We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Distributions and Composite Models for Size-Type Data

Yves Dominicy and Corinne Sinner

Additional information is available at the end of the chapter

Abstract

In the first part of this chapter, we present a sample of the best known and most used classical size distributions with their main statistical properties. In the second part, we introduce the concept of composite models and based on the size distributions of the first part, we describe those which already exist in the literature. In the last part of this chapter, we apply the described statistical size distributions and some of the composite models to two real data examples and compare their goodness-of-fit.

Keywords: size distributions, composite models, lognormal, Pareto, Weibull

1. Introduction

In statistical modeling, the continuous aim is to look for the probability law, which best describes the observations arising from a given field and which should represent the underlying data-generating process. The obtained probability distributions should possess desirable properties such as the flexibility of modeling different shapes and remain of a tractable form. This research avenue was initiated in the nineteenth century by famous mathematicians as Adolphe Quetelet, Sir Francis Galton, or Vilfredo Pareto, and since then it has never ceased. Nowadays, it still remains among the highly treated topics in statistics as shown by the large quantity of scientific papers recently published on the subject (see for instance the review papers [1] and [2]). The actual appeal for this topic is easily explained by the availability of large data sets in various scientific domains, making it essential and necessary to do further research on this subject.

In this chapter, we concentrate on probability distributions that analyze size-type data. By size distributions, we mean probability laws designed to model data that only take positive values. Positive observations appear naturally in different fields: survival analysis [3, 4], environmental science [5], network traffic modeling [6], economics [7, 8], hydrology [9], and actuarial

science [10]. Given the range of various domains of application, there exists a plethora of different size distributions and it is still a very active research area [11, 12].

The structure of this chapter is as follows. In Section 2, we review the most used and well-known size distributions, and state their main statistical properties. Section 3 introduces the notion of composite models and gives a small review of the composite models in the literature, based on the size distributions depicted in Section 2. In Section 4, we apply the described size distributions of Section 2 and some of the composite models of Section 3 to two real data sets, namely, an insurance data set and an Internet traffic data set. Finally, Section 5 concludes.

2. Review of size distributions

We describe here a sample of the best known and most used size distributions. We will state their probability density function (p.d.f) and their cumulative density function (c.d.f), show their moments and their quantile function, and give the estimators obtained via maximum likelihood estimation. More specifically, we take a closer look at the lognormal, Pareto, generalized Lomax, and generalized extreme value distributions.

2.1. The lognormal distribution

The English statistician Sir Francis Galton stated that in some situations it was preferable to measure the location of a distribution with the geometric mean instead of the arithmetic mean [13]. Indeed, laws of nature often behave in multiplicative ways; thus, the geometric mean becomes more appropriate as a measure of central tendency than the arithmetic mean. As a reply to Galton's request, the Scottish physician Donald McAlister established in 1879 a theory of the exponentiated (or multiplicative) normal distribution [14], this became to be known as the lognormal distribution.

Let X be a positive random variable (r.v.) such that $\log X =^D Y$ is normally distributed with parameters $\mu \in \mathbb{R}$ and $\sigma > 0$. The r.v. X then has a lognormal distribution, $X \sim LN(\mu, \sigma^2)$, with probability density function (p.d.f.)

$$f(x; \mu, \sigma^2) = \frac{1}{x\sqrt{2\pi}\sigma} e^{-\frac{(\log x - \mu)^2}{2\sigma^2}}, \ x > 0. \tag{1}$$

The location parameter $\mu \in \mathbb{R}$ and the scale parameter $\sigma > 0$ are characteristic for the r.v. $\log X$. However, by the exponential transformation, the geometric mean e^μ becomes a scale parameter, as depicted in **Figure 1**, and the multiplicative standard deviation e^σ appears as shape parameter impacting the skewness (see **Figure 2**).

If random variability enjoys multiplicative effects, as stated by Galton, then a lognormal distribution must be the result. This establishes the basis of the multiplicative central limit theorem, which asserts that the geometric means of nonlognormal random variables are approximated by a lognormal distribution.
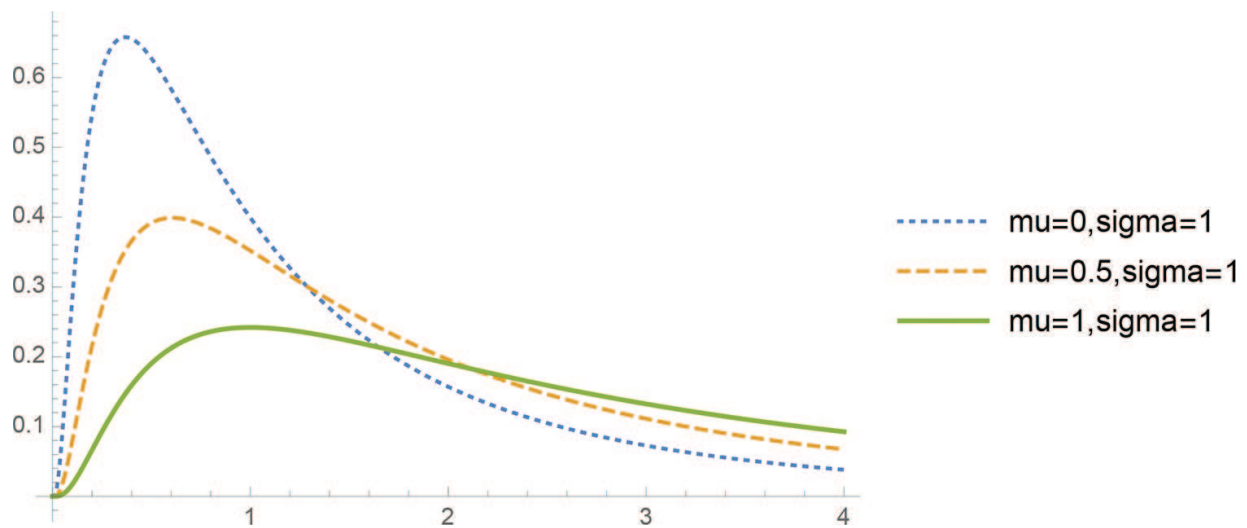
Figure 1. Density plots of the lognormal distribution with varying location parameter μ and fixed scale parameter σ.
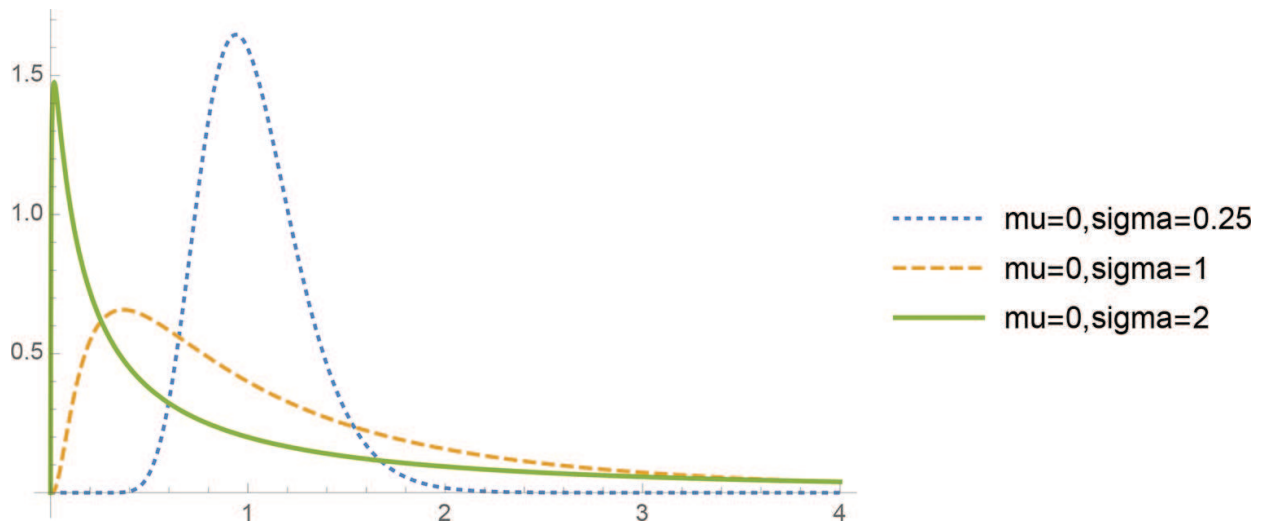

Figure 2. Density plots of the lognormal distribution with varying scale parameter σ and fixed location parameter μ.

The cumulative distribution function (c.d.f.) of the lognormal law is related to the c.d.f. of the normal distribution:

$$F(x; \mu, \sigma^2) = \Phi\left(\frac{\log x - \mu}{\sigma}\right), \ x > 0, \tag{2}$$

where $\Phi(.)$ represents the c.d.f. of a standard normal distribution.

The moments of order r are conveniently expressed as $E(X^r) = e^{r\mu + \frac{r^2\sigma^2}{2}}$. Hence, the mean is given by $E(X) = e^{\mu + \frac{\sigma^2}{2}}$, and the variance by $V(X) = e^{2\mu + \sigma^2}(e^{\sigma^2} - 1)$. The lognormal is a unimodal distribution and the unique mode is reached at $x_{mode} = e^{\mu - \sigma^2}$. By comparing the mean and the mode, we note that for a fixed μ, an increasing σ shifts the mode toward zero while the mean

increases. The quantile function is defined as $F^{-1}(y) = e^{\mu+\sigma\Phi^{-1}(y)}$, for $0 < y < 1$ and where $\Phi^{-1}(.)$ denotes the quantile function of a standard normal distribution.

Thanks to its relationship to the normal distribution, the likelihood function is given by

$$L(x_1, \ldots, x_n | \mu, \sigma^2) = \left(\prod_{i=1}^{n} \frac{1}{x_i}\right)\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\sum_{i=1}^{n} \frac{(\log x_i - \mu)^2}{2\sigma^2}}, \tag{3}$$

and hence the log-likelihood function can be expressed as

$$l(x_1, \ldots, x_n | \mu, \sigma^2) = -\sum_{i=1}^{n} \log x_i - \frac{n}{2}\log 2\pi - n\log\sigma - \sum_{i=1}^{n}\frac{(\log x_i - \mu)^2}{2\sigma^2}. \tag{4}$$

The maximum likelihood estimators for the mean and the scale are given by $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}\log x_i$ and $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(\log x_i - \hat{\mu})^2$, respectively.

The lognormal distribution is widely used to describe natural phenomena. In finance, the Black-Scholes model, which is a mathematical model containing derivative instruments, assumes the underlying derivative price to have a lognormal distribution [15]. In economics, income data are often modeled by a lognormal distribution [16], which can be easily explained as follows: a very low percentage of earners have very low income. To gain averaged revenue is frequent, whereas an elevated income is rare. In actuarial sciences, the law is assumed to fit well some types of insurance losses [17, 18]. In 1931, the French economist and engineer Robert Pierre Louis Gibrat stated that the firm size follows a lognormal distribution as its proportional growth rate is independent of its absolute size. Other applications can be found in biology [19, 20] or in linguistics to model the number of words in a sentence [21].

2.2. The Pareto distribution

The Italian economist and engineer Vilfredo Pareto observed in 1896 that in many populations the power law $cx^{-\alpha}$, for some constant $c > 0$ and some exponent $\alpha > 0$, was an appropriate approximation of the number of individuals with income exceeding a given threshold x_0 (see for instance [22, 23]). These power laws assume that small values of x are very frequent, while large occurrences are extremely rare. Their form implies that all power laws with a particular scaling exponent are equivalent up to constant factors since each is simply a scaled version of the others. This produces the linear relationship when logarithms are taken of both $f(x)$ and x, which denotes the signature of power laws. Such distributions are known as Pareto-type distributions.

The p.d.f. of a r.v. X having a Pareto (type I) distribution with parameters $\alpha > 0$ and $x_0 > 0$ is given by

$$f(x; \alpha, x_0) = \frac{\alpha}{x_0}\left(\frac{x_0}{x}\right)^{\alpha+1}, \ x \geq x_0. \tag{5}$$

The location parameter x_0 represents the lower bound of the data set and the shape parameter α is called the tail index or as well the Pareto index, and hence regulates the tail as can be seen in **Figure 3**. Note that a decreasing value of α implies a heavier tail.
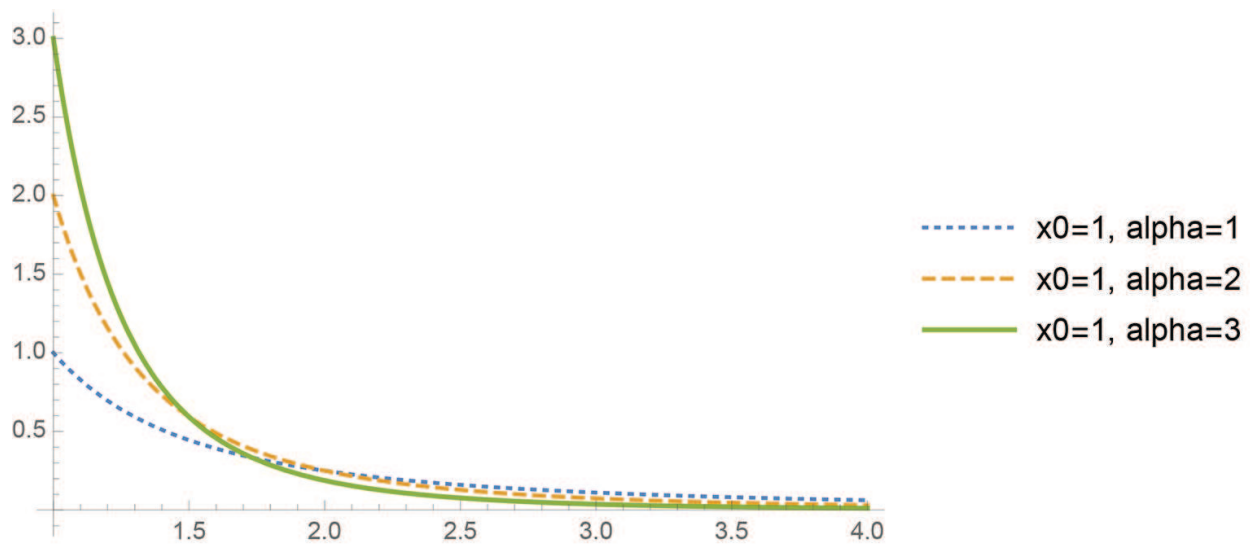
Figure 3. Density plots of the Pareto distribution with varying shape parameter α and fixed location parameter $x_0 = 1$.

The c.d.f. of the Pareto law is given by

$$F(x; \alpha, x_0) = 1 - \left(\frac{x_0}{x}\right)^{\alpha}, \; x \geq x_0. \tag{6}$$

For $\alpha > r$, the r-th moment of the Pareto distribution is given by $E(X^r) = \frac{\alpha x_0^r}{\alpha - r}$. The mean and the variance are then, respectively, $E(X) = \frac{\alpha x_0}{\alpha - 1}$ for $\alpha > 1$ and $V(X) = \frac{\alpha x_0^2}{(\alpha-1)^2(\alpha-2)}$ for $\alpha > 2$. The quantile function is expressed as $F^{-1}(y) = \frac{x_0}{(1-y)^{\frac{1}{\alpha}}}$, for $0 < y < 1$. Being an unimodal law, the Pareto distribution reaches its peak at $x_{mode} = x_0$. As x_0 represents the minimum value of x, its estimation is straightforward: $\hat{x}_0 = \min_{i=1,\ldots,n} x_i$. The likelihood function is given by

$$L(x_1, \ldots, x_n | \alpha, x_0) = \alpha^n x_0^{n\alpha} \prod_{i=1}^{n} \left(\frac{1}{x_i}\right)^{\alpha+1}$$

and to estimate the parameter α, we maximize the following log-likelihood function

$$l(x_1, \ldots, x_n | \alpha, x_0) = n \log \alpha + n\alpha \log x_0 - (\alpha + 1) \sum_{i=1}^{n} \log x_i \tag{7}$$

which yields the maximum likelihood estimator $\hat{\alpha} = \frac{n}{\sum_{i=1}^{n} \log \frac{x_i}{x_0}}$. Let us note that the maximum likelihood estimator of the tail index α corresponds to the popular Hill estimator [24], which is an estimator for the extreme value index in the extreme value theory. For a review on the Hill estimator, we refer the interested reader to reference [25]. Let us note that often the focus lies more on the power law probability distribution, which is a distribution whose density has approximately the form $L(x)x^{-\alpha}$, where $\alpha > 1$ and $L(x)$ is a slowly varying function. In many situations, it is convenient to assume a lower bound x_0 from which the law holds. Combining those two cases yields the Pareto-type distributions, or as well known in extreme value theory as distributions with regularly varying tails.

A generalization of the Pareto law is the so-called generalized Pareto distribution, and it regroups the Pareto type I, II, III, and IV distributions. The Pareto type IV contains the other types as special cases and hence as well other size distributions belonging to the different types as for instance the Lomax distributions [26]. This latter distribution belongs to the Pareto type II, and its p.d.f. is given by

$$f(x; \alpha, k) = \frac{\alpha}{k} \left(1 + \frac{x}{k}\right)^{-(\alpha+1)}, \; x > 0, \tag{8}$$

with shape parameter $\alpha > 0$ and scale parameter $k > 0$. It can be interpreted as a shifted Pareto type I distribution.

A generalization of the Pareto type I distribution is the Stoppa distribution [27], which comes from a power transformation of the Pareto c.d.f. and yields the following p.d.f.

$$f(x; \alpha, \delta, x_0) = \delta \alpha x_0^\alpha x^{-(\alpha+1)} \left(1 - \left(\frac{x}{x_0}\right)^{-\alpha}\right)^{\delta-1}, \; x > x_0, \tag{9}$$

with shape parameters $\alpha > 0$, $\delta > 0$ and location parameter $x_0 > 0$. If $\delta = 1$, we get the Pareto type I distribution. However, if the shape parameter $\delta > 1$, the Stoppa distribution presents a heavier tail than the classical Pareto law.

The Pareto distribution is often used to model fire losses in actuarial sciences [28, 29] as well as in reinsurance to approximate large losses. Originally, it was used to describe the income distribution and the allocation of wealth [22], but nowadays it is also used to model, for instance, areas burnt in forest fires or the file sizes of Internet traffic data [30]. Note that, in general, in empirical applications, the Pareto distribution does not fit for all the values but rather is used to fit their upper tail, i.e., large values. Hence, in order to fit a distribution to all the values, one often uses a composite model (see Section 3) which combines two distributions where one of both is the Pareto law.

2.3. The generalized Lomax distribution

The generalized Lomax (GL) distribution, also known as the exponentiated Lomax distribution, was introduced by Abdul-Moniem and Abdel-Hameed in 2012 [31] by powering the c.d.f. of the Lomax distribution to a positive real number.

The p.d.f. of a r.v. X following a generalized Lomax distribution with parameters $a > 0$, $b > 0$, and $k > 0$ corresponds to

$$f(x; a, b, k) = \frac{ab}{k} \left(1 + \frac{x}{k}\right)^{-(a+1)} \left(1 - \left(1 + \frac{x}{k}\right)^{-a}\right)^{b-1}, \; x > 0. \tag{10}$$

The shape parameter a regulates the heaviness of the tail, as can be seen in **Figure 4** and the shape parameter b controls the skewness (see **Figure 5**). The parameter k is a scale parameter as depicted in **Figure 6**.
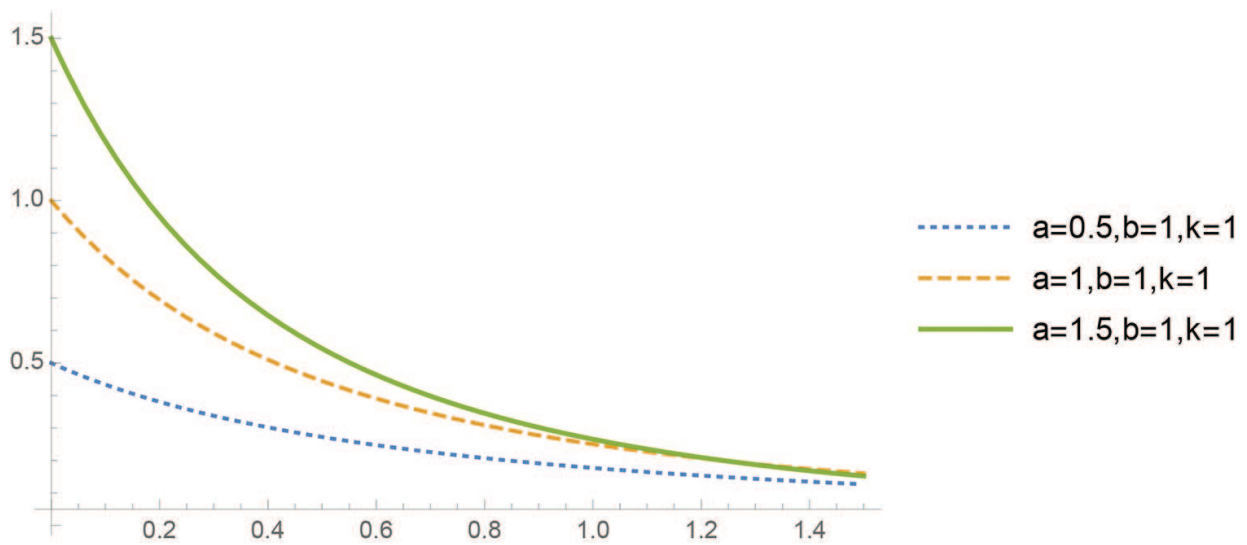
Figure 4. Density plots of the GL distribution with varying shape parameter a and fixed shape parameter b and scale parameter k.
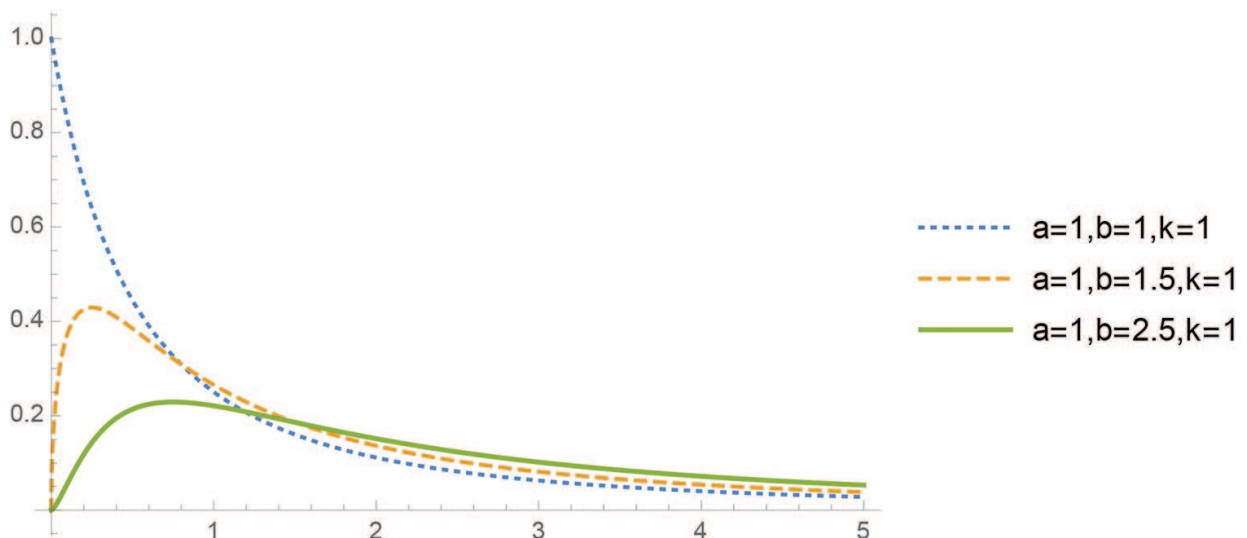

Figure 5. Density plots of the GL distribution with varying shape parameter b and fixed shape parameter a and scale parameter k.

The c.d.f. is expressed as

$$F(x;a,b,k) = \left(1-\left(1+\frac{x}{k}\right)^{-a}\right)^{b}, \quad x > 0. \tag{11}$$

The moments of order r are given by $E(X^r) = bk^r \sum_{i=1}^{r} \binom{r}{i}(-1)^i B\left(1-\frac{1}{a}(r-i),b\right)$, yielding

$E(X) = bkB\left(1-\frac{1}{a},b\right)-k$ for the mean and $V(X) = bk^2\left(B\left(1-\frac{2}{a},b\right) - b\left(B\left(1-\frac{1}{a},b\right)\right)^2\right)$ for the
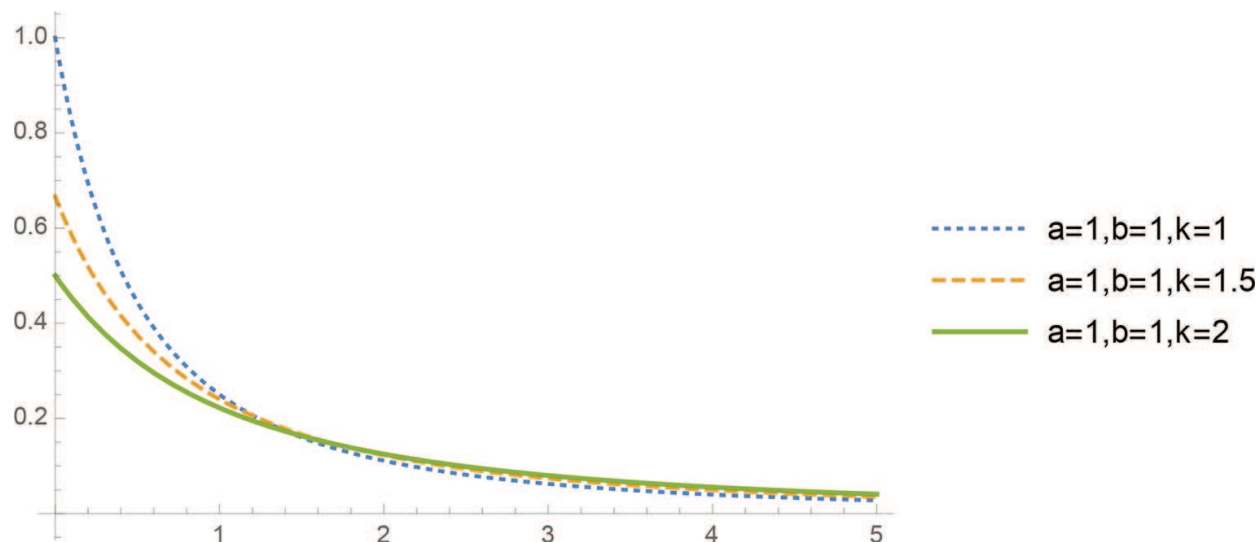
Figure 6. Density plots of the GL distribution with varying scale parameter k and fixed shape parameters a and b.

variance. The inverse c.d.f. is given by $F^{-1}(y) = k\left(\left(1 - y^{\frac{1}{b}}\right)^{-\frac{1}{a}} - 1\right)$, for $0 < y < 1$, and the unique mode is reached at $x_{mode} = k\left(\left(\frac{ab+1}{a+1}\right)^{\frac{1}{a}} - 1\right)$.

The likelihood function is given by

$$L(x_1, \ldots, x_n | a, b, k) = \left(\frac{ab}{k}\right)^n \prod_{i=1}^{n} \left(1 + \frac{x_i}{k}\right)^{-(a+1)} \prod_{i=1}^{n} \left(1 - \left(1 + \frac{x_i}{k}\right)^{-a}\right)^{b-1} \tag{12}$$

and hence the following log-likelihood function is obtained

$$l(x_1, \ldots, x_n | a, b, k) = n \log \frac{b}{k} - (a+1) \sum_{i=1}^{n} \log\left(1 + \frac{x_i}{k}\right) + (b-1) \sum_{i=1}^{n} \log\left(1 - \left(1 + \frac{x_i}{k}\right)^{-a}\right). \tag{13}$$

The calculated score functions are expressed by

$$\frac{\partial l(x_1, \ldots, x_n | a, b, k)}{\partial a} = \frac{n}{a} - \sum_{i=1}^{n} \log(1 + \frac{x_i}{k}) - a(b-1) \sum_{i=1}^{n} \frac{(1 + \frac{x_i}{k})^{-a} \log(1 + \frac{x_i}{k})}{1 - (1 + \frac{x_i}{k})^{-a}},$$

$$\frac{\partial l(x_1, \ldots, x_n | a, b, k)}{\partial b} = \frac{n}{b} + \sum_{i=1}^{n} \log\left(1 - \left(1 + \frac{x_i}{k}\right)^{-a}\right), \tag{14}$$

and

$$\frac{\partial l(x_1, \ldots, x_n | a, b, k)}{\partial k} = -\frac{n}{k} + \frac{a+1}{k^2} \sum_{i=1}^{n} \frac{x_i}{1 + \frac{x_i}{k}} - \frac{a(b-1)}{k^2} \sum_{i=1}^{n} \frac{(1 + \frac{x_i}{k})^{-(a+1)} x_i}{1 - (1 + \frac{x_i}{k})^{-a}}, \tag{15}$$

which have to be solved numerically by equating them to 0 in order to find the estimated parameters.
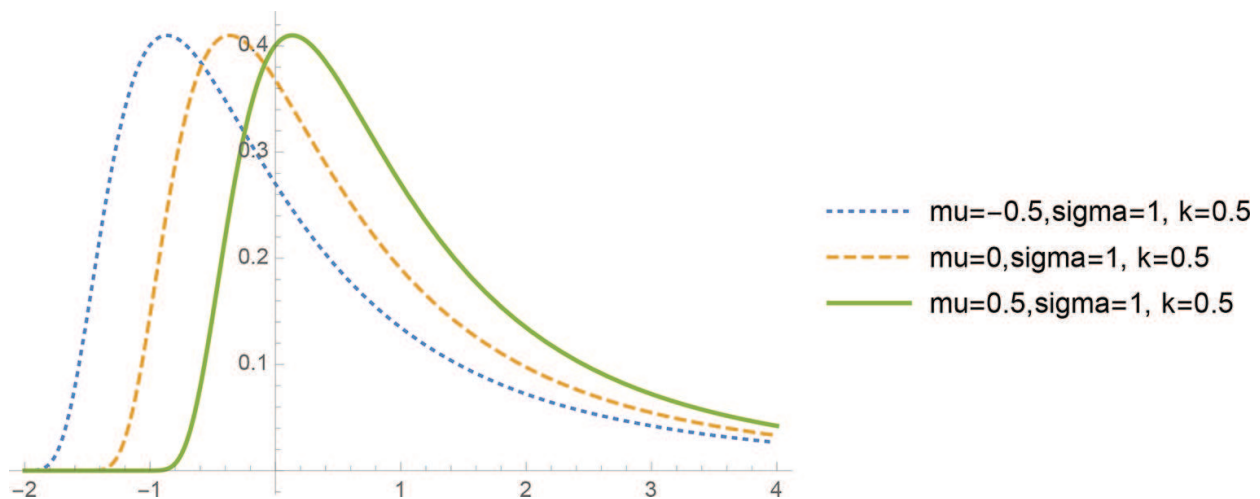
Figure 7. Density plots of the GEV distribution with varying location parameter μ.

The Lomax distribution is used to model income data, wealth allocation, and actuarial claim sizes [10]. The GL distribution is used to measure the breaking stress of carbon fibers [32], the survival times of patients getting chemotherapy treatment [33], and the number of successive failure of the air-conditioning system in airplanes [33].

2.4. The generalized extreme value distribution

The generalized extreme value (GEV) distribution is well known in extreme value theory as it combines the Gumbel, Fréchet, and Weibull distributions, which are also known as type I, II, and III extreme value distributions. The GEV distribution is the only possible limit distribution of properly normalized maxima of a sequence of independent and identically distributed random variables and this result arises from the central limit theorem of Fisher and Tippett [34]. Therefore, the GEV distribution is also known as Fisher-Tippett distribution in extreme value theory.

The GEV distribution has p.d.f.

$$f(x;\mu,\sigma,k) = \frac{1}{\sigma}\left(1 + k\left(\frac{x-\mu}{\sigma}\right)\right)^{-1-\frac{1}{k}} e^{-\left(1+k\left(\frac{x-\mu}{\sigma}\right)\right)^{-\frac{1}{k}}}, \tag{16}$$

if $1 + k\frac{x-\mu}{\sigma} > 0$, with location parameter $\mu \in \mathbb{R}$ (see **Figure 7**), scale parameter $\sigma > 0$ (see **Figure 8**), and shape parameter $k \in \mathbb{R}$, which governs the shape and the heaviness of the tail of the distribution, as can be seen in **Figure 9**.

Its c.d.f. is given by

$$F(x;\mu,\sigma,k) = e^{-\left(1+k\left(\frac{x-\mu}{\sigma}\right)\right)^{-\frac{1}{k}}}, \tag{17}$$

if $1 + k\frac{x-\mu}{\sigma} > 0$. The mean is defined as $E(X) = \mu - \frac{\sigma}{k} + \frac{\sigma}{k}\Gamma(1-k)$ and the variance as $V(X) = \frac{\sigma^2}{k^2}\left(\Gamma(1-2k) - \Gamma(1-k)^2\right)$, both expressed in terms of the Gamma function. The quantile
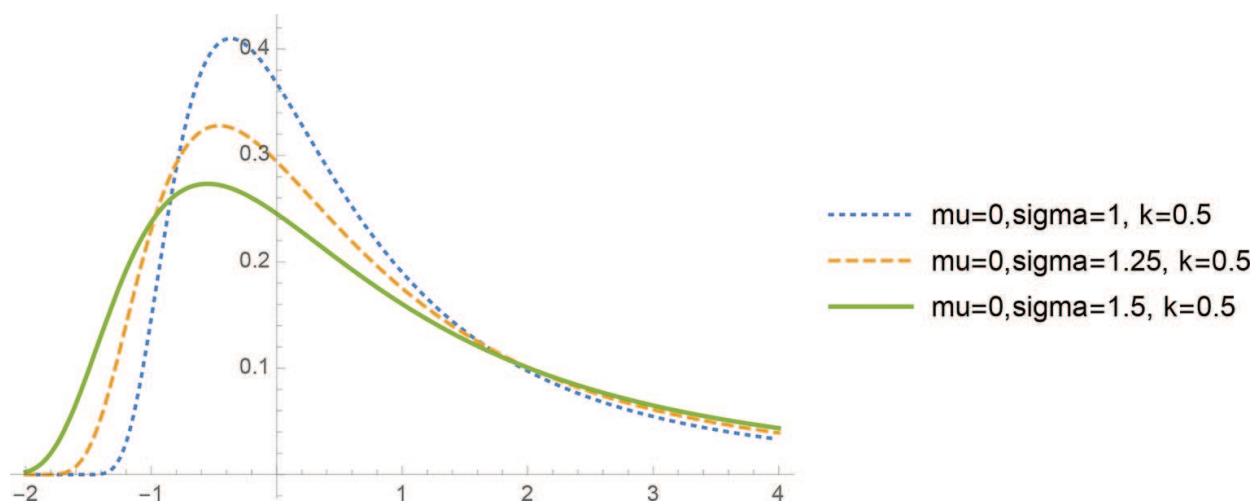
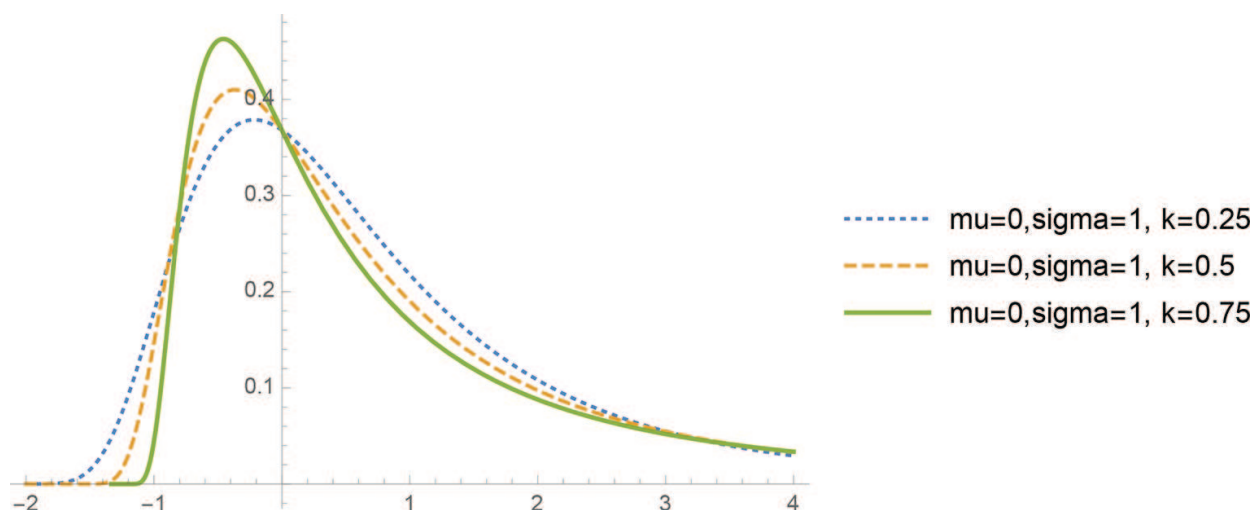**Figure 8.** Density plots of the GEV distribution with varying scale parameter σ.

Figure 9. Density plots of the GEV distribution with varying shape parameter k.

function is given by $F^{-1}(y) = \mu + k\sigma\left((-\log y)^{-k} - 1\right)$, for $0 < y < 1$ and the unique mode is reached at $x_{mode} = \mu + \frac{\sigma}{k}\left((1+k)^{-k} - 1\right)$. Depending on the value of the parameter k, the GEV reduces to one of the following special cases: If $k = 0$, we obtain the Gumbel distribution, if $k > 0$, we get the Fréchet distribution, and if $k < 0$, the Weibull distribution is obtained. The parameters of the GEV distributions are estimated using the maximum likelihood approach.

We will now focus more closely on one of the three GEV distributions, namely, the Weibull distribution, which got its name from the Swedish engineer and scientist Waloddi Weibull, who analyzed it in detail in 1951. We take a look at this law as it is often used for size-type data and it is considered as an alternative to the lognormal distribution for the construction of composite models (see Section 3). The Weibull distribution belongs to power laws with an exponential cut-off; this means it is a power law multiplied by an exponential function. In these distributions, the exponential decay term overpowers the power law behavior for very large values.

The p.d.f. of the Weibull distribution is given by

$$f(x; \sigma, \tau) = \frac{\tau}{\sigma} \left(\frac{x}{\sigma}\right)^{\tau-1} e^{-\left(\frac{x}{\sigma}\right)^{\tau}}, \; x \geq 0, \tag{18}$$

with shape parameter $\tau > 0$ governing the heaviness of the tail and scale parameter $\sigma > 0$.

The distribution has c.d.f.

$$F(x; \sigma, \tau) = 1 - e^{-\left(\frac{x}{\sigma}\right)^{\tau}}, \; x \geq 0. \tag{19}$$

The quantile function is given by $F^{-1}(y) = \sigma(-\log(1-y))^{\frac{1}{\tau}}$, for $0 < y < 1$. The r-th moment is given by $E(X^r) = \sigma^r \Gamma\left(1 + \frac{r}{\tau}\right)$. Hence, the expectation and the variance are expressed as $E(X) = \sigma\Gamma\left(1 + \frac{1}{\tau}\right)$ and $V(X) = \sigma^2 \left(\Gamma\left(1 + \frac{2}{\tau}\right) - \Gamma\left(1 + \frac{1}{\tau}\right)^2\right)$, respectively. The Weibull distribution is unimodal and for $\tau > 1$ it reaches the mode at $x_{mode} = \sigma\left(\frac{\tau-1}{\tau}\right)^{\frac{1}{a}}$ and for $\tau = 1$ the mode is reached at 0.

The parameters of the Weibull distribution are estimated via the maximum likelihood method. The corresponding likelihood and log-likelihood functions are given respectively by

$$L(x_1, \ldots, x_n | \sigma, \tau) = \frac{\tau^n}{\sigma^n} \prod_{i=1}^{n} \left(\frac{x_i}{\sigma}\right)^{\tau-1} e^{-\sum_{i=1}^{n}\left(\frac{x_i}{\sigma}\right)^{\tau}}, \tag{20}$$

and

$$l(x_1, \ldots, x_n | \sigma, \tau) = n \log \tau - n \, \tau \log \sigma + (\tau-1) \sum_{i=1}^{n} \log(x_i) - \sum_{i=1}^{n} \left(\frac{x_i}{\sigma}\right)^{\tau}. \tag{21}$$

The maximum likelihood estimator for the scale parameter σ, given τ, is $\hat{\sigma}^{\tau} = \frac{1}{n}\sum_{i=1}^{n} x_i^{\tau}$ and the maximum likelihood estimator for the shape parameter τ is given by an implicit function which has to be solved numerically: $\hat{\tau}^{-1} = \frac{\sum_{i=1}^{n} x_i^{\tau} \log(x_i)}{\sum_{i=1}^{n} x_i^{\tau}} - \frac{1}{n}\sum_{i=1}^{n} \log(x_i)$.

In risk management, finance, and insurance, the risk measure "Value at Risk" is assessed by considering the GEV distribution [35, 36]. They are as well used in hydrology [37, 38], telecommunications [39], and meteorology [40]. In material sciences, the Weibull is widely used thanks to its flexibility [41]. Other examples include wind speed distributions [42], forecasting technological change [43], the size of reinsurance claims [10], hydrology [9], and areas burnt in forest fires [44].

3. Composite models

Given the wealth of distinct size distributions, as can be seen from the small sample of size distributions described in the previous section, the practitioner is often confronted to the following question: Which size distribution shall he/she use in which situation? The variation

in the shapes, size distributions can take, for instance between the Pareto distribution and the lognormal distribution, renders the choice very complicated in practice.

For example, insurance companies face sometimes losses, which emerge from a combination of moderate and large claims. In order to model those large losses, the Pareto distribution seems to be the size distribution favored by practitioners. However, when losses consist of smaller values with high frequencies and larger losses with low frequencies, the lognormal or the Weibull distributions are preferred [45]. Nevertheless, no classical size distribution provides an acceptable fit for both small and large losses. On one hand, the Pareto fits well the tail, but on the other hand, lognormal and Weibull distributions produce an overall good fit, but fit badly the tail.

A solution to this dilemma comes from the composite parametric models introduced in 2005 by Cooray and Ananda [46]. The idea of the composite models is to join together two weighted distributions at a given threshold value. In statistical terms, let X be a r.v. and denote by $f_1(.)$ the p.d.f. of the first distribution and by $f_2(.)$ the p.d.f. of the second distribution. Let $F_1(.)$ and $F_2(.)$ be the corresponding c.d.f., respectively. Scollnik [47] noticed that the p.d.f. of a composite model can then be expressed as

$$f(x) = \begin{cases} c\, f_1^*(x), & -\infty < x \le \theta \\ (1-c)f_2^*(x), & \theta < x < \infty, \end{cases} \qquad (22)$$

where c is a normalization constant in [0,1], θ represents the threshold value, $f_1^*(x) = \frac{f_1(x)}{F_1(\theta)}$ for $-\infty < x \le \theta$, and $f_2^*(x) = \frac{f_2(x)}{1-F_2(\theta)}$ for $\theta < x < \infty$. In our setting, the considered composite models piece together two different size distributions with different shapes and tail-weights at a specific threshold. As size distributions are only for positive values, the p.d.f. of a composite model is rewritten as

$$f(x) = \begin{cases} c\, f_1^*(x), & 0 < x \le \theta \\ (1-c)f_2^*(x), & \theta < x < \infty, \end{cases} \qquad (23)$$

where $0 \le c \le 1$. The composite model can as well be interpreted as a two-component mixture model with mixing weights c and $(1-c)$. Hence, it can be seen as a convex sum of two density functions $f(x) = c\, f_1^*(x) + (1-c)\, f_2^*(x)$, as noted in [47].

As we have a threshold that cuts the composite model distribution into two, from a mathematical point of view, we need continuity and differentiability conditions at the threshold to yield a smooth density function. In order to make $f(x)$ continuous, the following condition $f(\theta^-) = f(\theta^+)$ is imposed and yields

$$c = \frac{f_2(\theta)F_1(\theta)}{f_2(\theta)F_1(\theta) + f_1(\theta)\left(1 - F_2(\theta)\right)}. \qquad (24)$$

The differential condition at the threshold value is given by $f'(\theta^-) = f'(\theta^+)$ and yields

$$c = \frac{f_2'(\theta)F_1(\theta)}{f_2'(\theta)F_1(\theta) + f_1'(\theta)\left(1 - F_2(\theta)\right)}. \tag{25}$$

If we combine the two results for the normalization constant c, we obtain the additional restriction for θ, i.e., $\frac{f_1(\theta)}{f_2(\theta)} = \frac{f_1'(\theta)}{f_2'(\theta)}$. Let us remark that in reference [48] they use a mode-matching procedure instead and state that it gives much simpler derivation of the model and allows for an easier implementation with any distribution which has a mode that has a closed form expression. Instead of having as threshold value θ, they use the modal value x_m. Denote by x_{m1} and x_{m2} the modes of the distributions used by the first and second components of the composite model, then the mode-matching conditions are $x_{m1} = x_{m2}$ and $f^*(x_{m1}) = f^*(x_{m2})$. The latter implies the continuity condition, and the former equality allows dropping the labels 1 and 2, which yields the following condition

$$c = \frac{f_2(x_m)F_1(x_m)}{f_2(x_m)F_1(x_m) + f_1(x_m)\left(1 - F_2(x_m)\right)}. \tag{26}$$

Remark that the derivative at the mode is 0, hence the differentiability condition is satisfied.

The c.d.f. of a composite model of size distributions is given by

$$F(x) = \begin{cases} c\, \dfrac{F_1(x)}{F_1(\theta)}, & 0 < x \le \theta \\ c + (1-c)\, \dfrac{F_2(x) - F_2(\theta)}{1 - F_2(\theta)}, & \theta < x < \infty. \end{cases} \tag{27}$$

The moments of the r-th order can be expressed using this formula $E_r(f) = cE_r(f_1^*) + (1-c)E_r(f_2^*)$.

Statistical inference for composite models is done using the classical maximum likelihood (ML) estimation approach. The ML estimation for composite models was first presented in [46] and as well in [49]. In order to apply the ML approach, we have to know the integer value m such that the unknown threshold parameter θ is in between the m-th and $m + 1$-th observation. If we assume somehow that we know the value of the integer m, we would be able to write out explicitly the likelihood function. However, unfortunately, we do not know the exact value of m and as m changes, the ML estimation changes. Therefore, the following ML estimation algorithm was proposed where we have s parameters ρ_i for $i = 1, \dots, s$. In a first step, for each integer $m = 1, \dots, n - 1$ we estimate the parameters as solution of the following ML system

$$\begin{cases} \dfrac{\partial \log L}{\partial \rho_i} = 0, \; i = 1, \dots, s, \\ \dfrac{\partial \log L}{\partial \theta} = 0. \end{cases} \tag{28}$$

If the inequality $x_m \le \hat{\theta} \le x_{m+1}$ holds, then the ML estimators can be denoted as $\hat{\theta}$ and $\hat{\rho}_i$ for $i = 1, \dots, s$. However, a second step is needed in case the first step does not provide any satisfying

result meaning that we are either in one of the following two settings $m = n$ or $m = 0$. This implies that the use of f_1 and f_2 are recommended for the likelihood function, respectively. For the ML procedure, one needs to check $n - 1$ intervals. Thus, the computing time strongly depends on the magnitude of n. For large n this leads to a complex system of equations that must be solved numerically.

In reference [50], the authors propose an alternative algorithm based on quantiles and a moment matching approach. In a first step, let us denote by q_1 and q_3 the first and third empirical quartiles of the data sample. We assume that $q_1 \leq \theta \leq q_3$. Then we use the method of moments to match the first $s - 1$ empirical moments with their theoretical counterparts, and we add two more equations from matching two quartiles

$$
\begin{cases}
c\dfrac{F_1(q_1)}{F_1(\theta)} = 0.25, \\[2mm]
c + (1 - c)\dfrac{F_2(q_3) - F_2(\theta)}{1 - F_2(\theta)} = 0.75.
\end{cases}
\tag{29}
$$

If no result is obtained, we move to a second step where we assume that the first and third quartiles are smaller than the threshold θ, and proceed like in the first step except using now the following two quartiles' equations

$$
\begin{cases}
c\dfrac{F_1(q_1)}{F_1(\theta)} = 0.25, \\[2mm]
c\dfrac{F_1(q_3)}{F_1(\theta)} = 0.75.
\end{cases}
\tag{30}
$$

If we still have no solution, we finally assume that the first and third quartiles are greater than θ and proceed again in a similar fashion as in the first step with the two equations

$$
\begin{cases}
c + (1 - c)\dfrac{F_2(q_1) - F_2(\theta)}{1 - F_2(\theta)} = 0.25, \\[2mm]
c + (1 - c)\dfrac{F_2(q_3) - F_2(\theta)}{1 - F_2(\theta)} = 0.75.
\end{cases}
\tag{31}
$$

Let us remark that those equations have to be solved numerically. Note that once we have a solution from this quantile and moment matching procedure, we can use the ML approach explained above to improve the result as now we have some *a priori* information on the parameter θ and hence on the integer m.

In general, in the area of size distributions, composite models comprise a lognormal or Weibull distribution up to a given threshold value and some form of the Pareto distribution thereafter. The obtained models are close in shape to the lognormal or Weibull law but with a thicker tail due to the Pareto distribution, see **Figure 10** and **Figure 11**.
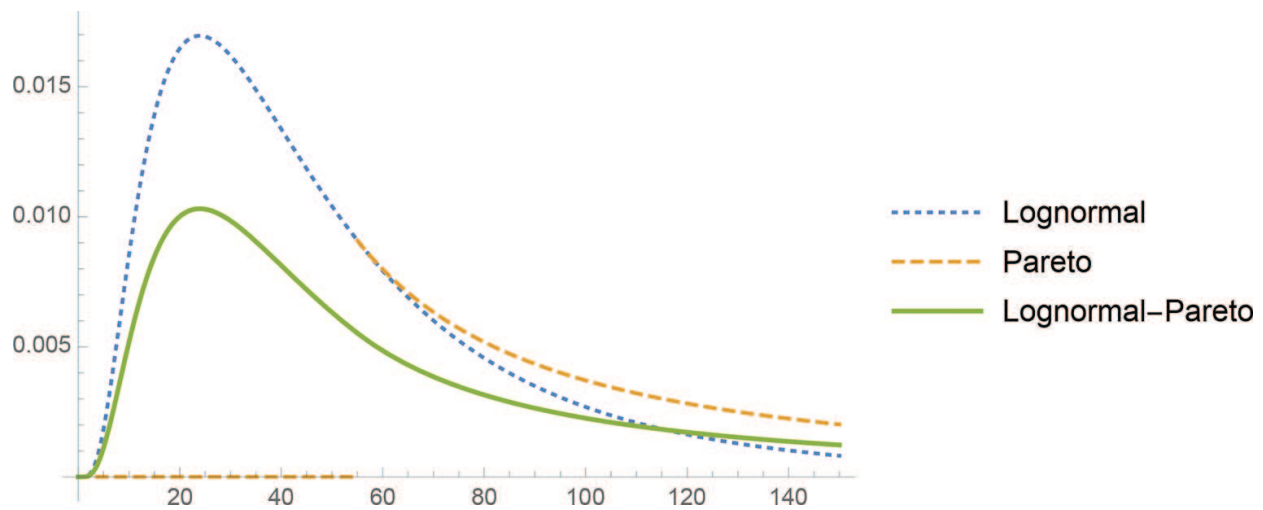
Figure 10. Density plot of the composite lognormal-Pareto model with $\theta = 0.55$ and $\alpha = 0.5$.
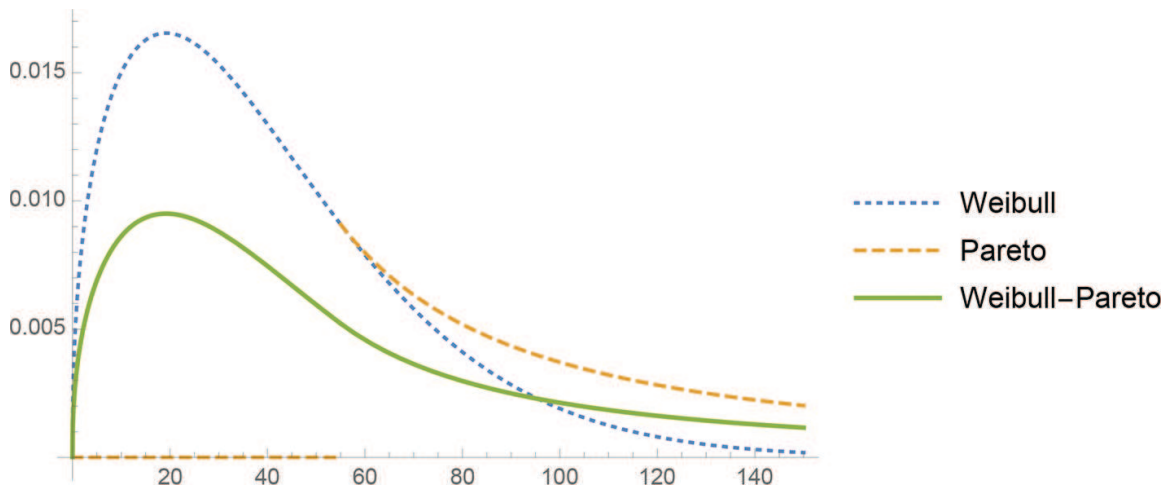

Figure 11. Density plot of the composite Weibull-Pareto model with $\theta = 0.55$ and $\tau = 1.42867$.

This research area for size distributions was initiated by Cooray and Ananda in 2005 [46], who proposed the composite lognormal-Pareto model. They suggested that this composite model may be better suited for insurers when confronted to smaller losses with high frequencies as well as for larger values with lower frequencies. The lognormal-Pareto composite model introduced in reference [46] has been further enhanced by Scollnik [47]. In that paper, the author noticed that the two-component composite model is very restrictive since it has fixed and *a priori* known mixing weights. Hence, he improved the model by using unrestricted mixing weights as coefficients in each component. In a similar way, the article [51] improves the composite Weibull-Pareto model proposed by reference [52]. Those are the composite models that will be described in more detail in the sequel. The papers [47] and [51] consider beside the classical Pareto distribution as well the Pareto type II distribution, known also as the Lomax distribution, as an alternative above the threshold value. In 2013, Teodorescu and Vernic [50] replace the lognormal distribution by any arbitrary continuous distribution, and they analyze in detail the composite Weibull-Pareto and the composite Gamma-Pareto models, and use as well the Lomax distribution as an alternative to the Pareto distribution above the

threshold point. The same authors suggested already the composite exponential-Pareto model [50]. More recently, reference [48] proposes a composite model based on the Stoppa distribution [27], which is a generalization of the Pareto law. More precisely, they propose the lognormal-Stoppa and Weibull-Stoppa composite models.

Let us now take a closer look at the composite lognormal-Pareto and Weibull-Pareto models. Given the general formulas above we can write the density for the composite lognormal-Pareto as

$$
f(x) = \begin{cases} c \dfrac{\dfrac{1}{x\sqrt{2\pi}\sigma} e^{-\dfrac{(\log x - \mu)^2}{2\sigma^2}}}{\Phi\left(\dfrac{\log\theta - \mu}{\sigma}\right)}, & 0 < x \le \theta \\[4ex] (1 - c)\dfrac{\alpha}{\theta}\left(\dfrac{\theta}{x}\right)^{\alpha+1}, & \theta < x < \infty, \end{cases}
\tag{32}
$$

with $0 \le c \le 1$ and $\Phi(.)$ denoting the c.d.f. of a standard normal distribution. In a similar way, the p.d.f. for the composite Weibull-Pareto can be written as

$$
f(x) = \begin{cases} c \dfrac{\dfrac{\tau}{\sigma}\left(\dfrac{x}{\sigma}\right)^{\tau-1} e^{-\left(\dfrac{x}{\sigma}\right)^{\tau}}}{1 - e^{-\left(\dfrac{\theta}{\sigma}\right)^{\alpha}}}, & 0 < x \le \theta \\[4ex] (1-c)\dfrac{\alpha}{\theta}\left(\dfrac{\theta}{x}\right)^{\alpha+1}, & \theta < x < \infty. \end{cases}
\tag{33}
$$

with $0 \le c \le 1$.

By verifying the continuity and differentiability conditions at the threshold point θ, we obtain for the composite lognormal-Pareto model:

$$
c = \frac{\dfrac{\alpha}{\theta}\left(\dfrac{\theta}{x}\right)^{\alpha+1}\Phi\left(\dfrac{\log\theta-\mu}{\sigma}\right)}{\dfrac{\alpha}{\theta}\left(\dfrac{\theta}{x}\right)^{\alpha+1}\Phi\left(\dfrac{\log\theta-\mu}{\sigma}\right) + \dfrac{1}{\theta\sqrt{2\pi}\sigma}e^{-\dfrac{(\log\theta-\mu)^2}{2\sigma^2}}}
\tag{34}
$$

and

$$
\alpha\sigma = \frac{\log\theta - \mu}{\sigma}.
\tag{35}
$$

These conditions guarantee that the p.d.f. of the composite lognormal-Pareto is continuous and smooth at the threshold value θ. The continuity and differentiability conditions at θ, for the composite Weibull-Pareto, yield:

$$c = \frac{\frac{\alpha}{\theta}\left(\frac{\theta}{x}\right)^{\alpha+1}\left(1 - e^{-\left(\frac{\theta}{\sigma}\right)^{\tau}}\right)}{\frac{\alpha}{\theta}\left(\frac{\theta}{x}\right)^{\alpha+1}\left(1 - e^{-\left(\frac{\theta}{\sigma}\right)^{\tau}}\right) + \frac{\tau}{\sigma}\left(\frac{\theta}{\sigma}\right)^{\tau-1}e^{-\left(\frac{\theta}{\sigma}\right)^{\tau}}} \tag{36}$$

and

$$\left(\frac{\theta}{\sigma}\right)^{\tau} = \frac{\alpha}{\tau} + 1. \tag{37}$$

These conditions guarantee the continuity and smoothness of the p.d.f. of the composite Weibull-Pareto at the threshold point θ.

The c.d.f. of the composite lognormal-Pareto and Weibull-Pareto is given, respectively, by

$$F(x) = \begin{cases} c\,\dfrac{\Phi\left(\dfrac{\log x - \mu}{\sigma}\right)}{\Phi\left(\dfrac{\log\theta - \mu}{\sigma}\right)}, & 0 < x \le \theta \\ c + (1 - c)\left(1 - \left(\dfrac{\theta}{x}\right)^{\alpha}\right), & \theta < x < \infty. \end{cases} \tag{38}$$

and

$$F(x) = \begin{cases} c\,\dfrac{1 - e^{-\left(\frac{x}{\sigma}\right)^{\tau}}}{1 - e^{-\left(\frac{\theta}{\sigma}\right)^{\tau}}}, & 0 < x \le \theta \\ c + (1 - c)\left(1 - \left(\dfrac{\theta}{x}\right)^{\alpha}\right), & \theta < x < \infty. \end{cases} \tag{39}$$

Finally, the moments of order r of the composite lognormal-Pareto and Weibull-Pareto are given by

$$E(X^r) = c\,\frac{e^{r\mu + \frac{r^2\sigma^2}{2}}}{\Phi\left(\frac{\log\theta - \mu}{\sigma}\right)} + (1 - c)\,\frac{\alpha\theta^r}{\alpha - r} \tag{40}$$

and

$$E(X^r) = c\,\frac{\sigma^r\Gamma\left(\frac{r}{\tau} + 1; \left(\frac{\theta}{\sigma}\right)^{\tau}\right)}{1 - e^{-\left(\frac{\theta}{\sigma}\right)^{\tau}}} + (1 - c)\,\frac{\alpha\theta^r}{\alpha - r} \tag{41}$$

for $\tau > r$, respectively.

To estimate the composite lognormal-Pareto and the composite Weibull-Pareto models, the algorithms described above are used.

4. Applications

In this section, we focus on two applications to real data sets, one from actuarial sciences, dealing with fire losses and one on Internet traffic data. We will analyze these two data sets with the size distributions seen in Section 2 and the two composite models, namely, the lognormal-Pareto and the Weibull-Pareto, seen in Section 3. In order to compare the distributions, we used the following three criteria:

1.   The maximum log-likelihood (MLL) value: the larger the value, the better the fit of the distribution to the data set.

2.   The Akaike information criterion (AIC):

$$AIC = 2p-2MLL,$$

where p represents the number of parameters to estimate. This criterion represents a measure of the relative quality of a distribution given a set of laws. The distribution with the lowest AIC value is preferred.

3.   The Bayesian information criterion (BIC):

$$BIC = p\log n-2MLL,$$

where n represents the length of the data set and p the number of parameters to estimate. This criterion is used to choose a distribution among a finite set of laws. The distribution with the lowest BIC is preferred.

The AIC and BIC give a trade-off between a reward for a good goodness-of-fit performance and a penalty for an increasing number of parameters to estimate. The BIC tends to favor more parsimonious models than does the AIC.

We carried out the calculations with *Wolfram Mathematica 10*. To calculate the MLL, AIC, and BIC values for the size distributions of Section 2, we used the function *NMaximize* with numerical maximization algorithm *Random Search method* enhanced with the option *Interior-Point*. For the composite lognormal-Pareto and the composite Weibull-Pareto models, we used the estimation algorithms described in Section 3.

4.1. Danish fire losses

In this example, we analyze a classical insurance data set. This is the set of Danish data on 2492 fire insurance losses in Danish Krone (DK) from the years 1980 to 1990 inclusive. The data set can be found in the "SMPracticals" add-on package for R, available from the CRAN website cran.r-project.org.

The comparison of the considered distributions using the three criteria explained above is presented in **Table 1**. The estimated values for the fitted distributions are given in **Table 2**.

Distribution	p	MLL	AIC	BIC
Lognormal	2	-4433.89	8871.78	8883.42
Pareto	2	−5675.09	11354.20	11365.80
GL	3	−3967.18	7940.36	7957.82
GEV	3	−3955.43	7916.86	7934.32
Lognormal-Pareto	2	−3877.84	7759.68	7771.32
Weibull-Pareto	2	−3959.78	7923.56	7935.20

Table 1. MLL, AIC, and BIC values for the Danish fire data set.

Distribution			
Lognormal	$\widehat{\mu} = 0.671854$	$\hat{\sigma} = 0.732317$	
Pareto	$\hat{\alpha} = 0.545817$	$\hat{x}_0 = 0.313404$	
GL	$\hat{\alpha} = 2.01251$	$\hat{b} = 435198$	$\hat{k} = 0.00227572$
GEV	$\widehat{\mu} = 1.42575$	$\hat{\sigma} = 0.712043$	$\hat{k} = 0.545094$
Lognormal-Pareto	$\hat{\alpha} = 1.43633$	$\hat{\theta} = 1.38513$	
Weibull-Pareto	$\hat{\tau} = 4.43613$	$\hat{\theta} = 1.46597$	

Table 2. Estimated values for the fitted distributions for the Danish fire data set.
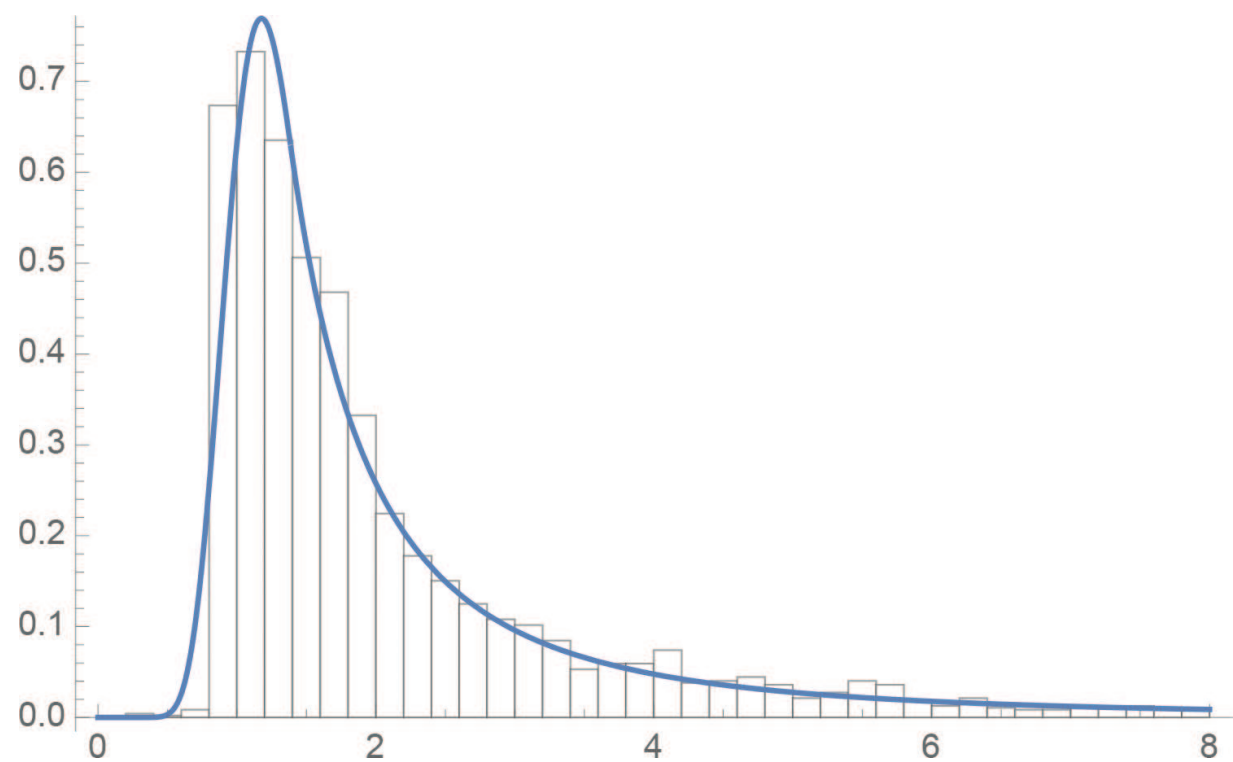

Figure 12. Histogram of the Danish fire data with the fitted density of the composite lognormal-Pareto model.

Distribution	p	MLL	AIC	BIC
Lognormal	2	−39582.2	79168.4	79180.5
Pareto	2	−43031.7	86067.4	86079.5
GL	3	−39581.7	79169.4	79187.6
GEV	3	−39608.4	79222.8	79241.0
Lognormal-Pareto	2	−40098.4	80200.8	80212.9
Weibull-Pareto	2	−42823.9	85651.8	85663.9

Table 3. MLL, AIC, and BIC values for the Internet traffic data.

With a MLL value of -3877.84 and only two parameters, yielding the values AIC = 7759.68 and BIC = 7771.32, the lognormal-Pareto model provides a better fit than the other models for the given data set. A visual conclusion of the fit can be seen in **Figure 12**.

As the data present a humped shape behavior for the lower values and tail behavior for the upper values; this example justifies the use and the necessity of the composite lognormal-Pareto model.

This data set has also been analyzed in reference [46] where the composite lognormal-Pareto model was introduced and reference [51] applied as well the Weibull-Pareto model to this data set. The results we obtain above coincide with their results.

4.2. Internet traffic data

In the second empirical illustration, we analyze Internet traffic data, which have already been analyzed from a Bayesian point of view in references [53] and [54]. This data set consists of 3143 transferred bytes/second within consecutive seconds.

Based on the MLL, the AIC, and BIC values represented in **Table 3**, we conclude that among the considered laws, the lognormal distribution performs the best fit, closely followed by the GL and the GEV distributions. The two considered composite models do not provide good fits for this example. The estimated values for the fitted densities are given in **Table 4**.

Figure 13 provides a visual proof of the goodness-of-fit of the lognormal distribution.

Distribution			
Lognormal	$\widehat{\mu} = 11.6518$	$\hat{\sigma} = 0.62067$	
Pareto	$\hat{\alpha} = 0.353628$	$\hat{x}_0 = 6795$	
GL	$\hat{\alpha} = 13.6735$	$\hat{b} = 4.08831$	$\hat{k} = 808429$
GEV	$\widehat{\mu} = 94465$	$\hat{\sigma} = 54467.5$	$\hat{k} = 0.204602$
Lognormal-Pareto	$\hat{\alpha} = 1.1.05077$	$\hat{\theta} = 85064.5$	
Weibull-Pareto	$\hat{\tau} = 1.12043$	$\hat{\theta} = 79366.3$	

Table 4. Estimated values for the fitted distributions for the Internet traffic data.
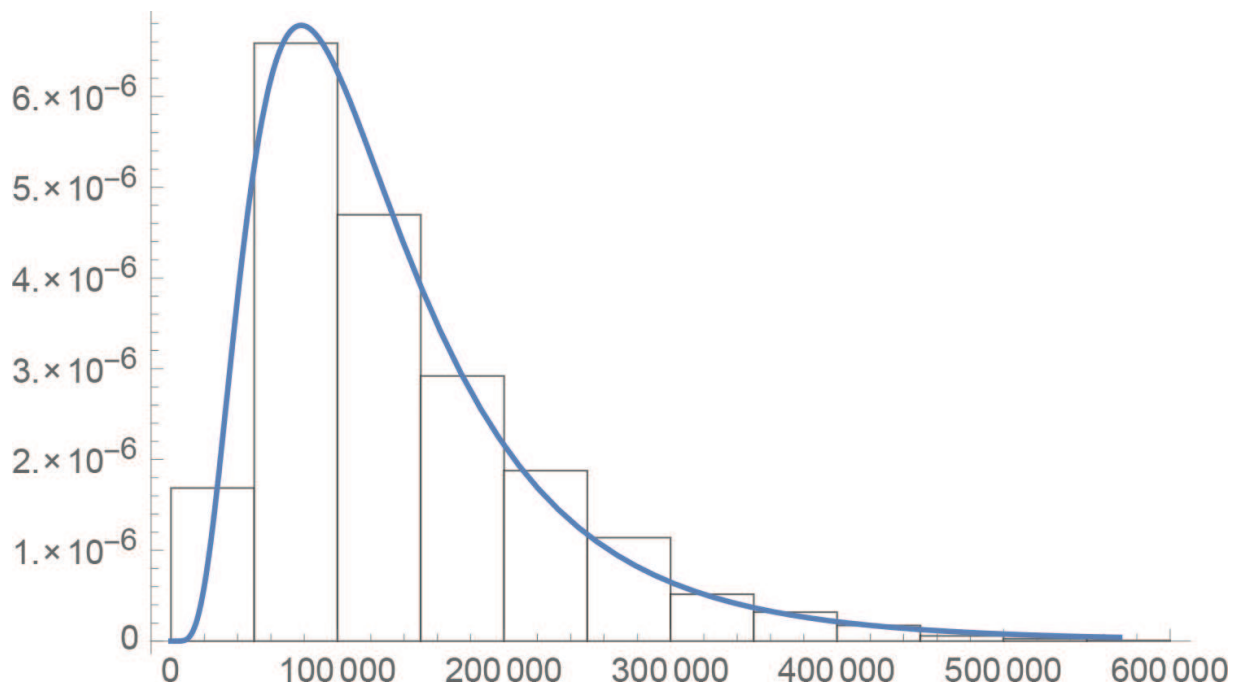
Figure 13. Histogram of the Internet traffic data with the fitted lognormal density.

5. Conclusion

To sum up, we review in this chapter the notion of size distributions by presenting the best known and most used ones. We further describe the general concept of composite models based on size distributions and present in more details the composite lognormal-Pareto and the composite Weibull-Pareto models. Besides providing their main statistical properties, we illustrate the size distributions and composite models by applying them to two real application examples to emphasize their use in practice. We compare the goodness-of-fit of the considered distributions using as criteria the MLL, AIC, and BIC. For the first data set dealing with fire losses we find that the composite lognormal-Pareto model performs the best, hinting at the usefulness of composite models in this research area. However, for the second data set on Internet traffic, the simple lognormal distribution outperforms the other size distributions and composite models. This shows how delicate the choice is for a practitioner when confronted with the question which distribution or model he/she should use on a given data set. The composite models are already quite flexible, but given the different shapes a data set can take, there is a quest for even more flexible distributions. In the literature, some families of distributions are proposed, which contain many of the classical size distributions and hence can model very diverse behaviors. The most popular one is the generalized beta distribution presented in reference [55], and very recently reference [56] introduces a new flexible distribution called the interpolating family of size distributions. Those distributions are quite flexible as they enable to model very distinct shapes and probably constitute the future avenue of research in the area of size distributions.

Author details

Yves Dominicy[1]* and Corinne Sinner[2]*

*Address all correspondence to: yves.dominicy@ulb.ac.be and corinne.sinner@ulb.ac.be

1 Université libre de Bruxelles, SBSEM, ECARES, Brussels, Belgium

2 Université libre de Bruxelles, Département de Mathématique, Brussels, Belgium

References

[1] Jones M.C. On families of distributions with shape parameters. International Statistical Review. 2015;**83**(2):175–192.

[2] Ley C. Flexible modelling in statistics: past, present and future. Journal de la Société Française de Statistique. 2015;**156**:76–96.

[3] Lawless J. Statistical Models and Methods for Lifetime Data. Wiley Series in Probability and Statistics. New York, USA: Wiley; 2003.

[4] Lee E., Wang J. Statistical Methods for Survival Data Analysis. Wiley Series in Probability and Statistics. New York, USA: Wiley; 2003.

[5] Marchenko Y., Genton M. Multivariate log-skew-elliptical distributions with applications to precipitation data. Environmetrics. 2010;**21**:318–340.

[6] Mitzenmacher M. A brief history of generative models for power law and lognormal distributions. Internet Mathematics. 2004;**1**(2):226–251.

[7] Eeckhout J. Gibrat's law for (All) cities. American Economic Review. 2004;**94**:1429–1451.

[8] Gabaix X. Power laws in economics: An introduction. Journal of Economic Perspectives. 2016;**30**:185–206.

[9] Clarke R. Estimating trends in data from the Weibull and a generalized extreme value distribution. Water Resources Research. 2002;**38**:25-1–25-10.

[10] Kleiber C., Kotz S. Statistical Size Distributions in Economics and Actuarial Sciences. New York: John Wiley & Sons; 2003.

[11] Asgharzadeh A., Nadarajah S., Sharafi F. Generalized inverse Lindley distribution with application to Danish fire insurance data. Communications in Statistics: Theory and Methods. Forthcoming.

[12] Ortega E.M.M., Lemonte A.J., Silva G.O., Cordeiro G.M. New flexible models generated by gamma random variables for lifetime modeling. Journal of Applied Statistics. 2015;**42**(10):2159–2179.

[13] Aitchison J., Brown J.A.C. The Lognormal Distribution. Cambridge: University Press; 1957.

[14] McAlister D. The law of the geometric mean. Proceedings of the Royal Society of London. 1879;**29**:367– 376.

[15] Black F., Scholes M. The pricing of options and corporate liabilities. The Journal of Political Economy. 1973;**81**(3):637– 654.

[16] Clementi F., Gallegati M. Power law tails in the Italian personal income distribution. Physical A: Statistical Mechanics and Its Applications. 2005;**350**(2):427– 438.

[17] Burnecki K., Kukla G., Weron R. Property insurance loss distributions. Physical A: Statistical Mechanics and Its Applications. 2000;**287**(1):269– 278.

[18] Burnecki K., Misiorek A., Cizek P., Härdle W.K. and Weron R. Loss distributions. In: Statistical Tools for Finance and Insurance. Berlin, Heidelberg: Springer; 2005. pp. 289– 317.

[19] Koch A.L. The logarithm in biology I. Mechanisms generating the log-normal distribution exactly. Journal of Theoretical Biology. 1966;**23**:276– 290.

[20] Limpert E., Stahel W.A., Abbt M. Log-normal distributions across the sciences: keys and clues. BioScience. 2001;**51**(5):341– 352.

[21] Tufféry S. Data mining et statistique décisionnelle. Paris, France: Technip; 2012.

[22] Pareto V. Cours d'économie politique. Geneva: Droz; 1896.

[23] Arnold B.C. Pareto Distributions. Fairland, Maryland, USA: International Co-operative Publishing House; 1983.

[24] Hill B.M. A simple general approach to inference about the tail of a distribution. Annals of Statistics. 1975;**3**:1163– 1174.

[25] Dominicy Y., Ilmonen P., Veredas D. Multivariate Hill Estimators. International Statistical Review. Forthcoming.

[26] Lomax K.S. Business failures: Another example of the analysis of failure data. Journal of the American Statistical Association. 1954;**49**:847– 852.

[27] Stoppa G. Proprieta campionarie di un nuovo modello Pareto generalizzato. Atti XXXV Riunione Scientica della Societa Italiana di Statistica, Padova: Cedam. 1990; pp. 137– 144.

[28] Benckert L.G., Sternberg I. An attempt to find an expression for the distribution of fire damage amount. Transactions of the 15th International Congress of Actuaries. 1957;**2**:288– 294.

[29] Andersson H. An analysis of the development of the fire losses in the north countries after the Second World War. Astin Bulletin. 1971;**6**:25– 30.

[30] Reed W.J., Jorgensen M. The double Pareto-lognormal distribution–A new parametric model for size distributions. Communications in Statistics: Theory and Methods. 2004;**33**(8):1733– 1753.

[31] Abdul-Moniem I.B., Abdel-Hameed H.F. On exponentiated Lomax distribution. International Journal of Mathematical Archive. 2012;**3**(5):2144– 2150.

[32] Shams T.M. The Kumaraswamy-generalized Lomax distribution. Middle-East Journal of Scientific Research. 2013;**17**(5):641– 646.

[33] Tahir M.H., Hussain M.A., Cordeiro G.M., Hamedani G.G., Mansoor M., Zubair M. The Gumbel-Lomax distribution: properties and applications. Journal of Statistical Theory and Applications. 2016;**15**(1):61– 79.

[34] Fisher R.A., Tippett L.H.C. Limiting forms of the frequency distribution of the largest or smallest member of a sample. Proceedings of the Cambridge Philosophical Society. 1928;**24**:180– 290.

[35] Embrechts P., Klü ppelberg C., Mikosch T. Modelling Extremal Events for Insurance and Finance. Berlin: Springer-Verlag; 1997.

[36] Guégan D., Hassani B.K. A mathematical resurgence of risk management: An extreme modeling of expert opinions. Frontiers in Finance and Economics. 2014;**11**(1):25– 45.

[37] Burke E.J., Perry R.H.J., Brown S.J. An extreme value analysis of UK drought and projections of change in the future. Journal of Hydrology. 2010;**388**:131– 143.

[38] Coles S. An Introduction to Statistical Modeling of Extreme Values. Berlin: Springer-Verlag; 2001.

[39] Finkenstädt B., Rootzén H. Extreme Values in Finance, Telecommunications and the Environment. London: Chapman & Hall/CRC; 2004.

[40] Jenkinson A.F. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. Quarterly Journal of the Royal Meteorological Society. 1955;**81**:158– 171.

[41] Lindquist E.S. Strength of materials and the Weibull distribution. Probabilistic Engineering Mechanics. 1994;**9**(3):191– 194.

[42] Manwell J.F., McGowan J.G., Rogers A.L. Wind Energy Explained: Theory, Design and Application. New York, USA: Wiley; 2009.

[43] Sharif N., Islam N. The Weibull distribution as a general model for forecasting technological change. Technological Forecasting and Social Change. 1980;**18**(3):247– 256.

[44] Alvarado-Celestino E. Large forest fires: An analysis using extreme value theory and robust statistics [thesis]. University of Washington, USA; 1992.

[45] Klugman S.A., Panjer H.H., Willmot G. Loss Models: From Data to Decisions. New York: Wiley; 2008.

[46] Cooray K., Ananda M.M.A. Modeling actuarial data with composite lognormal-Pareto model. Scandinavian Actuarial Journal. 2005;**5**:321– 334.

[47] Scollnik D.P.M. On composite lognormal-Pareto models. Scandinavian Actuarial Journal. 2007;**1**:20– 33.

[48] Calderin-Ojeda E., Kwok C.F. Modeling claims data with composite Stoppa models. Scandinavian Actuarial Journal. Forthcoming.

[49] Teodorescu S., Vernic R. Some composite exponential-Pareto models for actuarial prediction. Romanian Journal of Economic Forecasting. 2009;**12**(4):82– 100.

[50] Teodorescu S., Vernic R. On composite Pareto models. Mathematical Reports. 2013;**15** (65):11– 29.

[51] Scollnik D.P.M., Sun C. Modeling with Weibull-Pareto models. North American Actuarial Journal. 2012;**16**(2):260– 272.

[52] Ciumara R. An actuarial model based on composite Weibull-Pareto distribution. Mathematical Reports. 2006;**8**(58):404– 414.

[53] Ramirez-Cobo P., Lillo R.E., Wilson S., Wiper M.P. Bayesian inference for double Pareto lognormal queues. The Annals of Applied Statistics. 2010;4(3):1533– 1557.

[54] Rubio F.J., Steel M. Bayesian modelling of skewness and kurtosis with two-piece scale and shape distributions. Electronic Journal of Statistics. 2015;**9**:1884– 1912.

[55] McDonald J., Xu Y. A generalization of the beta distribution with applications. Journal of Econometrics. 1995;**66**:133– 152.

[56] Sinner C., Dominicy Y., Ley C., Trufin J., Weber P. An interpolating family of size distributions. Forthcoming.