Internal Modeling without copulas: the beauty of multivariate Thorin classes

O. Laverny [1,2]
E. Masiello [1], V. Maume-Deschamps [1] and D. Rullière [3]

[1] Institut Camille Jordan, UMR 5208, Université Claude Bernard Lyon 1, Lyon, France

[2] SCOR SE

[3] Mines Saint-Etienne, UMR CNRS 6158, LIMOS, Saint-etienne, France

June 23, 2022

Table of contents

Introduction and motivations

Introduction

A reinsurer such as SCOR observes many (positives) random variables that represent losses from several portfolios of risks.

A major topic is the evaluation and modeling of the dependency between those risks. Indeed, the properties of this dependency have a huge impact on the behavior of the total loss, and therefore on the capital management of the company.

Risks can be heavy-tailed, skewed, and are usually dependent with each other, including in the extremes.

The practical goal of our work was to produce relevant models of dependence structures targeted at the analysis of (re-)insurance portfolios.

Multivariate generalized Gamma convolutions

Multivariate generalized Gamma convolutions

Multivariate Gamma convolutions classes $\mathcal{G}_{d,n}$ and $\mathcal{G}_d$, motivations

Multivariate Gamma convolutions classes $\mathcal{G}_{d,n}$ and $\mathcal{G}_d$

Recall: Cumulant generating function: $K(\boldsymbol{t}) = \ln M(\boldsymbol{t}) = \ln\left(\mathbb{E}\left(e^{\langle \boldsymbol{t}, \boldsymbol{X}\rangle}\right)\right)$.

Definition (Multivariate Thorin Classes[1])

$\boldsymbol{X} \sim \mathcal{G}_{d,n}(\boldsymbol{\alpha}, \boldsymbol{s}) \Leftrightarrow K(\boldsymbol{t}) = -\sum_{i=1}^{n} \alpha_i \ln\left(1 - \langle \boldsymbol{s}_i, \boldsymbol{t}\rangle\right)$
$\boldsymbol{X} \sim \mathcal{G}_d(\nu) \Leftrightarrow K(\boldsymbol{t}) = -\int \ln\left(1 - \langle \boldsymbol{s}, \boldsymbol{t}\rangle\right) \nu(\partial \boldsymbol{s})$.

Property

$\mathcal{G}_1$ *is closed w.r.t (independent) sums and products of random variables !*

Example

All gammas, log-Normals, Paretos, α-stables, and sums and/or products of these are in $\mathcal{G}_1$.

No estimation procedures were available before our work.

[1] 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.

Motivations: interpretability of $\mathcal{G}_{d,n}$ models

Remark ($\boldsymbol{X} \in \mathcal{G}_{d,n}$ additive risk-factor representation)

$$\begin{pmatrix} X_1 \\ \dots \\ X_d \end{pmatrix} = \begin{pmatrix} s_{1,1} & \dots & \dots & s_{1,n} \\ \dots & \dots & \dots & \dots \\ s_{d,1} & \dots & \dots & s_{d,n} \end{pmatrix} \cdot \begin{pmatrix} G_1 \\ \dots \\ G_n \end{pmatrix}$$

(i) For all $i \in 1, ..., n$, $G_i \sim \mathcal{G}_{1,1}(\alpha_i, 1)$

(ii) $G_1, ..., G_n$ are independent.

Since $s_{i,j}$ can always be zero, by increasing n (typically $n >> d$) and using the infinite divisibility, we can approach any marginal in $\mathcal{G}_1$, and we have a wide variety of dependence structures (asymmetry, tail dependency, etc.).

Goal: Estimate ν from observations of the random vector $\boldsymbol{X}$.

Pb: It is a deconvolution problem, which is numerically hard (can also be interpreted as a solution-less moment problem for the Thorin measure, see MFK[2].)

[2] Justin Miles, Edward Furman, and Alexey Kuznetsov. "Risk Aggregation: A General Approach via the Class of Generalized Gamma Convolutions". In: *Variance* (2019).

Motivations: infinite divisibility of $\mathcal{G}_{d,n}$ models

Gamma distributions are divisible, and therefore so are their convolutions.

Definition (ID r.v. X)

(i) X is n-divisible if and only if there exists $X_1, ...X_n$ i.i.d. such that $X = X_1 + ... + X_n$

(ii) X is *ID* iff this can be done for all n.

(iii) X_i is called a *piece* of X.

(i) Dependencies are largely unobservable in reinsurance;

(ii) Experts are providing information as divisibility parameters: "20% of this is driven by the same risk factors as 10% of that..." *this means comonotone pieces across marginals.*

Multivariate generalized Gamma convolutions

The orthonormal Laguerre basis of $L_2(\mathbb{R}_+^d)$

The orthonormal Laguerre basis of $L_2(\mathbb{R}_+^d)$

Definition (Laguerre basis, Comte[3], Mabon[4] and Dussap[5])

For all $\boldsymbol{p} \in \mathbb{N}^d$, $\varphi_{\boldsymbol{p}}(\boldsymbol{x}) = \prod_{i=1}^{d} \varphi_{p_i}(x_i)$ where $\varphi_p(x) = \sqrt{2} \sum_{k=0}^{p} \binom{p}{k} \frac{(-2x)^k}{k!} e^{-x}$.

These functions form an orthonormal basis of $L_2(\mathbb{R}_+^d)$.

Therefore, every density f that is square-integrable can be expanded as :

$$f(\boldsymbol{x}) = \sum_{\boldsymbol{p} \in \mathbb{N}^d} a_{\boldsymbol{p}} \varphi_{\boldsymbol{p}}(\boldsymbol{x}) \text{ where } a_{\boldsymbol{p}} = \int \varphi_{\boldsymbol{p}}(\boldsymbol{x}) f(\boldsymbol{x}) d\boldsymbol{x}$$

Integrated square error loss: $\mathcal{L}(\alpha, \boldsymbol{s}) = \sum_{\boldsymbol{k} \leq \boldsymbol{m}} \left(\widehat{a_{\boldsymbol{k}}} - a_{\boldsymbol{k}}(\alpha, \boldsymbol{s})\right)^2$

[3] 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.

[4] Gwennaëlle Mabon. "Adaptive Deconvolution on the Non-Negative Real Line: Adaptive Deconvolution on R+". en. In: *Scandinavian Journal of Statistics* 44.3 (Sept. 2017), pp. 707–740. ISSN: 03036898.

[5] Florian Dussap. "Anisotropic multivariate deconvolution using projection on the Laguerre basis". In: *Journal of Statistical Planning and Inference* 215 (2021), pp. 23–46.

Definition (ε-well-behaved $\mathcal{G}_{d,n}(\boldsymbol{\alpha}, \boldsymbol{s})$)

$\mathcal{G}_{1,n}(\boldsymbol{\alpha}, \boldsymbol{s})$ is ε-w.b. $\iff |\boldsymbol{\alpha}| > 1$ and $\boldsymbol{s} \in \mathbf{0} \cup]\frac{\varepsilon}{2+\varepsilon}, \frac{2+\varepsilon}{\varepsilon}[^n$.

$\mathcal{G}_{d,n}(\boldsymbol{\alpha}, \boldsymbol{s})$ is ε-w.b. $\iff$ [Technical, useful but boring definition]

Property (well-behaved $\mathcal{G}_{d,n}(\boldsymbol{\alpha}, \boldsymbol{s})$)

$\mathcal{G}_{d,n}(\boldsymbol{\alpha}, \boldsymbol{s})$ is w.b. $\iff |\boldsymbol{\alpha}| > 1$ and $\forall I \subseteq \{1, \ldots, n\}$ such that $\sum_{i \in I} \alpha_i > \sum_{i \notin I} \alpha_i$, $Ker(\boldsymbol{s}_I) = \{\mathbf{0}\}$.

Example (Simple w.b. examples)

As soon as $|\nu| = |\boldsymbol{\alpha}| > 1$, the following are well-behaved:

(i) All univariate gamma convolution

(ii) All $\mathcal{G}_{d,n}$ with independent marginals and all invertible linear transformation of them.

(iii) All finite convolution of well-behaved gamma convolutions.

Analysis of the loss: Uniform exponential bound

Property (Exponential decay of Laguerre coefficients)

For any $\varepsilon' > 0$ and any dimension d, $\exists\, B(d, \varepsilon') \in]0, +\infty[$, such that Laguerre coefficients $(a_{\boldsymbol{k}})_{\boldsymbol{k} \in \mathbb{N}^d}$ of any d-variate ε-well-behaved gamma convolution, $\varepsilon > \varepsilon'$, verify:

$$|a_{\boldsymbol{k}}| \leq B(d, \varepsilon')(1 + \varepsilon')^{-|\boldsymbol{k}|}.$$

Corollary (Consistency)

The ISE loss $\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{s}) = \sum_{\boldsymbol{k} \leq \boldsymbol{m}} (\widehat{a}_{\boldsymbol{k}} - a_{\boldsymbol{k}}(\boldsymbol{\alpha}, \boldsymbol{s}))^2$ produces consistent estimators of $(\boldsymbol{\alpha}, \boldsymbol{s})$.

Proofs leverage analytics combinatorics is several variables on the function $R = exp \circ K \circ \boldsymbol{h}$, which happens to be the generating function of Laguerre coefficients, where $\boldsymbol{h}$ is a component wise Möbius transform.

Multivariate generalized Gamma convolutions

Investigations

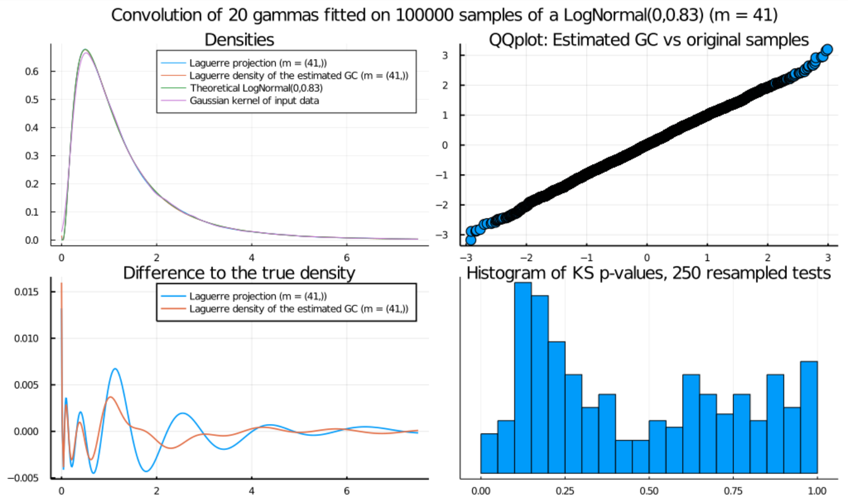


Figure 1: Results with 20 gammas.

Investigations: Multivariate log-Normal

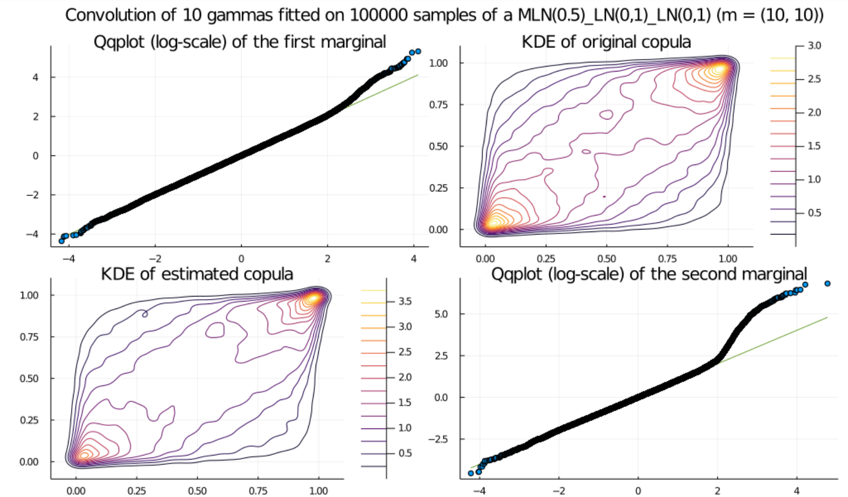


Figure 2: MLN case: $n = 10$

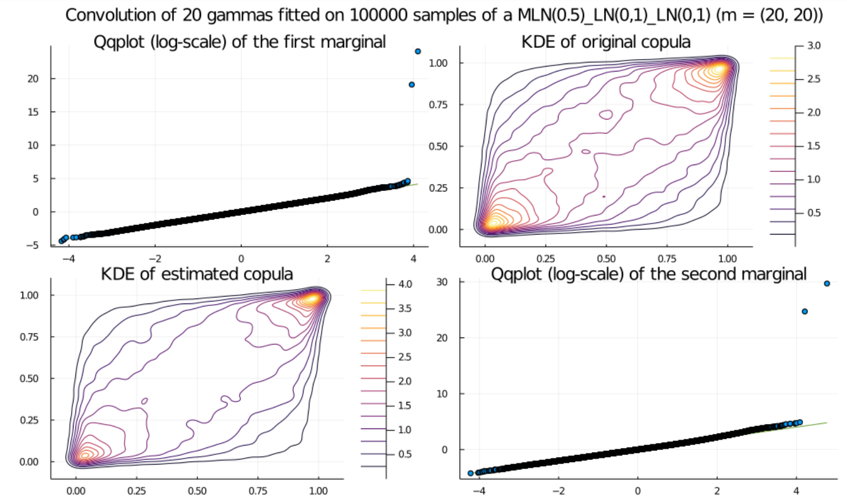


Figure 3: MLN case: $n = 20$

Tail dependency: A Clayton copula

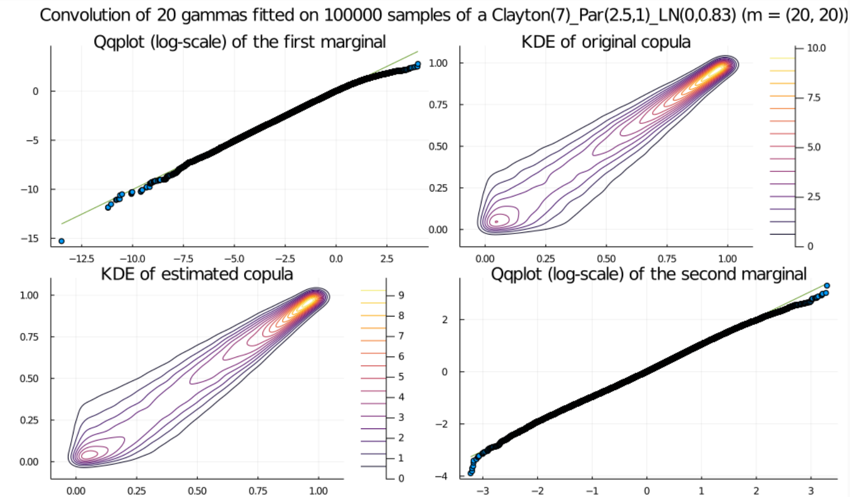


Figure 4: Clayton case: $n = 20$

Multivariate generalized Gamma convolutions

Take away

Take away (1/2)

(i) The univariate Thorin class is wide: LN, Pareto, α-stable, (some) Weibull...

(ii) The Multivariate analogue provides a flexible dependence structure.

(iii) Their estimation is a deconvolution, a hard inverse problem.

(iv) The final additive risk-factor model gives easy interpretation of parameters and easy aggregation schemes.

(v) A better approach for high dimensional cases might be possible...

Details and other illustrated applications are available in the paper[6] *and the open-source Julia package*[7]*.*

---

[6] OL, Esterina Masiello, Véronique Maume-Deschamps, and Didier Rullière. "Estimation of Multivariate Generalized Gamma Convolutions through Laguerre Expansions.". In: *Electronic Journal of Statistics* 15.2 (Jan. 2021). ISSN: 1935-7524.

[7] OL. *ThorinDistributions.jl*. Version v0.1. Mar. 2021.

High-dimensional Thorin measures

High-dimensional Thorin measures

Shifted moments, shifted cumulants and Laguerre coefficients

Another look at the Laguerre coefficients

Recall: $a_{\boldsymbol{k}} = \langle \varphi_{\boldsymbol{k}}, f \rangle = \mathbb{E}\left(\varphi_{\boldsymbol{k}}(\boldsymbol{X})\right) = \sqrt{2}^d \sum_{\boldsymbol{k} \leq \boldsymbol{p}} \binom{\boldsymbol{p}}{\boldsymbol{k}} \frac{(-2)^{|\boldsymbol{k}|}}{\boldsymbol{k}!} \mathbb{E}\left(X^{\boldsymbol{k}} e^{-|\boldsymbol{X}|}\right)$

Denote: $I \subset \mathbb{N}^d$ be an increasing index set, i.e., $\boldsymbol{j} \in I \implies \boldsymbol{i} \in I$ if $\boldsymbol{i} \leq \boldsymbol{j}$.

Definition (Shifted moments $\boldsymbol{\mu} = (\mu_{\boldsymbol{i}})_{\boldsymbol{i} \in I}$)

Denote $\mu_{\boldsymbol{k}} := \mathbb{E}\left(X^{\boldsymbol{k}} e^{-|\boldsymbol{X}|}\right)$. The mgf of the random vector writes:

$$M(\boldsymbol{t}) = \mathbb{E}\left(e^{\langle \boldsymbol{t}, \boldsymbol{X} \rangle}\right) = \sum_{\boldsymbol{k} \in \mathbb{N}^d} \mu_{\boldsymbol{k}} \frac{(\boldsymbol{t} - \boldsymbol{1})^{\boldsymbol{k}}}{\boldsymbol{k}!},$$

Bijection: There exists a left triangular and invertible matrix $\boldsymbol{A}$ s.t.

$$\boldsymbol{a} = (a_{\boldsymbol{i}})_{\boldsymbol{i} \in I} = \boldsymbol{A}\boldsymbol{\mu} \text{ and } \boldsymbol{\mu} = \boldsymbol{A}^{-1}\boldsymbol{a}.$$

Thorin's moments and their estimators

Definition (Thorin moments $\boldsymbol{\tau} = (\tau_{\boldsymbol{i}})_{\boldsymbol{i}\in I}$)

$$K(\boldsymbol{t}) = \ln M(\boldsymbol{t}) := \sum_{\boldsymbol{k}\in\mathbb{N}^d} \tau_{\boldsymbol{k}} \left(|\boldsymbol{k}| - 1\right)! \frac{(\boldsymbol{t}-\boldsymbol{1})^{\boldsymbol{k}}}{\boldsymbol{k}!}.$$

Bijection: There exists a function $\boldsymbol{B}$, based on Bell polynomials, s.t.

$$\boldsymbol{\mu} = \boldsymbol{B}(\boldsymbol{\tau}) \text{ and } \boldsymbol{\tau} = \boldsymbol{B}^{-1}(\boldsymbol{\mu}).$$

Estimation: from data $\boldsymbol{x} = (\boldsymbol{x}_1, ..., \boldsymbol{x}_N) \in \mathbb{R}_+^{N\times d}$ an N-sample of i.i.d. random vectors:

$$\hat{\boldsymbol{\mu}}(\boldsymbol{x}) = (\widehat{\mu_{\boldsymbol{k}}}(\boldsymbol{x}))_{\boldsymbol{k}\in I} = \left(\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{x}_i^{\boldsymbol{k}} e^{-|\boldsymbol{x}_i|}\right)_{\boldsymbol{k}\in I}$$

$$\hat{\boldsymbol{\tau}}(\boldsymbol{x}) = (\widehat{\tau_{\boldsymbol{k}}}(\boldsymbol{x}))_{\boldsymbol{k}\in I} = \boldsymbol{B}^{-1}\left(\hat{\boldsymbol{\mu}}(\boldsymbol{x})\right).$$

$$\hat{\boldsymbol{a}}(\boldsymbol{x}) = (\widehat{a_{\boldsymbol{k}}}(\boldsymbol{x}))_{\boldsymbol{k}\in I} = \boldsymbol{A}\hat{\boldsymbol{\mu}}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{B}(\hat{\boldsymbol{\tau}}(\boldsymbol{x})),$$

Thorin's moments of a Gamma convolution

Definition (Thorin moments of a Gamma convolution $\mathcal{G}_d(\nu)$)

(i) $\boldsymbol{\tau}(\nu) = \int \boldsymbol{\tau}(\delta_{\boldsymbol{s}})\nu(d\boldsymbol{s})$ *(linearity on $\mathcal{M}_+(\mathbb{R}_+^d)$)*

(ii) $\tau_{\mathbf{0}}(\delta_{\boldsymbol{s}}) = -\ln(1 + |\boldsymbol{s}|)$ *(cgf of a simple gamma in $-\mathbf{1}$)*

(iii) $\tau_{\boldsymbol{k}}(\delta_{\boldsymbol{s}}) = \left(\frac{\boldsymbol{s}}{1+|\boldsymbol{s}|}\right)^{|\boldsymbol{k}|}$ for $\boldsymbol{k} \neq 0$, $\boldsymbol{k} \in \mathbb{N}^d$.

Rem: τ are called Thorin moments as they are generalized moments of the Thorin measure (shape or scale dilemma). We are dealing with a multivariate moment problem *which might not have a solution. . . .*

High-dimensional Thorin measures

Better approach via random projections.

An approximated loss through random projections...

Problem: the loss $\mathcal{L}(\boldsymbol{x},\nu) = \|\hat{\boldsymbol{a}}(\boldsymbol{x}) - \boldsymbol{AB}(\boldsymbol{\tau}(\nu))\|_2^2$, although consistent, is too costly to work with when d gets large. E.g., for $I = \left\{\boldsymbol{k} \in \mathbb{N}^d, |\boldsymbol{k}| \leq m\right\}$ (isotropic loss), the number of coefficients $|I|$ to compute is exponential in d at fixed precision m !

Solution 1/2: Through a first-order Taylor expansion of the function $\boldsymbol{AB}$, we define:

$$\widehat{\mathcal{L}}(\boldsymbol{x},\nu) = \|\hat{\boldsymbol{\tau}}(\boldsymbol{x}) - \boldsymbol{\tau}(\nu)\|_{\nabla(\boldsymbol{x})}^2,$$

where $\nabla(\boldsymbol{x})$ is a Jacobian of the function $\boldsymbol{AB}$, taken in $\hat{\boldsymbol{\tau}}(\boldsymbol{x})$.

Solution 2/2: Then, through a univariate projection and re-integration we define:

$$\widetilde{\mathcal{L}}(\boldsymbol{x},\nu) := \int_{[0,1]^d} \widehat{\mathcal{L}}\left(\langle \boldsymbol{c}, \boldsymbol{x}\rangle, \nu_{\langle \boldsymbol{c}\rangle}\right) d\boldsymbol{c},$$

where

$$\nu_{\langle \boldsymbol{c}\rangle}(A) = \nu\left(\left\{\boldsymbol{x} \in \mathbb{R}_+^d : \langle \boldsymbol{c}, \boldsymbol{x}\rangle \in A\right\}\right) \; \forall A \subseteq \mathbb{R}_+.$$

. . . that is still consistent.

Theorem (Consistency of $\widetilde{\mathcal{L}}$.)

Let $\boldsymbol{x}$ be drawn from a well-behaved density $f \sim \mathcal{G}_d(\nu)$. The global minimizer

$$\tilde{\nu} := \underset{\nu : \mathcal{G}_d(\nu)\, w.b.}{\arg\min}\ \widetilde{\mathcal{L}}(\boldsymbol{x}, \nu)$$

ensures that

$$\|f - f_{\tilde{\nu}}\|_2^2 \xrightarrow[\substack{N \to \infty \\ I \to \mathbb{N}^d}]{a.s} 0.$$

Same result as before for $\mathcal{L}$. The proof leverages the theory of Grassmannian cubatures and some considerations about unisolvant sets for polynomials of bounded degree.

Question: Can we and will we reach a global minimizer ? Recall that the loss is clearly not convex and has a myriad of local minima. . .

A convergence result for the gradient flow.

Theorem (Global convergence of the gradient flow $\frac{\partial}{\partial t}\nu_t = -\nabla\widetilde{\mathcal{L}}(\boldsymbol{x}, \nu_t)$.)

For $\rho \in \mathcal{M}_+(\mathbb{R}_+^d)$ an absolutely continuous reference measure such that $\log \rho$'s density is Lipschitz, for any initial measure $\nu_0 \in \mathcal{M}_+(\mathbb{R}_+^d)$, there exists a constant C, dependent on the characteristics of the problem, such that if $W_\infty(\nu_0, \rho) \leq C$,

$$\exists\, \nu_\infty \in \underset{\nu \in \mathcal{M}_+(\mathbb{R}_+^d)}{\arg\min}\ \widetilde{\mathcal{L}}(\boldsymbol{x}, \nu) \text{ such that } W_\infty(\nu_t, \nu_\infty) \xrightarrow[t\to\infty]{} 0.$$

Furthermore, when $\nu_0 = \rho$, we achieve a precision ϵ, i.e., $W_\infty(\nu_t, \nu_\infty) \leq \epsilon$, provided the number of iteration is $t = \mathcal{O}\left(-\log(\epsilon)\right)$.

The proof leverages Chizat[8] and De Castro & Al[9].

[8] Lenaic Chizat. "Sparse optimization on measures with over-parameterized gradient descent". In: *Mathematical Programming* (2021), pp. 1–46.

[9] Yohann de Castro, Sébastien Gadat, Clément Marteau, and Cathy Maugis. "SuperMix: Sparse Regularization for Mixtures". In: *The annals of Statistics* 49.3 (2021), pp. 1779–1809.

High-dimensional Thorin measures

Examples

Estimation procedure

Algorithm 1: Estimation of Thorin measures via stochastic gradient descent on $\widetilde{\mathcal{L}}$.

Input: A dataset $\boldsymbol{x} \in \mathbb{R}^{N\times d}$, a number of Gammas $n \in \mathbb{N}$, a precision parameter $m \in \mathbb{N}$, a number of iterations $T \in \mathbb{N}$, and a learning rate $\eta \in \mathbb{R}_+$

Result: A Thorin measure ν_T that approximates the dataset $\boldsymbol{x}$ as a multivariate Gamma convolution.

Estimate standard deviations $\sigma_i = \mathrm{std}(\boldsymbol{x}_i)$ for all $i \in 1, ...d$, and standardize the marginals by dividing $\boldsymbol{x}_i$ by σ_i.

Initialize a measure $\nu_0 \in \mathcal{M}_+(\mathbb{R}_+^d)$ with n atoms and corresponding weights, chosen randomly through Gaussian noise, according to Chizat.

foreach $t \in 0, ..., T-1$ **do**

- Choose a random direction $\boldsymbol{c} \in [0,1]^d$.
- Compute the gradient $\boldsymbol{g}$ of $\widehat{\mathcal{L}}(\langle \boldsymbol{c}, \boldsymbol{x}\rangle, \nu_{t,\langle \boldsymbol{c}\rangle})$ with respect to ν_t (details missing).
- Let $\nu_{t+1} = \nu_t - \eta \boldsymbol{g}$

end

Rescale ν_T by $\sigma_1, ...\sigma_d$.

Return ν_T

Four-variate simulated data

For U_1, U_2, U_3 independent $\mathcal{U}([0,1])$ distributions and $Y_1, \ldots, Y_4$ independent log-Normal(0,1) distributions, we let the random vector $\boldsymbol{X}$ be defined by:

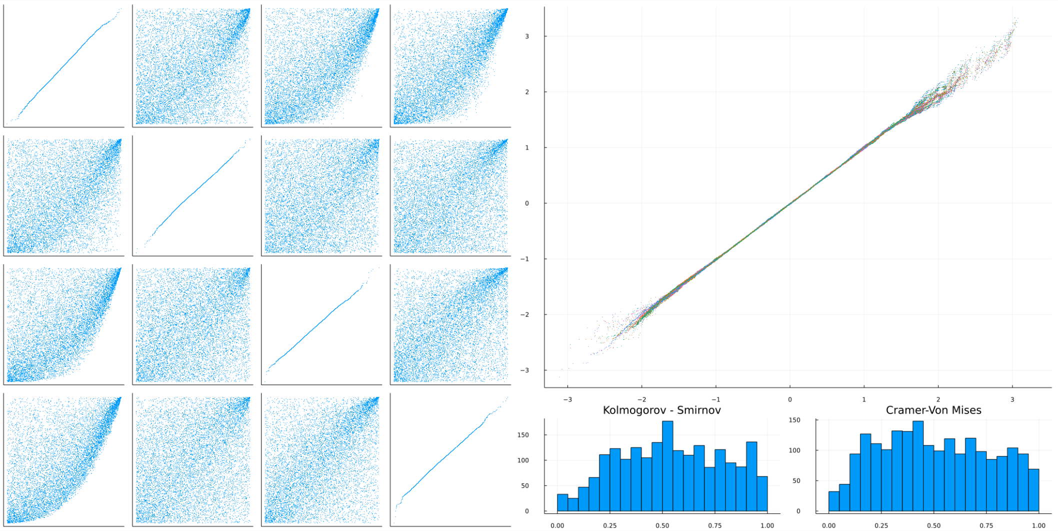
$$\boldsymbol{X} = \left(Y_1,\ Y_2 + U_1 Y_1^2,\ Y_3 + U_3 Y_1,\ Y_4 + Y_1^{1+\frac{U_3}{3}} \right),$$

We simulate a dataset $\boldsymbol{x} \in \mathbb{R}^{10000 \times 4}$ of $N = 10000$ i.i.d. samples from $\boldsymbol{X}$, and we ran the algorithm on it with 100 atoms.

We ended up with only 17 atoms at a $1e-16$ threshold on weights.

Remark: $\boldsymbol{X} \in \mathcal{G}_4$?

Four-variate simulated data



High-dimensional Thorin measures

Take away

Take away (2/2)

(i) The estimation of multivariate gamma convolutions is a d-variate moment problem, which is hard to solve when d gets large (tests at $d = 2000$ are OK.)

(ii) Random projections allow us to make the gradient cost essentially linear in the dimension, and still converges to a globally minimizing Thorin measure.

(iii) We obtained theoretical results ensuring the good behavior of our estimator.

(iv) We achieve sparse Thorin measure from dense proposals.

Details and examples can be found in the papers[10][11] *and the Julia package*[12].

[10] OL, Esterina Masiello, Véronique Maume-Deschamps, and Didier Rullière. "Estimation of Multivariate Generalized Gamma Convolutions through Laguerre Expansions.". In: *Electronic Journal of Statistics* 15.2 (Jan. 2021). ISSN: 1935-7524.

[11] OL. *Estimation of high dimensional Gamma convolutions through random projections*. 2022. DOI: 10.48550/ARXIV.2203.13741. URL: https://arxiv.org/abs/2203.13741.

[12] OL. *ThorinDistributions.jl*. Version v0.1. Mar. 2021.