Estimation of high dimensional gamma convolutions through random projections.

Thorin measures and Grassmannians cubatures...

O. Laverny [1,2]

April 20, 2021

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

[2] SCOR SE

Table of contents

Multivariate generalized gamma convolutions

Construction of the multivariate Thorin class

Consider that X is a univariate gamma distribution with shape $\alpha \in \mathbb{R}_+$ and scale $s \in \mathbb{R}_+$. By definition, the moment generating function of X is

$$M(t) := \mathbb{E}\left(e^{tX}\right) = (1 - ts)^{-\alpha}.$$

Construction of the multivariate Thorin class

Consider that X is a univariate gamma distribution with shape $\alpha \in \mathbb{R}_+$ and scale $s \in \mathbb{R}_+$. By definition, the cumulant generating function of X is

$$K(t) := \ln \mathbb{E}\left(e^{tX}\right) = -\alpha \ln(1 - ts).$$

Construction of the multivariate Thorin class

Consider that $\boldsymbol{X}$ is a multivariate gamma distribution with shape $\alpha \in \mathbb{R}_+$ and scales $\boldsymbol{s} \in \mathbb{R}_+^d$. By definition, the cumulant generating function of X is

$$K(\boldsymbol{t}) := \ln \mathbb{E}\left(e^{\langle \boldsymbol{t}, \boldsymbol{X} \rangle}\right) = -\alpha \ln\left(1 - \langle \boldsymbol{t}, \boldsymbol{s} \rangle\right).$$

Warn: This distribution is comonotonous. We have $\boldsymbol{X} = (s_1 X, ..., s_d X)$ with X a gamma distribution with shape α and unit scale.

Construction of the multivariate Thorin class

Consider that $\boldsymbol{X}$ is a multivariate gamma convolution with shapes $\boldsymbol{\alpha} \in \mathbb{R}_+^n$ and scales $\boldsymbol{s} \in \mathbb{R}_+^{n \times d}$. By definition, the cumulant generating function of X is

$$K(\boldsymbol{t}) := \ln \mathbb{E}\left(e^{\langle \boldsymbol{t}, \boldsymbol{X} \rangle}\right) = -\sum_{i=1}^{n} \alpha_i \ln\left(1 - \langle \boldsymbol{t}, \boldsymbol{s}_i \rangle\right).$$

Warn: This distribution can be absolutely continuous w.r.t. λ.

Construction of the multivariate Thorin class

Consider that $\boldsymbol{X}$ is a multivariate generalized gamma convolution with Thorin measure $\nu \in \mathcal{M}_+\left(\mathbb{R}_+^d\right)$. By definition, the cumulant generating function of X is

$$K(\boldsymbol{t}) := \ln \mathbb{E}\left(e^{\langle \boldsymbol{t}, \boldsymbol{X}\rangle}\right) = -\int \ln\left(1 - \langle \boldsymbol{t}, \boldsymbol{s}_i\rangle\right) \nu(\partial \boldsymbol{s}).$$

Good: This distribution can also be absolutely continuous w.r.t. λ, *under mild integration conditions*[1] *on* ν

We denote $\boldsymbol{X} \sim \mathcal{G}_d(\nu)$, and $\mathcal{G}_d$ is called the d-variate Thorin class.

[1] 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).

The d-variate Thorin class $\mathcal{G}_d$

Definition (Multivariate Thorin Classes[2])

$$\forall \nu \in \mathcal{M}_+(\mathbb{R}_+^d),\ \boldsymbol{X} \sim \mathcal{G}_d(\nu) \Leftrightarrow K(\boldsymbol{t}) := \ln \mathbb{E}\left(e^{\langle \boldsymbol{t}, \boldsymbol{X} \rangle}\right) = -\int \ln\left(1 - \langle \boldsymbol{s}, \boldsymbol{t} \rangle\right) \nu(\partial \boldsymbol{s}).$$

Prop: $\mathcal{G}_d$ is closed w.r.t (independent) sums and products, and contains many interesting marginals...

[2] 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.

A convolutive representation (and motivations)

Note that in the finitely atomic case, say $\boldsymbol{X} \sim \mathcal{G}_d(\sum_{i=1}^n \alpha_i \delta_{\boldsymbol{s}_i})$, for some $\boldsymbol{\alpha} \in \mathbb{R}_+^n$ and $\boldsymbol{s} \in \mathbb{R}_+^{n \times d}$, there exists independent Gamma random variables $G_i \sim \mathcal{G}_1(\alpha_i \delta_1)$, all having unit scale, such that:

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

Of course, we mostly consider $n \gg d$ (underdetermined case) with a sparse $\boldsymbol{s}$ matrix.

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

Pb: It is a deconvolution problem, which is numerically hard.

Motivation: Risk factor intepretation and need for infinite divisibility...

Shifted moments, shifted cumulants and Laguerre coefficients

Estimation idea: orthonormal expansion of the density.

The idea is the following:

(i) Find a suitable orthonormal basis of $\mathbb{R}_+^d$

(ii) Expand the density into this basis, with an appropriate truncature.

(iii) Compare theoretical and empirical coeficients to fit the parameters.

Bingo: The Laguerre basis[3] provides usable closed form expression for many usefull quantities here.

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

Laguerre coefficients a and shifted moments μ

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

$$\forall \boldsymbol{p} \in \mathbb{N}^d,\ \varphi_{\boldsymbol{p}}(\boldsymbol{x}) = \sqrt{2}^d \sum_{\boldsymbol{k} \leq \boldsymbol{p}} \binom{\boldsymbol{p}}{\boldsymbol{k}} \frac{(-2\boldsymbol{x})^{\boldsymbol{k}}}{\boldsymbol{k}!} e^{-|\boldsymbol{x}|}\ ; \qquad f(\boldsymbol{x}) = \sum_{\boldsymbol{k} \in \mathbb{N}^d} a_{\boldsymbol{k}} \varphi_{\boldsymbol{k}}(\boldsymbol{x})$$

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

Denote $\mu_{\boldsymbol{i}} := \mathbb{E}\left(X^{\boldsymbol{k}} e^{-|\boldsymbol{X}|}\right)$. Let $I \subset \mathbb{N}^d$ an increasing index set, $\boldsymbol{a} = (a_{\boldsymbol{i}})_{\boldsymbol{i} \in I}$ and $\boldsymbol{\mu} = (\mu_{\boldsymbol{i}})_{\boldsymbol{i} \in I}$.

Bijection: The relationship between $\boldsymbol{a}$ and $\boldsymbol{\mu}$ is linear. We encode this relationship through a (lower-triangular, invertible) matrix $\boldsymbol{A} \in \mathbb{R}^{|I| \times |I|}$ such that

$$\boldsymbol{a} = \boldsymbol{A}\boldsymbol{\mu} \text{ and } \boldsymbol{\mu} = \boldsymbol{A}^{-1}\boldsymbol{a}.$$

Thorin moments τ

Rem: Since $\mu_i := \mathbb{E}\left(\boldsymbol{X}^{\boldsymbol{k}} e^{-\langle \mathbf{1}, \boldsymbol{X}\rangle}\right)$, the mgf of the random vector writes:

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

Definition (Thorin moments τ)

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

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

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

Estimators of a, μ and τ

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

Definition (Monte-Carlo estimators)

$$\hat{\boldsymbol{\mu}}(\boldsymbol{x}) = (\widehat{\mu_{\boldsymbol{k}}}(\boldsymbol{x}))_{\boldsymbol{k} \in I} = \left(\frac{1}{N} \sum_{i=1}^{N} \boldsymbol{x}_i^{\boldsymbol{k}} e^{-|\boldsymbol{x}_i|} \right)_{\boldsymbol{k} \in I}$$
$$\hat{\boldsymbol{\tau}}(\boldsymbol{x}) = (\widehat{\tau_{\boldsymbol{k}}}(\boldsymbol{x}))_{\boldsymbol{k} \in I} = \boldsymbol{B}^{-1} \left(\hat{\boldsymbol{\mu}}(\boldsymbol{x}) \right).$$
$$\hat{\boldsymbol{a}}(\boldsymbol{x}) = (\widehat{a_{\boldsymbol{k}}}(\boldsymbol{x}))_{\boldsymbol{k} \in I} = \boldsymbol{A} \hat{\boldsymbol{\mu}}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{B}(\hat{\boldsymbol{\tau}}(\boldsymbol{x})),$$

Rem: biais, cv ?

Back to gamma convolutions...

Definition (Thorin moments of a Gamma convolution)

We denote the first Thorin moments of the $\mathcal{G}_d(\nu)$ distribution by:

$$\boldsymbol{\tau}(\nu) = (\tau_{\boldsymbol{k}}(\nu))_{\boldsymbol{k}\in I}\,.$$

Property

Denoting δ_x the Dirac measure at x, it holds:

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

(ii) $\boldsymbol{\tau}(\nu) = \int \boldsymbol{\tau}(\delta_{\boldsymbol{s}})\nu(d\boldsymbol{s})$, *where the equality and integration are intended componentwise.*

Rem: τ are called Thorin moments for a reason. We are dealing with a multivariate moment problem *which might not have a solution....*

Direct minimisation of the Laguerre loss.

The Laguerre loss

We use the integrated square error between densities, projected in the Laguerre basis:

$$\mathcal{L}(\boldsymbol{x}, \nu) = \|\hat{\boldsymbol{a}}(\boldsymbol{x}) - \boldsymbol{AB}(\boldsymbol{\tau}(\nu))\|_2^2.$$

Theorem (Consistency[4])

If $\boldsymbol{x}$ is drawn from an ϵ-well-behaved density $f \in \mathcal{G}_d$, any well-behaved estimator $\nu^\star$ such that $\mathcal{L}(\boldsymbol{x}, \nu^\star) \xrightarrow[N\to\infty]{a.s} 0$ ensures that

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

[4] Oskar Laverny, Esterina Masiello, Véronique Maume-Deschamps, and Didier Rullière. "Estimation of multivariate generalized gamma convolutions through Laguerre expansions.". In: *Electronic Journal of Statistics* 15.2 (2021), pp. 5158–5202. DOI: 10.1214/21-EJS1918. URL: https://doi.org/10.1214/21-EJS1918.

Dimensionality issues

The loss: $\mathcal{L}(\boldsymbol{x}, \nu) = \|\hat{\boldsymbol{a}}(\boldsymbol{x}) - \boldsymbol{AB}(\boldsymbol{\tau}(\nu))\|_2^2$.

Problem: $\mathcal{L}$ is too costly to work with when d gets large...

The vectors $\hat{\boldsymbol{a}}(\boldsymbol{x})$ and $\boldsymbol{AB}(\boldsymbol{\tau}(\nu))$ each consist of $|I|$ coefficients. If, e.g., $I = \left\{\boldsymbol{k} \in \mathbb{N}^d, |\boldsymbol{k}| \leq m\right\}$ is isotropic, the number of coefficents to compute is given by

$$D(m, d) = \sum_{i=1}^{m} \binom{i + d - 1}{d - 1}.$$

which is exponentially increasing in d and therefore unusable.

Better approach via random projections.

An approximated loss through random projections...

Through a first-order taylor expansion of the function $\boldsymbol{AB}$, we define:

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

where $\nabla(\boldsymbol{x})$ is a jacobian of the function $\boldsymbol{AB}$, taken in $\hat{\boldsymbol{\tau}}(\boldsymbol{x})$. Then, through a univariate projection and re-integration we define:

$$\begin{aligned}\widetilde{\mathcal{L}}(\boldsymbol{x}, \nu) :&= \int_{[0,1]^d} \widehat{\mathcal{L}}\left(\langle \boldsymbol{c}, \boldsymbol{x}\rangle, \nu_{\langle \boldsymbol{c}\rangle}\right) d\boldsymbol{c} \\ &= \int_{[0,1]^d} \|\hat{\boldsymbol{\tau}}(\langle \boldsymbol{c}, \boldsymbol{x}\rangle) - \boldsymbol{\tau}(\nu_{\langle \boldsymbol{c}\rangle})\|^2_{\nabla(\langle \boldsymbol{c}, \boldsymbol{x}\rangle)} \, d\boldsymbol{c},\end{aligned}$$

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

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

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 minimas...

Reframing the loss

Let $\mathcal{F} = L_2([0,1]^d, \mathbb{R}_+^m)$ be the space of square integrable functions from $[0,1]^d$ to $\mathbb{R}_+^m$. We consider the functional ψ defined as:

$$\psi\colon \begin{cases} \mathbb{R}_+^d \longrightarrow \mathcal{F} \\ \quad \boldsymbol{s} \longmapsto \psi(\boldsymbol{s})\colon \begin{cases} [0,1]^d \longrightarrow \mathbb{R}_+^m \\ \quad \boldsymbol{c} \longmapsto \boldsymbol{\tau}\left(\delta_{\langle \boldsymbol{s}, \boldsymbol{c} \rangle}\right). \end{cases} \end{cases}$$

We also consider, as our target,

$$\psi_{\boldsymbol{x}}\colon \begin{cases} [0,1]^d \longrightarrow \mathbb{R}_+^m \\ \quad \boldsymbol{c} \longmapsto \hat{\boldsymbol{\tau}}(\langle \boldsymbol{c}, \boldsymbol{x} \rangle) \end{cases} \in \mathcal{F}.$$

Reframing the loss

We endow $\mathcal{F}$ with the norm $\|\cdot\|_{\boldsymbol{x}}$ defined as:

$$\forall \psi \in \mathcal{F},\ \|\psi\|_{\boldsymbol{x}}^2 = \int_{[0,1]^d} \|\psi(\boldsymbol{c})\|_{\nabla(\langle \boldsymbol{c}, \boldsymbol{x} \rangle)}^2 d\boldsymbol{c},$$

This allow us to write our loss as:

$$\widetilde{\mathcal{L}}(\boldsymbol{x}, \nu) = \|\psi_{\boldsymbol{x}} - \int \psi(\boldsymbol{s})\nu(d\boldsymbol{s})\|_{\boldsymbol{x}}^2,$$

which shows that our loss is convex in $\int \psi(\boldsymbol{s})\nu(d\boldsymbol{s})$, but clearly not convex in the atoms and weights of an n-atomic measure ν. However Chizat[5] applies !

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

Fréchet-differentiability and gradient flow.

Property

The Fréchet differential of $\widetilde{\mathcal{L}}(\boldsymbol{x}, \cdot)$ at $\nu \in \mathcal{M}(\mathbb{R}_+^d)$ is represented by the function

$$\widetilde{\mathcal{L}}'_\nu : \begin{cases} \mathbb{R}_+^d \longrightarrow \mathbb{R} \\ \quad \boldsymbol{s} \longmapsto 2\langle \psi_{\boldsymbol{x}} - \displaystyle\int \psi(\boldsymbol{s})\nu(d\boldsymbol{s}), \psi(\boldsymbol{s}) \rangle_{\boldsymbol{x}}. \end{cases}$$

Definition (Gradient flow[6])

A gradient flow of $\widetilde{\mathcal{L}}(\boldsymbol{x}, \cdot)$ is an absolutely continuous curve $(\nu_t)_{t \geq 0}$ in the space $\mathcal{M}(\mathbb{R}_+^d)$ that satisfies

$$\frac{\partial}{\partial t}\nu_t = -\nabla \widetilde{\mathcal{L}}(\boldsymbol{x}, \nu_t)$$

Where the gradient $\nabla \widetilde{\mathcal{L}}(\boldsymbol{x}, \nu_t)$ is taken w.r.t. a given conic metric...

[6] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows: in metric spaces and in the space of probability measures*. Springer Science & Business Media, 2005.

A convergence result for the gradient flow.

Theorem (Global convergence of the gradient flow)

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 $t = \mathcal{O}\left(-\log(\epsilon)\right)$.

This result directly leverages Chizat's results. For general accelerated convex methods, $W_\infty(\nu_t, \nu_\infty) \leq \epsilon$ is achieved for $t = \mathcal{O}\left(\epsilon^{-1/d}\right)$ only, see[7] for more details.

[7] 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.

Takeaway: Initialize the Thorin measure all over $\mathbb{R}_+^{d+1}$, with a lot of atoms, to achieve global convergence.

Bonus: Same result with a lasso penalty on the measure, which is perfect for what we need (penalises the abbundance of sources), via a hyperparameter $\lambda > 0$ and a loss

$$\widetilde{\mathcal{L}}(\boldsymbol{x}, \nu) + \lambda |\nu|.$$

Any wanted hyperparameter searching method (with its associated cost) could be used to find λ.

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.

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

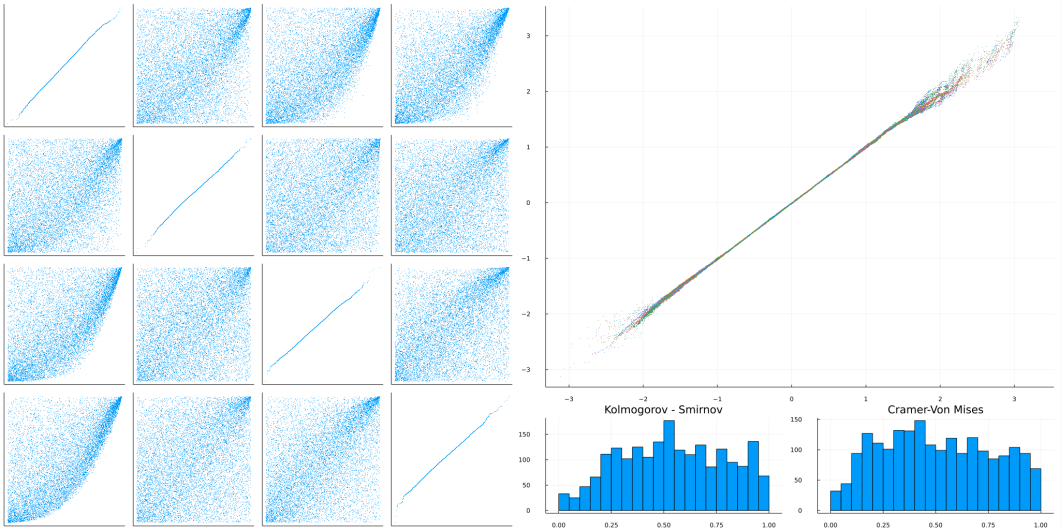
For U_1, U_2, U_3 independent $\mathcal{U}([0,1])$ distributions and $Y_1, ..., Y_4$ independent log-Normal(0,1) distributions, we let the random vector $\boldsymbol{X}$ be defined by:

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

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.

Four-variate simulated data



2000-dimensionnal multiplicative dataset

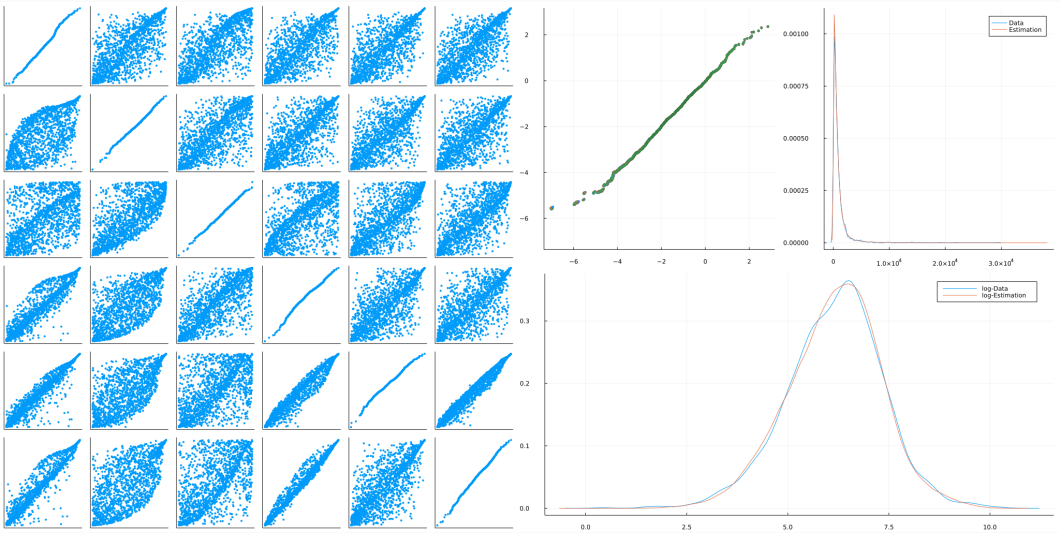
Let $d = 2000$, let $G \sim \mathcal{G}_1(\delta_1)$, $H \sim \mathcal{G}_1(2\delta_{\frac{1}{2}})$, and $Z_1, ..., Z_d \sim \mathcal{N}(0, 1)$ be all independent random variables, and let $\alpha_1, .., \alpha_d$ be fixed parameters in $[0, 1]$ (uniformly chosen). Construct the random vector $\boldsymbol{X} = (X_1, ..., X_d)$ as:

$$\boldsymbol{X} = \left(G\, e^{Z_i}\, H^{1+2\alpha_i} \right)_{i \in 1,...,d}.$$

We simulate a dataset $\boldsymbol{x} \in \mathbb{R}^{1500 \times 2000}$ of $N = 1500$ i.i.d. samples from $\boldsymbol{X}$.

We ran our algorithm with $n = 500$ atoms, and ended up with only 36 meaningfull ones.

2000-dimensionnal multiplicative dataset



Conclusion

Main takeways:

(i) Multivariate generalized gamma convolutions are fexible semi-parametric structures

(ii) They simplify divisions of positive random variables (in the infinite divisibility sense) by making this process parametrical

(iii) Estimating them can be reduced to a non-sparse d-variate moment problem, which can be very hard to solve.

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

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

(vi) The (OSS) Julia package `ThorinDistributions.jl` provides an implementation.

Thanks !