Estimation of multivariate gamma convolutions through Laguerre expansions

Under review at Electronic Journal of Statistics

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

May 18, 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

1. Multivariate Gamma convolutions classes $\mathcal{G}_{d,n}$ and $\mathcal{G}_d$
2. An efficient algorithm for Laguerre coefficients of $\mathcal{G}_{d,n}$ distributions.
3. Uniform exponential bound & consistent loss
4. Remark on implementation and complexity
5. Conclusion & Doggy bag

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])

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

$\boldsymbol{X} \sim \mathcal{G}_d(\nu) \Leftrightarrow K(\boldsymbol{t}) = -\int \ln\left(1 - \langle \boldsymbol{s}, \boldsymbol{t}\rangle\right) \nu(\partial \boldsymbol{s})$.

Prop: $\mathcal{G}_d$ is closed w.r.t (independent) sums and products !

Note: Estimation of ν from empirical cumulants, i.e derivatives of K, is a GMP, a hard inverse problem!

[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.

An efficient algorithm for Laguerre coefficients of $\mathcal{G}_{d,n}$ dists.

Definition (Laguerre basis of $L_2(\mathbb{R}_+^d)^2$)

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

$$\boldsymbol{a} = (a_{\boldsymbol{k}})_{\boldsymbol{k}\leq\boldsymbol{m}} :\ f(\boldsymbol{x}) = \sum_{\boldsymbol{k}\in\mathbb{N}^d} a_{\boldsymbol{k}} \varphi_{\boldsymbol{k}}(\boldsymbol{x})$$

$$\implies \boldsymbol{a}_{\boldsymbol{k}} = \sqrt{2}^d \sum_{\boldsymbol{k}\leq\boldsymbol{p}} \binom{\boldsymbol{p}}{\boldsymbol{k}} \frac{(-2)^{|\boldsymbol{k}|}}{\boldsymbol{k}!} \mathbb{E}\left(X^{\boldsymbol{k}} e^{-|\boldsymbol{X}|}\right)$$

Result: We achieved an algorithm that maps $(\alpha, \boldsymbol{s}) \longmapsto \boldsymbol{a}$ in quadratic time in $\#\boldsymbol{a}$ (and linear in n). *(through a fast Faa Di Bruno algo)*.

[2] Florian Dussap. “Anisotropic Multivariate Deconvolution Using Projection on the Laguerre Basis”. In: (2020).

Uniform exponential bound & consistent loss

Result: Under a simple taylor-made ε-non-degenerateness conditions, uniformly on all $(\boldsymbol{\alpha}, \boldsymbol{s})$ parametrization, we have an exponential bound:

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

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

Proofs leverage analytics combinatorics is several variables

Remark on implementation and complexity

The Loss we need to minimize is:

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

The package `ThorinDistributions.jl` is available on `github`.

Unfortunately, no time to go through applications examples !

Conclusion & Doggy bag

(i) Deconvolution is usually hard, even for $n = 2$.

(ii) $\mathcal{G}_1$ is wide: log-Normals, Paretos, α-stable, Weibull... and $\mathcal{G}_d$ dependence structures are very flexible: asymmetry, tail dependency, many shapes...

(iii) Interpretation as Additive risk factor models & dist. of the sum...

Details, more results and illustrated applications are availables in our paper[3] and Julia package repo[4].

Thanks !

[3]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].

[4]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.