



HAL
open science

Estimation of multivariate generalized gamma convolutions through Laguerre expansions

Oskar Laverny, Esterina Masiello, Véronique Maume-Deschamps, Didier Rullière

► **To cite this version:**

Oskar Laverny, Esterina Masiello, Véronique Maume-Deschamps, Didier Rullière. Estimation of multivariate generalized gamma convolutions through Laguerre expansions. 2021. hal-03160289v1

HAL Id: hal-03160289

<https://hal.science/hal-03160289v1>

Preprint submitted on 5 Mar 2021 (v1), last revised 9 Apr 2021 (v2)

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

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

Estimation of multivariate generalized gamma convolutions through Laguerre expansions.

Oskar Laverny^{*1,2,3}, Esterina Masiello^{1,3}, Véronique Maume-Deschamps^{1,3}, and Didier Rullière⁴

¹Université Claude Bernard Lyon 1, France

²SCOR SE, France

³Institut Camille Jordan UMR 5208, France

⁴Mines Saint-Étienne, UCA, CNRS, UMR 6158 LIMOS, Institut Henri Fayol, France

March 5, 2021

Abstract

The generalized gamma convolution class of distribution appeared in Thorin's work while looking for the infinite divisibility of the log-Normal and Pareto distributions. Although these distributions have been extensively studied in the univariate case, the multivariate case and the dependence structures that can arise from it have received little interest in the literature. Furthermore, only one projection procedure for the univariate case was recently constructed, and no estimation procedure are available. By expanding the densities of multivariate generalized gamma convolutions into a tensorized Laguerre basis, we bridge the gap and provide performant estimations procedures for both the univariate and multivariate cases. We provide some insights about performance of these procedures, and a convergent series for the density of multivariate gamma convolutions, which is shown to be more stable than Moschopoulos's and Mathai's univariate series. We furthermore discuss some examples.

1 Introduction

Olof Thorin introduced the Generalized Gamma Convolution (GGC) class of distributions in 1977 as a tool to study the infinite divisibility of the Pareto [50] and the log-Normal [51] distributions. The quite natural and pragmatic question of identifying if some distributions are infinitely divisible or not turned out to be theoretically fruitful, and notably gave rise to the concept of hyperbolic complete monotone functions. L. Bondesson, later, extended these concepts in a book [4] which is still today a good reference on the subject. An introduction to more recent literature about GGC can be found in the two surveys [23, 3], on the probabilistic side of the problem.

This class contains a lot of known distributions, including heavy-tailed distributions such a Pareto, log-Normal and α -stable distributions. This versatility makes the class appealing for the statistical practitioners working in various fields such as life and non-life insurance, reinsurance, anomaly or fraud detection, floods analysis and meteorology, etc. Other classes of general approximating distributions with high-order features can be considered. Mixtures of Erlangs [10, 28, 52], for example, have the property of being dense in L_2 [52] and allow for fast k -MLE estimation algorithms [41]. On the other hand, they lack crucial closure properties and interpretability that the generalized gamma convolutions have.

Surprisingly, on the statistical side, very little work can be found about estimation of generalized gamma convolutions. Only one projection procedure was published recently in [35, 17], that projects generalized gamma convolutions onto convolutions of a finite number of gammas. Although the resulting convergence result is stunning, this procedure cannot handle cases where the incoming density is not already inside the

*E-mail address: laverny@math.univ-lyon1.fr

class. This typically happens for any empirical distribution, due to sampling noise, even if the sampling density is in the class. Furthermore, we show that the theoretical background of the estimation procedure in [35, 17] is inherently univariate, and no direct extension to a multivariate case is possible.

The multivariate analogue is still an active research field, and was never considered as a statistical tool in the literature. In applications however statisticians might deal with marginals that are in this class. A way to handle dependence structure in this interpretable framework seems therefore appealing as it would allow studying functions of the random vector, like the sum of components, under dependence assumptions. Therefore, an estimation procedure for multivariate generalized gamma convolution that takes into account the dependence structure could be useful. The probabilistic background is already developed: based on an original idea from [7], Bondesson introduces a bivariate extension of the Thorin class, and gives some properties about these distributions. Later, [6] extend this concept to the multivariate case. More recently, [42, 43] proposes another equivalent framework allowing extension on any cone, with application to matrix gamma convolutions. On the other hand, no estimation procedure nor density exists. We discuss this multivariate class and show that the produced dependence structures that are highly flexible: non-exchangeability and tail dependence are possible.

Regarding tooling, the literature gives two series expansions for the density of univariate gamma convolutions: Mathai [30] proposes a density based on Kummer’s confluent geometric functions and Moschopoulos [37] refined it as a convergent gamma series, with an explicit truncation error. Both these densities have the problem on being based on the smallest scale parameter, and are not well conditioned when the smallest scale is too small. Unfortunately, this is typically the case when approximating a log-Normal or a Pareto distribution by a finite convolution of gammas. Therefore, no stable procedure for the density computation is available for the entirety of the parameter range. We are not aware of any density estimation available in the literature for the multivariate case.

We consider here the problem of estimation from samples of multivariate generalized gamma convolutions. By the introduction of a specific Laguerre basis, which was already considered for non-parametric deconvolution problems by [15, 29, 1], we manage to obtain a convergent series for the density. Classical deconvolution problems usually consider only one source of signal and one source of noise [31, 33, 54, 49]. Although we have here a finite number of signals to be estimated, the Laguerre expansion is still a useful tool. We show that this expansion is quite natural for our problem, as it includes and generalizes Moschopoulos’s univariate series. Using this basis, we bridge the gaps and provide a new stable algorithm for density computation in both the univariate and multivariate cases, as well as estimation procedures which handles both clean data (the density is given as a formal function) and dirty data (such as an empirical dataset in 64 bits precision).

After fixing notations, Section 2 covers some definitions, properties and algorithms to set up the stage. Section 3 considers the Laguerre expansions of densities in the multivariate Thorin class, provides some error bounds and develops an estimation procedure for these distributions. In Section 4, we investigate the numerical results of our estimation procedure on several examples. Section 5 concludes and gives leads for further work.

2 Background

As we deal with a lot of multivariate objects, we start by fixing some notations.

We use bold letters such as \mathbf{a} to designate generally indexable objects, e.g. vectors or matrices, and corresponding indexed and possibly unbolded letters, such as a_i , designate values in these objects. We use Cartesian indexing. For example, if we consider a row-major matrix \mathbf{r} , then \mathbf{r}_i denotes the i^{th} row of that matrix, $r_{i,j}$ the j^{th} value of that row and if $\mathbf{k} = (i, j)$, then $r_{\mathbf{k}} = r_{i,j}$.

We denote $|\mathbf{x}| = x_1 + \dots + x_d$ the sum of components of a vector \mathbf{x} . The product of factorials of component of an integer vector \mathbf{k} is denoted $\mathbf{k}! = k_1! \dots k_d!$, and we set $\mathbf{x}^{\mathbf{k}} = x_1^{k_1} \dots x_d^{k_d}$. We denote derivatives of a multivariate function f by $f^{(\mathbf{k})}(\mathbf{x}) = \frac{\partial}{\partial x_1^{k_1}} \dots \frac{\partial}{\partial x_d^{k_d}} f(\mathbf{x})$ and scalar products of vectors by $\langle \mathbf{x}, \mathbf{y} \rangle = x_1 y_1 + \dots + x_d y_d$. Finally, we denote $\binom{\mathbf{x}}{\mathbf{y}} = \prod_i \binom{x_i}{y_i}$ the products of binomial coefficients

Inequalities, sums and products between vectors of same shape are always intended componentwise, and standard broadcasting between objects of different shapes apply.

To characterize the distribution of random vectors, we might use several functions. For a (real-valued) random vector \mathbf{X} of dimension $d \geq 1$, the cumulative distribution function (cdf) of \mathbf{X} is $F(\mathbf{x}) = \mathbb{P}(\mathbf{X} \leq \mathbf{x})$. When they exist, the probability density function (pdf) of \mathbf{X} , $f(\mathbf{x}) = \frac{\partial}{\partial \mathbf{x}} F(\mathbf{x})$, its Laplace transform (lt) $L(\mathbf{t}) = \int e^{-\mathbf{t}\mathbf{x}} f(\mathbf{x}) \partial \mathbf{x}$, its moment generating function (mgf) $M(\mathbf{t}) = L(-\mathbf{t})$ or its cumulant generating function (cgf) $K(\mathbf{t}) = \ln(M(\mathbf{t}))$ can also be used to characterize the distribution. All these functions might be indexed by the random object to avoid ambiguity if needed.

We denote by $L_2(\mathbb{R}_+^d)$ the set of functions that are Lebesgue square-integrable onto \mathbb{R}_+^d , and by $L_2(\mathbb{R}_+, w)$ the set of square-integrable functions with respect to a weight function w .

The following subsections set the stage of our analysis by reviewing common tools from the literature.

2.1 Gamma convolutions classes

A random variable X is said to be Gamma distributed with shape $\alpha \in \mathbb{R}_+^*$, scale $s \in \mathbb{R}_+^*$ and rate $\frac{1}{s}$, which we denote $X \sim \mathcal{G}_{1,1}(\alpha, s)$, if it admits the cumulant generating function:

$$K(t) = \ln \mathbb{E}(e^{tX}) = -\alpha \ln(1 - st).$$

We denote by $\mathcal{G}_{1,1}$ this class of distributions. Recall that the cumulant generating function of the sum of independent random variables is the sum of their cumulant generating functions. Therefore, this expression shows that the gamma distributions are clearly infinite divisible with gamma distributed pieces (same scale and smaller shape). We now consider independent convolutions of these distributions:

Definition 2.1.1 (\mathcal{G}_1 and $\mathcal{G}_{1,n}$). *Let X be a random variable with support \mathbb{R}_+ .*

- (i) *X is a gamma convolution with shapes $\boldsymbol{\alpha} \in \mathbb{R}_+^n$ and scales $\mathbf{s} \in \mathbb{R}_+^n$, denoted $X \sim \mathcal{G}_{1,n}(\boldsymbol{\alpha}, \mathbf{s})$, if it has cgf:*

$$K(t) = -\sum_{i=1}^n \alpha_i \ln(1 - s_i t).$$

We denote this class of distribution $\mathcal{G}_{1,n}$.

- (ii) *X is an (untranslated) generalized gamma convolution with Thorin measure ν , denoted $X \sim \mathcal{G}_1(\nu)$, if its cgf writes:*

$$K(t) = -\int_0^\infty \ln(1 - st) \nu(\partial s),$$

where ν is such that:

$$\int_{[1,+\infty)} |\log(s)| \nu(\partial s) < \infty \text{ and } \int_{(0,1)} s \nu(\partial s) < \infty.$$

We denote by \mathcal{G}_1 this class of distributions.

The generalized gamma convolution class \mathcal{G}_1 , which is commonly called the one dimensional Thorin class, has received a lot of interest in the literature. These limits of convolutions are not dense in the set of positive continuous random variables, but they contain many interesting distributions, including the log-Normal and Pareto distributions as historically important cases [50, 51]. They have several interesting properties: noteworthy is the fact that the \mathcal{G}_1 class is closed with respect to independent sums or random variables, by construction, but also as a more recent result shows by independent products of random variables [3]. In fact, \mathcal{G}_1 can be defined as the smallest class that is closed through convolution and that contains gamma distributions. See [4] for an extensive study of these distributions¹.

¹Note that in [4], distributions were parametrized through rates $\beta = \frac{1}{s}$ instead. We chose here to parametrize by scales, as it simplifies some of our notations.

Remark that if the Thorin measure ν of $X \sim \mathcal{G}_1(\nu)$ has a finite discrete support of cardinal n , $X \sim \mathcal{G}_{1,n}(\boldsymbol{\alpha}, \mathbf{s})$ for $\boldsymbol{\alpha}, \mathbf{s}$ vectors such that $\nu = \sum_i \alpha_i \delta_{s_i}$ (where δ_x is the Dirac measure in x).

Mathai provides in [30] a series expression for the density of $\mathcal{G}_{1,n}$ random variables based on Kummer's confluent hypergeometric functions. Later, Moschopoulos [37] refines the result by providing the following gamma series:

Property 2.1.2 (Moschopoulos expansion). *For $X \sim \mathcal{G}_{1,n}(\boldsymbol{\alpha}, \mathbf{s})$, denoting w.l.o.g $s_1 = \min \mathbf{s}$, the density f of X is given by the following series:*

$$f(x) = C \sum_{k=0}^{\infty} \delta_k f_{\rho+k, \frac{1}{s_1}}(x)$$

where:

- $f_{a,b}$ is the density of a $\mathcal{G}_{1,1}(a, b)$ distribution,
- $\rho = |\boldsymbol{\alpha}|$ and $C = s_1^\rho \mathbf{s}^{-\boldsymbol{\alpha}}$,
- $\delta_0 = 0$ and for $k \geq 0$, $\delta_{k+1} = \frac{1}{k+1} \sum_{i=1}^{k+1} i \gamma_i \delta_{k+1-i}$ where $\gamma_{k+1} = \sum_{i=1}^n \alpha_i \frac{(1 - \frac{s_1}{s_i})^{k+1}}{(k+1)!}$

Furthermore, for $S = 1 - \frac{s_1}{\max \mathbf{s}}$, we have that $|\delta_{k+1}| \leq \frac{S^{k+1} \Gamma(\rho+k)}{\Gamma(\rho)(k+1)!}$, ensuring the uniform convergence of the series.

The dependence of this expansion onto the smallest scale parameter s_1 has a major drawback. When the smallest scale is close to zero, the corresponding truncated series is quite unstable. A simple show-case is the simulation of random numbers from a $\mathcal{G}_{1,2}((10, 10^{-3}), (1, 10^{-3}))$ distribution: the Moschopoulos implementation from the R package `coga` [22] gives a density that evaluates to 0 on all simulated random numbers, which is obviously wrong. Mathai's method, implemented in the same package, produces the same result. Sadly, as later implementation will show, parameters that correspond to approximations of useful distributions such as log-Normal, Weibull or Pareto usually have very small scales and trigger the same instability.

A likelihood approach to fit distributions in the Thorin class is therefore not practical, as there is no stable density, which might explain why there are currently no estimation procedure in the literature.

In [4, 6], Bondesson defined a class of multivariate convolution of gamma distributions based on the following idea from Cherian [7]. Suppose that $\mathbf{Z} = (Z_i)_{i \in \{0,1,2\}}$ are three independent gamma random variables with (respective) shapes α_i and scales s_i such that $s_0 = 1$ w.l.o.g. Then the cgf of the vector $(Z_0 + Z_1, Z_0 + Z_2)$ writes:

$$K(\mathbf{t}) = -\alpha_1 \ln(1 - s_1 t_1) - \alpha_2 \ln(1 - s_2 t_2) - \alpha_0 \ln(1 - s_1 t_1 + s_2 t_2).$$

This construction can be extended to what Bondesson called multivariate generalized gamma convolutions:

Definition 2.1.3 (\mathcal{G}_d and $\mathcal{G}_{d,n}$). *Let \mathbf{X} be a d -variate random vector with support \mathbb{R}_+^d .*

- (i) \mathbf{X} is a multivariate gamma convolution with shapes $\boldsymbol{\alpha} \in \mathbb{R}_+^n$ and scale matrix $\mathbf{s} = (s_{i,j}) \in \mathcal{M}(\mathbb{R}_+)_{n \times d}$, denoted by $\mathbf{X} \sim \mathcal{G}_{d,n}(\boldsymbol{\alpha}, \mathbf{s})$ if it has cumulant generating function:

$$K(\mathbf{t}) = \sum_{j=1}^n -\alpha_j \ln(1 - \langle \mathbf{s}_j, \mathbf{t} \rangle).$$

We denote by $\mathcal{G}_{d,n}$ this class of distributions.

- (ii) \mathbf{X} is an (untranslated) multivariate generalized gamma convolution with Thorin measure ν , denoted $\mathbf{X} \sim \mathcal{G}_d(\nu)$, if it is a weak limit of d -variates gamma convolutions. Its cgf writes:

$$K(\mathbf{t}) = - \int \ln(1 - \langle \mathbf{s}, \mathbf{t} \rangle) \nu(d\mathbf{s}),$$

for a positive measure ν on \mathbb{R}_+^d , the Thorin Measure, with suitable integration conditions (see [42]). We denote by \mathcal{G}_d this class of distributions.

Note that $\mathbf{X} \sim \mathcal{G}_{d,n}(\boldsymbol{\alpha}, \mathbf{s})$ if and only if there exists a vector \mathbf{Y} of n independent unit scale gamma random variable with shapes $\alpha_1, \dots, \alpha_n$ such that $\mathbf{X} = \mathbf{s}'\mathbf{Y}$. The i^{th} marginal X_i has distribution $\mathcal{G}_{1,n}(\boldsymbol{\alpha}, \mathbf{s}_i)$. We can interpret this as a decomposition of the random vector \mathbf{X} into an additive risk factor model with gamma distributed factors. Note that we allow scales to be zero, and some factors might therefore appear only in some marginals. As in the univariate case, we close the class by taking convolutional limits. For the analysis of these distributions, we refer to [4, 6], but also [42] which uses a slightly different but equivalent parametrization, allowing more generalization to other cones than \mathbb{R}_+^d .

We are here interested in the construction of distributions estimators in \mathcal{G}_d , with particular interest for estimators in $\mathcal{G}_{d,n}$. Indeed, $\mathcal{G}_{d,n}$ models have a structure that allows fast simulation, and that provides meaningful insight about the dependence structure, since a $\mathcal{G}_{d,n}$ distributions follows an additive risk factor model. On the other hand, distributions in \mathcal{G}_d with a non-atomic Thorin measure are hard to sample (see [5, 44] for potential solutions to this problem). They have no known closed-form density or distribution function, and even the cumulant generating function requires integration to be evaluated.

Before diving into statistical considerations and estimation of these distributions, we give here some properties about the dependence structure that can be achieved in the \mathcal{G}_d class. The following properties give more insights about the relation between the dependence structure of the random vector and the dependence structure of its Thorin measure:

Property 2.1.4 (The support of the Thorin measure). *Let $\mathbf{X} \sim \mathcal{G}_d(\nu)$, and let $\text{Sup}(\nu)$ be the support of ν . Then we have:*

(i) *The marginals of \mathbf{X} are mutually independent if and only if*

$$\text{Sup}(\nu) \subset S_{\perp} = \bigcup_{j=1}^d \left(\prod_{i=1}^{j-1} \{0\} \right) \times \mathbb{R}_+^* \times \left(\prod_{i=j+1}^d \{0\} \right).$$

In this case, $\nu(\mathbf{s}) = \sum_{j=1}^d \nu_j(s_j) \mathbb{1}_{s_k=0 \forall k \neq j}$ where ν_j denotes the Thorin measure of the j^{th} marginal, and the total masses of ν, ν_1, \dots, ν_d are such that $\nu(\mathbb{R}_+^d) = \nu(S_{\perp}) = \sum_{i=1}^d \nu_i(\mathbb{R}_+)$.

(ii) *The marginals of \mathbf{X} are comonotonous if and only if there exists a constant $\mathbf{c} \in \mathbb{R}_+^d, \|\mathbf{c}\| = 1$ such that*

$$\text{Sup}(\nu) \subset S_{\mathbf{c}} = \left\{ \mathbf{r} \in \mathbb{R}_+^d, \frac{\mathbf{r}}{\|\mathbf{r}\|} = \mathbf{c} \right\}.$$

In this case, the multivariate Thorin measure and its marginals have all the same total mass.

(iii) *The r.v. \mathbf{X} has a Lebesgue-square-integrable density if $D \geq d$, where D denotes the minimum integer such that there exists constants $\mathbf{c}_1, \dots, \mathbf{c}_D$ such that $\text{Sup}(\nu) \subset \bigcup_{i=1}^D S_{\mathbf{c}_i}$. On the other hand if $D < d, \mathbf{X}$ is a singular random vector with a D -dimensional support if $D < d$ (this is obviously the case ν is atomic with less than d atoms).*

Proof. For (i) and (ii), consider that $\mathbf{X} \sim \mathcal{G}_{d,n}(\boldsymbol{\alpha}, \mathbf{s})$. Then there exists a matrix \mathbf{Y} of gamma random variables $Y_{i,j} \sim \mathcal{G}_{1,1}(\alpha_j, s_{i,j})$ such that

$$X_i = Y_{i,1} + \dots + Y_{i,n},$$

where row vectors $\mathbf{Y}_{i,\cdot}$ are independent and column vectors $\mathbf{Y}_{\cdot,j}$ have two separate groups of marginals: a first comonotonous group (with gamma marginals) corresponding to scales $s_{i,j}$ that are strictly positive and a second group (with identically 0 marginals) corresponding to null scales. Note that one of these two groups might be empty. From this expansion, we can easily deduce the structure of maximal supports for ν that leads to independent or comonotonous random vectors \mathbf{X} . Consider non-atomic Thorin measures as limits of the previous case.

For (iii), the problem of regularity of \mathbf{X} and the existence of a density in L_2 , simply consider the rank of the (possibly non-square) matrix \mathbf{s} such that $\mathbf{X} = \mathbf{s}\mathbf{Z}$ where $Z_j \sim \mathcal{G}_{1,1}(\alpha_j, 1)$ are independent. If $D = d = n$, then \mathbf{X} is an invertible linear transformation of an independent, d -dimensional random vector, and therefore has a density in L_2 that we can express easily. If $D = d < n$, then \mathbf{X} is the convolution of the previous case with something else, and therefore also has a density in L_2 . Same goes for $D > d$. \square

Hence, the only comonotonous structure that is allowed inside the \mathcal{G}_d class is $X_i = c_i X_0$ for $X_0 \in \mathcal{G}_1$ and $\mathbf{c} > \mathbf{0}$. Between the independent and the comonotonous cases, there is a wide range of achievable dependence structures. Furthermore, since the class \mathcal{G}_d is closed w.r.t convex convolution, every (positive) value of dependence measures such as Kendall tau or Spearman rho are achievable. Some examples also exhibit highly asymmetrical dependence structure.

For the study of the relation between the dependence structure of \mathbf{X} and the dependence structure of ν , the Thorin measure proposed in Example 2.1.5 is of interest.

Example 2.1.5 (A curious distribution). *Let ν be a one-dimensional measure defined through $\nu(\partial x) = \mathbb{1}_{x \in [0,1]} dx$. Then $\mathcal{G}_1(\nu)$ has moment generating function:*

$$M(t) = e^{1 - \frac{(t-1)}{t} \ln(1-t)}$$

Moreover, if for all $n \in \mathbb{N}$, $G_{j,n} \sim \mathcal{G}_{1,1}\left(\frac{1}{n}, \frac{j}{n+1}\right)$ are independent, then $\lim_{n \rightarrow +\infty} \sum_{j=1}^n G_{j,n} \sim \mathcal{G}_1(\nu)$.

Proof. The second result comes from a discretization of the Thorin measure, from continuously uniform on $[0, 1]$ to uniform on $\left\{\frac{j}{n+1}, j \in \{1, \dots, n\}\right\}$. For the first result, note that ν satisfies the integration constraints given by Definition 2.1.1. Then, simply compute the cumulant generating function as:

$$\begin{aligned} K(-t) &= - \int_0^\infty \ln(1+rt) \mathbb{1}_{r \in [0,1]} \partial r = - \int_0^1 \ln(1+rt) \partial r = -\frac{1}{t} \int_1^{1+t} \ln(x) \partial x \\ &= - \frac{(1+t) \ln(1+t) - (1+t) - 1 \ln(1) + 1}{t} = 1 - \frac{1+t}{t} \ln(1+t). \end{aligned}$$

\square

This random variable has the nice property that its Thorin measure is uniform on $[0,1]$. It could be simulated more efficiently than by summing gammas through [5, 44] if we wanted. However, its use is limited by the following negative result:

Remark 2.1.6 (The no-bijection result). *Let $\mathbf{X} \sim \mathcal{G}_d(\nu)$ have marginal cdf F_1, \dots, F_d , and denote by G the cdf associated to the distribution of Example 2.1.5. Then, assuming that ν is absolutely continuous, the random vector*

$$\mathbf{Y} = (G^{-1}(F_i(X_i)))_{i \in 1, \dots, d}$$

has uniform marginal Thorin measures, but is not $\mathcal{G}_d(c_\nu)$ -distributed, where c_ν is the copula of ν .

There is in fact a bijection between the copula of the random vector and the copula of its Thorin measure, but conditionally on the marginals distributions. Therefore, an estimation scheme that separate the dependence structure from the marginals is not possible.

Before discussing the estimation of these distributions, the next subsection introduces some specific integrals, which correspond to derivatives of the moment generating function and the cumulant generating functions of a random vector.

2.2 Shifted moments and shifted cumulants

We introduce formally the shifted moments and the cumulants of a random vector, and present a known and useful result that maps moments to cumulants and vice-versa.

For \mathbf{X} a random vector of dimension d , we recall the following notations for, respectively, the moment generating function M and the cumulant generating function K of \mathbf{X} :

$$\forall \mathbf{t} \in \mathbb{C}^d, M(\mathbf{t}) = \mathbb{E} \left(e^{\langle \mathbf{t}, \mathbf{X} \rangle} \right) \text{ and } K(\mathbf{t}) = \ln M(\mathbf{t}),$$

We now introduce specific notations for the derivatives of these two functions.

Definition 2.2.1 (Shifted moments and cumulants). *For \mathbf{X} a random vector of dimension d , for $\mathbf{i} \in \mathbb{N}^d$, define the \mathbf{t} -shifted \mathbf{i}^{th} moments and cumulants as the \mathbf{i}^{th} derivatives of (respectively) M and K , taken in \mathbf{t} :*

$$\mu_{\mathbf{i}, \mathbf{t}} = M^{(\mathbf{i})}(\mathbf{t}) \text{ and } \kappa_{\mathbf{i}, \mathbf{t}} = K^{(\mathbf{i})}(\mathbf{t}).$$

They correspond to the moment and cumulants of an exponentially tilted version of the random vector \mathbf{X} , also called the Escher transform [16] of \mathbf{X} .

Although standard Monte-Carlo estimators for shifted moments are unbiased and easy to compute, shifted cumulants are a little harder to estimate from i.i.d samples of a random vector. We refer to [48, 47] for some unbiased estimators of multivariate cumulants, and [27] for an application to the estimation of stable laws. For switching from moments to cumulants, Property 2.2.2 gives a bijection between the two:

Property 2.2.2 (Bijection between moments and cumulants). *For a given shift $\mathbf{t} \in \mathbb{C}^d$ and $\mathbf{m} \in \mathbb{N}^d$, the sets $(\mu_{\mathbf{i}, \mathbf{t}})_{\mathbf{i} \leq \mathbf{m}}$ and $(\kappa_{\mathbf{i}, \mathbf{t}})_{\mathbf{i} \leq \mathbf{m}}$ can express one from one another through:*

$$\begin{aligned} \mu_{\mathbf{i}, \mathbf{t}} &= e^{\kappa_{\mathbf{0}, \mathbf{t}}} \sum_{\pi \in \Pi(\mathbf{i})} \prod_{B \in \pi} \kappa_{\mathbf{i}(B), \mathbf{t}} \\ \kappa_{\mathbf{i}, \mathbf{t}} &= \sum_{\pi \in \Pi(\mathbf{i})} (|\pi| - 1)! (-1)^{|\pi| - 1} (\mu_{\mathbf{0}, \mathbf{t}})^{-|\pi|} \prod_{B \in \pi} \mu_{\mathbf{i}(B), \mathbf{t}} \end{aligned}$$

where $\Pi(\mathbf{i}) = \mathcal{P}(B(\mathbf{i}))$ is the set of all partitions of the multi-set $B(\mathbf{i}) = \bigcup_{j=1}^d \bigcup_{p=1}^{i_j} \{j\}$, a partition π in $\Pi(\mathbf{i})$ is constituted of multi-sets B composed of integers, and $\mathbf{i}(B) = (\#\{\text{occurrences of } j \text{ in } B\})_{j \in \{1, \dots, d\}}$.

Proof. This is a direct application of the multivariate Faà Di Bruno's formula [9, 36, 39] to $M = \exp \circ K$ in one hand and $K = \ln \circ M$ on the other hand. \square

Remark that this property can be expressed in several other ways: most analysis in the literature uses multivariate Bell Polynomials [2, 11, 25, 32] for this task. For computational purposes, there also exists some practical functions that gives coefficients of these polynomials [53, 18], even for $d > 1$. However, in all generality, the number of coefficients to compute and store is exponential with \mathbf{m} making this method quite unpractical, as the Bell polynomials are combinatorial in nature. To switch from cumulants to moments, the good property of the exponential function of being its own derivative allows a recursive approach, which is described in full details in [34]. We adapted the notation and simplified the indexes conventions from [34], which allowed, after adding a slack variable to handle the dimensionality index, for an even faster implementation than the one from [34]. The expression in Algorithm 1 uses derivatives of the cgf and mgf as it is the case that we need, but the derivative of the exponential of any function could be computed through exactly

the same algorithm, knowing the derivatives of the exponent.

Algorithm 1: Recursive computation of μ from κ .

Input: $m \in \mathbb{N}^d$, Shifted Cumulants $(\kappa_{\mathbf{k},t})_{\mathbf{k} \leq m}$
Result: Shifted Moments $(\mu_{\mathbf{k},t})_{\mathbf{k} \leq m}$
Set $\mu_{\mathbf{0},t} = \exp(\kappa_{\mathbf{0},t})$
Set $\mu_{\mathbf{k},t} = 0$ for all $\mathbf{k} \neq \mathbf{0}$
foreach $\mathbf{k} : \mathbf{0} \neq \mathbf{k} \leq m$ **do**
 Set d as the index of the first k_i that is non-zero.
 Set $\mathbf{p} = \mathbf{k}$
 Set $p_d = p_d - 1$
 foreach $\mathbf{l} : \mathbf{l} \leq \mathbf{p}$ **do**
 $\mu_{\mathbf{k},t} += (\mu_{\mathbf{l},t}) (\kappa_{\mathbf{k}-\mathbf{l},t}) \binom{\mathbf{p}}{\mathbf{l}}$
 end
end
Return μ

Note that the main loop of Algorithm 1 must be done in the right order such that each $\mu_{\mathbf{k}}$ is computed before being used. This recursive formulation has stunning results and is at the moment the fastest way this computation can be done². Unfortunately, no equivalently fast procedure can be found the other way around, as the derivatives of the log function are more complicated.

Before diving into the estimation of multivariate gamma convolutions, we highlight in the next subsection the projection procedure from [17, 35].

2.3 Projection from \mathcal{G}_1 to $\mathcal{G}_{1,n}$

Suppose that we have a density in \mathcal{G}_1 . Then, for a certain $n \in \mathbb{N}$, the procedure provided by Miles, Furman & Kuznetsov in [35, 17], hereafter denoted MFK, gives expression for the shapes α and scales \mathbf{s} of a $\mathcal{G}_{1,n}(\alpha, \mathbf{s})$ that is exponentially close to the desired density. We will not describe the algorithm here, we refer to the original article for the details as the exposition is quite long, but we will comment on the results and use it as a comparison basis for later experiments.

A deep result about Padé approximation of Stieltjes functions supports the MFK procedure. Indeed, the first derivative of the cumulant generating function of a random variable in \mathcal{G}_1 writes (using the rate notation $r = \frac{1}{s}$ and the corresponding push-forward ν_r of the Thorin measure ν):

$$K^{(1)}(-t) = \int \frac{1}{r+t} \nu_r(\partial r).$$

This expression corresponds, when ν_r is discrete with n atoms, to a one-dimensional Stieltjes function. This Stieltjes function can be approximated from $(\mu_i)_{i \leq 2n}$ by $[m-1/m]$ standard Padé approximants, (see [12]), which denominator has zeros that are all simple, reals and negative. Applying a partial fraction decomposition to the Padé approximation is a direct way to estimate the rates and shapes of the gammas, since the obtained form for $K^{(1)}$, and therefore $K(t)$, matches the form given in Definition 2.1.1 for a $\mathcal{G}_{1,n}$ model.

The MFK algorithm has two major drawbacks. The first one is that the partial fraction decomposition of the Padé approximation, central in the procedure, is typical to univariate Stieltjes functions. Since the first derivative of the cgf is a Stieltjes function only in \mathcal{G}_1 , the method cannot be used to approximate a random variable that is not in the \mathcal{G}_1 class. Since equivalent results do not exist for multivariate Stieltjes functions, the methods is restricted to one-dimensional problems. The second one is that precision of computations needs to be at 1024 bits or higher for the algorithm to produce a result at all.

²The complexity is not exponentially increasing in the size of the arrays, on the contrary to naive implementations of Faà di Bruno's formula, see [34]

We do have an explanation for each of these drawbacks. In all generality, the Padé approximation of the Stieltjes function is a special case of a more general problem, a moment problem for the Thorin measure. Indeed, remark that:

$$\kappa_{k,t} = (k-1)! \int \frac{s^k}{(1+st)^k} \nu(\partial s),$$

and use the push-forward $x = \frac{s}{1+st}$, mapping \mathbb{R}_+ to the one-dimensional t -simplex $\Delta_1(t) = [0, \frac{1}{t}]$. You obtain a moment problem of the form:

$$\xi_k = \int_{\Delta_1(t)} x^k \xi(\partial x), \forall k \leq m,$$

where $\boldsymbol{\xi} = (\xi_k)_{k \leq m} = \left(\frac{\kappa_{k,t}}{(k-1)!} \right)_{k \leq m}$ is the vector of moments that the measure ξ needs to fit.

We can then obtain a solution with n atoms for the measures ξ and ν by solving this moment problem using classical tools, see e.g. [45] for an extensive discussion of potential algorithms. When the vector of cumulants $\boldsymbol{\kappa}$ comes from a density in the \mathcal{G}_1 class, the vector of moments $\boldsymbol{\xi}$ of the measure ξ is inside the moment cone of positives measure on $\Delta_1(t)$. We therefore obtain a proper measure ξ matching these moments. On the other hand, if $\boldsymbol{\xi}$ is *not* a sequence of moments of some measure on the simplex, which happens as soon as the random variable is *not* in \mathcal{G}_1 , then the moment problem becomes unsolvable. In this case, the approximation of the Stieltjes function fails to provide correct (within boundaries) atoms and weights for the measure ξ , and moreover for the Thorin measure.

Experiments from [35, 17] suggests that, if the inputted density is in \mathcal{G}_1 , the integrations of moments $\boldsymbol{\mu}$ and the derivation $\boldsymbol{\kappa}, \boldsymbol{\xi}$ from $\boldsymbol{\mu}$ needs to be carried out with at least 300 digits of accuracy (about 1024 bits) so that $\boldsymbol{\xi}$ is in the desired cone of moments. From an empirical dataset, even if the true distribution lies in \mathcal{G}_1 , evaluation of $\boldsymbol{\mu}$ through Monte-Carlo with enough precision is therefore impossible.

Last but not least, the results about the complete factorization of Padé approximants denominator of a Stieltjes function, which simplifies the moment problem, are not true for $d > 1$. The approach could nevertheless be carried out, as the corresponding moment problem still has a solution if the inputted cumulants are true cumulants from a \mathcal{G}_d distribution, which is an even stronger precision requirement as the dimension increases. This possibility is discussed in the next section.

3 An Estimator in $\mathcal{G}_{d,n}$

We attempt a multivariate extension of the MFK algorithm. In a multivariate setting, the Padé approximants of the cumulant generating function exists but has a denominator that is not known to be completely factorizable. Thus, we cannot take a partial fraction expansion to obtain the cumulant generating function in its canonical form to match the coefficients. Nevertheless, through the moment problem interpretation of MFK algorithm developed in Subsection 2.3, we can still express and try to solve the corresponding multivariate moment problem.

For $\mathbf{X} \sim \mathcal{G}_d(\nu)$, consider K the cumulant generating function of \mathbf{X} , given in Definition 2.1.3 as:

$$K(\mathbf{t}) = - \int \ln(1 - \langle \mathbf{s}, \mathbf{t} \rangle) \nu(\partial \mathbf{s}).$$

Assuming $|\mathbf{i}| \geq 1$ and $\mathbf{t} \leq \mathbf{0}$, we have that $\frac{\partial^{\mathbf{i}}}{\partial^{\mathbf{i}} \mathbf{t}} \ln(1 - \langle \mathbf{s}, \mathbf{t} \rangle) = -\mathbf{s}^{\mathbf{i}} (|\mathbf{i}| - 1)! (1 - \langle \mathbf{s}, \mathbf{t} \rangle)^{-|\mathbf{i}|}$, and therefore:

$$\begin{aligned} \kappa_{\mathbf{i}, \mathbf{t}} &= K^{(\mathbf{i})}(\mathbf{t}) = - \frac{\partial^{\mathbf{i}}}{\partial^{\mathbf{i}} \mathbf{t}} \int \ln(1 - \langle \mathbf{s}, \mathbf{t} \rangle) \nu(\partial \mathbf{s}) \\ &= \int \mathbf{s}^{\mathbf{i}} (|\mathbf{i}| - 1)! (-1)^{|\mathbf{i}|-1} (1 - \langle \mathbf{s}, \mathbf{t} \rangle)^{-|\mathbf{i}|} \nu(\partial \mathbf{s}) \\ &= (|\mathbf{i}| - 1)! \int \frac{\mathbf{s}^{\mathbf{i}}}{(1 - \langle \mathbf{s}, \mathbf{t} \rangle)^{|\mathbf{i}|}} \nu(\partial \mathbf{s}). \end{aligned} \tag{1}$$

Note that $\mathbf{t} \leq \mathbf{0}$ is inside the region of convergence of the function, and use now the (bijective, continuous) push-forward:

$$\mathbf{x} = \frac{\mathbf{s}}{1 - \langle \mathbf{s}, \mathbf{t} \rangle} \iff \mathbf{s} = \frac{\mathbf{x}}{1 + \langle \mathbf{x}, \mathbf{t} \rangle},$$

which goes from \mathbb{R}_+^d to the simplex $\Delta_d(-\mathbf{t}) = \{\mathbf{x} \in \mathbb{R}_+^d \text{ s.t. } \langle \mathbf{x}, -\mathbf{t} \rangle \leq 1\}$, yielding an equivalent problem of finding the measure ξ that solves the more standard moment problem:

$$\forall \mathbf{i} \leq \mathbf{m}, \xi_{\mathbf{i}} = \int_{\Delta_d(-\mathbf{t})} \mathbf{x}^{\mathbf{i}} \xi(\partial \mathbf{x}), \quad (2)$$

where we denoted $\xi = \left(\frac{\kappa_{\mathbf{i}, \mathbf{t}}}{(|\mathbf{i}|-1)!} \right)_{\mathbf{i} \leq \mathbf{m}}$ the moments that ξ needs to match according to Eq. (1).

Finding an n -atomic measure ξ solution to these equations is equivalent to finding the parametrization of a $\mathcal{G}_{d,n}$ distribution that fulfil the cumulant constraints. The corresponding Generalized Moment Problem is known as a hard problem in the literature, but can be solved through semi-definite positives relaxations, following e.g. [19, 40, 20]. Sadly, these algorithms and the corresponding literature focus on the first case, where ξ are, indeed, moments of a measure supported on the simplex $\Delta_d(-\mathbf{t})$, which is the case only if $\mathbf{X} \in \mathcal{G}_{d,n}$ and if ξ are computed with enough precision.

If the knowledge about the distribution of \mathbf{X} is given through empirical observations, with sampling noise, or if the true distributions is *not* in the $\mathcal{G}_{d,n}$ class, these moments will not belong to the moment cone, and the moment problem will have no solution. By projecting the moments onto the moment cone, we could force the known algorithm to provide a solution, but the projection onto this cone is not an easy task (see [13, 14] for details on moment cones).

The optimization of an n -atomic Thorin measure from samples of the random vector \mathbf{X} therefore rises several questions, both in univariate and multivariate cases. If we empirically compute cumulants, we will stand outside the moment cone and the moment problem will have no solution. Then, should we consider Equation 2 as a multi-objective optimization problem ? Do we seek a Pareto front, or can we consider that some moments are less important than others ? If we do, how do we weight the several objectives ?

These moments conditions could be weighted to produce a meaningful loss to minimize. For this purpose, we leverage a certain Laguerre basis of $L_2(\mathbb{R}_+^d)$ to expand the density of random vectors in $\mathcal{G}_{d,n}$, and to construct a coherent loss for this potentially impossible moment problem.

3.1 The tensorized Laguerre basis

The standard Laguerre polynomials [38] form an orthogonal basis of the set $L_2(\mathbb{R}_+, e^{-x})$ of square integrable functions with respect to the weight function $x \mapsto e^{-x}$. From these polynomials, we can extract the following orthonormal basis of the set $L_2(\mathbb{R}_+^d)$ of functions that are Lebesgue square-integrable on \mathbb{R}_+^d :

Definition 3.1.1 (Laguerre function). [8, 29, 15] For all $\mathbf{k} \in \mathbb{N}^d$, define the Laguerre functions $\varphi_{\mathbf{k}}$ with support \mathbb{R}_+^d by:

$$\varphi_{\mathbf{k}}(\mathbf{x}) = \prod_{i=1}^d \varphi_{k_i}(x_i) \text{ where for } k \in \mathbb{N}, \varphi_k(x) = \sqrt{2} e^{-x} \sum_{\ell=0}^k \binom{k}{\ell} \frac{(-2x)^\ell}{\ell!}.$$

For every function $f \in L^2(\mathbb{R}_+^d)$, for every $\mathbf{k} \in \mathbb{N}^d$, coefficients of f in the basis are denoted by:

$$a_{\mathbf{k}}(f) = \int f(\mathbf{x}) \varphi_{\mathbf{k}}(\mathbf{x}) \partial \mathbf{x},$$

and we have, since the basis is orthonormal, that $f(\mathbf{x}) = \sum_{\mathbf{k} \in \mathbb{N}^d} a_{\mathbf{k}}(f) \varphi_{\mathbf{k}}(\mathbf{x})$ and that $\|f\|_2^2 = \sum_{\mathbf{k} \in \mathbb{N}^d} a_{\mathbf{k}}^2(f)$.

Furthermore, for \mathbf{X} a random vector with density $f \in L_2(\mathbb{R}_+^d)$ and shifted moments $\boldsymbol{\mu}$, we have by simple plug-in of f and $\varphi_{\mathbf{k}}$ into the expression of $a_{\mathbf{k}}$ that:

$$a_{\mathbf{k}}(f) = \sqrt{2^d} \sum_{\boldsymbol{\ell} \leq \mathbf{k}} \binom{\mathbf{k}}{\boldsymbol{\ell}} \frac{(-2)^{|\boldsymbol{\ell}|}}{\boldsymbol{\ell}!} \boldsymbol{\mu}_{\boldsymbol{\ell}, -1} \quad (3)$$

We now denote $a_{\mathbf{k}}(\boldsymbol{\alpha}, \mathbf{s})$ the \mathbf{k}^{th} Laguerre coefficient of the $\mathcal{G}_{d,n}(\boldsymbol{\alpha}, \mathbf{s})$ distribution.

Property 3.1.2 ($\mathbf{X} \sim \mathcal{G}_{d,1}(\boldsymbol{\alpha}, \mathbf{s})$). *For $\mathbf{X} \sim \mathcal{G}_{d,1}(\boldsymbol{\alpha}, \mathbf{s})$, we have a bijection between the $d+1$ first Laguerre coefficients and the $d+1$ parameters, given through the following equations:*

$$a_{\mathbf{k}}(\boldsymbol{\alpha}, \mathbf{s}) = \sqrt{2^d} \sum_{\boldsymbol{\ell} \leq \mathbf{k}} \binom{\mathbf{k}}{\boldsymbol{\ell}} \frac{(-2\mathbf{s})^{\boldsymbol{\ell}}}{\boldsymbol{\ell}!} \frac{\Gamma(\boldsymbol{\alpha} + |\boldsymbol{\ell}|)}{\Gamma(\boldsymbol{\alpha})} (1 + |\mathbf{s}|)^{-\boldsymbol{\alpha} - |\boldsymbol{\ell}|},$$

on one hand, and

$$\alpha = \left\{ \frac{1}{c_2} - \frac{1}{\ln(c_1)} W_0 \left(\frac{\ln(c_1)}{c_2} e^{\frac{\ln(c_1)}{c_2}} \right) \right\}$$

$$s_i = \frac{a_0 - a_{\mathbf{1}(i)}}{2\alpha a_0} \text{ for all } i \in 1, \dots, d$$

reciprocally, where:

- $\mathbf{1}(i)$ is the d -variate vector with j^{th} component equal to $\mathbb{1}_{i=j}$, for all j ,
- $c_1 = a_0 \sqrt{2^{-d}}$ and $c_2 = \frac{d}{2} - \frac{1}{2a_0} \sum_{i=1}^d a_{\mathbf{1}(i)}$,
- W_0 is the zeroth branch of the Lambert function W defined by $y = W(x) \iff x = ye^y$.

Proof. For the first part, note that there exists $X_0 \sim \mathcal{G}_{1,1}(\alpha, 1)$ such that $\mathbf{X} = \langle \mathbf{s}, (X_0, \dots, X_0) \rangle$. Therefore, denoting by M the mgf of X_0 ,

$$\begin{aligned} \boldsymbol{\mu}_{\mathbf{k}, -1} &= \mathbb{E} \left(\mathbf{X}^{\mathbf{k}} e^{-\langle \mathbf{1}, \mathbf{X} \rangle} \right) = \mathbf{s}^{\mathbf{k}} \mathbb{E} \left(X_0^{|\mathbf{k}|} e^{-|\mathbf{s}| X_0} \right) \\ &= \mathbf{s}^{\mathbf{k}} M^{(|\mathbf{k}|)}(-|\mathbf{s}|) \\ &= \mathbf{s}^{\mathbf{k}} \frac{\Gamma(\boldsymbol{\alpha} + |\mathbf{k}|)}{\Gamma(\boldsymbol{\alpha})} (1 + |\mathbf{s}|)^{-\boldsymbol{\alpha} - |\mathbf{k}|} \end{aligned}$$

Therefore, the Laguerre coefficients $a_{\mathbf{k}}(\boldsymbol{\alpha}, \mathbf{s})$ of \mathbf{X} are given by:

$$a_{\mathbf{k}}(\boldsymbol{\alpha}, \mathbf{s}) = \sqrt{2^d} \sum_{\boldsymbol{\ell} \leq \mathbf{k}} \binom{\mathbf{k}}{\boldsymbol{\ell}} \frac{(-2\mathbf{s})^{\boldsymbol{\ell}}}{\boldsymbol{\ell}!} \frac{\Gamma(\boldsymbol{\alpha} + |\boldsymbol{\ell}|)}{\Gamma(\boldsymbol{\alpha})} (1 + |\mathbf{s}|)^{-\boldsymbol{\alpha} - |\boldsymbol{\ell}|}$$

Then, for the reverse assertion, consider the expression of the firsts coefficients:

$$a_0 = \sqrt{2^d} (1 - |\mathbf{s}|) \text{ and } a_{\mathbf{1}(i)} = a_0 \left(1 - 2\alpha \frac{s_i}{1 + |\mathbf{s}|} \right).$$

Use the simplex transformation $\mathbf{x} = \frac{\mathbf{s}}{1 + |\mathbf{s}|}$ to solve for \mathbf{s} as a function of $\boldsymbol{\alpha}$ in the second equation, and inject in the first. We obtain an equation of the form $ae^x + bx + c$ which is solved by the (zeroth branch of) the Lambert W function. \square

This example is quite peculiar since by Property 2.1.4 the random vector is singular. Note that the edge case $\exists i \in 1, \dots, d, r_i = 0$, in which the i^{th} marginal is identically zero, is included in the previous result.

We do not know if the same kind of inversion could be carried out analytically for the coefficients of $\mathcal{G}_{d,n}$ densities, since Laguerre coefficients are not as trivially expressed in all generality. Furthermore, this kind

of parametric inversion would not work for empirical datasets or densities outside the class, and is therefore not our focus here. However, we can efficiently compute coefficients from parameters of a $\mathcal{G}_{d,n}$ random vector through Algorithm 2, based on Algorithm 1:

Algorithm 2: Laguerre coefficients of $\mathcal{G}_{d,n}(\boldsymbol{\alpha}, \mathbf{s})$ random vectors.

Input: Shapes $\boldsymbol{\alpha} \in \mathbb{R}_+^d$, scales $\mathbf{s} \in \mathcal{M}_{n,d}(\mathbb{R}_+)$, and truncation threshold $\mathbf{m} \in \mathbb{N}^d$

Result: Laguerre coefficients $(a_{\mathbf{k}})_{\mathbf{k} \leq \mathbf{m}}$ of the $\mathcal{G}_{d,n}(\boldsymbol{\alpha}, \mathbf{s})$ density

Side-product: -1 -shifted cumulants $\boldsymbol{\kappa}$ and moments $\boldsymbol{\mu}$ up to order \mathbf{m} of the $\mathcal{G}_{d,n}(\boldsymbol{\alpha}, \mathbf{s})$ distribution.

Compute the simplex version of the scales $\mathbf{x}_i = \frac{\mathbf{s}_i}{1+|\mathbf{s}_i|}$ for all $i \in 1, \dots, n$.

Let $\kappa_{\mathbf{0}} = -\sum_{i=1}^n \alpha_i \ln(1 - |\mathbf{x}_i|)$

Let $a_{\mathbf{0}} = \mu_{\mathbf{0}} = \exp(\kappa_{\mathbf{0}})$

foreach $\mathbf{0} \neq \mathbf{k} \leq \mathbf{m}$ **do**

 Let $a_{\mathbf{k}} = \mu_{\mathbf{k}} = 0$

 Let d be the index of the first k_i that is non-zero.

 Let $\mathbf{p} = \mathbf{k}$

 Set $p_d = p_d - 1$

 Let $\kappa_{\mathbf{k}} = (|\mathbf{k}| - 1)! \sum_{i=1}^n \alpha_i \mathbf{x}_i^{\mathbf{k}}$ according to Eq. (1)

foreach $l \leq \mathbf{p}$ **do**

 Set $\mu_{\mathbf{k}} += (\mu_l) (\kappa_{\mathbf{k}-l}) \binom{\mathbf{p}}{\mathbf{l}}$ according to Algorithm 1

 Set $a_{\mathbf{k}} += \mu_l \binom{\mathbf{k}}{\mathbf{l}} \frac{(-2)^{|\mathbf{l}|}}{\mathbf{l}!}$ according to Eq. (3)

end

 Set $a_{\mathbf{k}} += \mu_{\mathbf{k}} \frac{(-2)^{|\mathbf{k}|}}{\mathbf{k}!}$

end

$\mathbf{a} = \sqrt{2}^d \mathbf{a}$

Return \mathbf{a}

The complexity of this algorithm is quadratic in the number of coefficients. Note that, as in Algorithm 1, computations need to be performed in the right order. The Laguerre coefficients $\mathbf{a} = (a_{\mathbf{k}})_{\mathbf{k} \leq \mathbf{m}}$ sometimes overflow the Float64 limits, but the implementation we provide in a Julia package `ThorinDistributions.jl`³ ensures that the computations do not overflow by using multiple precision arithmetic when needed. Furthermore, as in Algorithm 1, the use of Miatto's fast recursion gives this algorithm a good efficiency, even for reasonably large d, n, \mathbf{m} .

By reordering the coefficients and leveraging generalizations of Laguerre polynomials, we have the following link with Moschopoulos's density:

Remark 3.1.3 (Generalized Laguerre basis and Moschopoulos density). *For a random vector $X \sim \mathcal{G}_{d,n}(\boldsymbol{\alpha}, \mathbf{s})$ with Laguerre coefficients \mathbf{a} , the density f could be rewritten as a gamma series:*

$$f(\mathbf{x}) = \sum_{\mathbf{k} \in \mathbb{N}} b_{\mathbf{k}} f_{\mathbf{k}}(\mathbf{x}) \text{ where } f_{\mathbf{k}}(\mathbf{x}) = \frac{\mathbf{x}^{\mathbf{k}} e^{-|\mathbf{x}|}}{\mathbf{k}!} \text{ and } b_{\mathbf{k}} = \sqrt{2}^d (-2)^{|\mathbf{k}|} \sum_{\ell \geq \mathbf{k}} \binom{\ell}{\mathbf{k}} a_{\ell},$$

where $f_{\mathbf{k}}$ is the density of a random vector with independent $\mathcal{G}_{1,1}(k_i, 1)$ marginals. Laguerre's polynomials can be generalized by adding a parameter $\rho \geq 0$, which we took here to be 0. This alteration would reflect in the series by yielding densities of independent random vectors with $\mathcal{G}_{1,1}(k_i + \rho, 1)$ marginals instead. This hyper-parameter could be varied to produce different concurrent models, and could even be chosen dimension wise. Noteworthy is the fact that, in the one-dimensional case, if we chose ρ to be the total mass of the Thorin measure, we retrieve Moschopoulos series 2.1.2 as a Laguerre expansion.

Sadly, even if mixtures of gammas are easier to fit by k -MLE procedure [41, 46], we have no way of identifying the subspace of coefficients that would match a true convolution of gammas: as [4] noted in the univariate

³<https://github.com/lrnv/ThorinDistributions.jl>

case, generalized gamma convolutions can be expressed as mixtures of gammas, but there is no simple reverse mapping. Last but not least, if the expansion through generalized Laguerre polynomials with a parameter corresponding to the total mass of the Thorin measure would converge faster, we have no way of estimating this parameter beforehand from data, and for many useful cases it is infinite (log-Normals among others).

However, under very simple assumptions on the parameters of a $\mathcal{G}_{d,n}$ random vector, the Laguerre coefficients are exponentially decreasing. The Property 3.1.4 provides two useful bounds.

Property 3.1.4 (Error Bounds). *Let $X \sim \mathcal{G}_{d,n}(\boldsymbol{\alpha}, \mathbf{s})$ with $|\boldsymbol{\alpha}| \geq 1$. Denote $a_{\mathbf{k}}(\boldsymbol{\alpha}, \mathbf{s})$ the corresponding \mathbf{k}^{th} Laguerre coefficients.*

(i) (Tight bound) *There exists reals $\varepsilon(\boldsymbol{\alpha}, \mathbf{s}) > \mathbf{0}$ and $B(\boldsymbol{\alpha}, \mathbf{s}) > 0$, both constants with respect to \mathbf{k} , such that for all \mathbf{k} :*

$$|a_{\mathbf{k}}(\boldsymbol{\alpha}, \mathbf{s})| \leq B(\boldsymbol{\alpha}, \mathbf{s}) (1 + \varepsilon(\boldsymbol{\alpha}, \mathbf{s}))^{-\mathbf{k}},$$

(ii) (Uniform bound) *Let $\varepsilon > 0$. There exists a positive real constant $B(\varepsilon) < \infty$ such that, for all set of parameters $\boldsymbol{\alpha}, \mathbf{s}$ with all $s_{i,j}$ inside $\{\mathbf{0}\} \cup [\varepsilon, \frac{1}{\varepsilon}]$ and $|\boldsymbol{\alpha}| > 1$,*

$$|a_{\mathbf{k}}(\boldsymbol{\alpha}, \mathbf{s})| \leq B(\varepsilon)(1 + \varepsilon)^{-\mathbf{k}}.$$

Proof. For brevity, we denote in this proof $a_{\mathbf{k}} = a_{\mathbf{k}}(\boldsymbol{\alpha}, \mathbf{s})$ the Laguerre coefficients of \mathbf{X} , even if they do depend on the parameters. We follow the path of [24] to express the Laguerre coefficients $(a_{\mathbf{k}})_{\mathbf{k} \in \mathbb{N}^d}$ as the Taylor coefficients of a certain (multivariate, complex) function R . From the computation of the moment generating function $M(\mathbf{t})$ of the Laguerre expanded density $f(\mathbf{x}) = \sum_{\mathbf{k} \in \mathbb{N}^d} a_{\mathbf{k}} \varphi_{\mathbf{k}}(\mathbf{x})$, we have:

$$\begin{aligned} M(\mathbf{t}) &= \int_{\mathbb{R}_+^d} e^{\langle \mathbf{t}, \mathbf{x} \rangle} f(\mathbf{x}) \partial \mathbf{x} = \sqrt{2}^d \sum_{\mathbf{k} \in \mathbb{N}^d} a_{\mathbf{k}} \prod_{i=1}^d \sum_{j=0}^{k_i} \binom{k_i}{j} \frac{(-2)^j}{j!} \int_{\mathbb{R}_+} e^{(t_i-1)x_i} x_i^j \partial x_i \\ &= \sqrt{2}^d \sum_{\mathbf{k} \in \mathbb{N}^d} a_{\mathbf{k}} \prod_{i=1}^d \sum_{j=0}^{k_i} \binom{k_i}{j} \frac{(-2)^j}{(1-t_i)^{j+1}} \\ &= \sqrt{2}^d (1-\mathbf{t})^{-1} \sum_{\mathbf{k} \in \mathbb{N}^d} a_{\mathbf{k}} \left(1 - \frac{2}{1-\mathbf{t}}\right)^{\mathbf{k}} \\ &= \sqrt{2}^d (1-\mathbf{t})^{-1} \sum_{\mathbf{k} \in \mathbb{N}^d} a_{\mathbf{k}} \left(\frac{\mathbf{t}+1}{\mathbf{t}-1}\right)^{\mathbf{k}}, \end{aligned}$$

which implies that:

$$\sum_{\mathbf{k} \in \mathbb{N}^d} a_{\mathbf{k}} \left(\frac{\mathbf{t}+1}{\mathbf{t}-1}\right)^{\mathbf{k}} = \sqrt{2}^{-d} (1-\mathbf{t})^1 M(\mathbf{t}).$$

Let $h(t) = \frac{t+1}{t-1}$ for all $t \in \mathbb{C}$, and $\mathbf{h}(\mathbf{t}) = \frac{\mathbf{t}+1}{\mathbf{t}-1} = (h(t))_{t \in \mathbf{t}}$, so that \mathbf{h} applies h componentwise. Note that h is a Möbius transform which is its own inverse: $h(h(t)) = t$.

Using now the substitutions $\mathbf{y} = \mathbf{h}(\mathbf{t}) \iff \mathbf{t} = \mathbf{h}(\mathbf{y})$, which implies that $(1-\mathbf{t})^1 = 2^d (1-\mathbf{y})^{-1}$, to obtain:

$$R(\mathbf{y}) := \sum_{\mathbf{k} \in \mathbb{N}^d} a_{\mathbf{k}} \mathbf{y}^{\mathbf{k}} = \sqrt{2}^d M(\mathbf{h}(\mathbf{y})) (1-\mathbf{y})^{-1},$$

where we called R the function that has the Laguerre coefficients as Taylor coefficients.

We now study the regularity of the function $R : \mathbb{C}^d \rightarrow \mathbb{C}$. Denote by $\mathcal{V}_R(\mathbf{s})$ the singular variety of R , i.e. the set of points where R is not analytic, which depends on the scale matrix \mathbf{s} of the random vector (but not on the shapes as we show next).

From the singularities of $\mathbf{y} \mapsto (1 - \mathbf{y})^{-1}$ and from those of M , we have:

$$\mathcal{V}_R(\mathbf{s}) \subseteq \{\mathbf{y} : \exists j, y_j = 1\} \bigcup_{i=1}^n \{\mathbf{y} : \mathbf{h}(\mathbf{y}) \in \partial\Delta_d(\mathbf{s}_i)\}$$

where $\partial\Delta_d(\mathbf{s}_i) = \{\mathbf{x} : \langle \mathbf{x}, \mathbf{s}_i \rangle = 1\}$ is the $(d - 1)$ -dimensional frontier of the \mathbf{s}_i -simplex.

We now show that R is in fact regular at $\{\mathbf{y} : \exists j, y_j = 1\}$. By our substitution, this is equivalent to the regularity at infinity of the function $M(\mathbf{t})(1 - \mathbf{t})^1$, and since we supposed $|\boldsymbol{\alpha}| \geq 1$,

$$\lim_{\forall i, |t_i| \rightarrow \infty} |M(\mathbf{t})(1 - \mathbf{t})^1| = \lim_{\forall i, |t_i| \rightarrow \infty} \left| \frac{(1 - \mathbf{t})^1}{\prod_{i=1}^n (1 - \langle \mathbf{s}_i, \mathbf{t} \rangle)^{\alpha_i}} \right| < +\infty,$$

because the dominant term of the numerator is \mathbf{t}^1 and the dominant term of the denominator is proportional to $\mathbf{t}^{|\boldsymbol{\alpha}|}$. Therefore, $\mathbf{t} \mapsto |M(\mathbf{t})(1 - \mathbf{t})^1|$ is regular at infinity, which means that R is regular at $\{\mathbf{y} : \exists j, y_j = 1\}$. Hence,

$$\mathcal{V}_R(\mathbf{s}) \subseteq \bigcup_{i=1}^n \{\mathbf{y} : \mathbf{h}(\mathbf{y}) \in \partial\Delta_d(\mathbf{s}_i)\}.$$

We now claim that $\mathcal{V}_R(\mathbf{s})$ elements have coordinates that have modulus greater than one. For that, let $A_i = \{\mathbf{y} : \mathbf{h}(\mathbf{y}) \in \partial\Delta_d(\mathbf{s}_i)\}$. Note that A_i is the image of the simplex boundary by the function \mathbf{h} , since \mathbf{h} is its own inverse, i.e:

$$A_i = \mathbf{h}(\partial\Delta_d(\mathbf{s}_i)).$$

Moreover, the simplex boundary contains vectors with positives real coordinates, and imaginary coordinates that are orthogonal to \mathbf{s}_i . Since for all complex z such that $Re(z) > 0$, $|h(z)| > 1$, A_i contains only vectors with modulus greater than one, and so does $\mathcal{V}_R(\mathbf{s})$.

Denote now \mathbf{y}_0 the dominant singularity of R , with $|y_{0,i}| > 1$ for all i as we just showed.

Let $\boldsymbol{\varepsilon}(\mathbf{p}) = \left(\frac{|y_{0,i}| - 1}{p_i} \right)_{i=1, \dots, d}$ where $\mathbf{p} \in \mathbb{R}_+^d$, $\mathbf{p} > \mathbf{1}$ is a parameter. Note that $1 + \boldsymbol{\varepsilon}$ are the radii of a region of analyticity of R around the origin. Therefore, by Cauchy's inequality, see Theorem 2.2.7 in [21], we have that for all \mathbf{p} :

$$|a_{\mathbf{k}}(\boldsymbol{\alpha}, \mathbf{s})| = \frac{1}{\mathbf{k}!} \left| \frac{\partial^{\mathbf{k}} R}{\partial^{\mathbf{k}}} (0) \right| \leq (1 + \boldsymbol{\varepsilon}(\mathbf{p}))^{-\mathbf{k}} \sup_{D(\boldsymbol{\varepsilon}(\mathbf{p}))} |R|,$$

where $D(\boldsymbol{\varepsilon}(\mathbf{p})) = \{z \in \mathbb{C}^d : |z_i| < 1 + \varepsilon_i(\mathbf{p})\}$ is the centered polydisc with polyradius $1 + \boldsymbol{\varepsilon}(\mathbf{p})$. This bound can be minimized on \mathbf{p} to obtain wanted $B(\boldsymbol{\alpha}, \mathbf{s})$ and $\boldsymbol{\varepsilon}(\boldsymbol{\alpha}, \mathbf{s})$.

Now that the first bound is showed, we will show the second one by digging into the expression of the previous sup. Consider that we chose a \mathbf{p} and drop this parameter for clarity, $\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}(\mathbf{p})$.

Consider again the supremum $\sup_{\mathbf{y} \in D(\boldsymbol{\varepsilon})} |R(\mathbf{y})|$. Since $\mathbf{y} \in D(\boldsymbol{\varepsilon}) \iff \mathbf{h}(\mathbf{y}) \in \mathbf{h}(D(\boldsymbol{\varepsilon}))$, we have:

$$\sup_{\mathbf{y} \in D(\boldsymbol{\varepsilon})} |R(\mathbf{y})| = \sqrt{2}^{-d} \sup_{\mathbf{t} \in \mathbf{h}(D(\boldsymbol{\varepsilon}))} |M(\mathbf{t})(1 - \mathbf{t})^1|.$$

We neglect the $\sqrt{2}^{-d}$ in the following.

Now, the function \mathbf{h} has several interesting properties: \mathbf{h} is its own inverse, it maps circles in the complex planes to other circles in the complex plane, and maps real numbers to real numbers. It is also a symmetric function w.r.t the real axis, and therefore the space $\mathbf{h}(D(\boldsymbol{\varepsilon}))$ has a circle-shaped boundary in each dimension, and this circle is symmetric w.r.t the real line. Note that both $\mathbf{h}(1 + \varepsilon_i)$ and $\mathbf{h}(-1 - \varepsilon_i)$ are reals for all i ,

and are therefore the leftmost and rightmost points of these circles. We can compute the radii and centers of the circles, and we obtain the following expression for $\mathbf{h}(D(\varepsilon))$:

$$\mathbf{h}(D(\varepsilon)) = \{ \mathbf{z} \in \mathbb{C}^d : \forall i \in 1, \dots, d, |z_i - c_i| \geq r_i \},$$

for explicit centers and radii:

$$c_i = \frac{(1 + \varepsilon_i)^2 + 1}{(1 + \varepsilon_i)^2 - 1} \text{ and } r_i = \frac{2(1 + \varepsilon_i)}{(1 + \varepsilon_i)^2 - 1}.$$

Therefore, we are maximizing a complex function over the full \mathbb{C}^d except a polydisc, centered at a value in \mathbb{R}_+^d (symmetric along the real axis). These discs all contain the value 1, and therefore the point $\mathbf{1}$ is outside the domain we maximize on. We now switch to polar coordinates, centered on this polydisc: to each $\mathbf{t} \in \mathbf{h}(D(\varepsilon))$, we associate a vector of angles $\boldsymbol{\theta}$ and a vector of positives reals \mathbf{u} , such that: $\mathbf{t} = (c_j + (r_j + u_j)e^{i\theta_j})_{j \in 1, \dots, d}$. We rewrite the optimization problem accordingly:

$$\sup_{\mathbf{t} \in \mathbf{h}(D(\varepsilon))} |M(\mathbf{t})(1 - \mathbf{t})^{\mathbf{1}}| = \sup_{\substack{\mathbf{u} \geq \mathbf{0} \\ \boldsymbol{\theta} \in [0, 2\pi]^d}} \frac{|(1 - \mathbf{c} - (\mathbf{r} + \mathbf{u})e^{i\boldsymbol{\theta}})^{\mathbf{1}}|}{\prod_{j=1}^n \underbrace{|1 - \langle \mathbf{s}_j, \mathbf{c} + (\mathbf{r} + \mathbf{u})e^{i\boldsymbol{\theta}} \rangle|^{\alpha_j}}_{T_j}}.$$

Recall that the case where $|t_i| \rightarrow \infty$ has already been treated. We now show that, for \mathbf{t} finite, T_j is lower bounded for every j . The minimum of T_j is a distance between the frontier $\partial\Delta_d(\mathbf{s}_j)$ of the \mathbf{s}_j -simplex and the outside $\mathbf{h}(D(\varepsilon))$ of the (\mathbf{c}, \mathbf{r}) -polydisc. Since these sets are both closed, their intersection being empty is sufficient for the distance to be strictly positive. We now show this emptiness by contradiction. Suppose that $\mathbf{t} \in \partial\Delta_d(\mathbf{s}_j) \cap \mathbf{h}(D(\varepsilon))$. Then applying the function \mathbf{h} we have:

$$\mathbf{y} = \mathbf{h}(\mathbf{t}) \in \mathbf{h}(\partial\Delta_d(\mathbf{s}_j)) \cap D(\varepsilon).$$

Since $\mathbf{t} \in \partial\Delta_d(\mathbf{s}_j)$, all $\text{Re}(t_i) \geq 0$, and for $\text{Im}(t_i) < \infty$ we have:

$$|y_i| = |h(t_i)| = \sqrt{1 + \frac{4(\text{Re}(t_i) - 1)^2 + \text{Im}(t_i)^2}{\{\text{Re}(t_i) - 1\}^2 + \text{Im}(t_i)^2}} \geq 1 + \min(\text{Re}(t_i), \frac{1}{\text{Re}(t_i)}),$$

with equality only if $t_i = 0$ (otherwise this bound is quite loose, but practical).

Since at least one value in each scale vector $s_{j,1}, \dots, s_{j,d}$ must be strictly positive, at least one t_i must have strictly positive real part, smaller or equal to $\frac{1}{s_{j,i}}$ (for an $s_{j,i}$ strictly positive), and there therefore exists at least one y_i with modulus $|y_i| \geq 1 + \min_{s \in \mathbf{s}, s_i > 0} (s_i, \frac{1}{s_i})$.

On the other hand, $\mathbf{y} \in D(\varepsilon)$ implies that for all $i \in 1, \dots, d$, $|y_i| \leq 1 + \varepsilon_i$ by definition of $D(\varepsilon)$. It suffices to take ε as proposed to have a contradiction, making the intersection empty.

Since this intersection is empty, T_j is clearly lower-bounded, and the function cannot explode around $\mathbf{u} = 0$. Furthermore, the full denominator has a dominant term of the form $\mathbf{u}^{|\boldsymbol{\alpha}|}$. Since $|\boldsymbol{\alpha}| \geq 1$, the degree of the numerator is non-smaller than the degree of the denominator, and the limit when $|\mathbf{u}| \rightarrow \infty$ is still bounded by a constant that does not depend on $\boldsymbol{\alpha}$ but only on the scales. Hence, the sup is clearly upper-bounded by some $B(\varepsilon) < \infty$, which proves the second part of the result. \square

We note that a tighter uniform bound could be found by finding a tighter lower bound to $|h(s)|$, in the second part of the proof of Property 3.1.4. We stress that the remainder of the argument would essentially be the same.

Since we can now compute efficiently the Laguerre coefficients of density in $\mathcal{G}_{d,n}$, with assurance that they will decrease fast, we propose in the next subsection to discuss an estimation algorithm based on an Integrated Square Error loss that leverages these coefficients.

3.2 An incomplete but however consistent empirical loss

Expressing the density of $\mathcal{G}_{d,n}$ random vectors into the basis from Definition 3.1.1 allows us to truncate the basis and effectively compute an approximated density. From this, by evaluating empirical Laguerre coefficients from data, we will fit the parameters to data by minimizing the L_2 distance between coefficients (since the basis is orthonormal). For parameters $\boldsymbol{\alpha}, \mathbf{s}$, denote by $f_{\boldsymbol{\alpha}, \mathbf{s}}$ the density of the $\mathcal{G}_{d,n}(\boldsymbol{\alpha}, \mathbf{s})$ distribution, and f the density underlying the observations $\mathbf{x}_1, \dots, \mathbf{x}_N$.

We would like to minimize the integrated square error $\|f - f_{\boldsymbol{\alpha}, \mathbf{s}}\|_2^2 = \sum_{\mathbf{k} \in \mathbb{N}^d} (a_{\mathbf{k}}(f) - a_{\mathbf{k}}(\boldsymbol{\alpha}, \mathbf{s}))^2$.

Estimating $a_{\mathbf{k}}(f) = \int f(\mathbf{x}) \varphi_{\mathbf{k}}(\mathbf{x}) \partial \mathbf{x}$ by a simple Monte-Carlo plugin $\widehat{a}_{\mathbf{k}} = \frac{1}{N} \sum_{i=1}^N \varphi_{\mathbf{k}}(\mathbf{x}_i)$, we could compute an approximation of this loss. However, we cannot compute all couples of coefficients $\widehat{a}_{\mathbf{k}}$ and $a_{\mathbf{k}}(\boldsymbol{\alpha}, \mathbf{s})$ for all $\mathbf{k} \in \mathbb{N}^d$, and we are forced to truncate the basis.

Definition 3.2.1 ($\mathcal{G}_{d,n}$ parameter estimators). *For a dataset $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_N)$, we define parameters estimators of a $\mathcal{G}_{d,n}$ distribution that matches the observations as:*

$$(\widehat{\boldsymbol{\alpha}}, \widehat{\mathbf{s}}) = \arg \min_{\boldsymbol{\alpha} \geq 0, \mathbf{s} \geq 0} L_{\mathbf{m}}(\mathbf{x}, \boldsymbol{\alpha}, \mathbf{s}),$$

where the truncated loss is:

$$L_{\mathbf{m}}(\mathbf{x}, \boldsymbol{\alpha}, \mathbf{s}) = \sum_{\mathbf{k} \leq \mathbf{m}} (\widehat{a}_{\mathbf{k}} - a_{\mathbf{k}}(\boldsymbol{\alpha}, \mathbf{s}))^2,$$

and where the empirical Laguerre coefficients are given by:

$$\widehat{a}_{\mathbf{k}} = \frac{1}{N} \sum_{i=1}^N \varphi_{\mathbf{k}}(\mathbf{x}_i).$$

The loss $L_{\mathbf{m}}(\mathbf{x}, \boldsymbol{\alpha}, \mathbf{s})$ can be efficiently computed through Algorithm 2. It will be 0 if the first coefficients of the Laguerre expansions of f and of our estimator matches perfectly, i.e., assuming $f \in \mathcal{G}_{d,n}$ and \mathbf{m} is big enough, if we find the right Thorin measure. Note that if the goal was given theoretically, like in a projection from \mathcal{G}_d to $\mathcal{G}_{d,n}$, we could compute $\widehat{a}_{\mathbf{k}}$ by formal integration with high precision instead of Monte-Carlo. We will discuss this options in the next section. However, even in this case, the error that comes from the truncation of the basis cannot be computed.

To show the consistency of this loss, we use in the following the results from Property 3.1.4.

Property 3.2.2 (Consistency). *Consider that $f \sim \mathcal{G}_{d,n}(\boldsymbol{\alpha}_0, \mathbf{s}_0)$, $|\boldsymbol{\alpha}_0| \geq 1$, and denote \mathbf{x} a dataset generated from this density with N observations. Fix $\mathbf{m} \in \mathbb{N}^d$, and $\varepsilon > 0$ small enough so that $\mathbf{s}_0 \in \{\mathbf{0}\} \cup [\varepsilon, \varepsilon^{-1}]^n$. Suppose that we have a global minimizer*

$$(\widehat{\boldsymbol{\alpha}}, \widehat{\mathbf{s}}) = \arg \min_{\substack{|\boldsymbol{\alpha}| \geq 1 \\ \boldsymbol{\alpha} > 0 \\ \mathbf{s} \in \{\mathbf{0}\} \cup [\varepsilon, \frac{1}{\varepsilon}]^n}} L_{\mathbf{m}}(\mathbf{x}, \boldsymbol{\alpha}, \mathbf{s}),$$

which depends on the threshold \mathbf{m} and on the N observations \mathbf{x} . Then the following convergence in probability occurs:

$$\|f - f_{\widehat{\boldsymbol{\alpha}}, \widehat{\mathbf{s}}}\|_2^2 \xrightarrow[\mathbf{m} \rightarrow \infty]{N \rightarrow \infty} 0.$$

Proof. We will deduce the convergence in probability from the convergence in distribution since the limit is a constant. To show the result, we start by expressing the loss in the Laguerre basis, and we split the basis on indices smaller and greater than \mathbf{m} :

$$\begin{aligned} \|f - f_{\hat{\alpha}, \hat{s}}\|_2^2 &= \sum_{\mathbf{k} \in \mathbb{N}^d} (a_{\mathbf{k}} - a_{\mathbf{k}}(\hat{\alpha}, \hat{s}))^2 \\ &= \underbrace{\sum_{\mathbf{k} \leq \mathbf{m}} (a_{\mathbf{k}} - a_{\mathbf{k}}(\hat{\alpha}, \hat{s}))^2}_A + \underbrace{\sum_{\mathbf{k} > \mathbf{m}} (a_{\mathbf{k}} - a_{\mathbf{k}}(\hat{\alpha}, \hat{s}))^2}_B \end{aligned}$$

We discuss the A part first, for which we can treat each term in the summation independently as the sum is finite. A can be expanded further by plugging in the Monte-Carlo estimator $\widehat{a}_{\mathbf{k}} = \frac{1}{N} \sum_{i=1}^N \varphi_{\mathbf{k}}(\mathbf{x}_i)$ and expanding the square:

$$A = \underbrace{\sum_{\mathbf{k} \leq \mathbf{m}} (a_{\mathbf{k}} - \widehat{a}_{\mathbf{k}})^2}_{A_1} + 2 \underbrace{\sum_{\mathbf{k} \leq \mathbf{m}} (a_{\mathbf{k}} - \widehat{a}_{\mathbf{k}}) (\widehat{a}_{\mathbf{k}} - a_{\mathbf{k}}(\hat{\alpha}, \hat{s}))}_{A_2} + \underbrace{\sum_{\mathbf{k} \leq \mathbf{m}} (\widehat{a}_{\mathbf{k}} - a_{\mathbf{k}}(\hat{\alpha}, \hat{s}))^2}_{L_{\mathbf{m}}(\mathbf{x}, \hat{\alpha}, \hat{s})}.$$

Now, since $\widehat{a}_{\mathbf{k}} \xrightarrow[N \rightarrow \infty]{a.s.} a_{\mathbf{k}}$ for all \mathbf{k} by Monte-Carlo, $A_1 \xrightarrow[N \rightarrow \infty]{a.s.} 0$.

Furthermore, since we restricted the optimization to $|\hat{\alpha}| \geq 1$, each $a_{\mathbf{k}}(\hat{\alpha}, \hat{s})$ are bounded by Property 3.1.4. Now $|\widehat{a}_{\mathbf{k}}| \leq \sqrt{2^d}$ because $|\varphi_{\mathbf{k}}(\mathbf{x})| \leq \sqrt{2^d}$ whatever $\mathbf{x} \in \mathbb{R}_+^d$. Therefore, for any \mathbf{k} , $(\widehat{a}_{\mathbf{k}} - a_{\mathbf{k}}(\hat{\alpha}, \hat{s}))$ is bounded and $(a_{\mathbf{k}} - \widehat{a}_{\mathbf{k}}) \xrightarrow[N \rightarrow \infty]{} 0$, which makes $A_2 \xrightarrow[N \rightarrow \infty]{} 0$.

Last but not least, by definition of $\hat{\alpha}, \hat{s}$ as global minimizers, $L_{\mathbf{m}}(\mathbf{x}, \hat{\alpha}, \hat{s}) \leq L_{\mathbf{m}}(\mathbf{x}, \alpha_0, \mathbf{s}_0) \xrightarrow[N \rightarrow \infty]{} 0$ since the Monte-Carlo estimators are not biased, and since α_0, \mathbf{s}_0 are inside the ranges of optimizations. Therefore, $A \xrightarrow[N \rightarrow \infty]{} 0$.

Now consider the remainder B . By the second part of Property 3.1.4, applied to both the true model and the approximation, we have uniform exponentially decreasing bounds on both $a_{\mathbf{k}}$ and $a_{\mathbf{k}}(\hat{\alpha}, \hat{s})$ that only depends on ε . Therefore, the squared difference $(a_{\mathbf{k}} - a_{\mathbf{k}}(\hat{\alpha}, \hat{s}))^2$ is also exponentially bounded, uniformly on $\mathbf{k} > \mathbf{m}$, and converge to zero for \mathbf{k} increasing. Hence, $B \xrightarrow[\mathbf{m} \rightarrow \infty]{} 0$, which concludes the argument. \square

Although the consistency result restricts the parameter space, in practice we run our optimization procedures without this restriction, and simply check at the end that the resulting estimators are inside these bounds. Fixing ε so that a finite set of scales follows the restriction from 3.2.2 is always possible, and therefore only the fact that $|\alpha| \geq 1$ needs to be checked. This means that the produced Thorin measure as a mass greater than one, where distributions like log-Normal and Pareto have an infinite total mass for the Thorin measure.

We now have a reliable loss to minimize and an efficient algorithm to compute it, allowing us to estimate multivariate gamma convolutions from data. To test the approach, we propose in the next section to discuss some examples of potential applications.

4 Investigation of performance

In this section, we demonstrate the results of our algorithm on several examples.

Note that, through the recursive Faà di Bruno formula in Algorithm 2, we produce Laguerre coefficients as very high degree multinomials into α, μ_0 and ξ (the simplex projection of the scales). The loss will therefore have a myriad of local minima, making it highly non-convex and forcing us to use global minimization routines. We found the Particle Swarm [26, 55] global optimization routine to work particularly well on our test sets. All the implementation was gathered in the provided Julia package, along with the code corresponding to this investigation.

We start by the projection of known densities onto the $\mathcal{G}_{1,n}$ class, differentiating the cases when the known density is inside \mathcal{G}_1 or not. Then we discuss the estimation of $\mathcal{G}_{1,n}$ models from empirical data, with heavy tailed simulated data, and we finally look at the estimation of $\mathcal{G}_{d,n}$ models from multivariate empirical data, with a particular emphasis on the properties of the dependence structure underlying the data.

4.1 Projection from a given density

The two firsts examples are a log-Normal density and a Weibull density, that have the particularity of being respectively inside and outside \mathcal{G}_1 . We will also take a look at the Pareto case, which is inside \mathcal{G}_1 , but might be outside $\mathcal{L}_2(\mathbb{R}_+)$.

Only one projection algorithm is available in the literature that project \mathcal{G}_1 densities onto $\mathcal{G}_{1,n}$, finding a set of scales and shapes that matches the inputted distribution: the MFK procedure that we already discussed in Section 2.3.

We want to compare the two algorithm on fair grounds. Therefore, we modify a little our algorithm to use Laguerre coefficients based on the *theoretical* integrals $\mu_{k,-1}$, as MFK. The choice of the input distributions is also guided by [35, 17]. The experiment is as follows:

- We compute shifted moments $(\mu_{k,-1})_{k \leq 2n}$ of the inputted distribution through direct integration of the density, at 1024 bits of precision (to match the 300 digits that MFK requires).
- We use MFK to compute an approximation in $\mathcal{G}_{d,n}$ based on these moments.
- Through Eq 3, we compute Laguerre coefficients based on the same shifted moments $\mu_{k,-1}$, and we minimize the loss given through Algorithm 2 to obtain our approximation.
- We compute Kolmogorov-Smirnov (one-sample, exact) distances between $N = 10\,000$ simulations from the estimated $\mathcal{G}_{d,n}$ model to the theoretical distribution, on $B = 100$ different simulations.

4.1.1 Projection from \mathcal{G}_1 densities to $\mathcal{G}_{1,n}$: the log-Normal case

The log-Normal distribution is a fundamental distribution in the \mathcal{G}_1 class. The proof that it belongs to \mathcal{G}_1 is actually what historically initiated the study of the class by Thorin. To match MFK's experiment, we conduct our first comparison with a $LN(\mu = 0, \sigma = 0.83)$ density. The resampled Kolmogorov-Smirnov distances are summarized, for several n and for both algorithms, in Figure 1.

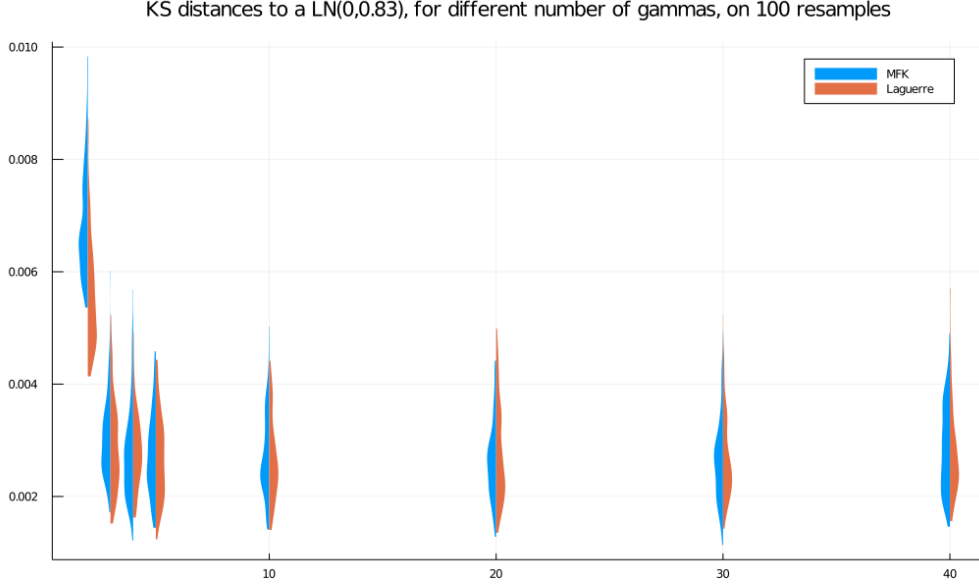


Figure 1: Violin densities of resampled KS distances (smaller is better). Blue: MFK projection, Orange: Laguerre Projection. In abscissa is the number of gammas $n \in \{2, 3, 4, 5, 10, 20, 30, 40\}$.

Although MFK has a very strong convergence results, it seems that for $n = 2$ gammas, our estimator performs better with the same information (theoretical shifted moments given with 1024 bits precision), and that for $n > 2$, the performance is overall the same. Indeed, if we look at estimated parameters, we see that in most cases the two algorithms found approximations that are close to each other. We reported the shapes and scales that both algorithm produced for $n = 2, 3, 4$ and 5 in Table 1.

Table 1: Estimated Parameters from both algorithms on the log-Normal example

	MFK		Laguerre	
	$\hat{\alpha}$	\hat{s}	$\hat{\alpha}$	\hat{s}
n=2	0.5688	1.5622	0.5458	1.6283
	2.4596	0.1941	2.4539	0.1999
n=3	0.2195	2.4902	0.2070	2.5781
	0.8934	0.6718	0.8919	0.6875
	2.6081	0.0972	2.6071	0.0987
n=4	0.0919	3.4076	0.0844	3.5307
	0.4596	1.2208	0.4555	1.2513
	1.0042	0.3737	1.0063	0.3792
	2.6956	0.0588	2.6957	0.0594
n=5	0.0398	4.3466	0.0346	4.5447
	0.2569	1.7963	0.2492	1.8576
	0.5708	0.7234	0.5721	0.7394
	1.0576	0.2399	1.0609	0.2428
	2.7574	0.0396	2.7582	0.0399

Although the produced estimators are close to each other on this example, our method has several clear benefits: the full density does not need to belong to \mathcal{G}_1 , and we do not require shifted moments up to 300 digits precision. In the next subsection, we describe the same experiment on a Weibull density that is known to make MFK simply fail, even with accurate enough computations.

4.1.2 Projection from densities outside \mathcal{G}_1 to $\mathcal{G}_{1,n}$: the Weibull case

Consider the Weibull distribution with shape $k > 0$. This distribution has a pdf given by:

$$f(x) = kx^{k-1} \exp\{-x^k\} \mathbf{1}_{x \geq 0}.$$

When $k > 1$, this distribution does not belong to \mathcal{G}_1 . In [35, 17], authors report negatives shapes for the MFK approximation in $\mathcal{G}_{1,2}$ of the Weibull with shape $k = \frac{3}{2}$, which is clearly a failure. This failure is occurring because the truncated moment problem that MFK is solving has no solution, as we detailed in Section 2.3, since the (normalized) cumulants ξ are not inside the moment cone.

However, the Laguerre basis is an orthonormal basis of $L_2(\mathbb{R}_+)$, which contains this density. The projection procedure detailed at the top of Subsection 4.1 still works correctly (i.e. produce meaningful shapes and scales that are positives, whatever the number of gammas we ask for). The Figure 2 displays a Violin plot of KS distances for several number of gammas:

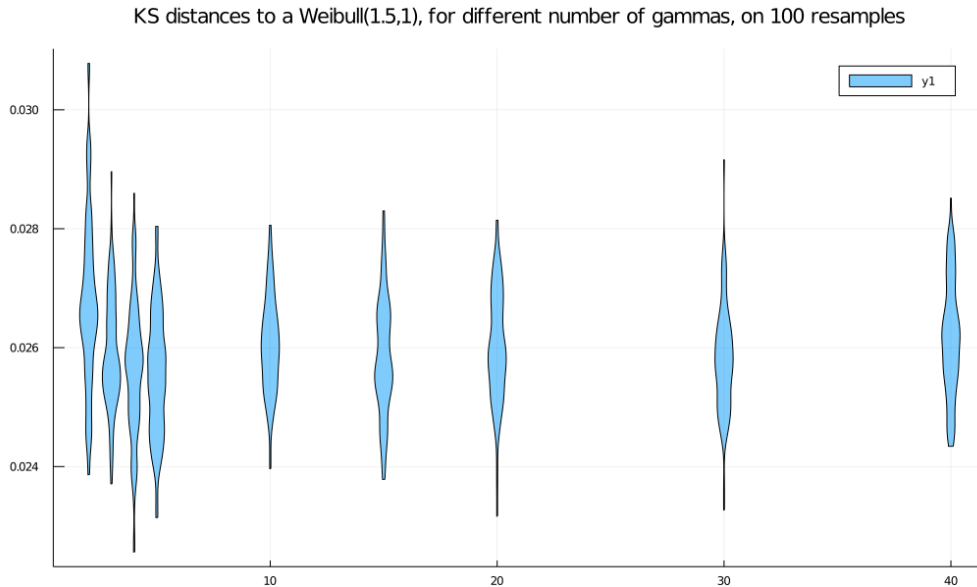


Figure 2: Violin densities of resampled KS distances for approximation of a Weibull($k = 1.5$). In abscissa is the number of gammas, n .

On Figure 2, we see that the KS distance is decreasing for an increasing number of gammas n , when n is small, but reaches a minimum quite quickly. Indeed, the distance is lower-bounded since the target is not in the class. Although none of the proposed approximations passes the KS test, we will revisit this example later from an estimation point of view to show that the projection is still an acceptable approximation for the inputted Weibull distribution.

We also note that the KS distance is not stable for increasing n : the algorithm has troubles to fix parameters to 0 if he does not need more precision. An additional penalty term in this direction might be a good solution to overcome this behavior.

Through these two first examples, we showed the projection possibilities of our algorithm, and compared them with MFK, the only existing method in the literature. We saw that for inputted densities inside the \mathcal{G}_1 class our method performs as good as MFK, but moreover that our method still works properly and produces meaningful results for examples outside the class, contrary to MFK which does not give any results. We will now talk about estimation from data, a problem that has never been solved for gamma convolutions, and showcase how our algorithm can find good representation for a given dataset, whatever the dimension of this dataset.

In the following, we will treat empirical dataset given in standard IEE 754 float64 precision.

4.2 Estimation from empirical data: univariate case

Before coming back in more details to the Weibull case, we wanted to perform some tests with empirical data coming from simulations of two heavy-tailed distributions: the log-Normal and the Pareto.

4.2.1 Heavy-tailed examples: log-Normals and Pareto cases

We consider first a dataset of 100'000 samples from a log-Normal distribution $LN(\mu = 0, \sigma = 0.83)$, in 64 bits precision.

From this sample, we compute empirical Laguerre coefficients, and then we optimize the set of shapes and scales of the estimated density to minimize the truncated L_2 loss between them and the ones produced through Algorithm 2. Following [17, 35], we choose the size of the basis as being $m = 2n + 1$, such that if we remove the first moment that is irrelevant, we have a number of moments equal to the number of parameters⁴. We ran the experiment for different number of gammas $n = 10, 20$, and 40. The corresponding results are respectively available in Figures 3, 4 and 5.

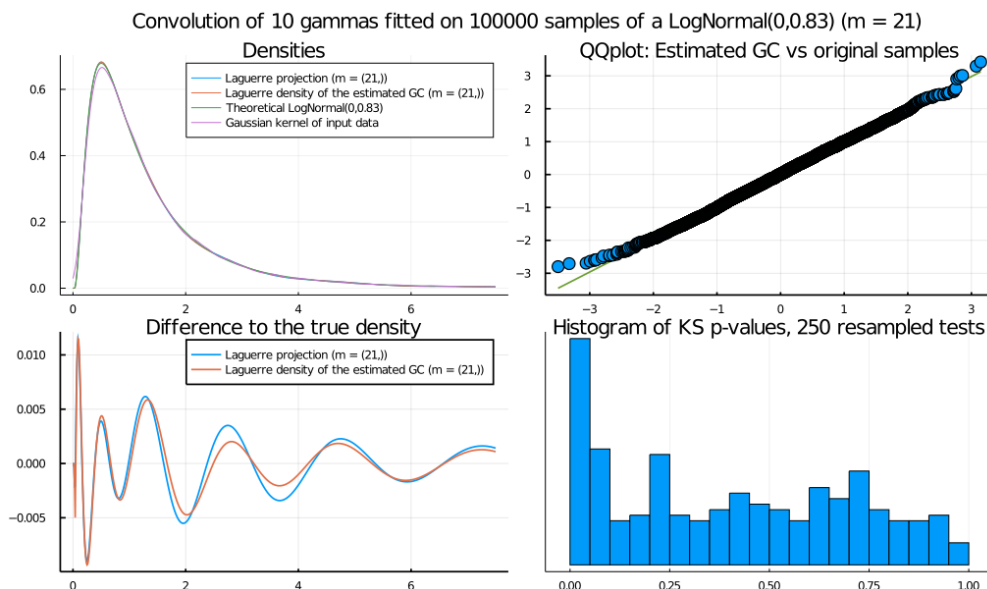


Figure 3: Log-Normal results with 10 gammas: Upper left, the comparison of the density approximations. Lower left, the difference of the approximated density to the true density. Upper right, a quantile-quantile plot of the approximation against the true distribution. Lower right, p-values of one-sample KS tests of simulations from the estimator against the true distribution.

We see on these graphs that the approximation is good enough to fool the Kolmogorov-Smirnov test. The absolute pointwise error between the Laguerre approximation of the density of the gamma convolution and the true log-Normal density is lower than 0.005 for $n = 10, 20$, and twice smaller for $n = 40$. A positive but surprising results is that, although the distance between the Laguerre approximation and the true density does not reduce with increasing m , the distance between the gamma approximation and the true density does.

The KS test setup is as follows: we simulate 250 resamples of 100'000-sized dataset from the estimated density, and compare it with a one-sample exact KS test to the true density (and not to the samples that the estimator was computed from). For all values of n , we observe close-to-uniform distribution of the p-values, which is a good result.

⁴Since Julia starts counting at 1 and not 0, we actually have m moments and not $m + 1$ as we did in the previous theoretical analysis.

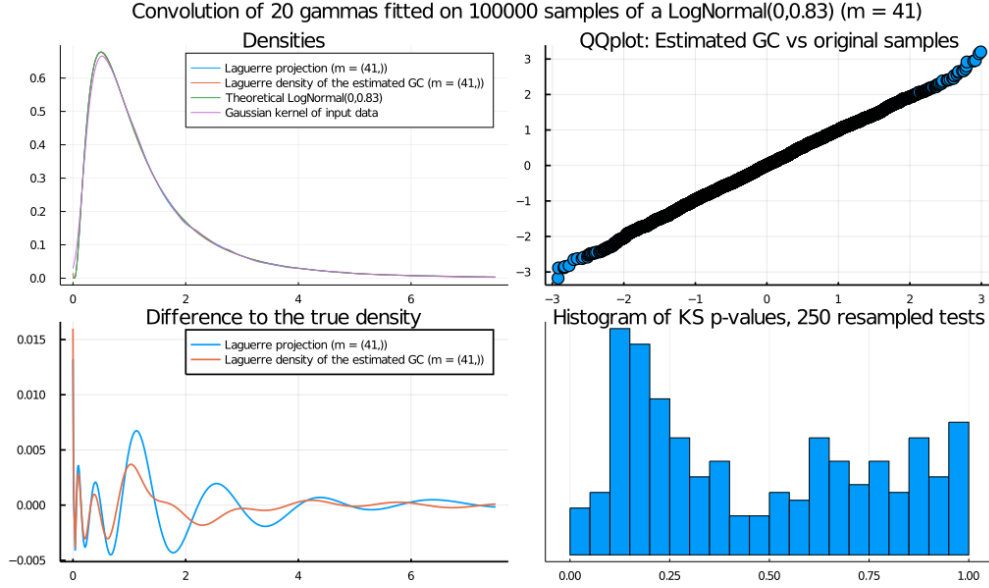


Figure 4: Log-Normal results with 20 gammas. Same legend as Figure 3.

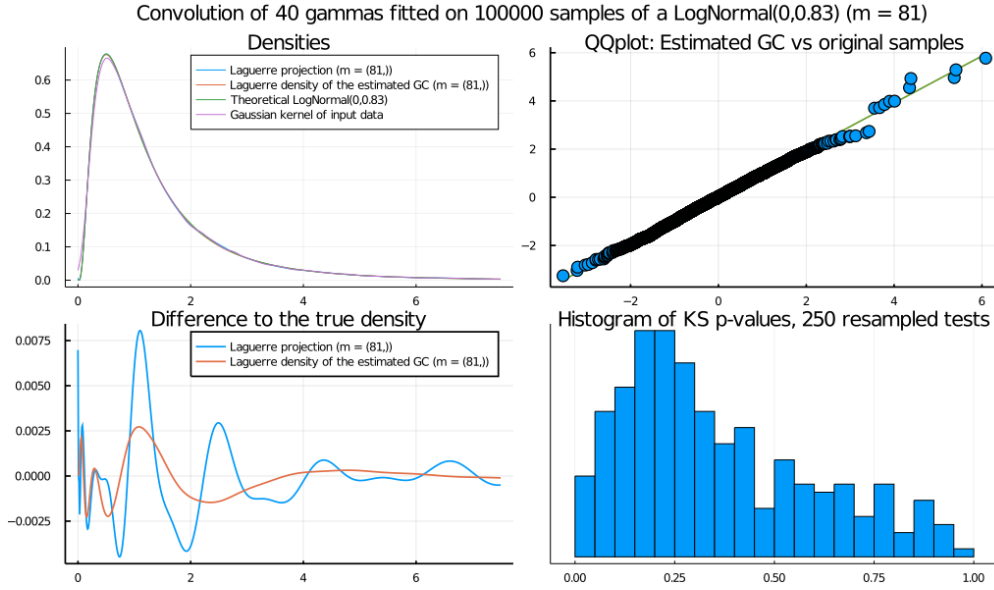


Figure 5: Log-Normal results with 40 gammas. Same legend as Figure 3.

We also treated several datasets simulated from Pareto distributions, with exactly the same experiment. Pareto distributions have a shape parameter $k > 0$ that influences the thickness of the tail: the distribution has a variance if and only if $k > 2$, an expectation if and only if $k > 1$ and the density belong to L_2 if and only if $k > 1/2$. Note that these inequalities are strict.

The fact that the distribution has no variance or no expectation is not really a problem for our procedure, since we use only shifted moments that all Pareto distributions have, but our convergence results requires that the density is square-integrable. Therefore, we propose to check several values of the shape parameter k , covering all cases, while varying the number n of gammas. Figure 6 summarizes the results by only showing quantile-quantile plots of the estimated distributions against the starting Pareto distributions. All the models

follow the same procedure as for the previous log-Normal estimators.

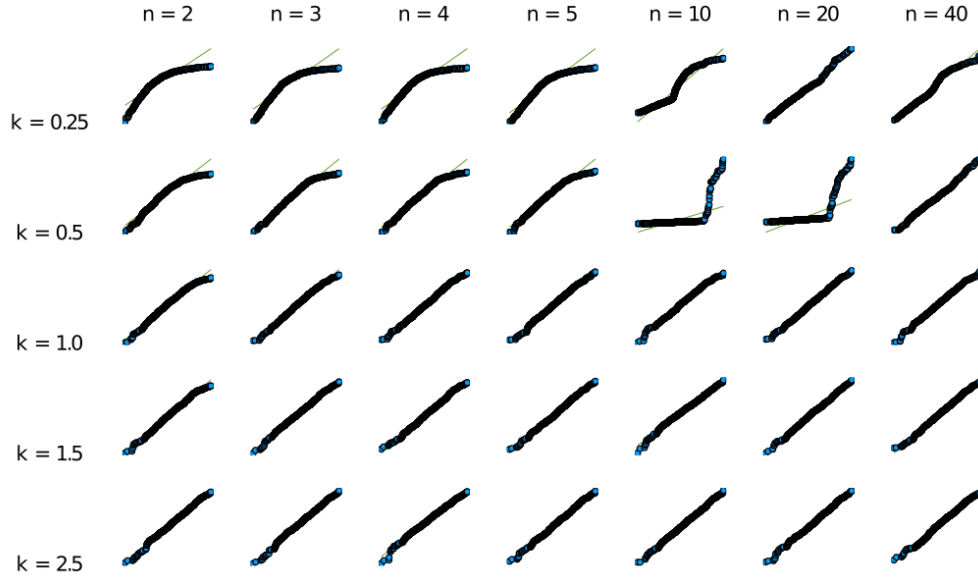


Figure 6: Quantile-quantile plots for Pareto experiments, with $N = 1000$ samples (log-scaled). We only show the 995 first points: 5 points are excluded in the tail for clarity. Each row correspond to a shape parameter for the Pareto, and each column to a number of gammas.

Note that even quantile-quantile plots between two simulations from the same Pareto distribution have a tendency to deviate in the extremes (due to the fat tail), which is why we excluded the last points. In each case, we also ran the same Kolmogorov-Smirnov testing procedure as for the log-Normal, and obtained very good uniformity of the p-values, except for $k = 0.25$ when n is too small. We see on the graph that increasing the number of gammas usually produces better results, but also that the heavier the tail (smaller shapes) the harder it is to get a good fit. Note that the KS distance is not the best distance to take into account the tails, which could explain why the distances from Figure 1 did not decrease for high values of n .

We see on these graphs that the method is accurate enough for reasonably high quantiles, and very good in the body of the distribution. The $k = 0.25$ case is not in the theoretical boundaries for the convergence result to take place, and from sample to sample we have to wait until at least 20 gammas to have an acceptable fit.

4.2.2 An example outside the class: the Weibull case

We revisit the Weibull($k = 1.5$) case, a distribution that is not inside the class.

We ran our algorithm on 100'000 samples from a Weibull with shape $k = \frac{3}{2}$, with successive number of gammas of 2, 10, 20, and 40. The corresponding graphs are shown in Figure 7. Here, we show only the difference on the density estimation and the quantile-quantile plot as one notable thing is that the Laguerre expansion of the density gets better and better as we augment the size of the basis (jointly with the number of gammas), but the estimator does not: the difference to the density is always the same, and always makes Kolmogorov-Smirnov reject the proposed estimator, although quantile-quantile plots are all OK.

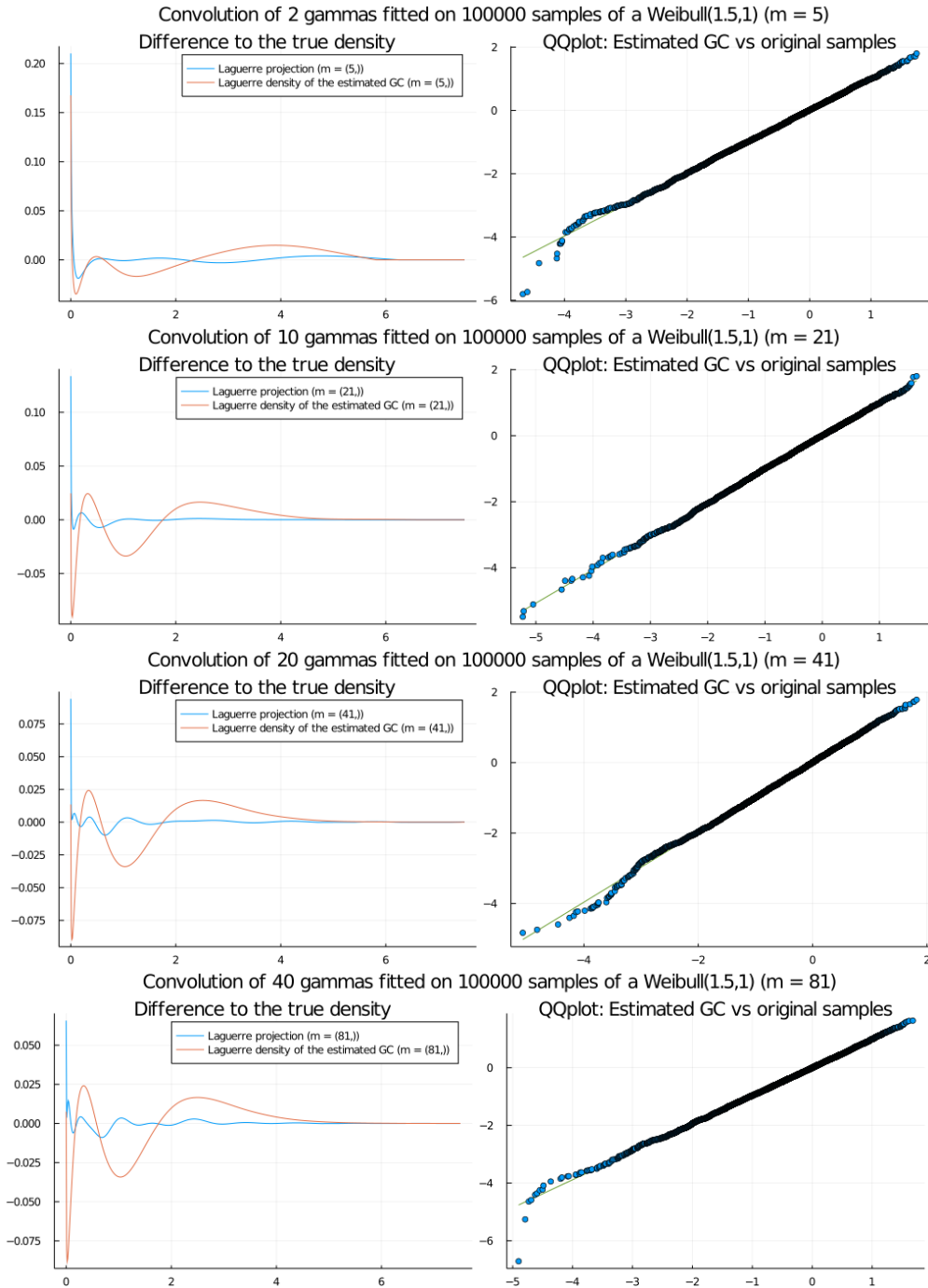


Figure 7: Weibull results with 2, 10, 20 and 40 gammas, respectively on each row. Same legend as Figure 3.

On these estimations, several things should be noted:

- The KS tests reject the hypothesis that the distributions are the same whatever the number of gamma, which was expected.
- The pointwise difference on the density of the estimator does not match the error of the Laguerre expansion (which is getting smaller as the basis grows) but stays the same whatever the size of the basis and the number of gammas. This error represents some kind of distance between the Weibull

distribution and the class.

- The quantile-quantile plot is on the other hand getting better and better as we add more gammas.

These three points shows that our procedure works even for data that are far from the \mathcal{G}_1 class, finding some orthogonal projection, relative to the basis, onto the class. This is a very good point for the procedure.

In the next examples, we will showcase some multivariate uses of the procedure, with a particular emphasis on the dependence structure of the simulated data and of the produced estimators.

4.3 Estimation from empirical data: multivariate case

Now that we extensively discussed the univariate capabilities of our algorithm, we turn ourselves to the multivariate case. As no known algorithm estimates multivariate gamma convolutions, we are not able to compare our results to other procedures. However, we can still check quantile-quantile plots for marginals and verify that the shape of the dependence structure matches the inputted one.

As use-cases of our models might vary, we propose two different examples: A multivariate log-Normal, with a Gaussian copula that induces no tail dependency, and a (survival) Clayton copula, including a strong tail dependency. Note that we reduced the exposition to estimation from empirical data, but formal integration to obtain $(\mu_{\mathbf{k},-1})_{\mathbf{k} \leq \mathbf{m}}$ from a formal density at any desired computational precision is also a perfectly fine use case.

4.3.1 The Multivariate log-Normal case

We do not know if the d -variate log-Normal distribution, defined as the componentwise exponentiation of a d -variate Gaussian random vector, is inside \mathcal{G}_d . However, we can still run our procedure and observe the resulting approximations.

A first result is given in Figure 8, that corresponds to a $\mathcal{G}_{2,10}$ estimation taken from 100'000 data points simulated from a $MLN(\mu = \mathbf{0}, \sigma_1 = \sigma_2 = 1, \rho = 0.5)$ with two standard log-Normal margins and a Gaussian copula with $\rho = 0.5$. The Figure 8 proposes quantile-quantile plots for the marginals, and a simple Gaussian kernel density estimator for the pseudo-observation on the original and estimated models.

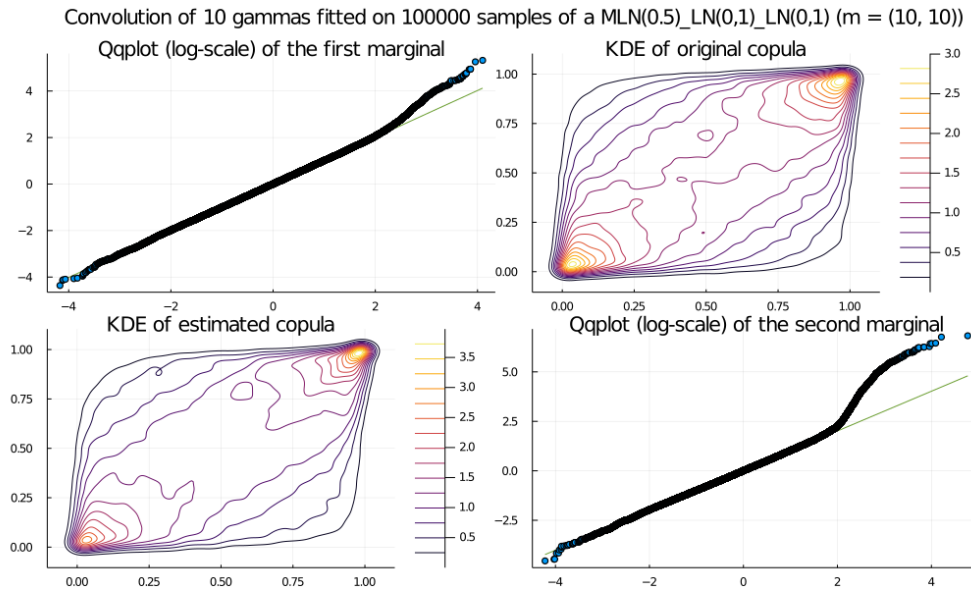


Figure 8: Multivariate log-Normal results with 10 gammas. Same legend as Figure 3.

We see on Figure 8 that the dependence structure was correctly reproduced by the approximation as a

convolution of 10 (comonotonous) gamma random vectors, but the marginal tails are not estimated correctly. Thankfully, pushing the number of multivariate gammas to $n = 20$ solves this issue, as shown in Figure 9:

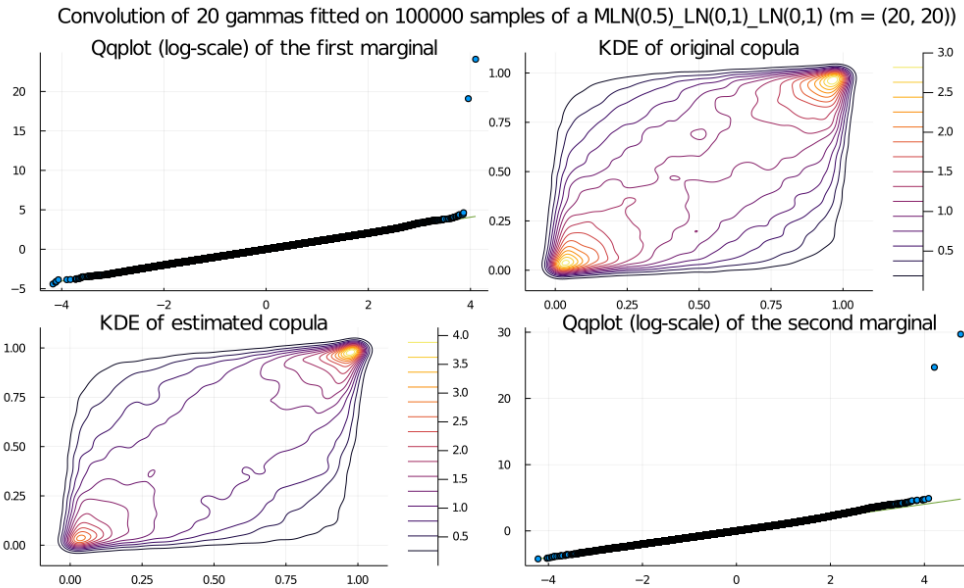


Figure 9: Multivariate log-Normal results with 20 gammas. Same legend as Figure 3.

On Figure 9, we see that the dependence structure is still very good, and that the problem of the marginal tails has been solved by adding a few more gammas. Note that since we have 100'000 points in the quantile-quantile plots, one point out of the line corresponds to a deviation at the 0.00001-quantiles, which are of little importance in many use cases.

Overall, marginals quantile-quantile plots are corrects when the number of gammas is big enough, and the Gaussian dependence structure is well fitted. The Gaussian copula is a copula that exhibit no tail dependency. We already know that our model can exhibit asymmetric behaviors, but the possibility of tail dependency property is also an important feature to have. Therefore, in the next example we run the model on datasets that have such a feature.

4.3.2 Tail dependency: a Clayton example

To convince ourselves that tail dependency is a property that our estimator can exhibit, we will try to fit data which dependency structure is given through a (survival) Clayton copula. This copula is a symmetric copula that exhibit upper tail dependency. We took here a parameter $\theta = 7$, yielding a very high upper tail dependence.

In Figure 10, you can see the results corresponding to data sampled from a Clayton copula, a Pareto($k = 2.5$) marginal and a $LN(0, 0.83)$ marginal. The conditions of the experiments are identically the previous subsection.

Convolution of 20 gammas fitted on 100000 samples of a Clayton(7)_Par(2.5,1)_LN(0,0.83) ($m = (20, 20)$)

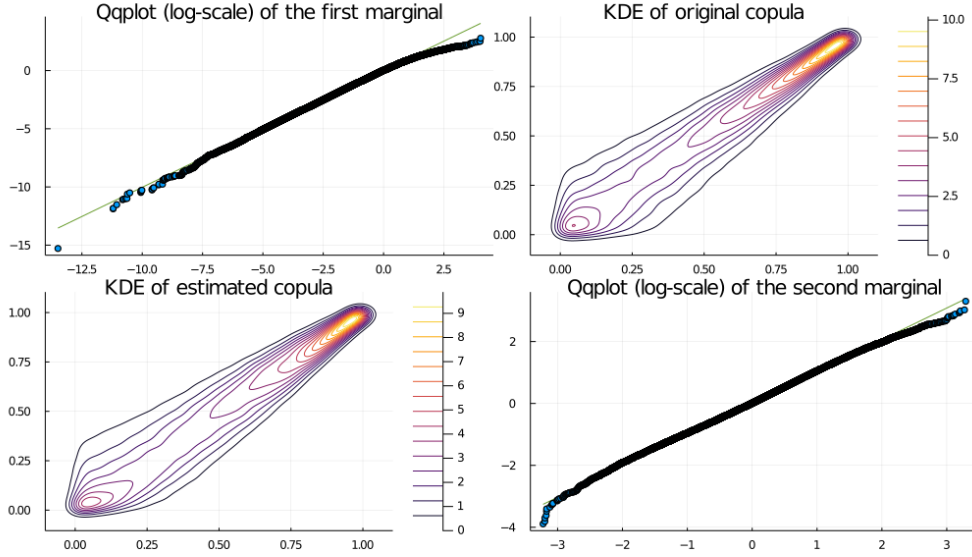


Figure 10: Results with 20 gammas. Same legend as Figure 3.

We note that the marginal upper tails are not perfectly fitted: this could probably be overcome by taking a greater number of gammas. However, we see that the shape of the estimated copula is not perfectly symmetric: there is no constraint on the model to produce a symmetric dependence structure, and therefore producing one is hard. The small misfit around the $(0.25, 0.02)$ quantile is probably due to this: the marginals are not on the same scale, and therefore the Laguerre coefficients with $k_1 \gg k_2$ are probably on a different scale than those with $k_2 \gg k_1$. The goodness of fit is therefore not symmetric.

5 Conclusion

Although the class of gamma convolutions has very appealing properties for the practitioner, the estimation of gamma convolutions from empirical data had not been considered in the literature. The closest algorithm we could find is the recent MFK algorithm, that we described in Subsection 2.3, which projects densities from the \mathcal{G}_1 class to the $\mathcal{G}_{1,n}$ class. Unfortunately, this algorithm cannot be extended for general estimation purposes.

Using a well-chosen Laguerre basis, we constructed a series expansion for the density of a distribution in $\mathcal{G}_{d,n}$. Coupled with a simple L_2 loss for the density, we were able to project densities from \mathcal{G}_d , $d > 1$ onto $\mathcal{G}_{d,n}$, which was not possible, but also to project any density onto $\mathcal{G}_{d,n}$, which was not possible either. Finally, through the same loss, we can estimate $\mathcal{G}_{d,n}$ models for any reasonable d, n , which was also not treated in the literature.

Although the convergence result is not as strong as MFK's, we saw that given the same information, shifted moments of the random variable with high precision, the performance of the two algorithms is overall the same, except maybe for small number of gammas where we beat MFK. For both algorithms, Figure 1 also shows that the fitness of the model seems to be decreasing for large numbers of gammas, and an adaptive estimator might be a suitable way out.

It is possible that a better loss can be found for estimation of these gamma convolutions. Furthermore, automatic penalization of the model could be done through the number of gammas, but also through sparsity of the Thorin measure, according to Property 2.1.4, yielding more independent factors when possible. Such a modification will also provide a clearer view at the additive risk factor structure.

Last but not least, many things are still to be discovered about multivariate Thorin classes. The estimation of multiplicative structures could yield interesting results, as Bondesson showed that the product of random

variables in \mathcal{G}_1 is also in \mathcal{G}_1 , but we still do not have a corresponding result for the multivariate case.

References

- [1] Rida Benhaddou, Marianna Pensky, and Rasika Rajapakshage. “Anisotropic Functional Laplace Deconvolution”. en. In: *Journal of Statistical Planning and Inference* 199 (Mar. 2019), pp. 271–285. ISSN: 03783758.
- [2] A. Bernardini, P. Natalini, and P.E. Ricci. “Multidimensional Bell Polynomials of Higher Order”. en. In: *Computers & Mathematics with Applications* 50.10-12 (Nov. 2005), pp. 1697–1708. ISSN: 08981221.
- [3] Lennart Bondesson. “A Class of Probability Distributions That Is Closed with Respect to Addition as Well as Multiplication of Independent Random Variables”. en. In: *Journal of Theoretical Probability* 28.3 (Sept. 2015), pp. 1063–1081. ISSN: 0894-9840, 1572-9230.
- [4] Lennart Bondesson. *Generalized Gamma Convolutions and Related Classes of Distributions and Densities*. en. Ed. by J. Berger et al. Vol. 76. Lecture Notes in Statistics. New York, NY: Springer New York, 1992. ISBN: 978-0-387-97866-6 978-1-4612-2948-3.
- [5] Lennart Bondesson. “On Simulation from Infinitely Divisible Distributions”. en. In: *Advances in Applied Probability* 14.4 (Dec. 1982), pp. 855–869. ISSN: 0001-8678, 1475-6064.
- [6] Lennart Bondesson. “On Univariate and Bivariate Generalized Gamma Convolutions”. en. In: *Journal of Statistical Planning and Inference* 139.11 (Nov. 2009), pp. 3759–3765. ISSN: 03783758.
- [7] KC Cherian. “A Bivariate Correlated Gamma-Type Distribution Function”. In: *Journal of the Indian Mathematical Society* 5 (1941), pp. 133–144.
- [8] Fabienne Comte and Valentine Genon-Catalot. “Adaptive Laguerre Density Estimation for Mixed Poisson Models”. en. In: *Electronic Journal of Statistics* 9.1 (2015), pp. 1113–1149. ISSN: 1935-7524.
- [9] G. M. Constantine and T. H. Savits. “A Multivariate Faà Di Bruno Formula with Applications”. en. In: *Transactions of the American Mathematical Society* 348.2 (1996), pp. 503–520. ISSN: 0002-9947, 1088-6850.
- [10] Hélène Cossette et al. “Moment-Based Approximation with Mixed Erlang Distributions”. In: *Variance* 10.1 (2016), pp. 161–182.
- [11] Djurdje Cvijović. “New Identities for the Partial Bell Polynomials”. en. In: *Applied Mathematics Letters* 24.9 (Sept. 2011), pp. 1544–1547. ISSN: 08939659.
- [12] M. S. Derevyagin and V. A. Derkach. “On the Convergence of Padé Approximations for Generalized Nevanlinna Functions”. en. In: *Transactions of the Moscow Mathematical Society* 68 (Oct. 2007), pp. 119–162. ISSN: 0077-1554, 1547-738X.
- [13] Philipp J di Dio and Konrad Schmüdgen. “The Multidimensional Truncated Moment Problem: The Moment Cone”. In: *arXiv preprint arXiv:1809.00584* (2018).
- [14] Philipp J. di Dio and Konrad Schmüdgen. “The Multidimensional Truncated Moment Problem: Atoms, Determinacy, and Core Variety”. en. In: *Journal of Functional Analysis* 274.11 (June 2018), pp. 3124–3148. ISSN: 00221236.
- [15] Florian Dussap. “Anisotropic Multivariate Deconvolution Using Projection on the Laguerre Basis”. In: (2020).
- [16] F Escher. “On the Probability Function in the Collective Theory of Risk”. In: *Skand. Aktuarie Tidskr.* 15 (1932), pp. 175–195.
- [17] Edward Furman, Daniel Hackmann, and Alexey Kuznetsov. “On Log-Normal Convolutions: An Analytical-Numerical Method With Applications to Economic Capital Determination”. en. In: *SSRN Electronic Journal* (2017). ISSN: 1556-5068.
- [18] Michael Hardy. “Combinatorics of Partial Derivatives”. en. In: *The Electronic Journal of Combinatorics* 13.1 (Jan. 2006), R1. ISSN: 1077-8926.
- [19] J. William Helton and Jiawang Nie. “A Semidefinite Approach for Truncated K-Moment Problems”. en. In: *arXiv:1105.0410 [math]* (Sept. 2012).
- [20] Didier Henrion and Jérôme Malick. “Projection Methods in Conic Optimization”. en. In: *Handbook on Semidefinite, Conic and Polynomial Optimization*. Ed. by Miguel F. Anjos and Jean B. Lasserre. Vol. 166. Boston, MA: Springer US, 2012, pp. 565–600. ISBN: 978-1-4614-0768-3 978-1-4614-0769-0.
- [21] Lars Hormander. *An Introduction to Complex Analysis in Several Variables*. Elsevier, 1973.

- [22] Chaoran Hu, Vladimir Pozdnyakov, and Jun Yan. “Density and Distribution Evaluation for Convolution of Independent Gamma Variables”. en. In: *Computational Statistics* 35.1 (Mar. 2020), pp. 327–342. ISSN: 0943-4062, 1613-9658.
- [23] Lancelot F. James, Bernard Roynette, and Marc Yor. “Generalized Gamma Convolutions, Dirichlet Means, Thorin Measures, with Explicit Examples”. en. In: *Probability Surveys* 5.0 (2008), pp. 346–415. ISSN: 1549-5787.
- [24] M. M. Kabardov and V. M. Ryabov. “Acceleration of the Convergence of the Laguerre Series in the Problem of Inverting the Laplace Transform”. en. In: *Computational Mathematics and Mathematical Physics* 49.4 (Apr. 2009), pp. 579–588. ISSN: 0965-5425, 1555-6662.
- [25] Taekyun Kim, Dae Kim, and Gwan-Woo Jang. “On Central Complete and Incomplete Bell Polynomials I”. en. In: *Symmetry* 11.2 (Feb. 2019), p. 288. ISSN: 2073-8994.
- [26] Hasan Koyuncu and Rahime Ceylan. “A PSO Based Approach: Scout Particle Swarm Algorithm for Continuous Global Optimization Problems”. en. In: *Journal of Computational Design and Engineering* 6.2 (Apr. 2019), pp. 129–142. ISSN: 2288-5048.
- [27] Annika Krutto. “Empirical Cumulant Function Based Parameter Estimation in Stable Laws”. en. In: *Acta et Commentationes Universitatis Tartuensis de Mathematica* 22.2 (Jan. 2019), pp. 311–338. ISSN: 2228-4699, 1406-2283.
- [28] Simon Lee and X Sheldon Lin. “Modeling Dependent Risks with Multivariate Erlang Mixtures”. In: *Astin Bulletin* 42.1 (2012), pp. 153–180.
- [29] Gwennaëlle Mabon. “Adaptive Deconvolution on the Non-Negative Real Line: Adaptive Deconvolution on \mathbb{R}^+ ”. en. In: *Scandinavian Journal of Statistics* 44.3 (Sept. 2017), pp. 707–740. ISSN: 03036898.
- [30] A. M. Mathai. “Storage Capacity of a Dam with Gamma Type Inputs”. en. In: *Annals of the Institute of Statistical Mathematics* 34.3 (Dec. 1982), pp. 591–597. ISSN: 0020-3157, 1572-9052.
- [31] Alexander Meister. *Deconvolution Problems in Nonparametric Statistics*. en. Vol. 193. Lecture Notes in Statistics. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009. ISBN: 978-3-540-87556-7 978-3-540-87557-4.
- [32] Vadim S Melman et al. “Realization of a Method for Calculating Bell Polynomials Based on Compositae of Generating Functions”. en. In: 10.4 (2018), p. 14.
- [33] Jerry M. Mendel. *Maximum-Likelihood Deconvolution*. en. Signal Processing and Digital Filtering. New York, NY: Springer New York, 1990. ISBN: 978-1-4612-7985-3 978-1-4612-3370-1.
- [34] Filippo M. Miatto. “Recursive Multivariate Derivatives of $\{f(X_1, \dots, X_n)\}$ of Arbitrary Order”. en. In: *arXiv:1911.11722 [cs, math]* (Nov. 2019).
- [35] Justin Miles, Edward Furman, and Alexey Kuznetsov. “Risk Aggregation: A General Approach via the Class of Generalized Gamma Convolutions”. In: *Variance* (2019).
- [36] Rumen L. Mishkov. “Generalization of the Formula of Faà Di Bruno for a Composite Function with a Vector Argument”. en. In: *International Journal of Mathematics and Mathematical Sciences* 24.7 (2000), pp. 481–491. ISSN: 0161-1712, 1687-0425.
- [37] P. G. Moschopoulos. “The Distribution of the Sum of Independent Gamma Random Variables”. en. In: *Annals of the Institute of Statistical Mathematics* 37.3 (Dec. 1985), pp. 541–544. ISSN: 0020-3157, 1572-9052.
- [38] Hussein Mustapha and Roussos Dimitrakopoulos. “Generalized Laguerre Expansions of Multivariate Probability Densities with Moments”. en. In: *Computers & Mathematics with Applications* 60.7 (Oct. 2010), pp. 2178–2189. ISSN: 08981221.
- [39] E. Di Nardo, G. Guarino, and D. Senato. “A New Algorithm for Computing the Multivariate Faà Di Bruno’s Formula”. en. In: *Applied Mathematics and Computation* 217.13 (Mar. 2011), pp. 6286–6295. ISSN: 00963003.
- [40] Jiawang Nie. “The A-Truncated K-Moment Problem”. en. In: *arXiv:1210.6930 [math]* (Oct. 2012).
- [41] Frank Nielsen. “ k -MLE: A Fast Algorithm for Learning Statistical Mixture Models”. en. In: *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (Mar. 2012), pp. 869–872.
- [42] Victor Pérez-Abreu and Robert Stelzer. “A Class of Infinitely Divisible Multivariate and Matrix Gamma Distributions and Cone-Valued Generalised Gamma Convolutions”. en. In: *arXiv:1201.1461 [math, stat]* (Jan. 2012).

- [43] Victor Pérez-Abreu and Robert Stelzer. “Infinitely Divisible Multivariate and Matrix Gamma Distributions”. en. In: *Journal of Multivariate Analysis* 130 (Sept. 2014), pp. 155–175. ISSN: 0047259X.
- [44] M. S. Ridout. “Generating Random Numbers from a Distribution Specified by Its Laplace Transform”. en. In: *Statistics and Computing* 19.4 (Dec. 2009), pp. 439–450. ISSN: 0960-3174, 1573-1375.
- [45] Konrad Schmudgen. *The Moment Problem*. en. Vol. 277. Graduate Texts in Mathematics. Cham: Springer International Publishing, 2017. ISBN: 978-3-319-64545-2 978-3-319-64546-9.
- [46] Olivier Schwander and Frank Nielsen. “Fast Learning of Gamma Mixture Models with K-MLE”. en. In: *Similarity-Based Pattern Recognition*. Ed. by David Hutchison et al. Vol. 7953. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 235–249. ISBN: 978-3-642-39139-2 978-3-642-39140-8.
- [47] Kevin D. Smith. “A Tutorial on Multivariate k -Statistics and Their Computation”. en. In: *arXiv:2005.08373 [math, stat]* (May 2020).
- [48] Peter J Smith. “A Recursive Formulation of the Old Problem of Obtaining Moments from Cumulants and Vice Versa”. en. In: (2020), p. 3.
- [49] David A Sprott. “Estimating the Parameters of a Convolution by Maximum Likelihood”. In: *Journal of the American Statistical Association* 78.382 (1983), pp. 457–460.
- [50] Olof Thorin. “On the Infinite Divisibility of the Pareto Distribution”. en. In: *Scandinavian Actuarial Journal* 1977.1 (Jan. 1977), pp. 31–40. ISSN: 0346-1238, 1651-2030.
- [51] Olof Thorin. “On the Infinite Divisibility of the Lognormal Distribution”. en. In: *Scandinavian Actuarial Journal* 1977.3 (Mar. 1977), pp. 121–148. ISSN: 0346-1238, 1651-2030.
- [52] Henk C Tijms. *A First Course in Stochastic Models*. John Wiley and sons, 2003.
- [53] CS Withers. “Expressions for Moments and Cumulants in Terms of Each Other Using Bell Polynomials”. In: *Proceedings of the 1994 New Zealand Statistical Association Conference, Massey University*. 1994, pp. 226–229.
- [54] Yongliang Zhang, Yu Chen, and David Day-Uei Li. “Optimizing Laguerre Expansion Based Deconvolution Methods for Analysing Bi-Exponential Fluorescence Lifetime Images”. en. In: *Optics Express* 24.13 (June 2016), p. 13894. ISSN: 1094-4087.
- [55] Zong-Jun Li. “Experiment of Standard Particle Swarm Optimization 2011 in Julia on the CEC2013 Benchmark Functions”. en. In: (2016).