MATTEX 2018,
CONFERENCE PROCEEDING, v. 1, pp. 43- 48

ON THE COMPOSITE LOGNORMAL-PARETO DISTRIBUTION WITH UNCERTAIN THRESHOLD*

SEZEN MUTALI (CIORABAI)

ABSTRACT: *An important challenge related to composite distribution is the estimation of their unknown threshold. For this purpose, several methods have been proposed in the literature, yielding different results for the same data. Therefore, in this paper we go further on and assume that the threshold is uncertain, allowing it to take values in a closed interval. Based on the arithmetic of intervals, we study the resulting model under continuity and differentiability conditions.*

KEYWORDS: *Composite Lognormal-Pareto Distribution, Uncertain Threshold, Interval arithmetic.*

1 Introduction

A spliced distribution is piecewise built from different distributions defined on mutually disjoint intervals. In particular, a composite distribution is usually built from two components: for example, the composite lognormal-Pareto distribution consists of a lognormal density up to a certain threshold and of a Pareto density from that threshold on. We note that a composite distribution can also be interpreted as a two-component mixture of two distributions singly truncated from above and, respectively, from below (their truncation point being the threshold itself), with corresponding mixing weights. Such distributions are useful to model, e.g., loss data arising in insurance because they are typically positively skewed. Therefore, the usual lognormal-Pareto distribution is well studied in the actuarial literature, starting from the paper of Cooray and Ananda [2]. An important challenge is the estimation of the unknown threshold. For this purpose, several methods have been proposed in the literature, yielding different results for the same data, usually necessitating the examination of all possible positions of the threshold relatively to the data; see, e.g., Teodorescu and Vernic [7] or the R package of Nadarajah and Bakar [4] and the references therein. Moreover, Pigeon and Denuit [5] assumed that the threshold is the outcome of a random variable (r.v.). In this paper, we go further on and assume that the threshold is uncertain, allowing it to take values in a closed interval. Based on the arithmetic of intervals, we study the resulting model under continuity and differentiability conditions. Such an assumption has the purpose to better capture the uncertainty encountered in the data set.

Therefore, the paper is structured in two more sections: in Section 2, we recall the basics of interval arithmetic and the form of the composite lognormal-Pareto distribution with its main characteristics; in Section 3, we introduce the threshold of interval type and study its effect on the parameters, the density, the distribution function and the expected value of the resulting lognormal-Pareto model. We end with a conclusions and further study section.

2 Preliminaries

2.1 Basics of interval arithmetic

Measurement uncertainties can be directly incorporated into calculation by means of the interval arithmetic, i.e., by using closed intervals on the real line represented as $[\underline{x}, \overline{x}] = \{x \in \mathbb{R} : \underline{x} \leq x \leq \overline{x}\}$. Consequently, an arithmetic for intervals and interval valued extensions of functions had

*Partially supported by Scientific Research Grant DDVU 02/91 of MON and Scientific Research Grant RD-08-75/2016 of Shumen University.

been defined with the purpose to provide rigorous enclosures of solutions in various applications (see, e.g., the books [1] and [3]). In the following, we recall several interval arithmetic operation:

1. addition: $[\underline{x},\overline{x}]+[\underline{y},\overline{y}]=[\underline{x}+\underline{y},\overline{x}+\overline{y}]$, $a+[\underline{x},\overline{x}]=[a+\underline{x},a+\overline{x}], a\in\mathbb{R}$;
2. subtraction: $[\underline{x},\overline{x}]-[\underline{y},\overline{y}]=[\underline{x}-\overline{y},\overline{x}-\underline{y}]$;
3. multiplication: $[\underline{x},\overline{x}]\cdot[\underline{y},\overline{y}]=[\underline{m},\overline{m}]$, where:

$$\underline{m}=\min\left(\underline{x}\underline{y},\underline{x}\overline{y},\overline{x}\underline{y},\overline{x}\overline{y}\right),\overline{m}=\max\left(\underline{x}\underline{y},\underline{x}\overline{y},\overline{x}\underline{y},\overline{x}\overline{y}\right);$$

$$a\cdot[\underline{x},\overline{x}]=\begin{cases}[a\underline{x},a\overline{x}] & ,a>0\\ [a\overline{x},a\underline{x}] & ,a<0\end{cases}.$$

2.2 Composite lognormal-Pareto distribution

We shall use the composite lognormal-Pareto density definition given by Scollnik [6]:

$$f(x;\mu,\sigma,a,\theta,r)=\begin{cases} r\frac{1}{x\sigma\sqrt{2\pi}\phi\left(\frac{\ln\theta-\mu}{\sigma}\right)}e^{-\frac{1}{2}\left(\frac{\ln x-\mu}{\sigma}\right)^2} & ,0<x\leq\theta\\ (1-r)\frac{a\theta^a}{x^{a+1}} & ,x>\theta,\end{cases}\tag{1}$$

where $\sigma,\theta,a>0$ and $\mu\in\mathbb{R}$. Also, φ and ϕ denote the standard normal density function and, respectively, distribution function. Imposing the continuity condition $f(\theta_-)=f(\theta_+)$ leads to:

$$r(\theta)=\left(1+\frac{e^{-\frac{1}{2}\left(\frac{\ln\theta-\mu}{\sigma}\right)^2}}{a\sigma\sqrt{2\pi}\phi\left(\frac{\ln\theta-\mu}{\sigma}\right)}\right)^{-1}.$$

We note that $r(\theta)$ is an increasing function in θ. Moreover, a differentiability condition is usually imposed, i.e., $f'(\theta_-)=f'(\theta_+)$, yielding:

$$\frac{\ln\theta-\mu}{\sigma}=a\sigma,\tag{2}$$

from where we obtain:

$$r=\left(1+\frac{\varphi(a\sigma)}{a\sigma\phi(a\sigma)}\right)^{-1}.\tag{3}$$

The distribution function of the composite lognormal-Pareto is given by:

$$F(x;\mu,\sigma,a,\theta,r)=\begin{cases} r\frac{\phi\left(\frac{\ln x-\mu}{\sigma}\right)}{\phi\left(\frac{\ln\theta-\mu}{\sigma}\right)} & ,0<x\leq\theta\\ r+(1-r)\left(1-\left(\frac{\theta}{x}\right)^a\right) & ,x>\theta.\end{cases}\tag{4}$$

The expected value of this composite distribution exists only for $a>1$ and is equal to:

$$\mathbb{E}(f)=\frac{r}{\phi(a\sigma)}e^{\mu+\frac{\sigma^2}{2}}+(1-r)\frac{a\theta}{a-1},a>1.\tag{5}$$

Remark 1. *According to the remark in the introduction, we can rewrite model (1) as a two-component mixture, which, in terms of r.v.s, is expressed as*

$$Z = IX + (1-I)Y,$$

where I is a Bernoulli (r) random variable, X is a lognormal distributed r.v. truncated from above in θ, while Y is a Pareto r.v. defined for $x > \theta$.

3 Composite lognormal-Pareto with interval type threshold

We now assume that the threshold θ is of interval type, i.e., $\theta \in [\underline{\theta}, \overline{\theta}]$. From condition (2) we note that if θ is of interval type then at least one of the other parameters must also be of interval type. Considering the meaning of the parameters, we assume that μ, the lognormal parameter related to the mean, is also of interval type, while the parameters σ, a are single values. Since $\mu = \ln\theta - a\sigma^2$, the corresponding interval for μ results from the interval of θ as $\mu \in [\underline{\mu}, \overline{\mu}]$, where $\underline{\mu} = \ln\underline{\theta} - a\sigma^2, \overline{\mu} = \ln\overline{\theta} - a\sigma^2$; therefore, we shall omit μ from the notation. From formula (3), we also note that since a, σ are now constant values, r is also constant, so we also omit it from the notation.

We start by noting that when θ is of interval type, we no longer have a proper distribution, but a bundle (or fascicle) of composite distributions. Let us now have a look as some density components of this bundle.

In Figure 1, we plotted the density (1) for $a = 0.5, \sigma = 1$ and $\theta \in \{2; 2.3; 2.5; 2.7; 3\} \subset [2, 3]$. In this case, $r = 0.4955$, while $m \in [0.1932, 0.5986]$.

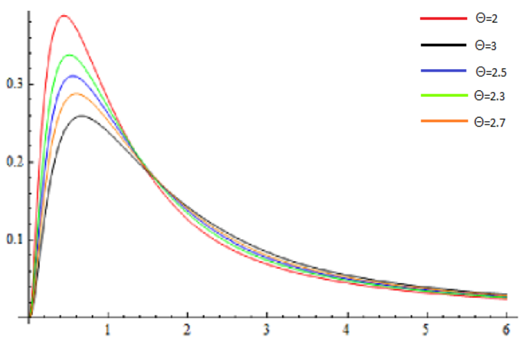

Figure 1: Density shapes for θ of interval type

We shall now see some properties of the resulting fascicle of distributions.

Proposition 1. *When $\theta \in [\underline{\theta}, \overline{\theta}]$ and $x > 0$ is fixed, it holds that*
i) The density $f(x; \sigma, a, \theta)$ is also of interval type;
ii) The distribution function $F(x; \sigma, a, \theta)$ is also of interval type;
iii) The expected value $\mathbb{E}(\sigma, a, \theta) = \mathbb{E}(f)$ is also of interval type.

Proof. i) In this case,

$$f(x;\sigma,a,\theta) = \begin{cases} r\dfrac{\varphi\left(\frac{\ln x - \ln\theta + a\sigma^2}{\sigma}\right)}{x\sigma\phi(a\sigma)} & ,0 < x \le \theta \\ (1-r)\frac{a\theta^a}{x^{a+1}} & ,x > \theta. \end{cases} \tag{6}$$

Since x is fixed and we vary $\theta \in [\underline{\theta}, \overline{\theta}]$, we have three cases:
Case 1: When $x < \underline{\theta} \leq \theta \leq \overline{\theta}$, we use formula:

$$f(x;\sigma,a,\theta) = r\frac{\varphi(\frac{\ln x - \ln\theta + a\sigma^2}{\sigma})}{x\sigma\phi(a\sigma)},$$

which is clearly continuous in θ, hence $f(x;\sigma,a,[\underline{\theta},\overline{\theta}])$ is also an interval.
Case 2: When $\underline{\theta} \leq \theta \leq \overline{\theta} < x$, we use formula:

$$f(x;\sigma,a,\theta) = (1-r)\frac{a\theta^a}{x^{a+1}},$$

which immediately yields $f(x;\sigma,a,[\underline{\theta},\overline{\theta}]) = \left[(1-r)\frac{a\underline{\theta}^a}{x^{a+1}},(1-r)\frac{a\overline{\theta}^a}{x^{a+1}}\right]$.
Case 3: When $\underline{\theta} \leq x \leq \overline{\theta}$, we have two subcases: if $\underline{\theta} \leq \theta \leq x$, then we use the second formula of equation (6) and obtain an interval, while if $x \leq \theta \leq \overline{\theta}$ we use the first formula of equation (6) and obtain another interval. We prove that these two intervals overlap. More precisely, we prove that we have equality at the limit $\theta = x$, i.e.,

$$r\frac{\varphi(\frac{\ln x - \ln x + a\sigma^2}{\sigma})}{x\sigma\phi(a\sigma)} = (1-r)\frac{ax^a}{x^{a+1}} \Leftrightarrow r\frac{\varphi(a\sigma)}{x\sigma\phi(a\sigma)} = (1-r)\frac{a}{x}. \tag{7}$$

To prove this we insert the formula (3) of r and obtain:

$$r\frac{\varphi(a\sigma)}{x\sigma\phi(\alpha\sigma)} = \frac{a\sigma\phi(a\sigma)}{a\sigma\phi(a\sigma)+\varphi(a\sigma)}\cdot\frac{\varphi(a\sigma)}{x\sigma\phi(\alpha\sigma)} = \frac{\varphi(a\sigma)}{a\sigma\phi(a\sigma)+\varphi(a\sigma)}\cdot\frac{a}{x},$$

while

$$(1-r)\frac{a}{x} = \left(1-\frac{a\sigma\phi(a\sigma)}{a\sigma\phi(a\sigma)+\varphi(a\sigma)}\right)\frac{a}{x} = \frac{\varphi(a\sigma)}{a\sigma\phi(a\sigma)+\varphi(a\sigma)}\cdot\frac{a}{x},$$

i.e., the equality (7) holds, which completes the proof of *(i)*.
ii) Based on the distribution function's formula:

$$F(x;\sigma,a,\theta) = \begin{cases} r\frac{\phi\left(\frac{\ln x - \ln\theta + a\sigma^2}{\sigma}\right)}{\phi(a\sigma)} & ,0 < x \leq \theta \\ r + (1-r)\left(1-\left(\frac{\theta}{x}\right)^a\right) & ,x > \theta, \end{cases}$$

the proof is very similar with the above one, hence we omit it.
iii) For $a > 1$, we rewrite:

$$\mathbb{E}(\sigma,a,\theta) = \frac{r}{\phi(a\sigma)}e^{\ln\theta - a\sigma^2 + \frac{\sigma^2}{2}} + (1-r)\frac{a\theta}{a-1},$$

which is continuous and increasing in θ and, clearly, $\mathbb{E}(\sigma,a,[\underline{\theta},\overline{\theta}]) = \left[\mathbb{E}(\sigma,a,\underline{\theta}),\mathbb{E}(\sigma,a,\overline{\theta})\right]$. This completes the proof. □

In Figure 2, we plotted the expected value for $\theta \in [1,3], \sigma = 1$ and various $a \in \{1.5, 2, 3\}$. In Figure 3, we plotted the expected value for $\theta \in [1,3], a = 2$ and various $\sigma \in \{0.5, 1, 1.3\}$. We note that both plots support the conclusion of *(iii)* in the above proposition.

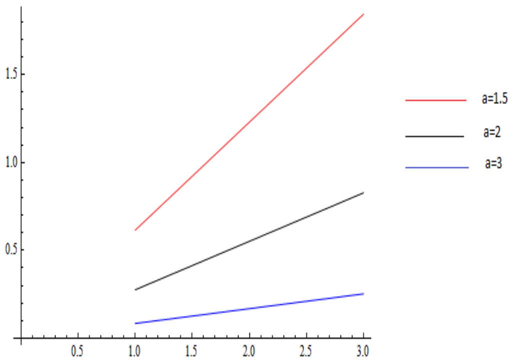

Figure 2: Expected value for θ of interval type and various a

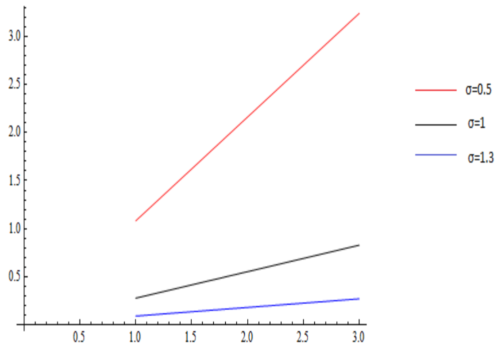

Figure 3: Expected value for θ of interval type and various σ

4 Conclusions

In this paper, we studied the effect of an interval type threshold on the composite lognormal-Pareto distribution and concluded that the density, the distribution function and the expected value also become of interval type. Because of the restrictions imposed by the continuity and differentiability conditions, we think that this study should be extended to the situation where the differentiability condition is relaxed. Moreover, as further work, we intend to study the same distribution under the assumption that the threshold is a fuzzy number and to apply the results on some real data.

Acknowledgements. The author is very grateful to the anonymous referee for the valuable comments.

REFERENCES:

[1] Alefeld, G. and Herzberger, J. (2012). Introduction to interval computation. Academic press.

[2] Cooray, K. and Ananda, M. M. (2005). Modeling actuarial data with a composite lognormal-Pareto model. Scandinavian Actuarial Journal, 2005(5), 321-334.

[3] Moore, R. E., Kearfott, R. B. and Cloud, M. J. (2009). Introduction to interval analysis. Siam.

[4] Nadarajah, S. and Bakar, S. (2013). CompLognormal: An R Package for Composite Lognormal Distributions. R journal, 5(2), 98-104.

[5] Pigeon, M. and Denuit, M. (2011). Composite Lognormal–Pareto model with random threshold. Scandinavian Actuarial Journal, 2011(3), 177-192.

[6] Scollnik, D. P. (2007). On composite lognormal-Pareto models. Scandinavian Actuarial Journal, 2007(1), 20-33.

[7] Teodorescu, S. and Vernic, R. (2013). On composite Pareto models. Mathematical Reports, 15(65), 11-29.

Sezen Mutali (Ciorabai),
Ovidius University of Constanta, Romania
E-mail: m.sezenn@yahoo.com