SPANISH JOURNAL OF STATISTICS
Vol. **2** No. 1 2020, Pages 7–21
doi:https://doi.org/10.37830/SJS.2020.1.02

REGULAR ARTICLE

Financial and Actuarial Properties of the Beta-Pareto as a Long-Tail Distribution

Emilio Gómez-Déniz[1], Enrique Calderín-Ojeda[2]
[1]Department of Quantitative Methods and TIDES Institute - University of Las Palmas de Gran Canaria
emilio.gomez-deniz@ulpgc.es
[2]Centre for Actuarial Studies, Department of Economics - The University of Melbourne, Australia
enrique.calderin@unimelb.edu.au

Received: December 18, 2020. Accepted: March 1, 2021.

Abstract: Undoubtedly, the single parameter Pareto distribution is one of the most attractive distribution in statistics; a power-law probability distribution that is found in a large number of real-world situations inside and outside the field of economics. Furthermore, it is usually used as a basis for excess of loss quotations as it gives a pretty good description of the random behaviour of large losses. In this paper, we provide properties of the Beta-Pareto distribution which can be useful in Economics, and in Financial and Actuarial fields, mainly related to the analysis of the tail of the distribution that makes it a candidate model for fitting actuarial data with extreme observations. As empirical applications two well-known data sources considered in general insurance are used to account for the suitability of the model.

Keywords: insurance, Beta-Pareto distribution, Danish and Norweigian data; Pareto distribution, right tail

MSC: 62E10, 62F10, 62P05

1 Introduction

Probability distributions such as the exponential, Pareto, gamma, lognormal and Weibull are frequently used in survival analysis, engineering applications and, specifically, in actuarial statistics to model losses in insurance and finance. Besides, other parametric families, e.g. Pareto and lognormal distributions are particularly appropriate to describe data that include large losses (see Boland, 2007, p. IX). More precisely, the study of the right tail of the distribution is an important issue in order to not underestimate the size of large claims. This is for example the case of the suitability of the Pareto distribution to describe fire claim data (Rolski et al., 1999, p. 49). This is also common in defaulted loans in banking sector. It is needless to say that, due to the simple form of its survival function, the Pareto distribution is commonly used in these scenarios.

© INE

It is well-known that the classical Pareto distribution (for a detailed discussion of the Pareto distribution see Arnold, 1983) with scale parameter $\sigma > 0$ and shape parameter $\theta > 0$ with probability density function

$$g(x) = \frac{\theta\sigma^\theta}{x^{\theta+1}}, \quad x > \sigma > 0, \ \theta > 0 \tag{1}$$

and survival function

$$\bar{G}(x) = \left(\frac{\sigma}{x}\right)^\theta, \quad x > \sigma > 0, \ \theta > 0 \tag{2}$$

has been proved to be useful as predicting tools in different socioeconomic contexts such as income (Mandelbrot, 1960), insurance (for applications of the Pareto model in rating property excess-of-loss reinsurance, the Pareto distribution has been used by Boyd, 1988, Hesselager, 1993 and Brazauskas and Serfling, 2003, among others), city size (Rosen and Resnick, 1980) and also in other fields as queue service (Harris, 1968). A thorough review of the reinsurance issue can be viewed in Albrecher et al. (2017). Perhaps, one of the most important characteristics of the Pareto distribution is that it produces a better extrapolation from the observed data when pricing high excess layers, in situations where there is little or no experience. In this regard, its efficacy dealing with inflation in claims and with the effect of deductibles and excess-of-loss levels for reinsurance has been demonstrated. Henceforward, a continuous random variable that follows the Pareto distribution with pdf as in (1) will be denoted as $X \sim Par(\theta, \sigma)$. Surely, one of the advantages of working with this probability distribution is, similarly to the exponential case, the simple form of its survival function which allows us to easily derive interesting properties. For example, it is straightforward to observe that if $X \sim Par(\theta, \sigma)$, then $\tau X \sim Par(\theta, \tau\sigma)$, $\tau > 0$. This property is useful when dealing with proportional reinsurance and also with claims inflation. Furthermore, if $X > Z$ we have that $X - Z \sim Par(\theta, Z)$. That is, the excess of X over Z is also Pareto (see Boland, 2007, p. 39).

In the last decades, a lot of attempts have been made to achieve generalizations of the classical Pareto distributions. Many of these new models try to obtain better fits to empirical data related to city populations and insurance losses. Some of them are the Stoppa's generalized Pareto distribution (see Stoppa, 1990 and Kleiber and Kotz, 2003); the Beta-Pareto distribution due to Akinsete et al. (2008); the Pareto positive stable distribution provided by Sarabia and Prieto (2009) and the recently proposals of Gómez-Déniz and Calderín (2014), Gómez-Déniz and Calderín (2015) and Ghitany et al. (2018). In general insurance settings and also in city size, mainly seeking to better adjust the right tail of the distribution, the recently proposed composite models have also made use of the Pareto distribution in their formulation and, therefore, can be considered as generalizations of the latter distribution (see Scollnik, 2007, Calderín-Ojeda and Kwok, 2016 and Calderín-Ojeda, 2016).

In actuarial settings, the single parameter Pareto distribution has been largely considered against other probability distributions, not only for its nice properties, but also for its appropriateness to describe the claims size. When modeling losses, there is widely concern on the frequencies and sizes of large claims, in particular, the study of the right tail of the distribution. On this subject, the single parameter Pareto distribution gives a good description of the random behaviour of large losses. See, for instance Boyd (1988) and Brazauskas and Serfling (2003), among others.

In this paper, we pay special attention to one generalization of the Pareto distribution, built from the scheme proposed by Jones (2004) and which was considered by Akinsete et al. (2008), the Beta-Pareto distribution. We will see that this distribution can be used as a basis for excess of loss quotations, and similarly to the Pareto distribution (see for instance, Rytgaard, 1990), providing a good description of the random behaviour of large losses.

In order to make the paper self-contained, some of the basic properties provided in Akinsete et al. (2008) are again reproduced here. Furthermore, new properties that are important in financial and actuarial applications are also provided. In particular, we give expressions for the limited expected values, integrated tail distribution and mean excess function, among others. Finally, the performance of the model is examined by using two well-known examples of real claims data in actuarial statistics.

The remainder of the paper is organized as follows. Basic background of the Beta-Pareto distribution is shown in Section 2. We pay special attention here to some of its more basic properties and the estimation of the parameters of the distribution by maximum likelihood method. Section 3 discusses properties related to the right tail of the distribution that are very relevant in the field of reinsurance. Two numerical applications are shown in Section 4 and conclusions are provided in the last Section.

2 Preliminaries

In an appealing paper Jones (2004) proposed a method to add more flexibility to a parent probability function by starting with a distribution function G (in that work author only considered symmetric distributions but the methodology is applicable to any distribution function) and generating the new one by adding two parameters in order to include skewness and vary the tail weight. The method is based on order statistics by using the classical Beta distribution. Specifically, for a probability density function $g(x)$ with distribution function $G(x)$ and survival function $\bar{G}(x) = 1-G(x)$, the author studied the family of probability distributions given by

$$f(x;\alpha,\beta) = \frac{1}{B(\alpha,\beta)} g(x)\,[G(x)]^{\alpha-1}\left[\bar{G}(x)\right]^{\beta-1}, \quad \alpha > 0, \beta > 0, \tag{3}$$

where $B(\cdot,\cdot)$ is the Euler Beta function.

When (3) is applied to (2), the probability density function of the Beta-Pareto distribution studied in Akinsete et al. (2008) is obtained with analytical expression given by

$$f(x) = \frac{1}{B(\alpha,\beta)} \frac{\theta}{x}\left(\frac{\sigma}{x}\right)^{\alpha\theta}\left[1-\left(\frac{\sigma}{x}\right)^{\theta}\right]^{\beta-1}, \quad x > \sigma. \tag{4}$$

This distribution includes a wide range of curve shapes as illustrated by the density plots shown in Figure 1.

Some special cases of the distribution provided in (4) are given below:

- If $\alpha = \beta = 1$ we get the classical Pareto distribution given in (1).
- The case $\beta = 1$ reduces to a $Par(\alpha\theta,\sigma)$.
- The case $\alpha = 1$ to the Stoppa distribution (see Stoppa, 1990 and Kleiber and Kotz, 2003).

Hereafter, a random variable X that follows the probability density function (4) will be denoted as $X \sim BP(\alpha,\beta,\theta,\sigma)$.

Simple computations show that the distribution is unimodal with modal value located at

$$x = \sigma\left[\frac{1+(\alpha+\beta-1)\theta}{1+\alpha\,\theta}\right]^{1/\theta}.$$

All moments of order $r > 0$ exist and they are given by,

$$E(X^r) = \frac{\sigma\Gamma(\alpha+\beta)\Gamma(\alpha-r/\theta)}{\Gamma(\alpha)\Gamma(\alpha+\beta-r/\theta)}.$$

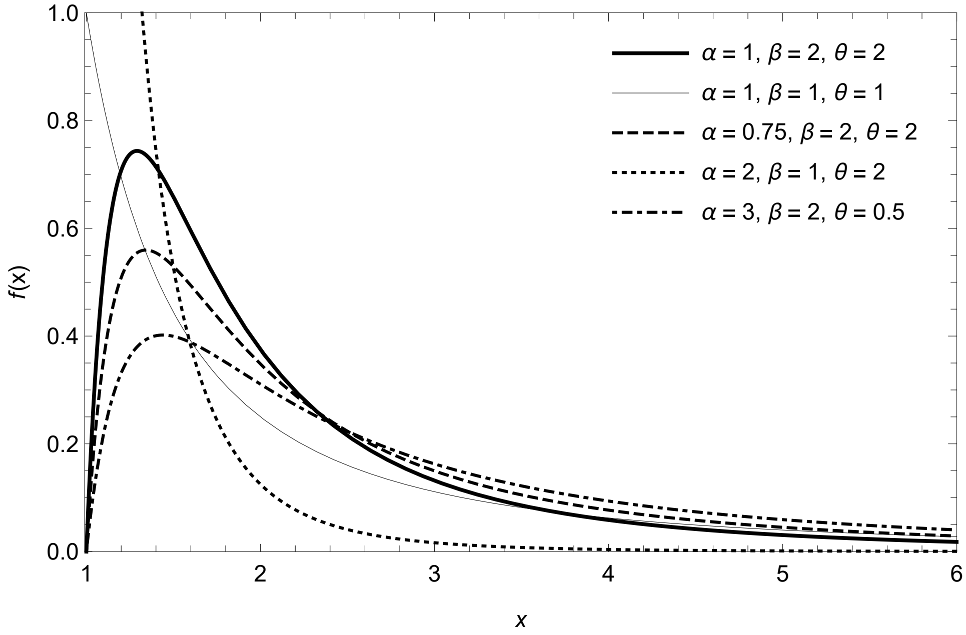

Figure 1: Graphs of the probability density function (4) for different values of parameter α, β and θ assuming in all the cases $\sigma = 1$.

In particular, the mean value is given by

$$E(X) = \frac{\sigma B(\alpha^*, \beta)}{B(\alpha, \beta)}, \quad \alpha > \frac{1}{\theta}, \tag{5}$$

where $\alpha^* = \alpha - 1/\theta$.

The variance is easily computed and it can be seen that the mean value increases with α and β (in this case when $\theta > 1$) and decreases with θ.

One of the advantage of this distribution is its simple form of its survival function, which is expressed in terms of the incomplete beta function ratio, a special function available in many statistical software and spreadsheet packages. That is, the survival function of the random variable following the probability distribution (4) results

$$\bar{F}(x) = I_{z(x)}(\alpha, \beta), \tag{6}$$

where $z(x) = (\sigma/x)^\theta$ and $I_u(\cdot,\cdot)$ represents the incomplete beta function ratio, given by

$$I_c(a,b) = \frac{1}{B(a,b)} \int_0^c t^{a-1}(1-t)^{b-1}\, dt.$$

Furthermore, the hazard rate function, $h(x) = f(x)/\bar{F}(x)$, has also a simple and closed-form expression.

Below, the hazard rate function has been plotted in Figure 2 for different values of the parameters α, β and θ and assuming again that $\sigma = 1$. It is observable that the hazard rate function is monotonically decreasing when $\beta \leq 1$. When $\beta > 1$ the hazard rate function has inverted-U shape.

Also, if $\beta < 1$ the distribution is log-convex, i.e. $(\log f(x))'' > 0$. Finally, closed-form expression for the entropy of the distribution can be viewed in Akinsete et al. (2008)

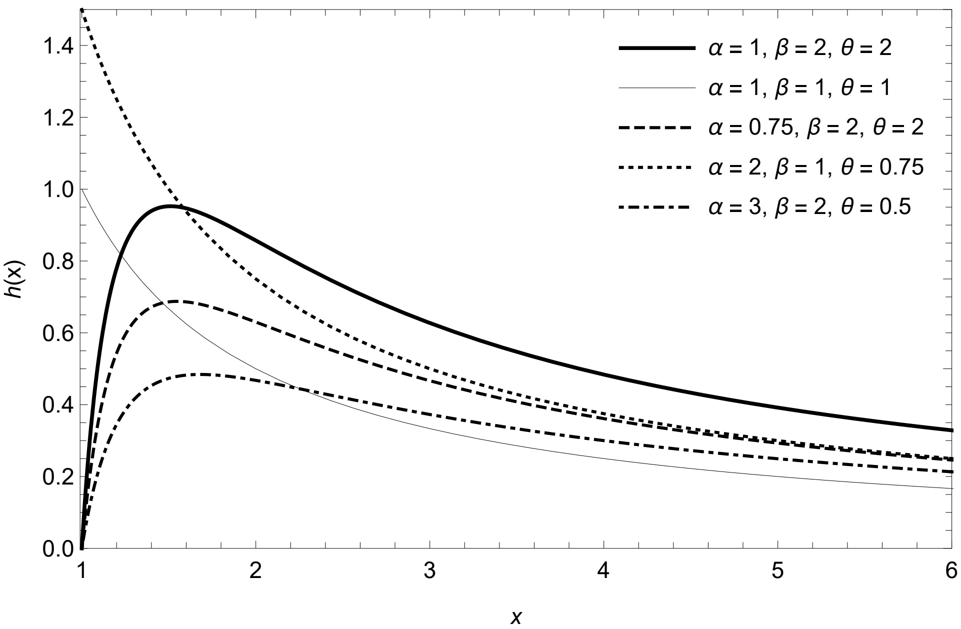

Figure 2: Graphs of the hazard rate function for different values of parameter α, β and θ, assuming again $\sigma = 1$.

2.1 Transformations

Let

$$X = \sigma(1 - Z)^{-1/\theta},$$

then, it is easy to see that the random variable Z follows a Beta distribution with parameters $\alpha > 0$ and $\beta > 0$. This change of variable facilitates computations on properties of the BP distribution studied here.

2.2 Estimation

In this subsection, we show how to estimate the parameters of the distribution. For that reason, let us assume that $\{x_1, x_2, \ldots, x_n\}$ is a random sample selected from the distribution (4) and also assume that $\sigma = \min\{x_i\}$, $i = 1, \ldots, n$. By using the first three moments, numerical computation can be carried out to obtain the moment estimates of the distribution. Alternatively, by using the maximum likelihood method, the likelihood function is given by

$$\begin{aligned} \ell(\boldsymbol{\omega}; \tilde{x}) &= n\left[\log\Gamma(\alpha+\beta) - \log\Gamma(\alpha) - \log\Gamma(\beta) + \log\theta + \alpha\theta\log\sigma\right] \\ &\quad -\alpha\theta\sum_{i=1}^{n}\log x_i + (\beta-1)\sum_{i=1}^{n}\log\left[1 - z(x_i)\right]. \end{aligned}$$

The maximum likelihood estimates (MLEs) $\widehat{\boldsymbol{\omega}} = (\widehat{\alpha}, \widehat{\beta}, \widehat{\theta})$, of the parameters $\boldsymbol{\omega} = (\alpha, \beta, \theta)$ are obtained by solving the score equations

$$\frac{\partial \ell(\boldsymbol{\omega};\tilde{x})}{\partial \alpha} = n\left[\psi(\alpha+\beta) - \psi(\alpha) + \theta \log \sigma\right] - \theta \sum_{i=1}^{n} \log x_i = 0, \tag{7}$$

$$\frac{\partial \ell(\boldsymbol{\omega};\tilde{x})}{\partial \beta} = n\left[\psi(\alpha+\beta) - \psi(\beta)\right] + \sum_{i=1}^{n} \log\left[1 - z(x_i)\right] = 0, \tag{8}$$

$$\begin{aligned}\frac{\partial \ell(\boldsymbol{\omega};\tilde{x})}{\partial \theta} &= n\left(\frac{1}{\theta} + \alpha \log \sigma\right) - \alpha \sum_{i=1}^{n} \log x_i \\ &+ (\beta - 1) \sum_{i=1}^{n} \left(\frac{\sigma}{x_i}\right)^{\theta} \frac{\log(\sigma/x_i)}{1 - z(x_i)} = 0,\end{aligned} \tag{9}$$

where $\psi(\cdot)$ gives the derivative of the digamma function (the logarithm of the gamma function).

Observe that from equation (7) we get

$$\theta = \frac{n\left[\psi(\alpha+\beta) - \psi(\alpha)\right]}{\sum_{i=1}^{n} \log x_i - n \log \sigma},$$

which can be plugged into equations (8) and (9) in order to derive system of equations which only depends on two parameters and that can be solved by a numerical method such as Newton-Raphson.

The second partial derivatives are as follows.

$$\begin{aligned}
\frac{\partial^2 \ell(\boldsymbol{\omega};\tilde{x})}{\partial \alpha^2} &= n\left[\psi_1(\alpha+\beta) - \psi_1(\alpha)\right], \\
\frac{\partial^2 \ell(\boldsymbol{\omega};\tilde{x})}{\partial \alpha \partial \beta} &= n\psi_1(\alpha+\beta), \\
\frac{\partial^2 \ell(\boldsymbol{\omega};\tilde{x})}{\partial \alpha \partial \theta} &= \log \theta - \sum_{i=1}^{n} \log x_i, \\
\frac{\partial^2 \ell(\boldsymbol{\omega};\tilde{x})}{\partial \beta^2} &= n\left[\psi_1(\alpha+\beta) - \psi_1(\beta)\right], \\
\frac{\partial^2 \ell(\boldsymbol{\omega};\tilde{x})}{\partial \beta \partial \theta} &= -\sum_{i=1}^{n} \left(\frac{\sigma}{x_i}\right)^{\theta} \frac{\log(\sigma/x_i)}{1 - z(x_i)}, \\
\frac{\partial^2 \ell(\boldsymbol{\omega};\tilde{x})}{\partial \theta^2} &= -\frac{n}{\theta^2} - (\beta - 1) \sum_{i=1}^{n} \frac{(\sigma/x_i)^{\theta} \log^2(\sigma/x_i)}{\left[1 - z(x_i)\right]^2}.
\end{aligned}$$

Once the parameters have been estimated the entries of the expected Fisher's information matrix, $\mathcal{I}(\widehat{\omega})$, can be approximated by the following expressions

$$
\begin{aligned}
\mathcal{I}_{11}(\widehat{\omega}) &= -n\Big[\psi_1(\widehat{\alpha}+\widehat{\beta})-\psi_1(\widehat{\alpha})\Big],\\
\mathcal{I}_{12}(\widehat{\omega}) &= -n\psi_1(\widehat{\alpha}+\widehat{\beta}),\\
\mathcal{I}_{13}(\widehat{\omega}) &\approx -\log\widehat{\theta}-\sum_{i=1}^{n}\log x_i,\\
\mathcal{I}_{22}(\widehat{\omega}) &= -n\Big[\psi_1(\widehat{\alpha}+\widehat{\beta})-\psi_1(\widehat{\beta})\Big],\\
\mathcal{I}_{23}(\widehat{\omega}) &\approx -\sum_{i=1}^{n}\left(\frac{\sigma}{x_i}\right)^{\widehat{\theta}}\frac{\log(\sigma/x_i)}{1-(\sigma/x_i)^{\widehat{\theta}}},\\
\mathcal{I}_{33}(\widehat{\omega}) &\approx \frac{n}{\widehat{\theta}^2}+(\widehat{\beta}-1)\sum_{i=1}^{n}\frac{(\sigma/x_i)^{\widehat{\theta}}\log^2(\sigma/x_i)}{\Big[1-(\sigma/x_i)^{\widehat{\theta}}\Big]^2}.
\end{aligned}
$$

3 Tail of the distribution and related issues

As it was previously mentioned, a random variable with non-negative support, such as the classic Pareto distribution, is commonly used in insurance to model the amount of claims (losses). In this sense, the size of the distribution tail is of vital importance in actuarial and financial scnearios, if it is desired that the chosen model allows to capture amounts sufficiently far from the start of the distribution support, that is, extreme values. Consequently, the use of heavy right-tailed distributions such as the Pareto, lognormal and Weibull (with shape parameter smaller than 1) distributions, among others, have been employed to model losses in motor third-party liability insurance, fire insurance or catastrophe insurance.

3.1 Right tail of the BP distribution

It is already known that any probability distribution, that is specified through its cumulative distribution function $F(x)$ on the real line, is heavy right-tailed (see Rolski et al., 1999) if $\limsup_{x\to\infty}(-\log\bar{F}(x)/x)=0$. Observe that $-\log\bar{F}(x)$ is the hazard function of $F(x)$. Next result shows that the BP is a heavy tail distribution.

Proposition 1. *The cumulative distribution function $F(x)$ of the Beta-Pareto distribution is a heavy tail distribution.*

Proof. We have that

$$
\begin{aligned}
\limsup_{x\to\infty}\frac{1}{x}\log\bar{F}(x) &= \limsup_{x\to\infty}\frac{1}{x}\log\left[\frac{1}{B(\alpha,\beta)}\int_0^{z(x)}t^{\alpha-1}(1-t)^{\beta-1}\right]\\
&= -\limsup_{x\to\infty}\frac{\theta\,[1-z(x)]^{\beta-1}}{\sigma B(\alpha,\beta)I_{z(x)}(\alpha,\beta)}\left(\frac{\sigma}{x}\right)^{\theta(\alpha+1)}\\
&= -\limsup_{x\to\infty}\frac{\theta}{xB(\alpha,\beta)}\left[\alpha+1-\frac{(\beta-1)\sigma^{\theta}}{1-z(x)}\right]=0,
\end{aligned}
$$

where we have applied twice L'Hospital rule and the Fundamental Theorem of Calculus. Hence the result. □

Corollary 1. *It is verified that* $\limsup_{x\to\infty} e^{sx}\bar{F}(x) = \infty$, $x > \sigma$, $s > 0$.

Proof. This is a direct consequence of Proposition 1. □

Therefore, as a long-tailed distribution is also heavy right-tailed, the Beta-Pareto distribution introduced in this manuscript is heavy right-tailed.

An important issue in extreme value theory is the regular variation (see Bingham, 1987 and Konstantinides, 2018). This is, a fexible description of the variation of some function according to the polynomial form of the type $x^{-\delta}+o(x^{-\delta})$, $\delta > 0$. This concept is formalized in the following definition.

Definition 1. *A distribution function (measurable function) is called regular varying at infinity with index* $-\delta$ *if it holds*

$$\lim_{x\to\infty} \frac{\bar{F}(\tau x)}{\bar{F}(x)} = \tau^{-\delta},$$

where $\tau > 0$ *and the parameter* $\delta \geq 0$ *is called the tail index.*

Next theorem establishes that the survival function given in (6) is a regular variation Lebesgue measure.

Proposition 2. *The survival function given in (6) is a survival function with regularly varying tails.*

Proof. Let us firstly consider the survival function given in (6). Then, after applying L'Hospital rule and Fundamental Theorem of Calculus we get

$$\begin{aligned}
\limsup_{x\to\infty} \frac{\bar{F}(\tau x)}{\bar{F}(x)} &= \limsup_{x\to\infty} \frac{\int_0^{z(\tau x)} t^{\alpha-1}(1-t)^{\beta-1}\,dt}{\int_0^{z(x)} t^{\alpha-1}(1-t)^{\beta-1}\,dt} \\
&= \limsup_{x\to\infty} \frac{t(\theta/\sigma)(\sigma/(\tau x))^{\theta(\alpha+1)}[1-z(\tau x)]^{\beta-1}}{(\theta/\sigma)(\sigma/x)^{\theta(\alpha+1)}[1-z(x)]^{\beta-1}} = \tau^{-\theta\alpha},
\end{aligned}$$

and taking into account that θ, $\alpha > 0$ the result follows. □

An immediate consequence of the previous result is the following (see Jessen and Mikosch, 2006).

Corollary 2. *If* $X, X_1, \ldots, X_n$ *are iid random variables with common survival function given by (6) and* $S_n = \sum_{i=1}^n X_i$, $n \geq 1$, *then*

$$\Pr(S_n > x) \sim \Pr(X > x) \quad as \;\; x \to \infty.$$

Thus, if $X, X_1, \ldots, X_n$ are iid random variables with common survival function given by (6) and $S_n = \sum_{i=1}^n X_i$, $n \geq 1$, then

$$\Pr(S_n > x) \sim \Pr(X > x) \quad \text{as} \;\; x \to \infty.$$

Therefore, if $P_n = \max_{i=1,\ldots,n} X_i$, $n \geq 1$, we have that

$$\Pr(S_n > x) \sim n\Pr(X > x) \sim \Pr(P_n > x).$$

This means that for large x the event $\{S_n > x\}$ is due to the event $\{P_n > x\}$. Therefore, exceedances of high thresholds by the sum S_n are due to the exceedance of this threshold by the largest value in the sample.

The integrated tail distribution or equilibrium distribution (see for example Yang, 2004), given by

$$F_I(x) = \frac{1}{E(X)} \int_{\sigma}^{x} \bar{F}(y)\, dy.$$

is an important concept that often appears in insurance and many other applied probability models. For the BP distribution studied in this work, the integrated tail distribution can be written as a closed-form expression as it is given in the following Proposition.

Proposition 3. *Let X be a random variable that follows the probability density function given in (4). Then, the integrated tail distribution of this random variable is given by*

$$F_I(x) = \frac{B(\alpha,\beta)}{B(\alpha^*,\beta)} \left[\frac{x}{\sigma} I_{z(x)}(\alpha,\beta) - I_1(\alpha,\beta)\right] + I_1(\alpha^*,\beta) - I_{z(x)}(\alpha^*,\beta). \tag{10}$$

Proof. First, we make the change of variable $u = (\sigma/y)^{\theta}$ by obtaining that

$$\int_{\sigma}^{x} \bar{F}(y)\, dy = -\frac{\sigma}{\theta} \int_{1}^{z(y)} u^{-1-1/\theta} I_u(\alpha,\beta)\, du$$

Now, using the indefinite integration of power functions of the beta incomplete ratio function given by[1]

$$\int u^{r-1} I_u(s,t)\, du = \frac{u^r}{r} I_u(s,t) - \frac{\Gamma(s+t)\Gamma(s+r)}{r\Gamma(s)\Gamma(s+t+r)} I_u(s+r,t)$$

and by using (5) and Fundamental Theorem of Calculus, we get the result after some computations. □

3.2 Actuarial tools

The surplus process of an insurance portfolio is defined as the wealth obtained by the premium payments minus the reimbursements made at the times of claims. When this process becomes negative (if ever), we say that ruin has occurred. Let $\{U(t)\}_{t\geq 0}$ be a classical continuous time surplus process, the surplus process at time t given the initial surplus $u = U(0)$, the dynamic of $\{U(t)\}_{t\geq 0}$ is given by

$$U(t) = u + ct - S(t),$$

where $S(t) = \sum_{i=1}^{N(t)} X_i$ is the aggregate claim amount up to time t and $S(t) = 0$ if $N(t) = 0$. Here, $u \geq 0$ is the insurer's initial risk surplus at $t = 0$ and $c = (1+\theta)\alpha\mu$ is the insurer's rate of premium income per unit time with loading factor $\rho \geq 0$. Here the random variables $\{X_i\}$ are independent and identically distributed random variables with $E(X_i) = \mu$.

Under the classical model of ruin theory (Yang, 2004) and assuming a positive security loading, ρ, for the claim size distributions with regularly varying tails it is known that by using (10), an approximation of the probability of ultimate ruin,

$$\psi(u) = \Pr\left[U(t) < 0 \text{ for some } t > 0 | U(0) = u\right].$$

[1] See The Wolfram functions site (`https://functions.wolfram.com`)

can be obtained. This asymptotic approximation of the ruin function is given by

$$\psi(u) \sim \frac{1}{\rho}\bar{F}_I(u), \quad u \to \infty,$$

where $\bar{F}_I(u) = 1 - F_I(u)$.

On the other hand, let the random variable X represent either a policy limit or reinsurance deductible (from an insurer's perspective); then the limited expected value function L of X with cdf $F(x)$, is defined by

$$L(x) = E[\min(X,x)] = \int_{\sigma}^{x} y\,dF(y) + x\bar{F}(x), \tag{11}$$

which is the expectation of the cdf $F(x)$ truncated at this point. In other words, it represents the expected amount per claim retained by the insured on a policy with a fixed amount deductible of x.

A variant of this last expression is given by

$$E[\min(N,\max(0,X-M))] = \int_{M}^{M+N} (x-M)f(x)\,dx + N\bar{F}(M+N), \tag{12}$$

which represents the expected cost per claim to the reinsurance layer when the losses excess of $M > \sigma$ subject to a maximum of $N > M$.

The following result, concerning to the classical Beta distribution, is useful to derive the Propositions which will be given later in order to calculate the limited expected value function for the Beta-Pareto distribution.

Proposition 4. *Let $h(y)$ the probability density function of a classical Beta distribution with parameters $\alpha > 0$ and $\beta > 0$. Then, it is verified that,*

$$\int_0^s (1-y)^r h(y)\,dy = \frac{1}{B(\alpha,\beta)} I_s(\alpha,\beta+r), \tag{13}$$

$$\int_s^{s+t} (1-y)^r h(y)\,dy = \frac{1}{B(\alpha,\beta)}\left[I_{s+t}(\alpha,\beta+r) - I_s(\alpha,\beta+r)\right]. \tag{14}$$

Proof. It is straightforward. □

Proposition 5. *Let X be a random variable denoting the individual claim size taking values only for individual claims greater than $M > \sigma$. Let us also assumed that X follows the probability density function (4). Then the expected cost per claim of the reinsurance layer when the losses excess of $M > \sigma$ is given by*

$$L(x) = \frac{\sigma B(\alpha^*,\beta)}{B(\alpha,\beta)}\left[1 - I_{z(M)}(\alpha^*,\beta)\right] + M I_{z(M)}(\alpha,\beta), \quad \alpha > \frac{1}{\theta}. \tag{15}$$

Proof. By taking (11), making the change of variable $u = 1 - z(y)$ and using (13) we get the result after some algebra. □

Proposition 6. *Let X be a random variable denoting the individual claim size taking values only for individual claims greater than $M > \sigma$. Let us also assumed that X follows the probability density function*

(4). Then the expected cost per claim of the reinsurance layer when the losses excess of $M > \sigma$ subject to a maximum of $N > M$ is given by

$$\begin{aligned} L(x) &= \frac{\sigma}{B(\alpha,\beta)}\Big\{B(\alpha^*,1+\beta)\Big[I_{z(M)}(\alpha^*,1+\beta)-I_{z(M+N)}(\alpha^*,1+\beta)\Big] \\ &\quad +B(1+\alpha^*,\beta)\Big[I_{z(M)}(1+\alpha^*,\beta)-I_{z(M+N)}(1+\alpha^*,\beta)\Big]\Big\} \\ &\quad +(M+N)I_{z(M+N)}(\alpha,\beta)-MI_{z(M)}(\alpha,\beta), \quad \alpha > \frac{1}{\theta}. \end{aligned}$$

Proof. The proof is similar to that in Propostion 5 but using now (12) and (14). □

3.3 Mean excess function

The failure rate of the integrated tail distribution, which is given by $\gamma_I(x) = \bar{F}(x)/\int_x^\infty \bar{F}(y)\,dy$ is also obtained in closed-form. Furthermore, the reciprocal of $\gamma_I(x)$ is the mean residual life that can be easily derived. For a claim amount random variable X, the mean excess function or mean residual life function is the expected payment per claim for a policy with a fixed amount deductible of $x > 0$, where claims with amounts less than or equal to x are completely ignored. Then,

$$e(x) = E(X-x|X>x) = \frac{1}{\bar{F}(x)}\int_x^\infty \bar{F}(u)\,du. \tag{16}$$

Next result gives the mean excess function of the BP distribution in a closed-form expression.

Proposition 7. *The mean excess function of the BP distribution is given by*

$$e(x) = \frac{\sigma^2 B(\alpha^*,\beta)I_{z(x)}(\alpha^*,\beta)}{B(\alpha,\beta)I_{z(x)}(\alpha,\beta)} - x. \tag{17}$$

Proof. Using the expression

$$e(x) = \frac{E(X)-L(x)}{\bar{F}(x)},$$

which relates the mean excess function given in (16) with the limited expected value function (see Hogg and Klugman, 1984, p. 59), the result follows by using and (5), (6), (15) and a some little algebra. □

Figure 3 shows the mean residual life function (16) for special cases of parameters. It can be seen that this function can be increasing, decreasing, unimodal or anti-unimodal.

4 Numerical application

Two well-known datasets in the actuarial literature will be used here to analyze hoe the BP distribution works. The first dataset deals with large losses in a fire insurance portfolio in Denmark. These dataset include 2157 losses over 1 million Danish Krone in the years 1980-1990. A detailed statistical analysis of this set of data can be seen in McNeil (1997) in Albrecher et al. (2017) and also in Embrechts et al. (1997). It can be found in the *R* package `CASdatasets` collected at *Copenhagen Reinsurance*. The second dataset is `norfire` comprises 9181 fire losses over the period 1972 to 1992 from an unknown Norwegian company. A priority of 500 thousands of Norwegian Krone (NKR) (if

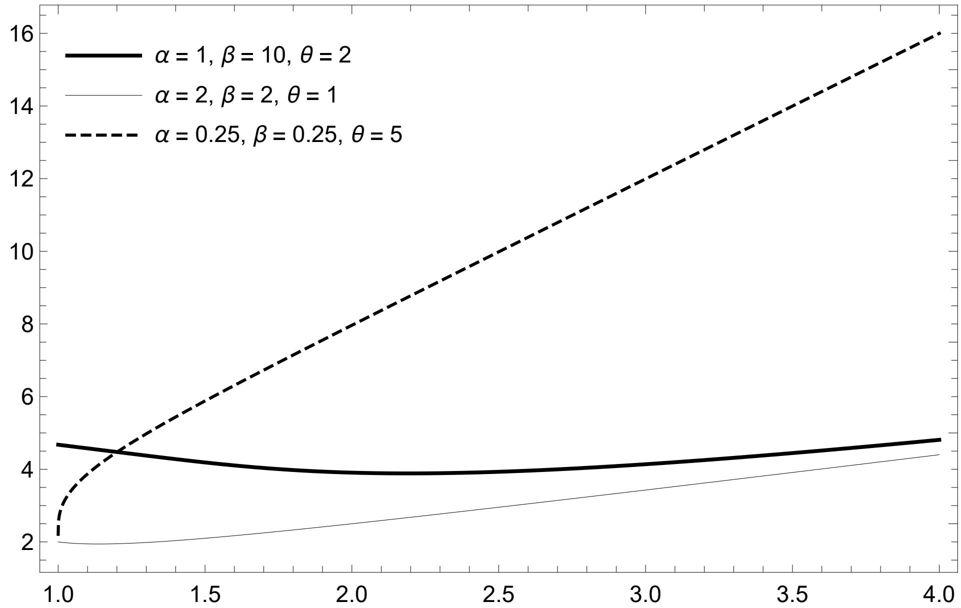

Figure 3: Mean residual life function of BP distribution for selected values of parameters when $\sigma = 1$.

this amount is exceeded, the reinsurer becomes liable to pay) was applied to obtain this dataset. This set of data is is also available in the R package `CASdatasets`.

Below in Table 1, parameter estimates and their corresponding p-values together with the negative value of the maximum likelihood function (NLL) evaluated at the maximum likelihood estimates for the two datasets considered are shown. Also the Kolmogorov-Smirnov (KS) and Anderson-Darling (AD) test statistics are displayed. As judged by the corresponding p-values, the BP distribution is not rejected for neither of the tests for the Danish dataset. However, for the Norwegian set of data, the BP distribution is rejected at the 5% significance level.

Danish dataset						
BP	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\theta}$	NLL	KS	AD
$\sigma = 0.999$	32.113	1.157	0.046	3341.895	0.027	0.00051
	(< 0.001)	(< 0.001)	(< 0.001)		(0.398)	(0.545)
Norwegian dataset						
BP	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\theta}$	NLL	KS	AD
$\sigma = 490$	77.090	1.549	0.020	20979.635	0.039	0.0011
	(< 0.001)	(< 0.001)	(< 0.001)		0.038	0.057

Table 1: Parameter estimates and their p-values (in brackets), negative of the maximum of the log likelihood function, Kolmogorov-Smirnov and Anderson-Darling test for the BP distribution.

These results are confirmed in Figure 4 where the empirical and theoretical cdf are plotted. It is observable that for the Danish dataset (left panel) the theoretical model adheres closer to the empirical data than for the Norwegian dataset.

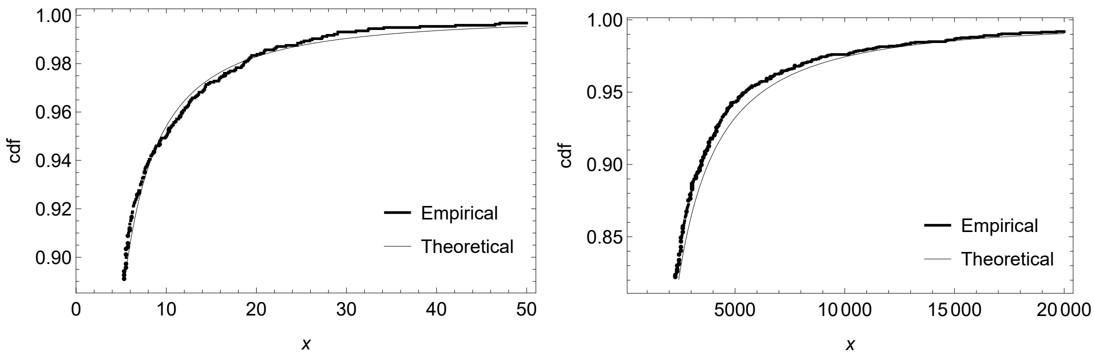

Figure 4: Empirical (thick line) and fitted cumulative distribution function for the Danish (left) and Norwegian (right) datasets.

In Figure 5, the limited expected value for the two sets of data have been plotted. It can be seen that when the policy limit x increases the theoretical model overestimates the empirical values for the Danish dataset. The converse occurs for the other set of data.

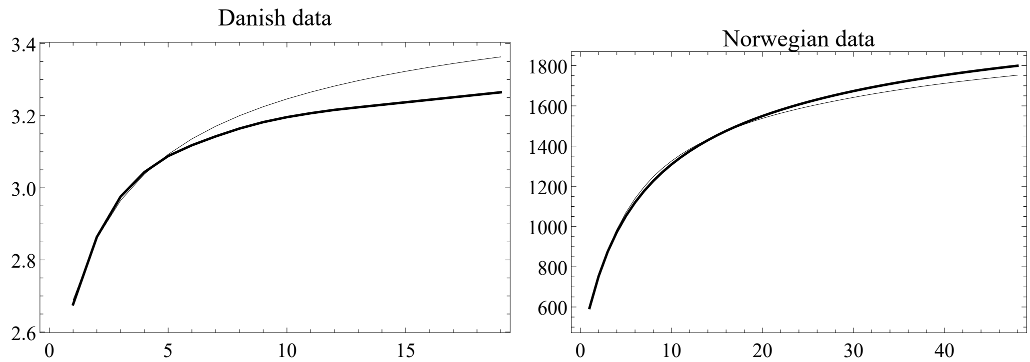

Figure 5: Empirical (thick line) and fitted limited expected values for the Danish (left) and Norwegian (right) datasets.

In Table 2, the tail value at risk (TVaR) (or first order tail moment), for different security levels has been calculated for the BP distribution. This risk measure describes the expected loss given that the loss exceeds the security level (quantile). These values have been calculated directly from the data. Empirical values have also been obtained. For the different risk levels it is discernible that the BP distribution overestimate the empirical TVaR values for the three security levels considered and the two sets of data.

5 Conclusions

In this work, the Beta-Pareto distribution, a generalization of the Pareto distribution that was introduced in the statistical literature not long time ago, has been extended and applied in financial and actuarial settings. In addition, several interesting properties related with the right-tail of the distribution were provided including the integrated tail distribution and the limited expected values among others. These properties, which had not been revealed until now, make the Beta-Pareto

	Risk Level α		
Risk Level	0.90	0.95	0.99
Danish dataset			
Empirical	15.637	24.305	61.376
BP	18.556	33.079	85.719
Norwegian dataset			
Empirical	9936.597	15635.295	42475.200
BP	11423.301	17576.914	53270.056

Table 2: Tail Value at Risk for different risk levels.

distribution a plausible alternative for applications in these fields. Additionally, its usefulness has been proven in its good performance against some well-known datasets usually considered in general insurance, improving the performance of other traditionally-used loss models.

Acknowledgments

The authors would like to express their gratitude to the Editor J.M. Sarabia for his kind invitation to write this paper and also for his valuable comments which have improved the content of this manuscript. EGD would also like to acknowledge partial financial support received from Ministerio de Economía, Industria y Competitividad. Agencia Estatal de Investigación (project ECO2017-85577-P).

References

Akinsete, A., F. Famoye, and C. Lee (2008). The beta-Pareto distribution. *Statistics 42*(6), 547–563.

Albrecher, H., J. Beirlant, and J.L. Teugels (2017). *Reinsurance: Actuarial and Statistical Aspects*. Wiley.

Arnold, B.C. (1983). *Pareto Distributions*. International Cooperative Publishing House, Silver Spring, MD.

Bingham, N.H. (1987). *Regular Variation*. Cambridge University Press, Cambridge.

Boland, P.J. (2007). *Statistical and Probabilistic Methods in Actuarial Science*. Chapman & Hall.

Boyd, A.V. (1988). Fitting the truncated Pareto distribution to loss distributions. *Journal of the Staple Inn Actuarial Society 31*, 151–158.

Brazauskas, V. and R. Serfling (2003). Favorable estimator for fitting Pareto models: a study using goodness-of-fit measures with actual data. *ASTIN Bulletin 33*(2), 365–381.

Calderín-Ojeda, E. (2016). The distribution of all French communes: A composite parametric approach. *Physica A 450*, 385–394.

Calderín-Ojeda, E. and C.F. Kwok (2016). Modeling claims data with composite stoppa models. *Scandinavian Actuarial Journal 2016*(9), 817–836.

Embrechts, P., C. Klüppelberg, and T. Mikosch (1997). *Modelling Extremal Events for Insurance and Finance*. Springer Verlag.

Ghitany, M.E., E. Gómez-Déniz, and S. Nadarajah (2018). A new generalization of the Pareto distribution and its application to insurance data. *Journal of Risk and Financial Management (Special Issue: Extreme Values and Financial Risk) 11*(10), 1–14.

Gómez-Déniz, E. and E. Calderín (2014). A suitable alternative to the Pareto distribution. *Hacettepe Journal of Mathematics and Statistics 43*(5), 843–860.

Gómez-Déniz, E. and E. Calderín (2015). On the use of the Pareto ArcTan distribution for describing city size in Australia and New Zealand. *Physica A 436*, 821–832.

Harris, C.M. (1968). The Pareto distribution as a queue service discipline. *Operations Research 16*(2), 307–313.

Hesselager, O. (1993). A class of conjugate priors with applications to excess-of-loss reinsurance. *ASTIN Bulletin 23*(1), 77–93.

Hogg, R. and S.A. Klugman (1984). *Loss Distributions*. John Wiley and Sons, New York.

Jessen, A.H. and T. Mikosch (2006). Regularly varying functions. *PUBLICATIONS DE L'INSTITUT MATHÉMATIQUE. Nouvelle série 80*(94), 171–192.

Jones, M.C. (2004). Families of distributions arising from distributions of order statistics. *Test 13*, 1–43.

Kleiber, C. and S. Kotz (2003). *Statistical Size Distributions in Economics and Actuarial Sciences*. John Wiley & Sons, Inc.

Konstantinides, D.G. (2018). *Risk Theory. A Heavy Tail Approach*. World Scientific Publishing.

Mandelbrot, B. (1960). The Pareto-Lévy law and the distribution of income. *International Economic Review 1*(2), 79–106.

McNeil, A.J. (1997). Estimating the tails of loss severity distribution using extreme value theory. *ASTIN Bulletin 27*(1), 117–137.

Rolski, T., H. Schmidli, V. Schmidt, and J. Teugel (1999). *Stochastic Processes for Insurance and Finance*. John Wiley & Sons.

Rosen, K.T. and M. Resnick (1980). The size distribution of cities: An examination of the Pareto law and primacy. *Journal of Urban Economics 8*(2), 165–186.

Rytgaard, M. (1990). Estimation in the Pareto distribution. *ASTIN Bulletin 20*(2), 201–216.

Sarabia, J.M. and F. Prieto (2009). The Pareto-positive stable distribution: A new descriptive model for city size data. *Physica A 388*, 4179–4191.

Scollnik, D. (2007). On composite lognormal-Pareto models. *Scandinavian Actuarial Journal 2007*(1), 20–33.

Stoppa, G. (1990). Proprietà campionarie di un nuovo modello Pareto generalizzato. *XXXV Riunione Scientifica della Società Italiana di Statistica, Padova: Cedam*, 137–144.

Yang, H. (2004). Crámer-Lundberg asymptotics. In *Encyclopedia of Actuarial Science*, pp. 1–6. Wiley.