Estimation of multivariate gamma convolutions

O. Laverny
Université Claude Bernard Lyon 1

Supervisor(s): Prof. V. Maume-Deschamps (UCBL), Dr. E. Masiello (UCBL) and Prof. D. Rullière (Ecoles des Mines de Saint-Etienne)

Ph.D. expected duration: Jun. 2019 - Jun. 2022

Adress: Institut Camille Jordan, 21 Avenue Claude Bernard, 69100 Villeurbanne

Email: laverny@math.univ-lyon1.fr

Abstract:

The class $\mathcal{G}_1$ of univariate generalized gamma convolution, defined as the class of weak limits of independent convolutions of gamma distributions, was first introduced by Thorin [7, 8] as a tool to show the infinite divisibility of log-normal and Pareto distributions. Although originating from this very practical question, its study led to many improvements of the theory, well summarized by Bondesson in [1] and is still an active field nowadays [4, 6]. By definition, $\mathcal{G}_1$ is closed under independent convolutions, but as it appears recently [2] it is also closed by independent products of random variables. Pareto, log-normal, α-stable, Weibull, and many other distributions are in this class, which makes it a nice framework for many applications fields such as climate events modeling, insurance, etc.

An analogue multivariate class $\mathcal{G}_d$ was constructed by Bondesson [2], following an old idea of Cherian [3], by convoluing comonotonous multivariate gamma distributions. Finally, we also consider $\mathcal{G}_{d,n}$, the subclass containing convolutions of at most n d-variates gamma. Although not much is known about $\mathcal{G}_d$, a random vector $\boldsymbol{X} \in \mathcal{G}_{d,n}$ follows an additive risk-factor stucture: there exists gamma random variables $Y_{i,j}$ such that $\forall\, i \in \{1, ..., d\}$,

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

where each vector $Y_{i,\cdot}$ has independent marginals, and each vector $Y_{\cdot,j}$ has comonotonous marginals. Since on one hand some $Y_{i,j}$ might be identically zero, as $0 \in \mathcal{G}_{1,1}$, and on the other hand every $Y_{i,j}$ is infinite divisible, by increasing n the model can achieve a wide variety of dependence structures and approach any marginal in $\mathcal{G}_1$.

A useful tool for the analysis of these models is the Thorin measure. It can be shown that the cumulant generating function of a random vector in $\mathcal{G}_d$ writes :

$$K(\boldsymbol{t}) = \ln\left(E\left[e^{\langle \boldsymbol{t}, \boldsymbol{X}\rangle}\right]\right) = -\int_{R_+^d} \ln\left(1 - \langle \boldsymbol{t}, \boldsymbol{\theta}\rangle\right) \nu\left(\partial \boldsymbol{\theta}\right),$$

where the Thorin measure ν is discrete with n atoms if $\boldsymbol{X} \in \mathcal{G}_{d,n}$. In this case, atoms and weights of ν are respectively scales and shapes of gamma vectors $Y_{\cdot,j}$. Therefore, the model parameters can be fully summarised by ν, which we would like to estimate.

The current literature contains no estimation procedure for distributions in $\mathcal{G}_d$ or $\mathcal{G}_{d,n}$. Noteworthy is the projection procedure from [5], that for a density in $\mathcal{G}_1$ finds an approximation in $\mathcal{G}_{1,n}$. This projection works only if the input density is exactly a density in the $\mathcal{G}_1$, whose Laplace transform derivatives must be evaluated with 300 digits precision, excluding e.g. empirical datasets.

In this work, we investigate the univariate and multivariate estimation of the Thorin measure. We focus on the subclass $\mathcal{G}_{d,n}$, as a continuous Thorin measure is not easy to work with: there is no

simulation procedure for the distribution, no known expressions for the density or the distribution function, and even the Laplace transform requires numerical integration for evaluation.

Through a projection in a Laguerre basis, we provide an estimation procedure that produces estimators in $\mathcal{G}_{d,n}$, for any finite d, n, be the distribution given through an exact density or an empirical dataset. This procedure is based on the fact that Laguerre coefficients can be efficiently computed from the parameters of the model, through a fast ad-hoc algorithm, based on a multivariate Faa di Bruno formula.

Moreover, we provide a series expansion for densities in $\mathcal{G}_{d,n}$, where the current literature gives densities relying on Mathai and Moschopoulos series, which work only when $d = 1$ and are known to be unstable or even dramatically failing for certain parameters ranges, noteworthy those which correspond to projection of log-normal, Pareto or Weibull distributions.

The estimation procedure can be modified to include a penalty on n, which tends to a sparser model, including a sparser dependence structure of the Thorin measure in the multivariate case.

References

[1] Lennart Bondesson. *Generalized Gamma Convolutions and Related Classes of Distributions and Densities*, volume 76 of *Lecture Notes in Statistics*. Springer New York, New York, NY, 1992.

[2] Lennart Bondesson. A Class of Probability Distributions that is Closed with Respect to Addition as Well as Multiplication of Independent Random Variables. *Journal of Theoretical Probability*, 28(3):1063–1081, September 2015.

[3] KC Cherian. A bivariate correlated gamma-type distribution function. *Journal of the Indian Mathematical Society*, 5:133–144, 1941.

[4] Wissem Jedidi and Thomas Simon. Further examples of GGC and HCM densities. *Bernoulli*, 19(5A):1818–1838, November 2013.

[5] Justin Miles, Edward Furman, and Alexey Kuznetsov. Risk aggregation: A general approach via the class of Generalized Gamma Convolutions. page 25, 2019.

[6] Tord Sjödin. On Mixtures of Gamma Distributions, Distributions with Hyperbolically Monotone Densities and Generalized Gamma Convolutions (GGC). *arXiv:1806.03926 [math]*, June 2018.

[7] Olof Thorin. On the infinite divisbility of the Pareto distribution. *Scandinavian Actuarial Journal*, 1977(1):31–40, January 1977.

[8] Olof Thorin. On the infinite divisibility of the lognormal distribution. *Scandinavian Actuarial Journal*, 1977(3):121–148, March 1977.

Short biography – Oskar Laverny is an actuary by formation, fond of statistics and code. Oskar's PhD is funded through a CIFRE grant between the UCBL and SCOR SE, a reinsurance company. The core of the PhD project is the development of new statistical and computational tools for the study of dependence structures and aggregation of risks in high-dimensional contexts, with direct applications to the risk management problems a re-insurer can face. Oskar has a taste for open-source software and all his work is publicly available on GitHub, be it the source code of his blog or R, Julia and Python packages he develops.