Pareto-normal Distributions: Theoretical Framework and Computational Issues

University of Nevada, Reno

Matthew A. Ohemeng and Tomasz J. Kozubowski

Department of Mathematics and Statistics, University of Nevada, Reno

Symposium on Data Science & Statistics, June 2-4, 2021

Abstract

We introduce a new, conditionally Gaussian, hierarchical stochastic model for heavy tailed data, which generalizes the Laplace probability distribution. We present basic properties of this model and discuss related computational issues. We also briefly consider inferential aspects of the model as well as its applications.

Motivation

- **Normal distribution** plays a prominent role in probability and statistics.
- It provides an **approximation to sums** which often come up in applications.
- Central limit effect: If $\{\mathbf{X}_i\}$ are IID with finite second moment and
$$a_n \sum_{i=1}^{n} (\mathbf{X}_i + \mathbf{b}_n) \xrightarrow{d} \mathbf{X} \quad \text{as n} \to \infty$$
then $\mathbf{X}$ has a multivariate normal distribution.
- **Limitations**: Lack of flexibility in terms of symmetry and distributional tails
- Does the normal distribution provide an approximation to the sums when the number of terms in the sum is random (not a deterministic constant)?
- A random summation frequently arises in applications too:
  - In insurance, the random sum can be the total claim for a period.

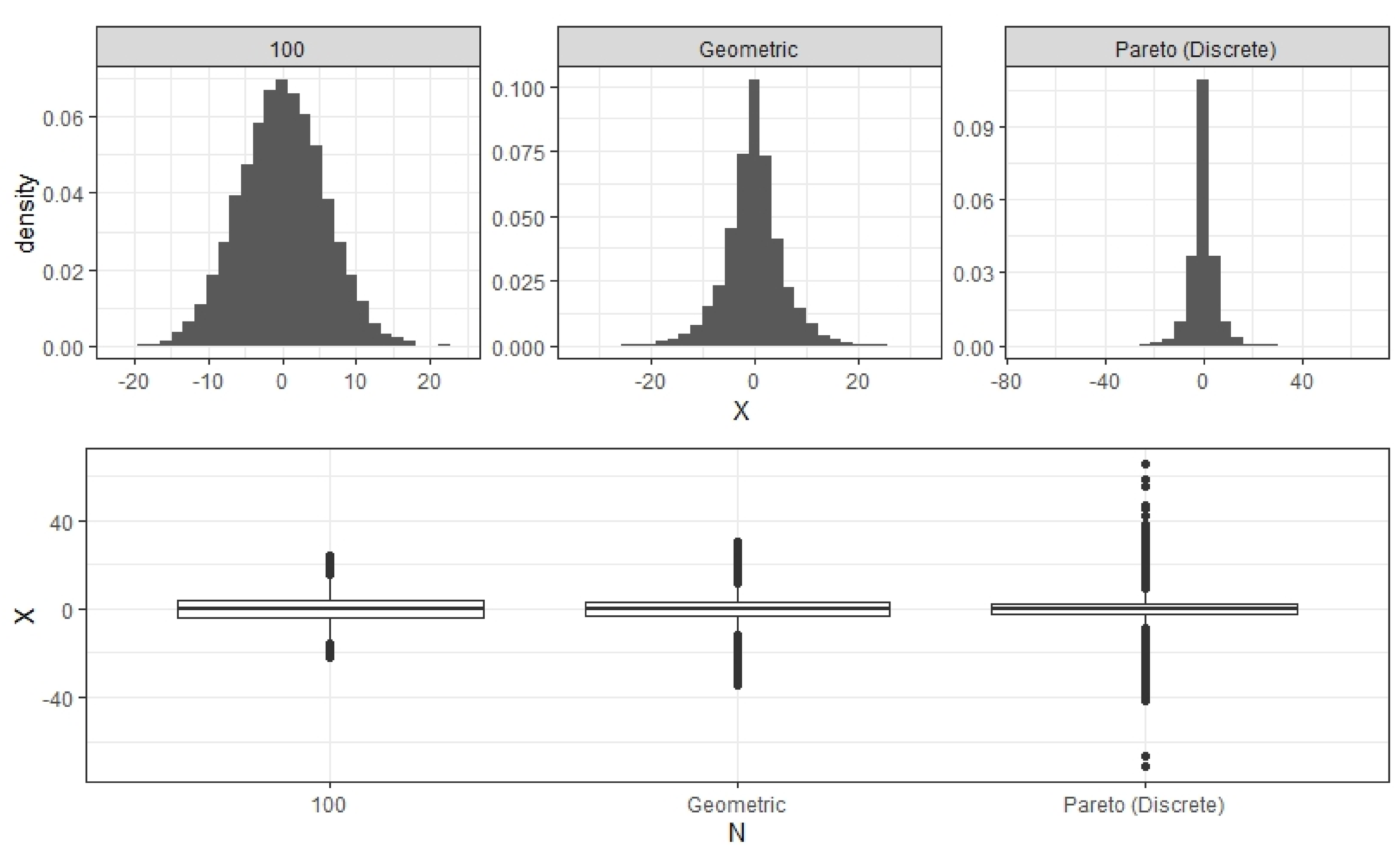

Figure 1: Empirical distributions of sums of uniform random variables $[-1, 1]$ using 10,000 simulations. Top: Histogram plots. Bottom: Boxplots. Left: Deterministic sum of 100. Middle: Random sum over Geometric, N where $\mathbb{E}(N) = 100$. Right: Random sum over Discrete Pareto, N where $\mathbb{E}(N) = 100$.

Random Summation Scheme

- An approach is to replace the **deterministic** n in the sum by a **random** number of terms which allows for skewness and heavy tails that goes beyond normality.
- Assume: $\{N_p,\, p \in (0,1)\}$ - family of positive, integer-valued random variables such that $N_p \xrightarrow{p} \infty$ and $pN_p \xrightarrow{d} W$ as $p \to 0$.
- W is a non-negative random variable.
- If $\{\mathbf{X}_i\}$ are IID with finite variance and
$$a_p \sum_{i=1}^{N_p} (\mathbf{X}_i + \mathbf{b}_p) \xrightarrow{d} \mathbf{Y} \text{ as } p \to 0,$$
then the stochastic representation of Y is given by
$$\mathbf{Y} \stackrel{d}{=} \mathbf{m}W + W^{1/2}\mathbf{X}$$
where $\mathbf{m} \in \mathbb{R}^d$, and $\mathbf{X}$ is multivariate normal $N_d(\mathbf{0}, \mathbf{\Sigma})$
- $\mathbf{Y}$ can be skewed and heavy-tailed.
- **Our proposal**: use a family $\{N_p,\, p \in (0,1)\}$ which allows heavy tails.
- The Discrete Pareto (DP), denoted by $DP(\alpha, p)$, gives rise to heavy tails and has the PDF
$$\mathbb{P}(N_p = k) = \left(\frac{1}{1-\alpha(k-1)\log(1-p)}\right)^{\frac{1}{\alpha}} - \left(\frac{1}{1-\alpha k \log(1-p)}\right)^{\frac{1}{\alpha}},$$
where $k \in \mathbb{N}, \alpha > 0, p \in (0,1)$.
- As $p \to 0$, we have $pN_p \xrightarrow{d} W$, where W has Pareto (Type II, aka Lomax) distribution denoted by $P(\alpha)$ and given by the SF
$$\mathbb{P}(W > x) = \left(\frac{1}{1+\alpha x}\right)^{\frac{1}{\alpha}}, \quad x > 0.$$
- **Note**: As $\alpha \to 0$, the $DP(\alpha, p)$ turns into the geometric distribution and the Pareto variable W becomes standard exponential $EXP(1)$ hence we obtain a geometric/exponential dual pair where $\mathbf{Y}$ is multivariate **asymmetric Laplace (AL)** distribution (see, e.g., Kotz et al., 2001).
- We proceed by breaking down W into two **independent components** E and T given by
$$W \stackrel{d}{=} \frac{E}{T}$$
where E is **standard exponential** and T is **gamma** with both shape/scale equal to $1/\alpha$.
- This shows that $\mathbf{Y}$ is a mixture of skew Laplace distributions.
- One can develop basic characteristics and properties of $\mathbf{Y}$ from their **AL** analogs.

Pareto-Normal case (PDF and CDF plots)

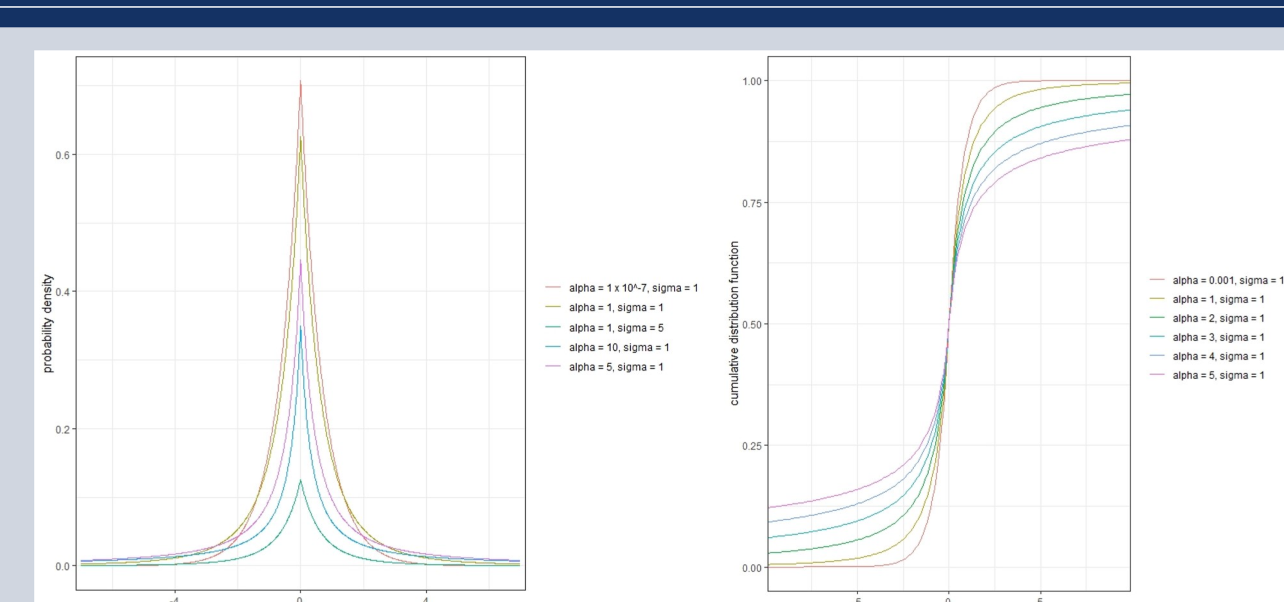

Figure 2: PDF and CDF plots when $W \sim P(\alpha)$ using Monte Carlo (MC) Simulations.

Parameter Estimation

- Using the method of moments estimation (MME) and the EM Algorithm approaches, the parameters were estimated with the following results.

True α	MME $\hat{\alpha}$ (MSE)	EM Algorithm $\tilde{\alpha}$ (MSE)
0.7	0.712071 (1.153833)	0.6695787 (0.067806)
0.7	0.6008105 (0.757494)	0.5678054 (0.065786)
0.7	0.6744689 (1.887113)	0.6240472 (0.072949)
1	0.767261 (0.791302)	0.9590148 (0.055194)
1	0.8240785 (1.144796)	0.9445769 (0.089651)
1	0.8855927 (1.495917)	0.8996565 (0.134264)
5	4.7322077 (0.92606)	4.3190482 (0.650525)
5	4.8769673 (0.816825)	4.1758659 (0.928343)
5	4.8570168 (0.95837)	3.5953683 (2.223461)
True σ	MME $\hat{\sigma}$ (MSE)	EM Algorithm $\tilde{\sigma}$ (MSE)
0.25	0.2738525 (0.008788)	0.2545343 (0.001037)
1	1.1108807 (0.133213)	1.046549 (0.021953)
10	11.399611 (15.939351)	10.391575 (2.420919)
0.25	0.2886389 (0.009714)	0.2558775 (0.00127)
1	1.1214201 (0.145042)	1.0235344 (0.025844)
10	11.2181388 (17.825127)	10.3378394 (4.07772)
0.25	0.2801455 (0.009462)	0.2472176 (0.002892)
1	1.1181092 (0.174252)	0.9889791 (0.053941)
10	11.541484 (20.864484)	9.5775141 (6.066282)

Table 1: Estimates using 150 observations with 1000 simulations.

Summary

A snap-shot of an on-going project leading to:
- Flexible univariate and multivariate probability distributions that incorporate skewness and heavy tails
- Laplace-like univariate distributions, highly peaked at the mode
- New class of Pareto - normal distributions
- Generalizations connected with mixtures of Laplace distributions
- Computational practical challenges with these models: approximation and estimation

References

1. Buddana, A. and Kozubowski, T.J. (2014). Discrete Pareto distribution. Economics Quality Control , 29 (2), 143-156.
2. Klebanov, L.B., Kozubowski, T.J. and Rachev, S.T. (2006). Ill-Posed Problems in Probability and Stability of Random Sums, Nova Science Publishers, New York.
3. Kotz, S., Kozubowski, T., and Podgorski, K. (2012). The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering, and Finance. Springer Science & Business Media.