 = \gamma = \alpha$.

With the reparametrization $\beta = 1/\lambda$, the MGF becomes

$$M_{X_1,X_2}(t_1, t_2|\alpha, \beta, \rho) = \Big\{(1 - \beta t_1)(1 - \beta t_2) - \rho^2\beta^2 t_1 t_2\Big\}^{-\alpha},$$

with $\rho^2\beta^2 t_1 t_2 < (1 - \beta t_1)(1 - \beta t_2)$, $\rho \in (-1, 1)$ and $\alpha, \beta > 0$. The associated cumulant-generating function is given by

$$K_{X_1,X_2}(t_1, t_2|\alpha, \beta, \rho) = \log(M_{X_1,X_2}(t_1, t_2|\alpha, \beta, \rho)) = -\alpha \log \Big\{(1 - \beta t_1)(1 - \beta t_2) - \rho^2\beta^2 t_1 t_2\Big\}.$$

We have

$$E[X_i] = \left.\frac{\partial K_{X_1,X_2}(t_1, t_2|\alpha, \beta, \rho)}{\partial t_1}\right|_{t_1=0,\ t_2=0} = \alpha\beta, \quad i = 1, 2,$$

$$Var(X_i) = \left.\frac{\partial^2 K_{X_1,X_2}(t_1, t_2|\alpha, \beta, \rho)}{\partial^2 t_i^2}\right|_{t_1=0,\ t_2=0} = \alpha\beta^2, \quad i = 1, 2,$$

and

$$Cov(X_1, X_2) = \left.\frac{\partial^2 K_{X_1,X_2}(t_1, t_2|\alpha, \beta, \rho)}{\partial t_1 \partial t_2}\right|_{t_1=0,\ t_2=0} = \alpha\beta^2\rho^2.$$

We obtain that $Corr(X_1, X_2) = \rho^2$. Then, α is the shape parameter, β is the scale parameter and ρ^2 is the correlation coefficient between X_1 and X_2. The characteristic function of the random vector (X_1, X_2) is given by

$$\varphi_{(X_1,X_2)}(t_1, t_2|\alpha, \beta, \rho) = \Big\{(1 - i\beta t_1)(1 - i\beta t_2) + \rho^2\beta^2 t_1 t_2\Big\}^{-\alpha}, \quad t_1,\ t_2 \in \mathbb{R}.$$

The characteristic function of Y is given by

$$\begin{aligned}\varphi_Y(t|\alpha, \beta, \rho) &= \varphi_{(X_1,X_2)}(t, -t|\alpha, \beta, \rho)\\ &= \Big\{(1 - i\beta t)(1 + i\beta t) - \rho^2\beta^2 t^2\Big\}^{-\alpha}\\ &= \Big\{1 + \beta^2(1 - \rho^2)t^2\Big\}^{-\alpha}, \quad \text{for all } \ t \in \mathbb{R}.\end{aligned}$$

The notation DGD(α, β, ρ) will be used for double gamma difference distribution. Its MGF is given by

$$M_Y(t|\alpha, \beta, \rho) = \Big\{1 - \beta^2(1 - \rho^2)t^2\Big\}^{-\alpha}, \quad \beta^2(1 - \rho^2)t^2 < 1.$$

As in Augustyniak and Doray (2012), here are some properties.

– All of the moments of the DGD family are finite since the moment-generating function exists.

– Given that the characteristic function of the DGD family is real and even, the densities in this family are symmetric with respect to 0. Odd moments are equal to 0.

– $E[Y^{2k}] = \frac{(2k)![\beta^2(1-\rho^2)]^k}{k!} \prod_{j=0}^{k-1}(\alpha + j), \quad k = 1, 2, \ldots$ The proof is similar to the one provided by Augustyniak and Doray (2012).

– The variance of Y is equal to $2\alpha(1-\rho^2)\beta^2$, and its kurtosis is equal to $(3 + 3/\alpha)$. The family of DGD is leptokurtic (the kurtosis is greater than 3 since $\alpha > 0$).

– The characteristic function is infinitely divisible: if $Y_1, \ldots, Y_n$ are i.i.d. random variables from DGD$(\alpha/n, \beta, \rho)$, then the characteristic function of $Y = Y_1 + \cdots + Y_n$ is

$$\varphi_Y(t|\alpha, \beta, \rho) = \varphi_{Y_1}(t|\alpha/n, \beta, \rho) \cdots \varphi_{Y_n}(t|\alpha/n, \beta, \rho) = \left\{1 + \beta^2(1-\rho^2)t^2\right\}^{-\alpha}.$$

– The family of DGD is closed under the scale operation: if Y is a DGD(α, β, ρ) random variable and $a > 0$, then aY is a DGD$(\alpha, a\beta, \rho)$ random variable.

We are interested in parameter estimation. In fact, the densities associated with the variance gamma and DGD$(\alpha, \beta, 0)$ families are known, see Seneta (2004) and Klar (2015). The one of DGD(α, β, ρ), $\rho^2 > 0$ can be derived. The common point between these densities is that it is possible (but not easy) to derive the maximum likelihood estimator of the parameters, see Vaidyanathan and Vani Lakshmi (2015). An alternative method is the empirical characteristic function (ECF) estimation. This method has been investigated or has been used by Feuerverger and McDunnough (1981a), Feuerverger (1990), Carrasco et al. (2002), Yu (2004), Augustyniak and Doray (2012) and Klar (2015). The empirical characteristic function of Y is given by

$$\psi_n(t|\mathbf{y}) = \frac{1}{n}\sum_{j=1}^{n} e^{ity_j},$$

where y_j, $j = 1, \ldots, n$, are i.i.d. observations from Y, and $\mathbf{y} = (y_1, y_2, \ldots, y_n)$. We consider two ways to conduct estimation using ECF method. Firstly, the continuous ECF method where estimators are obtained by minimizing the function

$$g(\theta|y) = \int_{\mathbb{R}} |\varphi_Y(t|\theta) - \psi_n(t|\mathbf{y})|^2 w(t|\theta)\mathrm{d}t,$$

where θ is the vector of parameters, and $w(\cdot|\theta)$ is a continuous weight function. The challenge of this method relates to the choice of the weight function. In practice, an exponential weight function is often used. Although the exponential weight function has the numerical advantages of being tractable, the resulting ECF estimator is in general less efficient than the maximum likelihood estimator, see Yu (2004). An optimal weight function in the sense that the resulting estimator attains the maximum likelihood efficiency is given by Feuerverger and McDunnough (1981b) with

$$w^*(t) = \frac{1}{2\pi}\int \exp(-ity)\frac{\partial \log f(y|\theta)}{\partial\theta}\mathrm{d}y,$$

where $f(\cdot|\theta)$ is the density function of Y. This optimal choice is not easy to apply in the case of DGD and VG families. A second possibility is the discrete ECF method which consists of choosing q discrete points $t_1, \ldots, t_q$ and minimizing

$$\left[\frac{1}{n}\sum_{j=1}^{n} Z(y_j|\theta)\right]' W_n \left[\frac{1}{n}\sum_{j=1}^{n} Z(y_j|\theta)\right],$$

where,

$$Z(y_j|\theta) = (Re[C(t_1|y_j,\theta)], \ldots, Re[C(t_q|y_j,\theta)], Im[C(t_1|y_j,\theta)], \ldots, Im[C(t_q|y_j,\theta)]),$$

$$C(t|y_j,\theta) = \exp(ity_j) - \varphi_Y(t|\theta),$$

$Re[\cdot]$ and $Im[\cdot]$ are the real and imaginary parts of a complex number, and W_n is a positive semi-definite weighting matrix which converges to a positive definite matrix W_0 almost surely, see Yu (2004) for more details. Augustyniak and Doray (2012) used this approach for the estimation of the parameters in the case of DGD$(\alpha, \beta, 0)$. Augustyniak (2008), in the conclusion of his thesis, was wondering how would the continuous ECF estimators perform compared to their results in the discrete setting. We are going to estimate the parameters in the case of DGD(α, β, ρ), firstly by using the continuous ECF method with a weight function being the density function of a normal random variable, and secondly with the discrete ECF method. We will compare the results of the two approaches.

The rest of this paper is organized as follows. In Section 3.2, we present the best equivariant estimator of β when α and ρ are known. In Section 3.3, we derive the density function of a random variable Y following DGD(α, β, ρ). Then, we explain how to generate observations from a bivariate gamma (X_1, X_2) such that $Y = X_1 - X_2 \sim$ DGD(α, β, ρ). The Section 3.4 is reserved for estimation of the parameters. We explain how to compute the continuous and discrete ECF estimators. We use the control variate method in the continuous case in order to reduce the variance of the estimators. Finally, in Section 3.5, we provide simulation results for different situations.

3.2. Best equivariant estimator of β

Let $X_i \sim$ Gamma(α, β), $i = 1, 2$, be two independent random variables. We suppose that α is known. Let $Y = X_1 - X_2$ and f be the density of X_i, $i = 1, 2$. The density of Y is given by

$$\begin{aligned} h(y|\beta) &= \int_{-\infty}^{+\infty} f(t)f(t-y)\mathrm{d}t \\ &= \int_{-\infty}^{+\infty} \frac{1}{\Gamma^2(\alpha)\beta^{2\alpha}} t^{\alpha-1}(t-y)^{\alpha-1} \exp\left(-\frac{(2t-y)}{\beta}\right) I_{(0,\infty)}(t) I_{(0,\infty)}(t-y)\mathrm{d}t. \end{aligned}$$

For $y \geq 0$, we have

$$\begin{aligned} h(y|\beta) &= \int_{y}^{\infty} \frac{1}{\Gamma^2(\alpha)\beta^{2\alpha}} t^{\alpha-1}(t-y)^{\alpha-1} \exp\left(-\frac{(2t-y)}{\beta}\right) \mathrm{d}t \\ &= y^{2\alpha-1} \int_{1}^{\infty} \frac{1}{\Gamma^2(\alpha)\beta^{2\alpha}} [u(u-1)]^{\alpha-1} \exp\left(-\frac{y(2u-1)}{\beta}\right) \mathrm{d}u \\ &= \frac{1}{2} y^{2\alpha-1} \int_{1}^{\infty} \frac{1}{\Gamma^2(\alpha)\beta^{2\alpha}} \frac{[v^2-1]^{\alpha-1}}{4^{\alpha-1}} \exp\left(-\frac{yv}{\beta}\right) \mathrm{d}v. \end{aligned}$$

For $y \leq 0$, we have

$$h(y|\beta) = \int_{0}^{\infty} \frac{1}{\Gamma^2(\alpha)\beta^{2\alpha}} t^{\alpha-1}(t-y)^{\alpha-1} \exp\left(-\frac{(2t-y)}{\beta}\right) \mathrm{d}t$$

$$= \int_{-y}^{\infty} \frac{1}{\Gamma^2(\alpha)\beta^{2\alpha}} (x+y)^{\alpha-1} x^{\alpha-1} \exp\left(-\frac{2x+y}{\beta}\right) \mathrm{d}x$$

$$= \frac{1}{2}(-y)^{2\alpha-1} \int_{1}^{\infty} \frac{1}{\Gamma^2(\alpha)\beta^{2\alpha}} \frac{[v^2-1]^{\alpha-1}}{4^{\alpha-1}} \exp\left(\frac{yv}{\beta}\right) \mathrm{d}v.$$

Then,

$$h(y|\beta) = \frac{1}{2^{2\alpha-1}} \frac{1}{\Gamma^2(\alpha)\beta^{2\alpha}} |y|^{2\alpha-1} \int_{1}^{\infty} [v^2-1]^{\alpha-1} \exp\left(-\frac{|y|v}{\beta}\right) \mathrm{d}v$$

$$= \frac{1}{\Gamma^2(\alpha)} \frac{1}{\beta} \left(\frac{|y|}{2\beta}\right)^{2\alpha-1} \int_{1}^{\infty} [v^2-1]^{\alpha-1} \exp\left(-\frac{|y|v}{\beta}\right) \mathrm{d}v.$$

If $Y_1, \ldots, Y_n$ are independent and identically distributed random variables with density function $h(\cdot|\beta)$, then the best equivariant estimator of the parameter β, for the loss function

$$L(\beta, d) = \frac{(d-\beta)^2}{\beta^2},$$

is given by the Pitman formula

$$\widehat{\beta} = \frac{\int_0^\infty \frac{1}{\beta^2} \prod_{i=1}^n h(y_i|\beta) \mathrm{d}\beta}{\int_0^\infty \frac{1}{\beta^3} \prod_{i=1}^n h(y_i|\beta) \mathrm{d}\beta} \quad \text{(see Lehmann and Casella, 1998).}$$

Let $d(\cdot)$, be the function given by

$$d(\gamma) = \int_0^\infty \frac{1}{\beta^{\gamma+1}} \prod_{i=1}^n h(y_i|\beta) \mathrm{d}\beta$$

$$= \int_0^\infty \omega^{\gamma+1} \prod_{i=1}^n h(y_i|\frac{1}{\omega}) \mathrm{d}\omega$$

$$= \int_0^\infty \int_{[1,\infty]^n} \frac{\omega^{2\alpha n+\gamma-1}}{\Gamma^{2n}(\alpha)} \prod_{i=1}^n \left(\left|\frac{y_i}{2}\right|^{2\alpha-1} [v_i^2-1]^{\alpha-1} \right) \exp\left\{-\omega \sum_{i=1}^n |y_i| v_i \right\} \mathrm{d}v_1 \ldots \mathrm{d}v_n \mathrm{d}\omega$$

$$= \int_{[1,\infty]^n} \frac{1}{\Gamma^{2n}(\alpha)} \prod_{i=1}^n \left(\left|\frac{y_i}{2}\right|^{2\alpha-1} [v_i^2-1]^{\alpha-1} \right) \left\{ \int_0^\infty \omega^{2\alpha n+\gamma-1} \exp\left(-\omega \sum_{i=1}^n |y_i| v_i \right) \mathrm{d}\omega \right\} \mathrm{d}v_1 \ldots \mathrm{d}v_n$$

$$= \int_{[1,\infty]^n} \frac{1}{\Gamma^{2n}(\alpha)} \prod_{i=1}^n \left(\left|\frac{y_i}{2}\right|^{2\alpha-1} [v_i^2-1]^{\alpha-1} \right) \frac{\Gamma(2\alpha n+\gamma)}{(\sum_{i=1}^n |y_i| v_i)^{2\alpha n+\gamma}} \mathrm{d}v_1 \ldots \mathrm{d}v_n.$$

We are going to calculate the values of the function $d(\cdot)$ numerically by using a Markov Chain Monte Carlo approximation with the Gibbs sampler. The function

$$d(\gamma) = \int_{[1,\infty]^n} \frac{\Gamma(2\alpha n+\gamma)}{\Gamma^{2n}(\alpha)} \frac{\left(\prod_{i=1}^n \left(\frac{|y_i|^2}{4}(v_i^2-1)\right)\right)^{\alpha-1}}{\left(\sum_{i=1}^n |y_i|v_i\right)^{2n(\alpha-1)+\gamma}} \frac{\prod_{i=1}^n \frac{|y_i|}{2}}{\left(\sum_{i=1}^n |y_i|v_i\right)^{2n}} \mathrm{d}v_1 \ldots \mathrm{d}v_n$$

is proportional to the expectation

$$\Gamma(2\alpha n+\gamma) E\left[\frac{\left(\prod_{i=1}^n \left(\frac{|y_i|^2}{4}(V_i^2-1)\right)\right)^{\alpha-1}}{\left(\sum_{i=1}^n |y_i|V_i\right)^{2n(\alpha-1)+\gamma}}\right]$$

where $(V_1,\ldots,V_n)$ is a random vector having the joint density function proportional to

$$\frac{1}{\left(\sum_{i=1}^n |y_i|v_i\right)^{2n}}, \quad v_i \geq 1,\ i=1,\ldots,n.$$

The best equivariant estimator of β, with the loss function given above, is

$$\begin{aligned}\widehat{\beta} &= d(1)/d(2)\\ &= \frac{1}{2\alpha n+1} \frac{E\left[\frac{\left(\prod_{i=1}^n \left(\frac{|y_i|^2}{4}(V_i^2-1)\right)\right)^{\alpha-1}}{\left(\sum_{i=1}^n |y_i|V_i\right)^{2n(\alpha-1)+1}}\right]}{E\left[\frac{\left(\prod_{i=1}^n \left(\frac{|y_i|^2}{4}(V_i^2-1)\right)\right)^{\alpha-1}}{\left(\sum_{i=1}^n |y_i|V_i\right)^{2n(\alpha-1)+2}}\right]}.\end{aligned}$$

We will compute this expression numerically by the following formula

$$\widehat{\beta} = \frac{1}{2\alpha n+1} \frac{\frac{1}{m}\sum_{t=1}^m \frac{\left(\prod_{i=1}^n \left[\frac{|y_i|^2}{4}\left(\left(v_i^{(t)}\right)^2-1\right)\right]\right)^{\alpha-1}}{\left(\sum_{i=1}^n |y_i|v_i^{(t)}\right)^{2n(\alpha-1)+1}}}{\frac{1}{m}\sum_{t=1}^m \frac{\left(\prod_{i=1}^n \left[\frac{|y_i|^2}{4}\left(\left(v_i^{(t)}\right)^2-1\right)\right]\right)^{\alpha-1}}{\left(\sum_{i=1}^n |y_i|v_i^{(t)}\right)^{2n(\alpha-1)+2}}}, \tag{3.1}$$

where m is a large integer, $(v_1^{(t)},\ldots,v_n^{(t)})$, $t=1,\ldots,m$ are observations generated by the Gibbs sampler. The conditional density of V_j given $(V_1,\ldots,V_{j-1},V_{j+1},\ldots,V_n) =$

$(v_1, \ldots, v_{j-1}, v_{j+1}, \ldots, v_n)$ is proportional to

$$\frac{1}{\left(|y_j|v_j + \sum_{i=1, i\neq j}^{n} |y_i|v_i\right)^{2n}}, \quad v_j \geq 1.$$

In order to generate the Markov chain, we find the following survival function

$$\begin{aligned}
&P(V_j \geq v_j | (V_1, \ldots, V_{j-1}, V_{j+1}, \ldots, V_n) = (v_1, \ldots, v_{j-1}, v_{j+1}, \ldots, v_n)) \\
&= \frac{\int_{v_j}^{\infty} \frac{1}{\left(|y_j|u + \sum_{i=1, i\neq j}^{n} |y_i|v_i\right)^{2n}} \mathrm{d}u}{\int_{1}^{\infty} \frac{1}{\left(|y_j|u + \sum_{i=1, i\neq j}^{n} |y_i|v_i\right)^{2n}} \mathrm{d}u} \\
&= \frac{\frac{1}{(2n-1)|y_j|\left(|y_j|v_j + \sum_{i=1, i\neq j}^{n} |y_i|v_i\right)^{2n-1}}}{\frac{1}{(2n-1)|y_j|\left(|y_j| + \sum_{i=1, i\neq j}^{n} |y_i|v_i\right)^{2n-1}}} \\
&= \left(\frac{|y_j| + \sum_{i=1, i\neq j}^{n} |y_i|v_i}{|y_j|v_j + \sum_{i=1, i\neq j}^{n} |y_i|v_i}\right)^{2n-1}.
\end{aligned}$$

We have

$$\left(\frac{|y_j| + \sum_{i=1, i\neq j}^{n} |y_i|v_i}{|y_j|v_j + \sum_{i=1, i\neq j}^{n} |y_i|v_i}\right)^{2n-1} = u \Longleftrightarrow v_j = \frac{\left(|y_j| + \sum_{i=1, i\neq j}^{n} |y_i|v_i\right) u^{-\frac{1}{2n-1}} - \sum_{i=1, i\neq j}^{n} |y_i|v_i}{|y_j|}$$

Let us notice that, if F is a cumulative distribution function and $\overline{F}$ is the associated survival functions, then

$$\begin{aligned}
F^{-1}(u) = x &\Longleftrightarrow F(x) = u \\
&\Longleftrightarrow \overline{F}(x) = 1 - u \\
&\Longleftrightarrow \overline{F}^{-1}(1 - u) = x.
\end{aligned}$$

Furthermore, if $U \sim U(0, 1)$ then $1 - U \sim U(0, 1)$. Thus, the random variable $U = \overline{F}_X(X)$ has a uniform distribution on $[0, 1]$. According to the inverse transform sampling, $X = \overline{F}_X^{-1}(U)$ has the distribution F when $U \sim U(0, 1)$. For an observation u from $U([0, 1])$ and with

$v_1, \ldots, v_{j-1}, v_{j+1}, \ldots, v_n$ known, the observation v_j is given by the following formula

$$v_j = \frac{\left(|y_j| + \sum_{i=1, i\neq j}^{n} |y_i| v_i\right) u^{-\frac{1}{2n-1}} - \sum_{i=1, i\neq j}^{n} |y_i| v_i}{|y_j|}. \tag{3.2}$$

Then, to generate the Markov chain, we can generate observations from $U(0,1)$ and use Formula (3.2).

Algorithm for computing $\widehat{\beta}$ via the Gibbs sampler

– Generate the initial value of the vector $\mathbf{v}$, $(v_1^{(0)}, \ldots, v_n^{(0)})$, from a uniform distribution.

– At iteration $t = 1, 2, \ldots, m$, given $\mathbf{v}^{(t-1)} = (v_1^{(t-1)}, \ldots, v_n^{(t-1)})$, generate

$$\begin{aligned}
&v_1^{(t)} \sim f_{V_1|V_{-1}}(v_1 | v_2^{(t-1)}, \ldots, v_n^{(t-1)}),\\
&v_2^{(t)} \sim f_{V_2|V_{-2}}(v_2 | v_1^{(t)}, v_3^{(t-1)}, \ldots, v_n^{(t-1)}),\\
&\vdots\\
&v_k^{(t)} \sim f_{V_k|V_{-k}}(v_k | v_1^{(t)}, \ldots, v_{k-1}^{(t)}, v_{k+1}^{(t-1)} \ldots, v_n^{(t-1)}),\\
&\vdots\\
&v_n^{(t)} \sim f_{V_k|V_{-k}}(v_n | v_1^{(t)}, \ldots, v_{n-1}^{(t)}).
\end{aligned}$$

– Compute $\widehat{\beta}$ using Formula 3.1.

3.3. Difference of two independent or positively correlated gamma random variables

3.3.1. Density function

Let

$$K_n(w) = \frac{\sqrt{\pi}}{\Gamma(n+1/2)} \left(\frac{w}{2}\right)^n \int_1^{+\infty} e^{-wx}(x^2-1)^{n-1/2} \mathrm{d}x, \text{ for } n > -1/2 \text{ and } w > 0, \tag{3.3}$$

be one representation of the modified Bessel function of the second kind.

Lemma 3.1. *The density function of $Y \sim DGD(\alpha, \beta, \rho)$ is given by*

$$f_Y(y|\alpha,\beta,\rho) = \begin{cases} \sum_{k=0}^{+\infty} \frac{\rho^{2k}}{k!\sqrt{\pi}\Gamma(\alpha)} \frac{1}{\beta^\alpha[\beta(1-\rho^2)]^{k+1/2}} \left(\frac{|y|}{2}\right)^{k+\alpha-1/2} K_{k+\alpha-1/2}\left(\frac{|y|}{\beta(1-\rho^2)}\right) & \text{if } y \neq 0,\\ \sum_{k=0}^{+\infty} \frac{\Gamma(2(\alpha+k)-1)}{k!\Gamma(k+\alpha)\Gamma(\alpha)} \rho^{2k} \frac{(1-\rho^2)^{\alpha-1}}{\beta} \left(\frac{1}{2}\right)^{2(\alpha+k)-1} & \text{if } y = 0, \end{cases}$$

where $K_n(\cdot)$ is defined in Equation (3.3).

Proof. The density of (X_1, X_2) is given by (see Chen et al., 2014)

$$\begin{aligned} & f_{X_1,X_2}(x_1, x_2|\alpha, \beta, \rho) \\ & \quad = f_{X_1}(x_1|\alpha, \beta) f_{X_2}(x_2|\alpha, \beta) \frac{\exp\{-\rho^2(x_1 + x_2)/(\beta(1-\rho^2))\}}{(1-\rho^2)^\alpha} {}_0F_1\left(;\alpha; \frac{\rho^2}{\beta^2(1-\rho^2)^2} x_1 x_2\right), \end{aligned} \tag{3.4}$$

where ${}_0F_1(;\alpha;z) = \Gamma(\alpha)\sum_{k=0}^{\infty}(z^k/(k!\Gamma(\alpha+k)))$ is the confluent hypergeometric limit function. This density can be rewritten as

$$\begin{aligned} & f_{X_1,X_2}(x_1, x_2|\alpha, \beta, \rho) \\ & \quad = f_{X_1}(x_1|\alpha, \beta) f_{X_2}(x_2|\alpha, \beta) \frac{\exp\{-\rho^2(x_1 + x_2)/(\beta(1-\rho^2))\}}{(1-\rho^2)^\alpha} {}_0F_1\left(;\alpha; \frac{\rho^2}{\beta^2(1-\rho^2)^2} x_1 x_2\right) \\ & \quad = \frac{x_1^{\alpha-1} e^{-x_1/\beta}}{\Gamma(\alpha)\beta^\alpha} \frac{x_2^{\alpha-1} e^{-x_2/\beta}}{\Gamma(\alpha)\beta^\alpha} \frac{\exp\{-\rho^2(x_1 + x_2)/(\beta(1-\rho^2))\}}{(1-\rho^2)^\alpha} {}_0F_1\left(;\alpha; \frac{\rho^2}{\beta^2(1-\rho^2)^2} x_1 x_2\right) \\ & \quad = \exp\left(-\frac{x_1 + x_2}{\beta(1-\rho^2)}\right) \frac{x_1^{\alpha-1} x_2^{\alpha-1}}{\Gamma^2(\alpha)\{\beta^2(1-\rho^2)\}^\alpha} {}_0F_1\left(;\alpha; \frac{\rho^2}{\beta^2(1-\rho^2)^2} x_1 x_2\right), \end{aligned}$$

where ${}_0F_1(;\alpha;z) = \Gamma(\alpha)\sum_{k=0}^{\infty}(z^k/(k!\Gamma(\alpha+k)))$ is the confluent hypergeometric limit function, with $x_1 \geq 0$, $x_2 \geq 0$, $\alpha > 0$, $\beta > 0$ and $\rho \in (-1, 1)$. Let $Y = X_1 - X_2$ and $Z = X_1$. The density function of the random variable Y is given by

$$\begin{aligned} & f_Y(y|\alpha, \beta, \rho) \\ & = \int_{\mathbb{R}} f_{Z,Y}(z, y|\alpha, \beta, \rho) \mathrm{d}z \\ & = \int_{0\vee y}^{+\infty} f_{X_1,X_2}(z, z-y|\alpha, \beta, \rho) \mathrm{d}z \\ & = \int_{0\vee y}^{+\infty} \exp\left(-\frac{z + (z-y)}{\beta(1-\rho^2)}\right) \frac{[z(z-y)]^{\alpha-1}}{\Gamma^2(\alpha)\{\beta^2(1-\rho^2)\}^\alpha} {}_0F_1\left(;\alpha; \frac{\rho^2}{\beta^2(1-\rho^2)^2} z(z-y)\right) \mathrm{d}z \\ & = \exp\left(\frac{y}{\beta(1-\rho^2)}\right) \int_{0\vee y}^{+\infty} \exp\left(-\frac{2z}{\beta(1-\rho^2)}\right) \frac{[z(z-y)]^{\alpha-1}}{\Gamma^2(\alpha)\{\beta^2(1-\rho^2)\}^\alpha} \sum_{k=0}^{+\infty} \left\{\frac{\rho^2 z(z-y)}{\beta^2(1-\rho^2)^2}\right\}^k \frac{\Gamma(\alpha)}{k!\Gamma(k+\alpha)} \mathrm{d}z \end{aligned}$$

$$= \exp\left(\frac{y}{\beta(1-\rho^2)}\right)\sum_{k=0}^{+\infty}\frac{\{\rho^2/[\beta^2(1-\rho^2)^2]\}^k}{k!\Gamma(k+\alpha)\Gamma(\alpha)\{\beta^2(1-\rho^2)\}^\alpha}\int_{0\vee y}^{+\infty}[z(z-y)]^{k+\alpha-1}\exp\left(-\frac{2z}{\beta(1-\rho^2)}\right)\mathrm{d}z. \tag{3.5}$$

Let

$$I = \int_{0\vee y}^{+\infty}[z(z-y)]^{k+\alpha-1}\exp\left(-\frac{2z}{\beta(1-\rho^2)}\right)\mathrm{d}z.$$

For $y > 0$ and $z = uy$, we have

$$I = \int_1^{+\infty}[y^2u(u-1)]^{k+\alpha-1}\exp\left(-\frac{2uy}{\beta(1-\rho^2)}\right)y\mathrm{d}u.$$

Let $v = 2u-1$, we have $u(u-1) = [(2u-1)^2-1]/4$ and

$$\begin{aligned} I &= \int_1^{+\infty}\left[\left(\frac{y}{2}\right)^2(v^2-1)\right]^{k+\alpha-1}\exp\left(-\frac{2}{\beta(1-\rho^2)}\frac{v+1}{2}y\right)\frac{y}{2}\mathrm{d}v \\ &= \left(\frac{y}{2}\right)^{2(k+\alpha)-1}\exp\left(-\frac{y}{\beta(1-\rho^2)}\right)\int_1^{+\infty}\left[v^2-1\right]^{k+\alpha-1}\exp\left\{-\frac{vy}{\beta(1-\rho^2)}\right\}\mathrm{d}v. \end{aligned}$$

For $y < 0$ and $z = (1-u)y$, we have

$$I = \int_1^{+\infty}[y^2u(u-1)]^{k+\alpha-1}\exp\left(-\frac{2y(1-u)}{\beta(1-\rho^2)}\right)(-y)\mathrm{d}u.$$

Let $v = 2u-1$, we have

$$\begin{aligned} I &= \int_1^{+\infty}\left[\left(-\frac{y}{2}\right)^2(v^2-1)\right]^{k+\alpha-1}\exp\left(-\frac{2y}{\beta(1-\rho^2)}\left(1-\frac{v+1}{2}\right)\right)\left(-\frac{y}{2}\right)\mathrm{d}v \\ &= \left(-\frac{y}{2}\right)^{2(k+\alpha)-1}\exp\left(-\frac{y}{\beta(1-\rho^2)}\right)\int_1^{+\infty}\left[v^2-1\right]^{k+\alpha-1}\exp\left(\frac{vy}{\beta(1-\rho^2)}\right)\mathrm{d}v. \end{aligned}$$

Thus, for $y \neq 0$,

$$\begin{aligned} I &= \left(\frac{|y|}{2}\right)^{2(k+\alpha)-1}\exp\left(-\frac{y}{\beta(1-\rho^2)}\right)\int_1^{+\infty}\left[v^2-1\right]^{k+\alpha-1}\exp\left(-\frac{v|y|}{\beta(1-\rho^2)}\right)\mathrm{d}v \\ &= \left(\frac{|y|}{2}\right)^{2(k+\alpha)-1}\exp\left(-\frac{y}{\beta(1-\rho^2)}\right)\frac{\Gamma(k+\alpha)}{\sqrt{\pi}}\left(\frac{2\beta(1-\rho^2)}{|y|}\right)^{k+\alpha-1/2}K_{k+\alpha-1/2}\left(\frac{|y|}{\beta(1-\rho^2)}\right) \\ &= \exp\left(-\frac{y}{\beta(1-\rho^2)}\right)\left(\frac{\beta(1-\rho^2)|y|}{2}\right)^{k+\alpha-1/2}\frac{\Gamma(k+\alpha)}{\sqrt{\pi}}K_{k+\alpha-1/2}\left(\frac{|y|}{\beta(1-\rho^2)}\right). \end{aligned}$$

For $y \neq 0$, the density of Y is given by

$$f_Y(y|\alpha,\beta,\rho) = \sum_{k=0}^{+\infty} \frac{\{\rho^2/[\beta^2(1-\rho^2)^2]\}^k}{k!\sqrt{\pi}\Gamma(\alpha)\{\beta^2(1-\rho^2)\}^\alpha} \left(\frac{\beta(1-\rho^2)|y|}{2}\right)^{k+\alpha-1/2} K_{k+\alpha-1/2}\left(\frac{|y|}{\beta(1-\rho^2)}\right)$$
$$= \sum_{k=0}^{+\infty} \frac{\rho^{2k}}{k!\sqrt{\pi}\Gamma(\alpha)} \frac{1}{\beta^\alpha[\beta(1-\rho^2)]^{k+1/2}} \left(\frac{|y|}{2}\right)^{k+\alpha-1/2} K_{k+\alpha-1/2}\left(\frac{|y|}{\beta(1-\rho^2)}\right).$$

For $y = 0$, we have from Formula (3.5),

$$f_Y(0|\alpha,\beta,\rho) = \sum_{k=0}^{+\infty} \frac{\{\rho^2/[\beta^2(1-\rho^2)^2]\}^k}{k!\Gamma(k+\alpha)\Gamma(\alpha)\{\beta^2(1-\rho^2)\}^\alpha} \int_0^{+\infty} [z]^{2(k+\alpha-1)} \exp\left(-\frac{2z}{\beta(1-\rho^2)}\right) \mathrm{d}z.$$

Let

$$I_3 = \int_0^{+\infty} [z]^{2(k+\alpha-1)} \exp\left(-\frac{2z}{\beta(1-\rho^2)}\right) \mathrm{d}z \quad \text{and} \quad u = \frac{2z}{\beta(1-\rho^2)},$$

we have

$$I_3 = \int_0^{+\infty} e^{-u} \left(\frac{\beta(1-\rho^2)}{2}u\right)^{2(k+\alpha-1)} \frac{\beta(1-\rho^2)}{2}\mathrm{d}u$$
$$= \left(\frac{\beta(1-\rho^2)}{2}\right)^{2(k+\alpha)-1} \int_0^{+\infty} e^{-u}u^{2(k+\alpha-1)}\mathrm{d}u$$
$$= \left(\frac{\beta(1-\rho^2)}{2}\right)^{2(k+\alpha)-1} \Gamma(2(\alpha+k)-1).$$

Then,

$$f_Y(0|\alpha,\beta,\rho) = \sum_{k=0}^{+\infty} \frac{\{\rho^2/[\beta^2(1-\rho^2)^2]\}^k}{k!\Gamma(k+\alpha)\Gamma(\alpha)\{\beta^2(1-\rho^2)\}^\alpha} \left(\frac{\beta(1-\rho^2)}{2}\right)^{2(k+\alpha)-1} \Gamma(2(\alpha+k)-1)$$
$$= \sum_{k=0}^{+\infty} \frac{\Gamma(2(\alpha+k)-1)}{k!\Gamma(k+\alpha)\Gamma(\alpha)} \rho^{2k} \frac{(1-\rho^2)^{\alpha-1}}{\beta} \left(\frac{1}{2}\right)^{2(\alpha+k)-1}.$$

□

3.3.2. Generation of observations from a bivariate gamma distribution

Lemma 3.2. *Let (X_1, X_2) be a random vector from $DGD(\alpha,\beta,\rho)$.*

1. *The conditional distribution of X_2 given $X_1 = x_1$ is a mixture distribution of $Gamma(\alpha+k, \beta(1-\rho^2))$, where k is a nonnegative integer, and weights come from*

a Poisson distribution with mean $\rho^2 x_1/(\beta(1-\rho^2))$:

$$f_{X_2|X_1}(x_2|x_1,\alpha,\beta,\rho) = \sum_{k=0}^{\infty} e^{-\rho^2 x_1/(\beta(1-\rho^2))} \frac{\left\{\frac{\rho^2}{\beta(1-\rho^2)} x_1\right\}^k}{k!} \frac{x_2^{\alpha+k-1} e^{-x_2/(\beta(1-\rho^2))}}{\Gamma(\alpha+k)\{\beta(1-\rho^2)\}^{\alpha+k}}.$$

2. *The density of (X_1, X_2) can be written as a mixture density. The mixture components are the product of two Gamma$(\alpha+k, \beta(1-\rho^2))$ densities and the weights are given by the probability mass function of a generalized negative binomial distribution with parameters $\alpha > 0$ and $\rho^2 \in (0,1)$. Therefore,*

$$f_{X_1,X_2}(x_1,x_2|\alpha,\beta,\rho)$$
$$= \sum_{k=0}^{\infty} \frac{\Gamma(\alpha+k)}{k!\Gamma(\alpha)} (\rho^2)^k (1-\rho^2)^\alpha \left[\frac{x_1^{\alpha+k-1} \exp\left(\frac{-x_1}{\beta(1-\rho^2)}\right)}{\Gamma(\alpha+k)[\beta(1-\rho^2)]^{\alpha+k}}\right] \times \left[\frac{x_2^{\alpha+k-1} \exp\left(\frac{-x_2}{\beta(1-\rho^2)}\right)}{\Gamma(\alpha+k)[\beta(1-\rho^2)]^{\alpha+k}}\right].$$

Proof. The density of the random vector (X_1, X_2), given by the Formula (3.4). The density of $X_2|X_1 = x_1$ is given by

$$f_{X_2|X_1}(x_2|x_1,\alpha,\beta,\rho)$$
$$= f_{X_2}(x_2|\alpha,\beta) \frac{\exp\{-\rho^2 x_1/(\beta(1-\rho^2)) - \rho^2 x_2/(\beta(1-\rho^2))\}\Gamma(\alpha)}{(1-\rho^2)^\alpha} \sum_{k=0}^{\infty} \frac{\left\{\frac{\rho^2}{\beta^2(1-\rho^2)^2} x_1 x_2\right\}^k}{k!\Gamma(\alpha+k)}$$
$$= e^{-\rho^2 x_1/(\beta(1-\rho^2))} \sum_{k=0}^{\infty} \frac{\left\{\frac{\rho^2}{\beta(1-\rho^2)} x_1\right\}^k}{k!} \frac{\left\{\frac{x_2}{\beta(1-\rho^2)}\right\}^k}{\Gamma(\alpha+k)} \frac{e^{-\rho^2 x_2/(\beta(1-\rho^2))}\Gamma(\alpha)}{(1-\rho^2)^\alpha} f_{X_2}(x_2|\alpha,\beta)$$
$$= e^{-\rho^2 x_1/(\beta(1-\rho^2))} \sum_{k=0}^{\infty} \frac{\left\{\frac{\rho^2}{\beta(1-\rho^2)} x_1\right\}^k}{k!} \frac{\left\{\frac{x_2}{\beta(1-\rho^2)}\right\}^k}{\Gamma(\alpha+k)} \frac{e^{-\rho^2 x_2/(\beta(1-\rho^2))}\Gamma(\alpha)}{(1-\rho^2)^\alpha} \frac{x^{\alpha-1} e^{-x_2/\beta}}{\Gamma(\alpha)\beta^\alpha}$$
$$= \sum_{k=0}^{\infty} e^{-\rho^2 x_1/(\beta(1-\rho^2))} \frac{\left\{\frac{\rho^2}{\beta(1-\rho^2)} x_1\right\}^k}{k!} \frac{x_2^{\alpha+k-1} e^{-x_2/(\beta(1-\rho^2))}}{\Gamma(\alpha+k)\{\beta(1-\rho^2)\}^{\alpha+k}}.$$

Then, the distribution of $X_2|X_1 = x_1$ is a mixture distribution of Gamma$(\alpha+k, \beta(1-\rho^2))$ where the weights are the probabilities of Poisson distribution with mean $\rho^2 x_1/\{\beta(1-\rho^2)\}$. Moreover,

$$
\begin{aligned}
&f_{X_1,X_2}(x_1,x_2|\alpha,\beta,\rho)\\
&\quad= f_{X_1}(x_1|\alpha,\beta)f_{X_2}(x_2|\alpha,\beta)\frac{\exp\{-\rho^2(x_1+x_2)/(\beta(1-\rho^2))\}}{(1-\rho^2)^\alpha}\,{}_0F_1\left(;\alpha;\frac{\rho^2}{\beta^2(1-\rho^2)^2}x_1x_2\right)\\
&\quad= \frac{x_1^{\alpha-1}e^{-x_1/\beta}}{\Gamma(\alpha)\beta^\alpha}\frac{x_2^{\alpha-1}e^{-x_2/\beta}}{\Gamma(\alpha)\beta^\alpha}\frac{\exp\{-\rho^2(x_1+x_2)/(\beta(1-\rho^2))\}}{(1-\rho^2)^\alpha}\,{}_0F_1\left(;\alpha;\frac{\rho^2}{\beta^2(1-\rho^2)^2}x_1x_2\right)\\
&\quad= \exp\left(-\frac{x_1+x_2}{\beta(1-\rho^2)}\right)\frac{x_1^{\alpha-1}x_2^{\alpha-1}}{\Gamma^2(\alpha)\{\beta^2(1-\rho^2)\}^\alpha}\,{}_0F_1\left(;\alpha;\frac{\rho^2}{\beta^2(1-\rho^2)^2}x_1x_2\right)\\
&\quad= \exp\left(-\frac{x_1+x_2}{\beta(1-\rho^2)}\right)\frac{x_1^{\alpha-1}x_2^{\alpha-1}}{\Gamma^2(\alpha)\{\beta^2(1-\rho^2)\}^\alpha}\sum_{k=0}^{\infty}\frac{\Gamma(\alpha)}{k!\Gamma(\alpha+k)}\left(\frac{\rho^2}{\beta^2(1-\rho^2)^2}x_1x_2\right)^k\\
&\quad= \sum_{k=0}^{\infty}\frac{\Gamma(\alpha)\rho^{2k}(1-\rho^2)^\alpha}{k!\Gamma(\alpha+k)\Gamma^2(\alpha)\{\beta^2(1-\rho^2)^2\}^{\alpha+k}}x_1^{\alpha+k-1}x_2^{\alpha+k-1}\exp\left(-\frac{x_1+x_2}{\beta(1-\rho^2)}\right)\\
&\quad= \sum_{k=0}^{\infty}\frac{\Gamma(\alpha+k)}{k!\Gamma(\alpha)}(\rho^2)^k(1-\rho^2)^\alpha\left[\frac{x_1^{\alpha+k-1}\exp\left(\frac{-x_1}{\beta(1-\rho^2)}\right)}{\Gamma(\alpha+k)[\beta(1-\rho^2)]^{\alpha+k}}\right]\times\left[\frac{x_2^{\alpha+k-1}\exp\left(\frac{-x_2}{\beta(1-\rho^2)}\right)}{\Gamma(\alpha+k)[\beta(1-\rho^2)]^{\alpha+k}}\right].
\end{aligned}
$$

□

To generate (x_1,x_2) from $f_{(X_1,X_2)}(\cdot,\cdot|\alpha,\beta,\rho)$, we use the conditional distribution method (see Devroye, 1986), that is, generate x_1 with density $f_{X_1}(\cdot|\alpha,\beta)$ and x_2 with $f_{X_2|X_1}(\cdot|x_1,\alpha,\beta,\rho)$. To generate from $f_{X_2|X_1}(\cdot|x_1,\alpha,\beta,\rho)$, we use the mixture method.

Corollary 3.1. *To generate (x_1,x_2) from $f_{(X_1,X_2)}(\cdot,\cdot|\alpha,\beta,\rho)$, we can use one of the following algorithms:*

a) 1- *Generate* x_1 *from the Gamma*(α,β)

2- *Generate* k *from Poisson*$(\rho^2 x_1/\{\beta(1-\rho^2)\})$

3- *Generate* x_2 *from Gamma*$(\alpha+k,\beta(1-\rho^2))$

4- *Return* (x_1,x_2).

b) 1- *Generate* k *from Negative Binomial* $(\alpha;\rho^2)$

2- *Generate* x_1 *and* x_2 *independenly from Gamma*$(\alpha+k;\beta(1-\rho^2))$

3- *Return* (x_1,x_2).

3.4. Parameter estimation

Let $\alpha = 1/\lambda$ and $\beta = \sqrt{\lambda\theta}$. We want to estimate λ and θ. The characteristic function of the random variable Y with the new parametrization is given by

$$\phi(t|\lambda, \theta, \rho) = (1 + \lambda\theta(1-\rho^2)t^2)^{-1/\lambda},$$

its variance is $2\theta(1-\rho^2)$ and its kurtosis is $3(1+\lambda)$. There is a problem of identifiability. For example, $\phi(t|\lambda, \theta = 2, \rho = \sqrt{2}/2) = \phi(t|\lambda, \theta = 3/2, \rho = \sqrt{3}/3)$. We assume in the following that ρ is known. Estimators from the method of moments (MOM) are given by

$$\widehat{\lambda} = \frac{n\sum_{i=1}^{n} y_i^4}{3\left(\sum_{i=1}^{n} y_i^2\right)^2} \quad \text{and} \quad \widehat{\theta} = \frac{\sum_{i=1}^{n} y_i^2}{2n(1-\rho^2)}.$$

In this section, we explain how to compute estimates of λ and θ using the continuous empirical characteristic function (ECF) estimation method. We also briefly review the discrete approach.

3.4.1. Continuous ECF method

The empirical characteristic function (ECF) estimation method with a normal weight function consists in finding the arguments $(\widehat{\lambda}, \widehat{\theta})$ minimizing the function

$$\begin{aligned} g(\lambda, \theta, \rho|\mathbf{y}) &= \int_{\mathbb{R}} |\psi_n(t|\mathbf{y}) - \varphi_Y(t|\lambda, \theta, \rho)|^2 \sqrt{\lambda\theta}\phi_{\sigma^2}\left(\sqrt{\lambda\theta}t\right) \mathrm{d}t \\ &= \int_{\mathbb{R}} \left| \psi_n\left(\frac{t}{\sqrt{\lambda\theta}}\middle|\mathbf{y}\right) - \varphi_Y\left(\frac{t}{\sqrt{\lambda\theta}}\middle|\lambda, \theta, \rho\right) \right|^2 \phi_{\sigma^2}(t)\, \mathrm{d}t, \end{aligned}$$

where ϕ_{σ^2} is the density function of the normal random variable with mean 0 and variance σ^2, and $|x|^2 = x \cdot \overline{x}$, where $\overline{x}$ is the complex conjugate of x. We have

$$\begin{aligned} &g(\lambda, \theta, \rho|\mathbf{y}) \\ &= \int_{\mathbb{R}} \left(\frac{1}{n}\sum_{j=1}^{n} e^{\frac{ity_j}{\sqrt{\lambda\theta}}} - \{1 + (1-\rho^2)t^2\}^{-1/\lambda} \right) \left(\frac{1}{n}\sum_{j=1}^{n} e^{\frac{-ity_j}{\sqrt{\lambda\theta}}} - \{1 + (1-\rho^2)t^2\}^{-1/\lambda} \right) \phi_{\sigma^2}(t)\mathrm{d}t \\ &= \int_{\mathbb{R}} \left(\frac{1}{n^2}\sum_{k=1}^{n}\sum_{j=1}^{n} e^{\frac{it(y_k - y_j)}{\sqrt{\lambda\theta}}} - \frac{1}{n}\sum_{k=1}^{n} \left(e^{\frac{ity_k}{\sqrt{\lambda\theta}}} + e^{\frac{-ity_k}{\sqrt{\lambda\theta}}} \right) \{1 + (1-\rho^2)t^2\}^{-1/\lambda} + \{1 + (1-\rho^2)t^2\}^{-2/\lambda} \right) \phi_{\sigma^2}(t)\mathrm{d}t. \end{aligned}$$

Since

$$\int_{\mathbb{R}} \sin(ta)\phi_{\sigma^2}(t)\mathrm{d}t = 0, \text{ for all } a \text{ in } \mathbb{R},$$

we have

$$\begin{aligned}\int_{\mathbb{R}} \cos(ta)\phi_{\sigma^2}(t)\mathrm{d}t &= \int_{\mathbb{R}} \left(\cos(ta) + i\sin(ta)\right)\phi_{\sigma^2}(t)\mathrm{d}t \\ &= \int_{\mathbb{R}} \exp(ita)\phi_{\sigma^2}(t)\mathrm{d}t \\ &= \exp(-\sigma^2 a^2/2), \text{ for all } a \text{ in } \mathbb{R}.\end{aligned}$$

Then the function g becomes,

$$\begin{aligned} g(\lambda,\theta,\rho|\mathbf{y}) = \frac{1}{n^2}\sum_{k=1}^{n}\sum_{j=1}^{n}\exp\left(\frac{-\sigma^2(y_k-y_j)^2}{2\lambda\theta}\right) + \int_{\mathbb{R}}\{1+(1-\rho^2)t^2\}^{-2/\lambda}\phi_{\sigma^2}(t)\mathrm{d}t \\ -\frac{2}{n}\sum_{k=1}^{n}\int_{\mathbb{R}}\cos\left(\frac{ty_k}{\sqrt{\lambda\theta}}\right)\{1+(1-\rho^2)t^2\}^{-1/\lambda}\phi_{\sigma^2}(t)\mathrm{d}t. \qquad (3.6)\end{aligned}$$

The second and the last terms of the function g can be computed numerically by using a Monte-Carlo method. Let $t_1, t_2, \ldots, t_M$ be M observations from the normal distribution with mean 0 and variance σ^2. We have

$$\begin{aligned}\sum_{k=1}^{n}\int_{\mathbb{R}}\cos\left(\frac{ty_k}{\sqrt{\lambda\theta}}\right)\{1+(1-\rho^2)t^2\}^{-1/\lambda}\phi_{\sigma^2}(t)\mathrm{d}t &= \sum_{k=1}^{n} E\left[\cos\left(\frac{Ty_k}{\sqrt{\lambda\theta}}\right)\{1+(1-\rho^2)T^2\}^{-1/\lambda}\right], \\ &\qquad \text{with } T \sim N(0,\sigma^2) \\ &\approx \frac{1}{M}\sum_{k=1}^{n}\sum_{\ell=1}^{M}\cos\left(\frac{t_\ell y_k}{\sqrt{\lambda\theta}}\right)\{1+(1-\rho^2)t_\ell^2\}^{-1/\lambda}.\end{aligned}$$

Moreover,

$$\begin{aligned}\int_{\mathbb{R}}\{1+(1-\rho^2)t^2\}^{-2/\lambda}\phi_{\sigma^2}(t)\mathrm{d}t &= E\left[\{1+(1-\rho^2)T^2\}^{-2/\lambda}\right] \\ &\approx \frac{1}{M}\sum_{\ell=1}^{M}\{1+(1-\rho^2)t_\ell^2\}^{-2/\lambda}.\end{aligned}$$

Thus, we will minimize the following function

$$g_1(\lambda,\theta,\rho|\mathbf{y},\mathbf{t}) = \frac{1}{n^2}\sum_{k=1}^{n}\sum_{j=1}^{n}\exp\left(\frac{-\sigma^2(y_k-y_j)^2}{2\lambda\theta}\right) + \frac{1}{M}\sum_{\ell=1}^{M}\{1+(1-\rho^2)t_\ell^2\}^{-2/\lambda}$$

$$- \frac{2}{nM} \sum_{k=1}^{n} \sum_{\ell=1}^{M} \cos\left(\frac{t_\ell y_k}{\sqrt{\lambda\theta}}\right) \{1 + (1-\rho^2)t_\ell^2\}^{-1/\lambda},$$

with M large and with $t_1, \ldots, t_M$ being i.i.d. observations from $N(0, \sigma^2)$, and $\mathbf{t} = (t_1, t_2, \ldots, t_M)$. We can use the control variates method in order to reduce the variance of our estimator. Let us notice that,

$$\exp\{(1-\rho^2)t^2\} = 1 + (1-\rho^2)t^2 + o(t^2), \quad \text{as } t \longrightarrow 0.$$

Thus,

$$\exp\{-(1-\rho^2)t^2/\lambda\} \approxeq \{1 + (1-\rho^2)t^2\}^{-1/\lambda}, \quad \text{as } t \longrightarrow 0.$$

For $T \sim N(0, \sigma^2)$, we have

$$\begin{aligned}
&E\left[\cos\left(\frac{Ty_k}{\sqrt{\lambda\theta}}\right)\exp(-(1-\rho^2)T^2/\lambda)\right] \\
&= \int_{\mathbb{R}} \cos\left(\frac{ty_k}{\sqrt{\lambda\theta}}\right)\exp(-(1-\rho^2)t^2/\lambda)\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2\sigma^2}t^2\right)\mathrm{d}t \\
&= \int_{\mathbb{R}} \cos\left(\frac{ty_k}{\sqrt{\lambda\theta}}\right)\frac{1}{\sqrt{1+2(1-\rho^2)\sigma^2/\lambda}}\frac{1}{\sqrt{2\pi\frac{\sigma^2}{1+2(1-\rho^2)\sigma^2/\lambda}}}\exp\left(-\frac{1}{2}\frac{1}{\sigma^2/\{1+2(1-\rho^2)\sigma^2/\lambda\}}t^2\right)\mathrm{d}t \\
&= \frac{1}{\sqrt{1+2(1-\rho^2)\sigma^2/\lambda}}\int_{\mathbb{R}} \exp\left(\frac{ity_k}{\sqrt{\lambda\theta}}\right)\frac{1}{\sqrt{2\pi\frac{\sigma^2}{1+2(1-\rho^2)\sigma^2/\lambda}}}\exp\left(-\frac{1}{2}\frac{1}{\sigma^2/\{1+2(1-\rho^2)\sigma^2/\lambda\}}t^2\right)\mathrm{d}t \\
&= \frac{1}{\sqrt{1+2(1-\rho^2)\sigma^2/\lambda}}\exp\left(-\frac{\sigma^2 y_k^2}{2\{1+2(1-\rho^2)\sigma^2/\lambda\}\lambda\theta}\right)
\end{aligned}$$

and

$$\begin{aligned}
&E\left[\exp\{-2(1-\rho^2)T^2/\lambda\}\right] \\
&= \int_{\mathbb{R}} \exp\{-2(1-\rho^2)t^2/\lambda\}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2\sigma^2}t^2\right)\mathrm{d}t \\
&= \frac{1}{\sqrt{1+4(1-\rho^2)\sigma^2/\lambda}}\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\frac{\sigma^2}{1+4(1-\rho^2)\sigma^2/\lambda}}}\exp\left(-\frac{1}{2}\frac{1}{\sigma^2/\{1+4(1-\rho^2)\sigma^2/\lambda\}}t^2\right)\mathrm{d}t \\
&= \frac{1}{\sqrt{1+4(1-\rho^2)\sigma^2/\lambda}}.
\end{aligned}$$

The function to minimize becomes

$$
\begin{aligned}
&g_2(\lambda,\theta,\rho|\mathbf{y},\mathbf{t})\\
&= \frac{1}{M}\left\{\sum_{\ell=1}^{M}\{1+(1-\rho^2)t_\ell^2\}^{-2/\lambda} - C_2(\lambda,\rho)\sum_{\ell=1}^{M}\left[\exp\{-2(1-\rho^2)t_\ell^2/\lambda\} - \frac{1}{\sqrt{1+4(1-\rho^2)\sigma^2/\lambda}}\right]\right\}\\
&-\frac{2}{nM}\left\{\sum_{k=1}^{n}\sum_{\ell=1}^{M}\cos\left(\frac{t_\ell y_k}{\sqrt{\lambda\theta}}\right)\{1+(1-\rho^2)t_\ell^2\}^{-1/\lambda} - C_1(\lambda,\theta,\rho)\sum_{k=1}^{n}\sum_{\ell=1}^{M}\left[\cos\left(\frac{t_\ell y_k}{\sqrt{\lambda\theta}}\right)\exp\{-(1-\rho^2)t_\ell^2/\lambda\}\right.\right.\\
&\left.\left.-\frac{1}{\sqrt{1+2(1-\rho^2)\sigma^2/\lambda}}\exp\left(-\frac{\sigma^2 y_k^2}{2\{1+2(1-\rho^2)\sigma^2/\lambda\}\lambda\theta}\right)\right]\right\} + \frac{1}{n^2}\sum_{k=1}^{n}\sum_{j=1}^{n}\exp\left(\frac{-\sigma^2(y_k-y_j)^2}{2\lambda\theta}\right),
\end{aligned}
$$

where the functions C_1 and C_2 are defined as follows

$$
\begin{aligned}
&C_1(\lambda,\theta,\rho|\mathbf{y},\mathbf{t})\\
&= \frac{Cov\left(\sum_{k=1}^{n}\sum_{\ell=1}^{M}\cos\left(\frac{T_\ell y_k}{\sqrt{\lambda\theta}}\right)\{1+(1-\rho^2)T_\ell^2\}^{-1/\lambda}, \sum_{k=1}^{n}\sum_{\ell=1}^{M}\cos\left(\frac{T_\ell y_k}{\sqrt{\lambda\theta}}\right)\exp\{-(1-\rho^2)T_\ell^2/\lambda\}\right)}{Var\left(\sum_{k=1}^{n}\sum_{\ell=1}^{M}\cos\left(\frac{T_\ell y_k}{\sqrt{\lambda\theta}}\right)\exp\{-(1-\rho^2)T_\ell^2/\lambda\}\right)}\\
&= \frac{Cov\left(\sum_{k=1}^{n}\cos\left(\frac{T_1 y_k}{\sqrt{\lambda\theta}}\right)\{1+(1-\rho^2)T_1^2\}^{-1/\lambda}, \sum_{k=1}^{n}\cos\left(\frac{T_1 y_k}{\sqrt{\lambda\theta}}\right)\exp\{-(1-\rho^2)T_1^2/\lambda\}\right)}{Var\left(\sum_{k=1}^{n}\cos\left(\frac{T_1 y_k}{\sqrt{\lambda\theta}}\right)\exp\{-(1-\rho^2)T_1^2/\lambda\}\right)}
\end{aligned}
$$

and

$$
\begin{aligned}
C_2(\lambda,\rho|\mathbf{t}) &= \frac{Cov\left(\sum_{\ell=1}^{M}\{1+(1-\rho^2)T_\ell^2\}^{-2/\lambda}; \sum_{\ell=1}^{M}\exp\{-2(1-\rho^2)T_\ell^2/\lambda\}\right)}{Var\left(\sum_{\ell=1}^{M}\exp\{-2(1-\rho^2)T_\ell^2/\lambda\}\right)}\\
&= \frac{Cov\left(\{1+(1-\rho^2)T_1^2\}^{-2/\lambda}; \exp\{-2(1-\rho^2)T_1^2/\lambda\}\right)}{Var(\exp\{-2(1-\rho^2)T_1^2/\lambda\})}.
\end{aligned}
$$

Let

$$
\begin{aligned}
A_\ell &= \sum_{k=1}^{n}\cos\left(\frac{t_\ell y_k}{\sqrt{\lambda\theta}}\right)\{1+(1-\rho^2)t_\ell^2\}^{-1/\lambda},\\
B_\ell &= \sum_{k=1}^{n}\cos\left(\frac{t_\ell y_k}{\sqrt{\lambda\theta}}\right)\exp\{-(1-\rho^2)t_\ell^2/\lambda\},\\
D_\ell &= \{1+(1-\rho^2)t_\ell^2\}^{-2/\lambda}
\end{aligned}
$$

and

$$E_\ell = \exp\{-2(1-\rho^2)t_1^2/\lambda\},$$

where t_ℓ, $\ell = 1, \ldots, M$, are i.i.d. observations from $N(0, \sigma^2)$, and $\mathbf{t} = (t_1, t_2, \ldots, t_M)$. The function C_1 and C_2 will be estimated by

$$\widehat{C_1}(\lambda, \theta, \rho|\mathbf{y}, \mathbf{t}) = \frac{\sum_{\ell=1}^{M}(A_\ell - \overline{A})(B_\ell - \overline{B})}{\sum_{\ell=1}^{M}(B_\ell - \overline{B})^2},$$

and

$$\widehat{C_2}(\lambda, \rho|\mathbf{t}) = \frac{\sum_{\ell=1}^{M}(D_\ell - \overline{D})(E_\ell - \overline{E})}{\sum_{\ell=1}^{M}(E_\ell - \overline{E})^2},$$

where

$$\overline{A} = \frac{1}{M}\sum_{\ell=1}^{M} A_\ell, \quad \overline{B} = \frac{1}{M}\sum_{\ell=1}^{M} B_\ell, \quad \overline{D} = \frac{1}{M}\sum_{\ell=1}^{M} D_\ell \quad \text{and} \quad \overline{E} = \frac{1}{M}\sum_{\ell=1}^{M} E_\ell.$$

Lemma 3.3. *The last two terms of the function g in Equation (3.6), can be expressed analytically by the following formula*

$$\frac{2}{n}\sum_{k=1}^{n}\int_{\mathbb{R}} \cos\left(\frac{ty_k}{\sqrt{\lambda\theta}}\right)\{1 + (1-\rho^2)t^2\}^{-1/\lambda}\phi_{\sigma^2}(t)\mathrm{d}t$$
$$= \frac{2}{n}\sum_{k=1}^{n}\sum_{i_1=0}^{\infty}\sum_{i_2=0}^{\infty}\sum_{i_3=0}^{\infty} \frac{a_{i_1,i_2,i_3}(\lambda,\theta,\rho|y_k)e^{-\frac{\sigma^2 y_k^2}{2\lambda\theta}}}{\sigma^4\Gamma^2(1/\lambda)(1-\rho^2)^{1/\lambda}}\left(\frac{\sqrt{2}}{\sigma}\right)^{i_1+i_2+2(i_3+1/\lambda-2)}\Gamma\left(\frac{i_1+i_3+1/\lambda}{2}\right)\Gamma\left(\frac{i_2+i_3+1/\lambda}{2}\right)$$

and

$$\int_{\mathbb{R}}\{1 + (1-\rho^2)t^2\}^{-1/\lambda}\phi_{\sigma^2}(t)\mathrm{d}t$$
$$= \sum_{i_1=0}^{\infty}\sum_{i_2=0}^{\infty}\sum_{i_3=0}^{\infty} \frac{b_{i_1,i_2,i_3}(\rho)}{\sigma^4\Gamma^2(1/\lambda)(1-\rho^2)^{1/\lambda}}\left(\frac{\sqrt{2}}{\sigma}\right)^{i_1+i_2+2(i_3+1/\lambda-2)}\Gamma\left(\frac{i_1+i_3+1/\lambda}{2}\right)\Gamma\left(\frac{i_2+i_3+1/\lambda}{2}\right)$$

where,

$$a_{i_1,i_2,i_3}(\lambda,\theta,\rho|y_k) = \frac{\sigma^{2i_3}}{i_1!i_2!i_3!}\left(\frac{1}{\sqrt{1-\rho^2}} + \frac{\sigma^2 y_k}{\sqrt{\lambda\theta}}\right)^{i_1}\left(\frac{1}{\sqrt{1-\rho^2}} - \frac{\sigma^2 y_k}{\sqrt{\lambda\theta}}\right)^{i_2}(-1)^{i_1+i_2}$$

and

$$b_{i_1,i_2,i_3}(\rho) = \frac{\sigma^{2i_3}}{i_1!i_2!i_3!}\left(\frac{-1}{\sqrt{1-\rho^2}}\right)^{i_1+i_2}.$$

PROOF. 1. Let

$$I_1 = \int_{\mathbb{R}} \cos\left(\frac{ty_k}{\sqrt{\lambda\theta}}\right)\{1 + (1-\rho^2)t^2\}^{-1/\lambda}\phi_{\sigma^2}(t)\mathrm{d}t.$$

We have

$$\{1+(1-\rho^2)t^2\}^{-1/\lambda} = (1-i\sqrt{1-\rho^2}t^2)^{-1/\lambda}(1+i\sqrt{1-\rho^2}t^2)^{-1/\lambda}$$
$$= \left(\int_0^\infty e^{itw_1}\frac{w_1^{1/\lambda-1}e^{-w_1/\sqrt{1-\rho^2}}}{\Gamma(1/\lambda)(\sqrt{1-\rho^2})^{1/\lambda}}\mathrm{d}w_1\right)\left(\int_0^\infty e^{-itw_2}\frac{w_2^{1/\lambda-1}e^{-w_2/\sqrt{1-\rho^2}}}{\Gamma(1/\lambda)(\sqrt{1-\rho^2})^{1/\lambda}}\mathrm{d}w_2\right).$$

Since $\int_{\mathbb{R}} \sin\left(\frac{ty_k}{\sqrt{\lambda\theta}}\right)\{1+(1-\rho^2)t^2\}^{-1/\lambda}\phi_{\sigma^2}(t)\mathrm{d}t = 0$,

$$\begin{aligned}
I_1 &= \int_{\mathbb{R}} e^{\left(\frac{ity_k}{\sqrt{\lambda\theta}}\right)}\{1+(1-\rho^2)t^2\}^{-1/\lambda}\phi_{\sigma^2}(t)\mathrm{d}t \\
&= \int_{\mathbb{R}}\int_0^\infty\int_0^\infty e^{\left(\frac{ity_k}{\sqrt{\lambda\theta}}\right)} e^{itw_1}\frac{w_1^{1/\lambda-1}e^{-w_1/\sqrt{1-\rho^2}}}{\Gamma(1/\lambda)(\sqrt{1-\rho^2})^{1/\lambda}} e^{-itw_2}\frac{w_2^{1/\lambda-1}e^{-w_2/\sqrt{1-\rho^2}}}{\Gamma(1/\lambda)(\sqrt{1-\rho^2})^{1/\lambda}}\phi_{\sigma^2}(t)\mathrm{d}w_1\mathrm{d}w_2\mathrm{d}t \\
&= \int_0^\infty\int_0^\infty \frac{w_1^{1/\lambda-1}w_2^{1/\lambda-1}e^{-(w_1+w_2)/\sqrt{1-\rho^2}}}{\Gamma^2(1/\lambda)(1-\rho^2)^{1/\lambda}}\int_{\mathbb{R}} e^{it(w_1-w_2+y_k/\sqrt{\lambda\theta})}\phi_{\sigma^2}(t)\mathrm{d}t\mathrm{d}w_1\mathrm{d}w_2 \\
&= \int_0^\infty\int_0^\infty \frac{w_1^{1/\lambda-1}w_2^{1/\lambda-1}e^{-(w_1+w_2)/\sqrt{1-\rho^2}}}{\Gamma^2(1/\lambda)(1-\rho^2)^{1/\lambda}} e^{-\left(\frac{y_k}{\sqrt{\lambda\theta}}+w_1-w_2\right)^2\frac{\sigma^2}{2}}\mathrm{d}w_1\mathrm{d}w_2 \\
&= \int_0^\infty\int_0^\infty \frac{w_1^{1/\lambda-1}w_2^{1/\lambda-1}e^{-(w_1+w_2)/\sqrt{1-\rho^2}}}{\Gamma^2(1/\lambda)(1-\rho^2)^{1/\lambda}} e^{-\left(\frac{y_k^2}{\lambda\theta}+w_1^2+w_2^2\right)\frac{\sigma^2}{2}} e^{-\sigma^2\left(\frac{w_1y_k}{\sqrt{\lambda\theta}}-\frac{w_2y_k}{\sqrt{\lambda\theta}}-w_1w_2\right)}\mathrm{d}w_1\mathrm{d}w_2.
\end{aligned}$$

Let $I_2 = e^{-(w_1+w_2)/\sqrt{1-\rho^2}}e^{-\sigma^2\left(\frac{w_1y_k}{\sqrt{\lambda\theta}}-\frac{w_2y_k}{\sqrt{\lambda\theta}}-w_1w_2\right)}$. We have

$$\begin{aligned}
I_2 &= e^{-w_1\left(\frac{1}{\sqrt{1-\rho^2}}+\frac{\sigma^2y_k}{\sqrt{\lambda\theta}}\right)}e^{-w_2\left(\frac{1}{\sqrt{1-\rho^2}}-\frac{\sigma^2y_k}{\sqrt{\lambda\theta}}\right)}e^{\sigma^2w_1w_2} \\
&= \left\{\sum_{i_1=0}^\infty\left(\frac{1}{\sqrt{1-\rho^2}}+\frac{\sigma^2y_k}{\sqrt{\lambda\theta}}\right)^{i_1}\frac{(-w_1)^{i_1}}{i_1!}\right\}\left\{\sum_{i_2=0}^\infty\left(\frac{1}{\sqrt{1-\rho^2}}-\frac{\sigma^2y_k}{\sqrt{\lambda\theta}}\right)^{i_2}\frac{(-w_2)^{i_2}}{i_2!}\right\}\left\{\sum_{i_3=0}^\infty\frac{(\sigma^2w_1w_2)^{i_3}}{i_3!}\right\} \\
&= \sum_{i_1=0}^\infty\sum_{i_2=0}^\infty\sum_{i_3=0}^\infty\frac{\sigma^{2i_3}}{i_1!i_2!i_3!}\left(\frac{1}{\sqrt{1-\rho^2}}+\frac{\sigma^2y_k}{\sqrt{\lambda\theta}}\right)^{i_1}\left(\frac{1}{\sqrt{1-\rho^2}}-\frac{\sigma^2y_k}{\sqrt{\lambda\theta}}\right)^{i_2}(-1)^{i_1+i_2}w_1^{i_1+i_3}w_2^{i_2+i_3} \\
&= \sum_{i_1=0}^\infty\sum_{i_2=0}^\infty\sum_{i_3=0}^\infty a_{i_1,i_2,i_3}(\lambda,\theta,\rho|y_k)w_1^{i_1+i_3}w_2^{i_2+i_3},
\end{aligned}$$

where

$$a_{i_1,i_2,i_3}(\lambda,\theta,\rho|y_k) = \frac{\sigma^{2i_3}}{i_1!i_2!i_3!}\left(\frac{1}{\sqrt{1-\rho^2}}+\frac{\sigma^2y_k}{\sqrt{\lambda\theta}}\right)^{i_1}\left(\frac{1}{\sqrt{1-\rho^2}}-\frac{\sigma^2y_k}{\sqrt{\lambda\theta}}\right)^{i_2}(-1)^{i_1+i_2}.$$

Thus,

$$I_1 = \int_0^{\infty}\int_0^{\infty} \frac{w_1^{1/\lambda-1} w_2^{1/\lambda-1}}{\Gamma^2(1/\lambda)(1-\rho^2)^{1/\lambda}} e^{-\left(\frac{y_k^2}{\lambda\theta}+w_1^2+w_2^2\right)\sigma^2/2} \sum_{i_1=0}^{\infty}\sum_{i_2=0}^{\infty}\sum_{i_3=0}^{\infty} a_{i_1,i_2,i_3}(\lambda,\theta,\rho|y_k) w_1^{i_1+i_3} w_2^{i_2+i_3} \mathrm{d}w_1 \mathrm{d}w_2$$

$$= \sum_{i_1=0}^{\infty}\sum_{i_2=0}^{\infty}\sum_{i_3=0}^{\infty} \frac{a_{i_1,i_2,i_3}(\lambda,\theta,\rho|y_k) e^{-\frac{\sigma^2 y_k^2}{2\lambda\theta}}}{\Gamma^2(1/\lambda)(1-\rho^2)^{1/\lambda}} \left(\int_0^{\infty} w_1^{i_1+i_3+1/\lambda-1} e^{-\sigma^2 w_1^2/2} \mathrm{d}w_1\right)\left(\int_0^{\infty} w_2^{i_2+i_3+1/\lambda-1} e^{-\sigma^2 w_2^2/2}\mathrm{d}w_2\right).$$

In using the change of variable $t = \sigma^2 w_1^2/2$, we have

$$\begin{aligned}\int_0^{\infty} w_1^{i_1+i_3+1/\lambda-1} e^{-\sigma^2 w_1^2/2}\mathrm{d}w_1 &= \int_0^{\infty} \left(\frac{\sqrt{2t}}{\sigma}\right)^{i_1+i_3+1/\lambda-1} e^{-t}\frac{1}{\sigma^2\sqrt{2t}}\mathrm{d}t \\ &= \frac{1}{\sigma^2}\left(\frac{\sqrt{2}}{\sigma}\right)^{i_1+i_3+1/\lambda-1} \int_0^{+\infty} t^{(i_1+i_3+1/\lambda)/2-1} e^{-t}\mathrm{d}t \\ &= \frac{1}{\sigma^2}\left(\frac{\sqrt{2}}{\sigma}\right)^{i_1+i_3+1/\lambda-2} \Gamma\left(\frac{i_1+i_3+1/\lambda}{2}\right).\end{aligned}$$

Similarly,

$$\int_0^{\infty} w_1^{i_2+i_3+1/\lambda-1} e^{-\sigma^2 w_1^2/2}\mathrm{d}w_1 = \frac{1}{\sigma^2}\left(\frac{\sqrt{2}}{\sigma}\right)^{i_2+i_3+1/\lambda-2} \Gamma\left(\frac{i_2+i_3+1/\lambda}{2}\right).$$

Finally,

$$I_1 = \sum_{i_1=0}^{\infty}\sum_{i_2=0}^{\infty}\sum_{i_3=0}^{\infty} \frac{a_{i_1,i_2,i_3}(\lambda,\theta,\rho|y_k) e^{-\frac{\sigma^2 y_k^2}{2\lambda\theta}}}{\sigma^4\Gamma^2(1/\lambda)(1-\rho^2)^{1/\lambda}} \left(\frac{\sqrt{2}}{\sigma}\right)^{i_1+i_2+2(i_3+1/\lambda-2)} \Gamma\left(\frac{i_1+i_3+1/\lambda}{2}\right)\Gamma\left(\frac{i_2+i_3+1/\lambda}{2}\right).$$

2. The expression of $\int_{\mathbb{R}}\{1+(1-\rho^2)t^2\}^{-1/\lambda}\phi_{\sigma^2}(t)\mathrm{d}t$ follows from the one of $\int_{\mathbb{R}}\cos\left(\frac{ty_k}{\sqrt{\lambda\theta}}\right)\{1+(1-\rho^2)t^2\}^{-1/\lambda}\phi_{\sigma^2}(t)\mathrm{d}t$ by taking $y_k = 0$. □

3.4.2. Discrete ECF method

The discrete ECF method was used by Augustyniak and Doray (2012). For more details and references, the reader can refer to their paper. In the following, we summarize the method. The method consists in choosing the arguments which minimize a quadratic distance between the theoretical characteristic function and the empirical characteristic function. We will use only the real part because the imaginary one does not depend on the parameters, since the theoretical characteristic function of the DGD family is real. Let $t_1, \cdots, t_k >$

0, $\mathbf{Z}_n = [\psi_n^{\text{Re}}(t_1), \cdots, \psi_n^{\text{Re}}(t_k)]'$ and $\mathbf{Z}(\lambda, \theta|\rho^2) = [\phi_Y(t_1|\lambda, \theta, \rho^2), \cdots, \phi_Y(t_k|\lambda, \theta, \rho^2)]'$, with $\psi_n^{\text{Re}}(t)$ being the real part of $\psi_n(t)$. The points are chosen positively because the functions $\psi_n^{\text{Re}}(t)$ and $\phi_Y(t|\lambda, \theta, \rho^2)$ are symmetrical. The estimator is obtained by minimizing the quadratic distance between $\mathbf{Z}_n$ and $\mathbf{Z}(\lambda, \theta|\rho^2)$ given by

$$d(\lambda, \theta|\rho^2) = [\mathbf{Z}_n - \mathbf{Z}(\lambda, \theta|\rho^2)]'Q(\lambda, \theta|\rho^2)[\mathbf{Z}_n - \mathbf{Z}(\lambda, \theta|\rho^2)],$$

where $Q(\lambda, \theta|\rho^2)$ is a positive definite matrix. With the choice $Q(\lambda, \theta|\rho^2) = I$, the identity matrix, we obtained the ordinary least square estimator (OLS). A good choice of this matrix is $Q(\lambda, \theta|\rho^2) = \Sigma^{-1}(\lambda, \theta|\rho^2)$ (Luong and Thompson, 1987) where $\Sigma(\lambda, \theta|\rho^2) = (\sigma_{ij})$ is the variance-covariance of $\mathbf{Y}_n(\lambda, \theta|\rho^2) = \sqrt{n}[\mathbf{Z}_n - \mathbf{Z}(\lambda, \theta|\rho^2)]$, with

$$\sigma_{ij} = \frac{1}{2}(\phi_Y(t_i + t_j|\lambda, \theta, \rho^2) + \phi_Y(t_i - t_j|\lambda, \theta, \rho^2)) - \phi_Y(t_i|\lambda, \theta, \rho^2)\phi_Y(t_j|\lambda, \theta, \rho^2).$$

This matrix depends on parameters which are unknown. The iterative procedure is used to estimate parameter in this part as suggested by Luong and Doray (2002). Firstly, obtain $(\widehat{\lambda}, \widehat{\theta})$ by using $Q(\lambda, \theta) = I$. Then, Estimate $(\widehat{\lambda^*}, \widehat{\theta^*})$ by choosing $Q(\lambda, \theta) = \widehat{\Sigma}^{-1}$ with $\widehat{\Sigma} = \Sigma(\widehat{\lambda}, \widehat{\theta})$. This procedure can be repeated with Σ estimated at each step. The estimator $(\widehat{\lambda^*}, \widehat{\theta^*})$ is defined as the convergent vector value of the procedure, and will be called weighted least square estimator (WLS).

3.5. Simulation studies

3.5.1. Best equivariant estimator of β

In Section 3.2, Formula (3.1) is used as a numerical approximation of the best equivariant estimator of β when α is known. We used $m = 1000$, the number of observations generated for approximations. For the study, we have generated respectively 100, 500 and 1000 samples of size 30, 50, 100, 500, 1000 and 2000 from Y, the difference of two independent gamma random variables with parameters $\alpha = 1$ and $\beta = 1$. Table 3. I, gives estimations of the mean and standard error of the best equivariant estimator in different cases.

Table 3. I. Best equivariant estimates of β based on K samples of size n ($\beta = 1$)

n	$K = 100$		$K = 500$		$K = 1000$	
	$E[\widehat{\beta}]$	$s.e.(\widehat{\beta})$	$E[\widehat{\beta}]$	$s.e.(\widehat{\beta})$	$E[\widehat{\beta}]$	$s.e.(\widehat{\beta})$
30	0.95765	0.17084	0.96854	0.17528	0.96413	0.17492
50	0.97840	0.14010	0.97940	0.14635	0.97752	0.14000
100	0.98051	0.08896	0.98595	0.10292	0.98666	0.09985
500	0.99937	0.03879	0.99992	0.04369	0.99870	0.04415
1000	1.00228	0.02734	1.00114	0.03158	0.99893	0.03161
2000	1.00014	0.02215	0.99959	0.02262	0.99975	0.02229

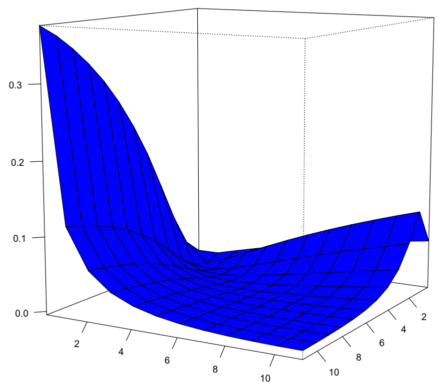
3.5.2. Continuous ECF method

Here we are going to show the results of the calculation with the estimator of moments (MOM), the standard ECF estimator (STAND) and the control variate ECF estimator (C.V.) derived in Section 3.4.1 for the parameters λ and θ. We generated 100 samples of size 100, 500 and 1000 from Y, the difference of two correlated gamma random variables with $\rho^2 = 0.5$. We performed estimations using different values of the parameter σ^2 (100, 1000, 2500 and 5000) introduced in the continuous ECF estimation method. The obtained results are presented in Table 3. II. In the calculations using the function optim in R, the methods C.V. and STAND happen to be numerically unstable. For example, in the case $n = 500$ and $\sigma^2 = 100$, the continuous ECF methods have failed, giving large standard errors and means too far from the true value of the estimators. A plot of the function $g(\cdot, \cdot, \rho|\mathbf{y})$, for many values of ρ and $\mathbf{y}$, indicates that there is a region $\{(a(s), b(s)), s \in \mathbb{R}\}$, such that $||(a(s), b(s)) - (\widehat{\lambda}, \widehat{\theta})||$ can be large while $\mid g(a(s), b(s), \rho \mid \mathbf{y}) - g(\widehat{\lambda}, \widehat{\theta} \mid \mathbf{y}) \mid$ is almost equal to zero, see Figure 3.1. With large values of σ^2 (greater or equal to 2500), we did not have this problem. However, we have tried the case $\sigma^2 = 10000$, and the program failed due to large values. The case where $\rho^2 = 0$ is quite similar to the results we have here. The estimator from the method of moments performed always better than the STAND and C.V. estimators.

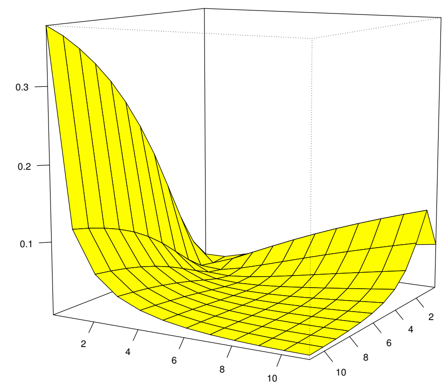
3.5.3. Discrete ECF method

We used 100 samples of size 100 (500 and 1000) from Y, with $\rho^2 = 0.5$. We computed estimator based on method of moments (MOM), ordinary least square (OLS) and weighted least square estimators (WSL) presented in Section 3.4.2. For the simulation, we used the

Figure 3.1. Function $g(\cdot, \cdot, \rho|\mathbf{y})$ for $n = 500$, $\sigma^2 = 100$ and $\rho^2 = 0.5$, when $\lambda = 1$ and $\theta = 1$.



(a) Standard ECF.



(b) Control variates ECF.

set of points

$$\left\{\frac{3}{10},\ 2\frac{3}{10},\ 3\frac{3}{10},\ \ldots, 9\frac{3}{10},\ 3\right\}$$

which was recommended by Augustyniak and Doray (2012). The table 3. III gives some simulation results. The WLS performed better than OLS, which did better than the MOM estimator as in the case of independence presented by Augustyniak and Doray (2012). Finally, we can see from the simulations that the discrete ECF method performs better than the method of moments, which performs better than the continuous ECF method with the weight function that we used. All our simulations were completed using R 3.3.1.

3.6. Conclusion

In this work we extended the family of distribution used by Augustyniak and Doray (2012) by considering the difference of two positively correlated or independent and identically distributed gamma random variables. Its distribution is symmetric, leptokurtic and complicated to handle. An easy way for estimation is to use continuous or discrete empirical characteristic function method. Based on simulation studies, we tried to compare these two approaches. The challenge with the continuous one remains the choice of the weight function. We used the density of a normal distribution, but the results were not satisfactory. The

Table 3. II. Estimates from continuous ECF method based on 100 samples of size n with $\rho^2 = 0.5$

n	σ^2	Method	$\lambda = 1$		$\theta = 1$	
			$E[\widehat{\lambda}]$	s.e.$(\widehat{\lambda})$	$E[\widehat{\theta}]$	s.e.$(\widehat{\theta})$
100	100	MOM	0.78691	0.89484	1.03016	0.20545
		STAND	1.10609	1.63572	1.27668	1.18207
		V.C.	223.51409	2171.06470	32.66594	313.59340
	1000	MOM	0.78691	0.89484	1.03016	0.20545
		STAND	0.98034	0.70847	1.22868	0.70887
		V.C.	0.92458	0.62587	1.26634	1.30552
	2500	MOM	0.82487	0.81938	1.02281	0.22536
		STAND	0.96256	0.67809	1.25273	0.69910
		V.C.	0.92721	0.60893	1.15078	0.54938
	5000	MOM	0.81781	0.59510	0.98784	0.22963
		STAND	0.87535	0.67015	1.02637	0.53021
		V.C.	0.88295	0.64136	1.03833	0.63750
500	100	MOM	0.94586	0.40320	0.99383	0.10364
		STAND	17.61137	106.26000	10.75970	78.72481
		V.C.	11.08325	100.55638	10.00512	89.50502
	1000	MOM	0.94586	0.40320	0.99383	0.10364
		STAND	0.84741	0.62867	1.03279	0.44525
		V.C.	155.76687	1548.40320	60.22907	592.08170
	2500	MOM	0.94586	0.40320	0.99383	0.10364
		STAND	0.90424	0.61111	1.07443	0.43649
		V.C.	0.93987	0.44821	1.01876	0.36426
	5000	MOM	0.94586	0.40320	0.99383	0.10364
		STAND	0.90401	0.65158	1.08193	0.48272
		V.C.	0.94924	0.49130	1.04944	0.39414
1000	100	MOM	0.89198	0.32340	0.98526	0.06588
		STAND	0.87504	0.59283	0.99184	0.35999
		V.C.	0.92480	0.42579	0.98658	0.24351
	1000	MOM	0.89198	0.32340	0.98526	0.06588
		STAND	0.88198	0.60815	1.04359	0.40495
		V.C.	0.90047	0.46193	0.98755	0.28422
	2500	MOM	0.98463	0.31319	0.99724	0.07190
		STAND	0.81565	0.67088	1.01753	0.42482
		V.C.	0.89976	0.50868	0.99723	0.32052
	5000	MOM	0.96102	0.31686	1.00829	0.07383
		STAND	0.85919	0.65671	1.05682	0.45742
		V.C.	0.89293	0.51685	1.03895	0.33196

discrete approach performed well all the time. One problem encountered here was identifiability of parameters. With this model it is not possible to estimate simultaneously the three parameters. We supposed that the correlation coefficient ρ^2 was known. We also computed

Table 3. III. Estimates from discrete ECF method based on 100 samples of size n with $\rho^2 = 0.5$

n	Method	$\lambda = 1$		$\theta = 1$	
		$E[\widehat{\lambda}]$	s.e.$(\widehat{\lambda})$	$E[\widehat{\theta}]$	s.e.$(\widehat{\theta})$
100	MOM	0.72671	0.61777	0.98848	0.20402
	OLS	1.04234	0.49908	1.04600	0.33359
	WLS	0.97661	0.43923	0.99841	0.21943
500	MOM	0.93122	0.36511	0.97668	0.09592
	OLS	0.98750	0.23374	0.98118	0.11962
	WLS	0.99036	0.19989	0.97941	0.09606
1000	MOM	0.91167	0.26190	0.99057	0.07245
	OLS	0.97756	0.15628	0.99154	0.08298
	WLS	0.97560	0.13390	0.99186	0.07094

the best equivariant estimator of the scale parameter β in the case of independence and known α. In this case the results were quite good.

Acknowledgment

The author Romain KADJE KENMOGNE wishes to thank Institut des sciences Mathématiques, Centre de Recherches Mathématiques, Faculté des Études Supérieures et Postdoctorales of the University of Montreal, and Fondation J. Armand Bombardier for their financial support.

The second author wishes to thank NSERC for their financial support.

References

Maciej Augustyniak. Une famille de distributions symétriques et leptocurtiques représentée par la différence de deux variables aléatoires gamma, 2008.

Maciej Augustyniak and Louis G. Doray. Inference for a leptokurtic symmetric family of distributions represented by the difference of two gamma variates. *J. Stat. Comput. Simul.*, 82(11): 1621–1634, 2012.

Marine Carrasco, Jean-Pierre Florens, et al. Efficient gmm estimation using the empirical characteristic function. Technical report, Institut d'Économie Industrielle (IDEI), Toulouse, 2002.

Li-Shya Chen, I-Shiang Tzeng, and Chien-Tai Lin. Bivariate generalized gamma distributions of Kibble's type. *Statistics*, 48(4):933–949, 2014.

Luc Devroye. *Nonuniform random variate generation*. Springer-Verlag, New York, 1986.

Andrey Feuerverger. An efficiency result for the empirical characteristic function in stationary time-series models. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, 18(2): 155–161, 1990.

Andrey Feuerverger and Philip McDunnough. On the efficiency of empirical characteristic function procedures. *Journal of the Royal Statistical Society. Series B (Methodological)*, 43(1):20–27, 1981a.

Andrey Feuerverger and Philip McDunnough. On some Fourier methods for inference. *Journal of the American Statistical Association*, 76(374):379–387, 1981b.

Bernhard Klar. A note on gamma difference distributions. *J. Stat. Comput. Simul.*, 85(18):3708–3715, 2015.

E. L. Lehmann and George Casella. *Theory of point estimation*. Springer Texts in Statistics. Springer-Verlag, New York, second edition, 1998.

A. Luong and M. E. Thompson. Minimum-distance methods based on quadratic distances for transforms. *Canadian Journal of Statistics*, 15(3):239–251, 1987.

Andrew Luong and Louis G. Doray. General quadratic distance methods for discrete distributions definable recursively. *Insurance: Mathematics and Economics*, 30(2):255 – 267, 2002.

Dilip B Madan and Eugene Seneta. The variance gamma (vg) model for share market returns. *Journal of Business*, pages 511–524, 1990.

Eugene Seneta. Fitting the variance-gamma model to financial data. *J. Appl. Probab.*, 41A:177–187, 2004. Stochastic methods and their applications.

V. S. Vaidyanathan and R. Vani Lakshmi. Parameter estimation in multivariate gamma distribution. *Stat. Optim. Inf. Comput.*, 3(2):147–159, 2015.

Jun Yu. Empirical characteristic function estimation and its applications. *Econometric Reviews*, 23 (2):93–123, 2004.

Chapitre 4

CONCLUSION

Dans cette thèse, nous avons essayé premièrement de proposer des méthodes d'estimation performantes et de les comparer avec les méthodes existantes. Plus précisément, nous nous sommes intéressés à la comparaison des estimateurs issus de la vraisemblance des rangs et de la pseudo-vraisemblance, au rapport de deux variables de lois normales, et à la différence de deux variables aléatoires positivement corrélées ou indépendantes et identiquement distribuées de loi gamma.

La principale contribution du premier article est le développement de la méthode d'estimation basée sur la vraisemblance des rangs. Nous avons montré dans le cadre de l'estimation des paramètres que l'estimateur issu de cette méthode est meilleur comparé à celui donné par la pseudo-vraisemblance. L'application avec la famille de copules de Farlie-Gumbel-Morgenstern donne des résultats assez similaires.

Après plusieurs décennies où des auteurs abordent le sujet du rapport de deux variables aléatoires de lois normales, il a été possible en utilisant une paramétrisation différente, d'arriver à de nouveaux résultats. L'expression de la densité de ce rapport comme un mélange de densité nous a permis de définir une nouvelle famille de densité qui généralise les densités de Student de degrés de liberté impairs. Des propriétés et des résultats de convergence intéressants ont été établis. Il est mentionné dans la littérature que la distribution du rapport de deux variables aléatoires de lois normales peut sous certaines conditions être proche d'une loi normale. Nous avons pu développer le résultat de convergence assurant la convergence en loi vers la loi normale lorsque l'un des paramètres tend vers l'infini. Nous sommes allés plus loin en trouvant une borne supérieure de cette approximation. Finalement, dans des problèmes d'estimation où l'estimateur est le rapport de deux variables de lois normales, l'une des difficultés vient du fait que les moments n'existent pas et par conséquent, il devient difficile de juger de la qualité de cet estimateur à l'aide de mesures telles que l'erreur des moindres carrés ou des moindres valeurs absolues. Nous contournons cette difficulté en définissant un critère de comparaison approprié. Dans le cadre de l'estimation du rapport de deux moyennes des variables aléatoires de lois normales, nous utilisons une méthode bayésienne en prenant comme estimateur la médiane a posteriori du paramètre estimé. L'estimateur proposé semble mieux performer que l'estimateur usuel, qui est le rapport des moyennes échantillonnales.

Dans le troisième article, nous avons pu étendre les travaux de Augustyniak et Doray (2012) en considérant la différence de deux variables aléatoires positivement corrélées et identiquement distribuées de loi gamma. Cette famille a des propriétés similaires à celles qu'ils ont obtenues dans le cas des variables indépendantes. La fonction de densité associée à cette famille est donnée et nous trouvons deux algorithmes simples permettant de générer des observations de deux variables aléatoires positivement corrélées et identiquement distribuées de loi gamma. L'estimation par la méthode des fonctions caractéristiques empiriques a été utilisée. La difficulté dans l'approche continue reste le choix de la fonction de poids. La fonction de poids optimale proposée dans la littérature n'est pas pratique dans des cas comme celui-ci, où l'expression de la fonction de densité est complexe. Nous avons obtenu des estimations peu satisfaisantes en utilisant la densité d'une loi normale comme fonction de poids. Finalement, l'approche discrète est appropriée dans ce problème, car elle a été facile à mettre en œuvre et a produit des estimations satisfaisantes. Une limite observée est l'identifiabilité des paramètres, ce qui rend le modèle invalide pour estimer simultanément le coefficient de corrélation et le paramètre d'échelle. Nous avons aussi trouvé le meilleur estimateur équivariant du paramètre d'échelle dans le cas de l'indépendance.

Annexe A

QUELQUES CALCULS DU PREMIER ARTICLE

Calculation of the functions $c_{\ell,n}$, $\ell \leq n$ in Section 1.3 can be done analytically by using combinatorics. Let $B_n = \{1, 2, \ldots, n+1\}$, $n \geq 1$ and consider the function h_n on B_n, $n \geq 1$ given by

$$h_n(i) = 2i - (n+1).$$

For $n \geq 1$ and $1 \leq i \leq n$,

$$c_{1,n}(i) = \sum_{j=1}^{n+1} (-1)^{\mathbb{1}(j>i)} = h_n(i).$$

For $n \geq 2$ and $1 \leq i_1 < i_2 \leq n$, we have

$$\begin{aligned}
c_{2,n}(i_1, i_2) &= \sum_{\{j_1,j_2 \in B_n :\, j_1 \neq j_2\}} (-1)^{\mathbb{1}(j_1>i_1)+\mathbb{1}(j_2>i_2)} \\
&\quad +2 \sum_{\{j_1,j_2 \in B_n :\, j_1 = j_2\}} (-1)^{\mathbb{1}(j_1>i_1)+\mathbb{1}(j_2>i_2)} \\
&= \sum_{\{j_1,j_2 \in B_n\}} (-1)^{\mathbb{1}(j_1>i_1)+\mathbb{1}(j_2>i_2)} + \sum_{\{j \in B_n\}} (-1)^{\mathbb{1}(j>i_1)+\mathbb{1}(j>i_2)} \\
&= \Big\{ h_n(i_1) h_n(i_2) - h_n(i_2 - i_1) \Big\}.
\end{aligned}$$

For $n \geq 3$ and $1 \leq i_1 < i_2 < i_3 \leq n$, we have

$$\begin{aligned}
c_{3,n}(i_1, i_2, i_3) &= \sum_{\{j_1,j_2,j_3 \in B_n :\, j_1 \neq j_2, j_2 \neq j_3, j_1 \neq j_3\}} (-1)^{\mathbb{1}(j_1>i_1)+\mathbb{1}(j_2>i_2)+\mathbb{1}(j_3>i_3)} \\
&\quad +2 \sum_{\{j_1,j_2,j_3 \in B_n :\, j_1 = j_2, j_2 \neq j_3\}} (-1)^{\mathbb{1}(j_1>i_1)+\mathbb{1}(j_2>i_2)+\mathbb{1}(j_3>i_3)} \\
&\quad +2 \sum_{\{j_1,j_2,j_3 \in B_n :\, j_1 = j_3, j_1 \neq j_2\}} (-1)^{\mathbb{1}(j_1>i_1)+\mathbb{1}(j_2>i_2)+\mathbb{1}(j_3>i_3)}
\end{aligned}$$

$$+2 \sum_{\{j_1,j_2,j_3 \in B_n:\, j_2=j_3, j_1\neq j_2\}} (-1)^{\mathbb{1}(j_1>i_1)+\mathbb{1}(j_2>i_2)+\mathbb{1}(j_3>i_3)}$$

$$+6 \sum_{\{j_1,j_2,j_3 \in B_n:\, j_1=j_2=j_3\}} (-1)^{\mathbb{1}(j_1>i_1)+\mathbb{1}(j_2>i_2)+\mathbb{1}(j_3>i_3)}$$

$$= \sum_{\{j_1,j_2,j_3 \in B_n\}} (-1)^{\mathbb{1}(j_1>i_1)+\mathbb{1}(j_2>i_2)+\mathbb{1}(j_3>i_3)}$$

$$+ \sum_{\{j_1,j_2,j_3 \in B_n:\, j_1=j_2\}} (-1)^{\mathbb{1}(j_1>i_1)+\mathbb{1}(j_2>i_2)+\mathbb{1}(j_3>i_3)}$$

$$+ \sum_{\{j_1,j_2,j_3 \in B_n:\, j_1=j_3\}} (-1)^{\mathbb{1}(j_1>i_1)+\mathbb{1}(j_2>i_2)+\mathbb{1}(j_3>i_3)}$$

$$+ \sum_{\{j_1,j_2,j_3 \in B_n:\, j_2=j_3\}} (-1)^{\mathbb{1}(j_1>i_1)+\mathbb{1}(j_2>i_2)+\mathbb{1}(j_3>i_3)}$$

$$+2 \sum_{\{j_1,j_2,j_3 \in B_n:\, j_1=j_2=j_3\}} (-1)^{\mathbb{1}(j_1>i_1)+\mathbb{1}(j_2>i_2)+\mathbb{1}(j_3>i_3)}$$

$$= h_n(i_1)h_n(i_2)h_n(i_3)$$

$$-h_n(i_1)h_n(i_3-i_2)$$

$$-h_n(i_2)h_n(i_3-i_1)$$

$$-h_n(i_3)h_n(i_2-i_1)$$

$$+2h_n(i_1-i_2+i_3)$$

$$= h_n(i_1)h_n(i_2)h_n(i_3)$$

$$-\{h_n(i_1)h_n(i_3-i_2)+h_n(i_2)h_n(i_3-i_1)+h_n(i_3)h_n(i_2-i_1)\}$$

$$+2h_n(i_1-i_2+i_3).$$

For $n \geq 4$ and $1 \leq i_1 < i_2 < i_3 < i_4 \leq n$ we have

$$c_{4,n}(i_1,i_2,i_3,i_4) = \sum_{\{j_1,j_2,j_3,j_4 \in B_n:\, j_1,\ldots,j_4 \text{ all distinct }\}} (-1)^{\mathbb{1}(j_1>i_1)+\mathbb{1}(j_2>i_2)+\mathbb{1}(j_3>i_3)+\mathbb{1}(j_4>i_4)}$$

$$+2 \sum_{\{j_1,j_2,j_3,j_4 \in B_n:\ \text{with a pair }\}} (-1)^{\mathbb{1}(j_1>i_1)+\mathbb{1}(j_2>i_2)+\mathbb{1}(j_3>i_3)+\mathbb{1}(j_4>i_4)}$$

$$+6 \sum_{\{j_1,j_2,j_3,j_4 \in B_n:\ \text{with three of a kind }\}} (-1)^{\mathbb{1}(j_1>i_1)+\mathbb{1}(j_2>i_2)+\mathbb{1}(j_3>i_3)+\mathbb{1}(j_4>i_4)}$$

$$+4 \sum_{\{j_1,j_2,j_3,j_4 \in B_n:\ \text{with two pairs }\}} (-1)^{\mathbb{1}(j_1>i_1)+\mathbb{1}(j_2>i_2)+\mathbb{1}(j_3>i_3)+\mathbb{1}(j_4>i_4)}$$

$$+24 \sum_{\{j_1,j_2,j_3,j_4 \in B_n:\ \text{with four of a kind}\}} (-1)^{\mathbb{1}(j_1>i_1)+\mathbb{1}(j_2>i_2)+\mathbb{1}(j_3>i_3)+\mathbb{1}(j_4>i_4)}$$

$$= A + B + 2C + D + 6E,$$

where

$$\begin{aligned} A &= \sum_{\{j_1,j_2,j_3,j_4 \in B_n\}} (-1)^{\mathbb{1}(j_1>i_1)+\mathbb{1}(j_2>i_2)+\mathbb{1}(j_3>i_3)+\mathbb{1}(j_4>i_4)} \\ &= h_n(i_1)h_n(i_2)h_n(i_3)h_n(i_4), \end{aligned}$$

$$\begin{aligned} B = & \sum_{\{j_1,j_2,j_3,j_4 \in B_n:\ j_1=j_2\}} (-1)^{\mathbb{1}(j_1>i_1)+\mathbb{1}(j_2>i_2)+\mathbb{1}(j_3>i_3)+\mathbb{1}(j_4>i_4)} \\ & + \sum_{\{j_1,j_2,j_3,j_4 \in B_n:\ j_1=j_3\}} (-1)^{\mathbb{1}(j_1>i_1)+\mathbb{1}(j_2>i_2)+\mathbb{1}(j_3>i_3)+\mathbb{1}(j_4>i_4)} \\ & + \sum_{\{j_1,j_2,j_3,j_4 \in B_n:\ j_1=j_4\}} (-1)^{\mathbb{1}(j_1>i_1)+\mathbb{1}(j_2>i_2)+\mathbb{1}(j_3>i_3)+\mathbb{1}(j_4>i_4)} \\ & + \sum_{\{j_1,j_2,j_3,j_4 \in B_n:\ j_2=j_3\}} (-1)^{\mathbb{1}(j_1>i_1)+\mathbb{1}(j_2>i_2)+\mathbb{1}(j_3>i_3)+\mathbb{1}(j_4>i_4)} \\ & + \sum_{\{j_1,j_2,j_3,j_4 \in B_n:\ j_2=j_4\}} (-1)^{\mathbb{1}(j_1>i_1)+\mathbb{1}(j_2>i_2)+\mathbb{1}(j_3>i_3)+\mathbb{1}(j_4>i_4)} \\ & + \sum_{\{j_1,j_2,j_3,j_4 \in B_n:\ j_3=j_4\}} (-1)^{\mathbb{1}(j_1>i_1)+\mathbb{1}(j_2>i_2)+\mathbb{1}(j_3>i_3)+\mathbb{1}(j_4>i_4)} \\ = & -[h_n(i_2-i_1)h_n(i_3)h_n(i_4) + h_n(i_3-i_1)h_n(i_2)h_n(i_4) + h_n(i_4-i_1)h_n(i_2)h_n(i_3) \\ & + h_n(i_3-i_2)h_n(i_1)h_n(i_4) + h_n(i_4-i_2)h_n(i_1)h_n(i_3) + h_n(i_4-i_3)h_n(i_1)h_n(i_2)], \end{aligned}$$

$$\begin{aligned} C = & \sum_{\{j_1,j_2,j_3,j_4 \in B_n:\ j_1=j_2=j_3\}} (-1)^{\mathbb{1}(j_1>i_1)+\mathbb{1}(j_2>i_2)+\mathbb{1}(j_3>i_3)+\mathbb{1}(j_4>i_4)} \\ & + \sum_{\{j_1,j_2,j_3,j_4 \in B_n:\ j_1=j_2=j_4\}} (-1)^{\mathbb{1}(j_1>i_1)+\mathbb{1}(j_2>i_2)+\mathbb{1}(j_3>i_3)+\mathbb{1}(j_4>i_4)} \\ & + \sum_{\{j_1,j_2,j_3,j_4 \in B_n:\ j_1=j_3=j_4\}} (-1)^{\mathbb{1}(j_1>i_1)+\mathbb{1}(j_2>i_2)+\mathbb{1}(j_3>i_3)+\mathbb{1}(j_4>i_4)} \\ & + \sum_{\{j_1,j_2,j_3,j_4 \in B_n:\ j_2=j_3=j_4\}} (-1)^{\mathbb{1}(j_1>i_1)+\mathbb{1}(j_2>i_2)+\mathbb{1}(j_3>i_3)+\mathbb{1}(j_4>i_4)} \\ = & \ h_n(i_3-i_2+i_1)h_n(i_4) + h_n(i_4-i_2+i_1)h_n(i_3) \end{aligned}$$

$$+h_n(i_4 - i_3 + i_1)h_n(i_2) + h_n(i_4 - i_3 + i_2)h_n(i_1),$$

$$\begin{aligned} D &= \sum_{\{j_1,j_2,j_3,j_4 \in B_n:\ j_1=j_2, j_3=j_4\}} (-1)^{\mathbb{1}(j_1>i_1)+\mathbb{1}(j_2>i_2)+\mathbb{1}(j_3>i_3)+\mathbb{1}(j_4>i_4)} \\ &\quad + \sum_{\{j_1,j_2,j_3,j_4 \in B_n:\ j_1=j_3, j_2=j_4\}} (-1)^{\mathbb{1}(j_1>i_1)+\mathbb{1}(j_2>i_2)+\mathbb{1}(j_3>i_3)+\mathbb{1}(j_4>i_4)} \\ &\quad + \sum_{\{j_1,j_2,j_3,j_4 \in B_n:\ j_1=j_4, j_2==j_3\}} (-1)^{\mathbb{1}(j_1>i_1)+\mathbb{1}(j_2>i_2)+\mathbb{1}(j_3>i_3)+\mathbb{1}(j_4>i_4)} \\ &= h_n(i_2 - i_1)h_n(i_4 - i_3) + h_n(i_3 - i_1)h_n(i_4 - i_2) + h_n(i_4 - i_1)h_n(i_3 - i_2), \end{aligned}$$

$$\begin{aligned} E &= \sum_{\{j_1,j_2,j_3,j_4 \in B_n:\ j_1=j_2=j_3=j_4\}} (-1)^{\mathbb{1}(j_1>i_1)+\mathbb{1}(j_2>i_2)+\mathbb{1}(j_3>i_3)+\mathbb{1}(j_4>i_4)} \\ &= -h_n(i_4 - i_3 + i_2 - i_1). \end{aligned}$$