Estimation of multivariate generalized gamma convolutions through Laguerre expansions

The beauty of multivariate Thorin classes

Accepted at EJS

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

October 22, 2021

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

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

[3] SCOR SE

Table of contents

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

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

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 currently avaliables.

[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.

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

$\boldsymbol{X} \in \mathcal{G}_{d,n}$ follows an additive risk-factor structure: there exists gamma random variables $G_1, ...G_n$ such that:

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

(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.).

A generalized moment problem : Miles, Furman and Kuznetsov

Miles, Furman and Kuznetsov[2] exhibit in the univariate case a moment problem for the Thorin measure. A multivariate analogue of their train of thoughts might be the following:

For $\boldsymbol{X} \sim \mathcal{G}_d(\nu)$, $K(\boldsymbol{t}) = -\int \ln(1 - \langle \boldsymbol{s}, \boldsymbol{t}\rangle)\,\nu(\partial \boldsymbol{s})$. Derive $\boldsymbol{i}$ times on $\boldsymbol{t}$ to obtain:

$$\kappa_{\boldsymbol{i},\boldsymbol{t}} = (|\boldsymbol{i}| - 1)! \int \frac{\boldsymbol{s}^{\boldsymbol{i}}}{(1 - \langle \boldsymbol{s}, \boldsymbol{t}\rangle)^{|\boldsymbol{i}|}} \nu(\partial \boldsymbol{s})$$

The substitution $\boldsymbol{x} = \frac{\boldsymbol{s}}{1-\langle \boldsymbol{s},\boldsymbol{t}\rangle} \iff \boldsymbol{s} = \frac{\boldsymbol{x}}{1+\langle \boldsymbol{x},\boldsymbol{t}\rangle}$ gives: $\xi_{\boldsymbol{i}} = \int\limits_{\Delta_d(-\boldsymbol{t})} \boldsymbol{x}^{\boldsymbol{i}} \xi(\partial \boldsymbol{x})$, which can be seen as a d-variate generalized moment problem on the $(-\boldsymbol{t})$-simplex.

Solving this moment problem for any $\boldsymbol{t}$ (e.g. $-\boldsymbol{1}$) provides a $\mathcal{G}_d$ parametrization corresponding to the shifted cumulants $\kappa_{\boldsymbol{i},\boldsymbol{t}}$.

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

Unfortunately, no solution

Generalized moment problems have solvers through Lassere's hierarchy of SDPs, see Helton & Nie[3]. **To even obtain a solution, the input data needs to be inside the cone of moments.**

(i) The cumulants $\kappa_{\boldsymbol{i},\boldsymbol{t}}$ needs to come from a true $\mathcal{G}_d$ distribution so that $\boldsymbol{\xi}$ is in the cone of moments (300 digits of tanh-sinh integration per moment if $d = 1$ according to MFK, probably more as d increase)

(ii) **For empirical cumulants, no way to define a coherent loss:** Should we take a Pareto front ? Give more weight to smaller moments ?

A least square approach could work for inexact cumulants, but we do not have a way to weight the objectives. **Also this breaks Lassere's hierachy simplicity..**

$\Longrightarrow$ We need another approach.

[3] J. William Helton and Jiawang Nie. "A Semidefinite Approach for Truncated K-Moment Problems". en. In: *arXiv:1105.0410 [math]* (Sept. 2012), Jiawang Nie. "The A-Truncated K-Moment Problem". en. In: *arXiv:1210.6930 [math]* (Oct. 2012), 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.

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

Definition (Laguerre basis, see Comte[4], Mabon[5] and Dussap[6])

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

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

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

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

[4] 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.

[5] 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.

[6] Florian Dussap. "Anisotropic Multivariate Deconvolution Using Projection on the Laguerre Basis". In: (2020).

Algorithm 1: Laguerre coefficients of $\mathcal{G}_{d,n}(\boldsymbol{\alpha}, \boldsymbol{s})$ random vectors

Input: Shapes $\boldsymbol{\alpha} \in \mathbb{R}_+^d$, scales $\boldsymbol{s} \in \mathcal{M}_{n,d}(\mathbb{R}_+)$, and truncation threshold $\boldsymbol{m} \in \mathbb{N}^d$
Result: Laguerre coefficients $(a_{\boldsymbol{k}})_{\boldsymbol{k} \leq \boldsymbol{m}}$of the $\mathcal{G}_{d,n}(\boldsymbol{\alpha}, \boldsymbol{s})$ density
Compute the simplex version of the scales $\boldsymbol{x}_i = \frac{\boldsymbol{s}_i}{1+|\boldsymbol{s}_i|}$ for all $i \in 1, ..., n$.
Let $\kappa_{\boldsymbol{0}} = -\sum_{i=1}^{n} \alpha_i \ln(1 - |\boldsymbol{x}_i|)$ and $a_{\boldsymbol{0}} = \mu_{\boldsymbol{0}} = \exp(\kappa_{\boldsymbol{0}})$
foreach $\boldsymbol{0} \neq \boldsymbol{k} \leq \boldsymbol{m}$ **do**
- Let $a_{\boldsymbol{k}} = \mu_{\boldsymbol{k}} = 0$, d be the index of the first k_i that is non-zero, $\boldsymbol{p} = \boldsymbol{k}$ and set $p_d = p_d - 1$.
- Let $\kappa_{\boldsymbol{k}} = (|\boldsymbol{k}| - 1)! \sum_{i=1}^{n} \alpha_i \boldsymbol{x}_i^{\boldsymbol{k}}$
- **foreach** $\boldsymbol{l} \leq \boldsymbol{p}$ **do**
  - Set $\mu_{\boldsymbol{k}} \mathrel{+}= (\mu_{\boldsymbol{l}})(\kappa_{\boldsymbol{k}-\boldsymbol{l}})\binom{\boldsymbol{p}}{\boldsymbol{l}}$ according to efficient Faà di Bruno's algorithm from Miatto[a]
  - Set $a_{\boldsymbol{k}} \mathrel{+}= \mu_{\boldsymbol{l}} \binom{\boldsymbol{k}}{\boldsymbol{l}} \frac{(-2)^{|\boldsymbol{l}|}}{\boldsymbol{l}!}$
- **end**
- Set $a_{\boldsymbol{k}} \mathrel{+}= \mu_{\boldsymbol{k}} \frac{(-2)^{|\boldsymbol{k}|}}{\boldsymbol{k}!}$

end
$\boldsymbol{a} = \sqrt{2}^d \boldsymbol{a}$
Return $\boldsymbol{a}$

[a] Filippo M. Miatto. "Recursive Multivariate Derivatives of \$e^{f(X_1,\dots, X_n)}\$ of Arbitrary Order". en. In: *arXiv:1911.11722 [cs, math]* (Nov. 2019).

Analysis of the loss: The ε-w.b. condition

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 $\boldsymbol{s} \in]\frac{\varepsilon}{2+\varepsilon}, \frac{2+\varepsilon}{\varepsilon}[^n$.

$\mathcal{G}_{d,n}(\boldsymbol{\alpha}, \boldsymbol{s})$ 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 $\forall I \subseteq \{1, \dots, n\}$ 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 w.b. :

(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:

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

Implication: The ISE loss $L(\boldsymbol{\alpha}, \boldsymbol{s}) = \sum_{\boldsymbol{k} \leq \boldsymbol{m}} \left(\widehat{a_{\boldsymbol{k}}} - a_{\boldsymbol{k}}(\boldsymbol{\alpha}, \boldsymbol{s})\right)^2$ 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 componentwise Möbius transform.

Remark on implementation and complexity

We need to minimize a quadratic loss on the coefficients computed through Algorithm 1.

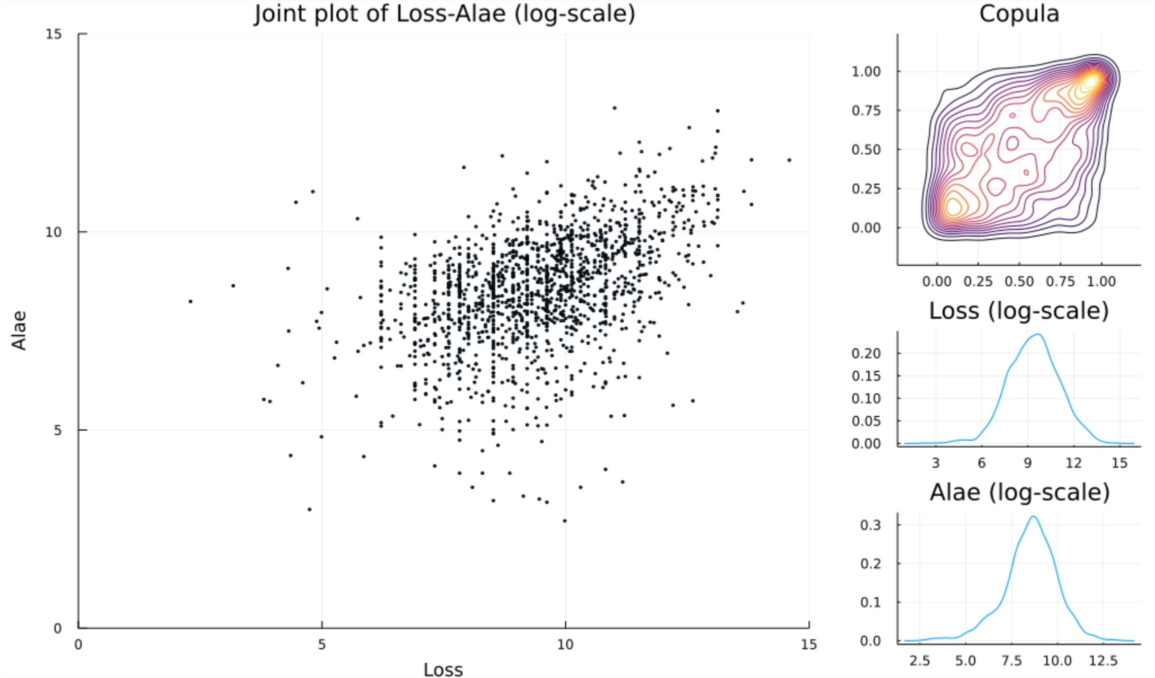
Therefore, our loss is:

$$\begin{aligned} &\text{Combinatorial} &&\Longrightarrow \text{Arbitrary precision} && \\ &\text{Highly recursive} &&\Longrightarrow \text{Compiled code} &&\Longrightarrow \texttt{Julia} \\ &\text{Highly non-convex} &&\Longrightarrow \text{Global optimization} && \end{aligned}$$

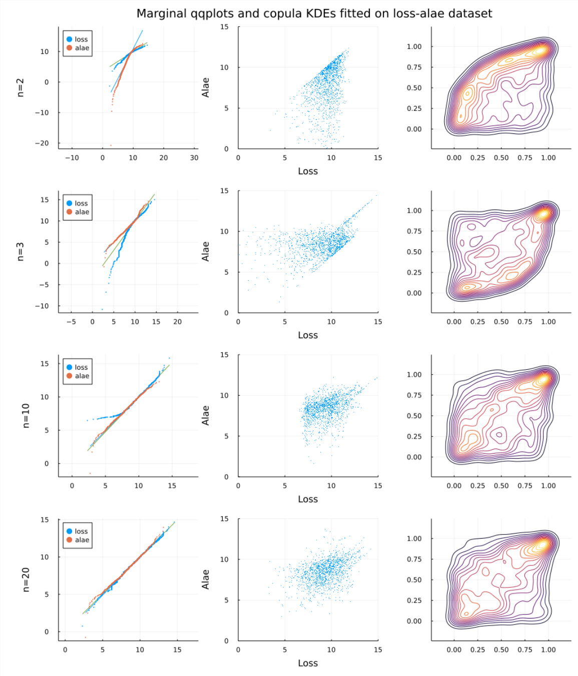
Our package `ThorinDistributions.jl`[7], available on `github`, implements this loss efficiently.

[7] Oskar Laverny. *lrnv/ThorinDistributions.jl: ThorinDistributions.jl v0.1*. Version v0.1. Mar. 2021. DOI: 10.5281/zenodo.4644109. URL: https://doi.org/10.5281/zenodo.4644109.

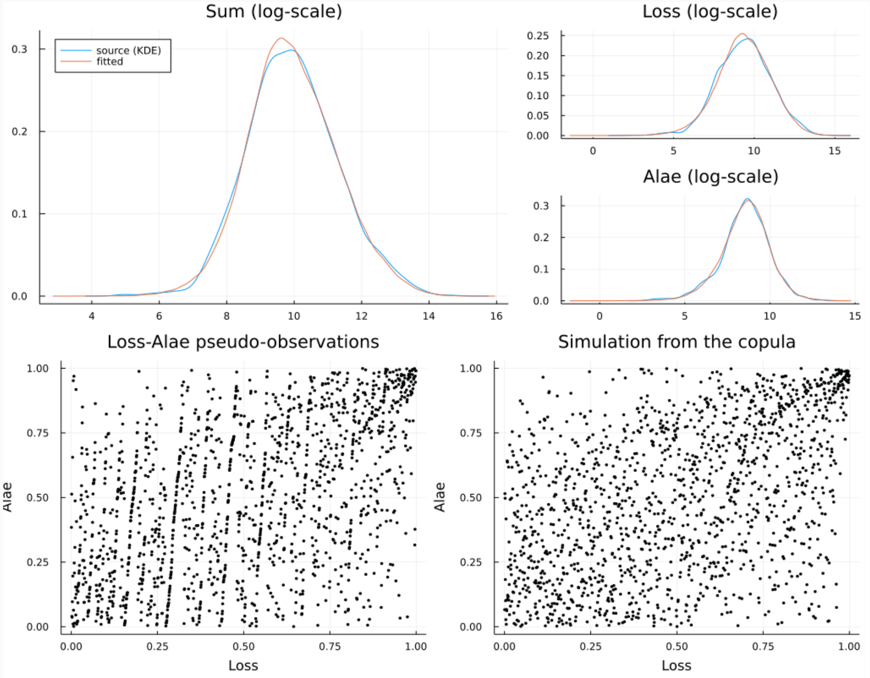
Crunching numbers on Loss-Alae: the dataset



Crunching numbers on Loss-Alae: Results for several numbers of gammas



Crunching numbers on Loss-Alae: Results for $n = 20$



Drawbacks of the approach and potential solutions

The total number of coefficients to compute is $\prod_{i=1}^{d}(m_i+1)$. This is quite unpractical when d gets large, and therefore restricts the use of the algorithm to $d < 5$ or 6 to have a performant procedure.

However, there is a way out of the curse: If $X_i \sim \mathcal{G}_{1,n_i}(\boldsymbol{\alpha}_i, \boldsymbol{s}_i)$, denoting $\mathbb{N}_{\boldsymbol{n}}^d = \left\{\boldsymbol{p} \in \mathbb{N}^d : \forall i, 0 \leq p_i \leq n_i\right\}$ and $N = \prod_{i=1}^{d}(n_i+1)$ its cardinal, any multivariate gamma convolutions with these marginals is a $\{\mathcal{G}_{d,N}(\boldsymbol{a}, \boldsymbol{S})\}$ such that:

(i) $\boldsymbol{S} = \left\{(s_{1,i_1}, ..., s_{d,i_d}) : \forall \boldsymbol{i} \in \mathbb{N}_{\boldsymbol{n}}^d\right\}$, with $s_{i,0} = 0$ for all i.

(ii) $\boldsymbol{a} \in \mathbb{R}_+^{\mathbb{N}_{\boldsymbol{n}}^d}$ and for all i, j, $\sum_{\boldsymbol{p} \in \mathbb{N}_{\boldsymbol{n}}^d} a_{\boldsymbol{p}} \mathbf{1}_{p_i = j} = \alpha_{i,j}$. (this constraint is linear).

Last, for a given constant $\boldsymbol{c}$ to be chosen, $\langle \boldsymbol{c}, \boldsymbol{X}\rangle \sim \mathcal{G}_{1,N}\left(\boldsymbol{a}, \{\langle \boldsymbol{c}, \boldsymbol{s}\rangle : \boldsymbol{s} \in \boldsymbol{S}\}\right)$.

Estimating $\langle \boldsymbol{c}, \boldsymbol{X}\rangle$ conditionally on the marginals is therefore enough.

Conclusion

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

(ii) The Multivariate analogue provides an asymmetrical dependence structure which include tail dependency, and can take a lot of different shapes.

(iii) Deconvolution is a hard inverse problem, and estimation of these distributions is complicated.

(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 (working on it).

Details and other illustrated applications are available in our paper[8] *and Julia package repo*[9].

Thanks !

[8] Oskar Laverny, Esterina Masiello, Véronique Maume-Deschamps, and Didier Rullière. *Estimation of multivariate generalized gamma convolutions through Laguerre expansions*. 2021. arXiv: 2103.03200 [math.ST].

[9] Oskar Laverny. *lrnv/ThorinDistributions.jl: ThorinDistributions.jl v0.1*. Version v0.1. Mar. 2021. DOI: 10.5281/zenodo.4644109. URL: https://doi.org/10.5281/zenodo.4644109.