PARETO CONFLUENT HYPERGEOMETRIC DISTRIBUTION

Jailson Araujo Rodrigues
Instituto Federal da Bahia
jailsondearaujo@yahoo.com.br

Jaime dos Santos Filho
Instituto Federal da Bahia
jaime@ifba.edu.br

Lucas Monteiro Chaves
Universidade Federal de Lavras
lucas@dex.ufla.br

RESUMO

Neste trabalho é introduzida uma distribuição com cinco parâmetros, denominada distribuição hipergeométrica confluente Pareto. Essa nova distribuição gerada a partir de uma variável aleatória hipergeométrica confluente inclui, como casos especiais, algumas importantes distribuições, como a beta Pareto, a Pareto exponenciada e a Pareto. Algumas das principais propriedades dessa distribuição são deduzidas, incluindo o momento de ordem n, média, variância, coeficiente de assimetria, coeficiente de curtose e a entropia de Rényi. A estimação dos parâmetros utilizando o método da máxima verossimilhança e o método dos momentos também é discutida.

ABSTRACT

In this note, is introduced a five-parameter distribution, so-called the Pareto confluent hypergeometric distribution. This new distribution generated from a confluent hypergeometric random variable includes some important distributions as special case, such as beta Pareto, Pareto exponentiated and Pareto. Some of the main properties of this distribution are derived, including, nth moment, mean, variance, skewness, kurtosis and Rényi entropy. The estimation of parameters using the methods of moments and maximum likelihood is also discussed.

Palavras-chave: Distribuição Gama, distribuição beta, entropia de Rényi.

1 INTRODUCTION

The Pareto distribution is widely used in various areas of applied sciences, including, survival analysis, hydrology, telecommunications.

In recent years, several authors have proposed generalizations of existing distributions. This is justified because the traditional distributions often do not provide good fit in relation to the real data set studied. For example, Mudholkar [10] studied the exponentiated Weibull distribution. Eugene [5] introduced and studied the beta normal distribution. Nadarajah [11] proposed the beta Gumbel distribution. Nadarajah [12] studied the beta exponential distribution. Ali [3] presented the exponentiated Pareto distribution. Lee [9] proposed the beta Weibull distribution. Akinsete [2] introduced the beta Pareto distribution. Khan [8] proposed the beta inverse Weibull distribution. Silva [16] studied the beta modified Weibull distribution. Pascoa [15] introduced the Kumaraswamy generalized gamma distribution and Paranaíba [14] presented the beta Burr XII.

In this article, is presented a new five-parameter distribution, so-called the Pareto confluent hypergeometric distribution. Some of the main properties of this distribution are derived.

The paper is organized as follows. In Section 2 is defined the Pareto confluent hypergeometric distribution and some special sub-models are discussed. The nth moment, mean, variance, skewness and kurtosis are derived in Section 3. In Section 4 is derived the Rényi entropy. Finally, in Section 5, the estimation of parameters using the methods of moments and maximum likelihood is discussed.

The calculations of this note involve the gamma function defined by

$$\Gamma(a) = \int_0^\infty t^{a-1} \exp(-t) dt \tag{1}$$

the incomplete gamma function defined by

$$\gamma(a,x) = \int_0^x t^{a-1} \exp(-t) dt \tag{2}$$

the beta function defined by

$$B(a,b) = \int_0^1 t^{a-1} (1-t)^{b-1} dt \tag{3}$$

and the confluent hypergeometric function defined by

$${}_1F_1(a;c;x) = \sum_{j=0}^{\infty} \frac{(a)_j}{(c)_j j!} x^j \tag{4}$$

where $(x)_j = (x)(x+1)\ldots(x+j-1)$ denotes the Pochhammer symbol. The properties of these special functions are given by Oldham [13].

We also need the following important lemmas.

Lemma 1: *(Equation (3.381.1), Gradshteyn [7]). For $\nu > 0$,*

$$\int_0^u x^{\nu-1} \exp(-\mu x) dx = \mu^{-\nu} \gamma(\nu, \mu u).$$

Lemma 2: *(Equation (3.383.1), Gradshteyn [7]). For $\mu > 0$ and $\nu > 0$,*

$$\int_0^u x^{\nu-1} (u-x)^{\mu-1} \exp(\beta x) dx = B(\mu,\nu) u^{\mu+\nu-1} {}_1F_1(\nu; \mu+\nu; \beta u).$$

2 THE MODEL

The confluent hypergeometric distribution presented by Gordy [6] has probability density function (pdf) defined by

$$f(x) = \frac{x^{a-1}(1-x)^{b-1} \exp(-cx)}{B(a,b)\, {}_1F_1(a; a+b; -c)} \tag{5}$$

where $1 > x > 0$, $a > 0$, $b > 0$ and $+\infty > c > -\infty$.

The works of Rodrigues [1] and Cordeiro [4] can be used to construct a new class of generalized distribution: If G denotes the cumulative distribution (cdf) of a random variable, then a generalized class of distribution can be defined by

$$F(x) = \frac{1}{B(a,b)\ {}_1F_1(a; a+b; -c)} \int_0^{G(x)} t^{a-1}(1-t)^{b-1} \exp(-ct)\, dt \tag{6}$$

Now consider the Pareto distribution with cdf defined by

$$G(x) = 1 - \left(\frac{x}{s}\right)^{-k} \tag{7}$$

where $x \geq s$, $k > 0$ and $s > 0$. Inserting the cdf (7) in (6), we obtain the cdf of Pareto confluent hypergeometric (PCH) distribution

$$F(x) = \frac{1}{B(a,b)\ {}_1F_1(a; a+b; -c)} \int_0^{1-\left(\frac{x}{s}\right)^{-k}} t^{a-1}(1-t)^{b-1} \exp(-ct)\, dt \tag{8}$$

for $x \geq s$, $a > 0$, $b > 0$, $+\infty > c > -\infty$, $k > 0$ and $s > 0$. The pdf and hazard rate functions associated with (8) are

$$f(x) = \frac{k s^{bk} x^{-(bk+1)} \left[1 - (x/s)^{-k}\right]^{a-1} \exp\left[c\,(x/s)^{-k}\right]}{\exp(c) B(a,b)\ {}_1F_1(a; a+b; -c)} \tag{9}$$

and

$$\lambda(x) = \frac{k s^{bk} x^{-(bk+1)} \left[1 - (x/s)^{-k}\right]^{a-1} \exp\left\{-c\left[1 - (x/s)^{-k}\right]\right\}}{B(a,b)\ {}_1F_1(a; a+b; -c) - \displaystyle\int_0^{1-(x/s)^{-k}} t^{a-1}(1-t)^{b-1} \exp(-ct)\, dt} \tag{10}$$

If $\mid z \mid < 1$ and $b > 0$ is real non-integer, we have the series representation

$$(1-z)^{b-1} = \sum_{j=0}^{\infty} \frac{(-1)^j \Gamma(b) z^j}{\Gamma(b-j) j!} \tag{11}$$

Using (1) and the Lemma 1, the cdf of Pareto confluent hypergeometric distribution (8) can be written as

$$F(x) = \frac{\Gamma(b)}{B(a,b)\ {}_1F_1(a; a+b; -c)} \sum_{j=0}^{\infty} \frac{(-1)^j\ \gamma\left(a+j, c\left[1 - (x/s)^{-k}\right]\right)}{c^{a+j} \Gamma(b-j) j!} \tag{12}$$

When $b > 0$ is a interger, the above sum stops at $b-1$.

The Pareto confluent hypergeometric distribution represents a generalization of some well-known distributions. Clearly, the Pareto distribution is a sub-model when $a = b = 1$ and $c = 0$. For $c = 0$, (9) is referred to as beta Pareto distribution, see Akinsete [2]. If $b = 1$ and $c = 0$ the reduced model becomes the exponentiated Pareto distribution intorduced by Ali [3].

Some of the possible shapes of the PCH density (9) and hazard rate function (10) are illustrated in Figure 1.

3 MOMENTS

Theorem 1: *If X has the PCH pdf (9), then its nth moment can be written as*

$$E(X^n) = \frac{s^n B\left(a, b - \frac{n}{k}\right)\ {}_1F_1\left(a; a+b-\frac{n}{k}; -c\right)}{B(a,b)\ {}_1F_1(a; a+b; -c)} \tag{13}$$

for $n \geq 1$ and $b > \frac{n}{k}$.

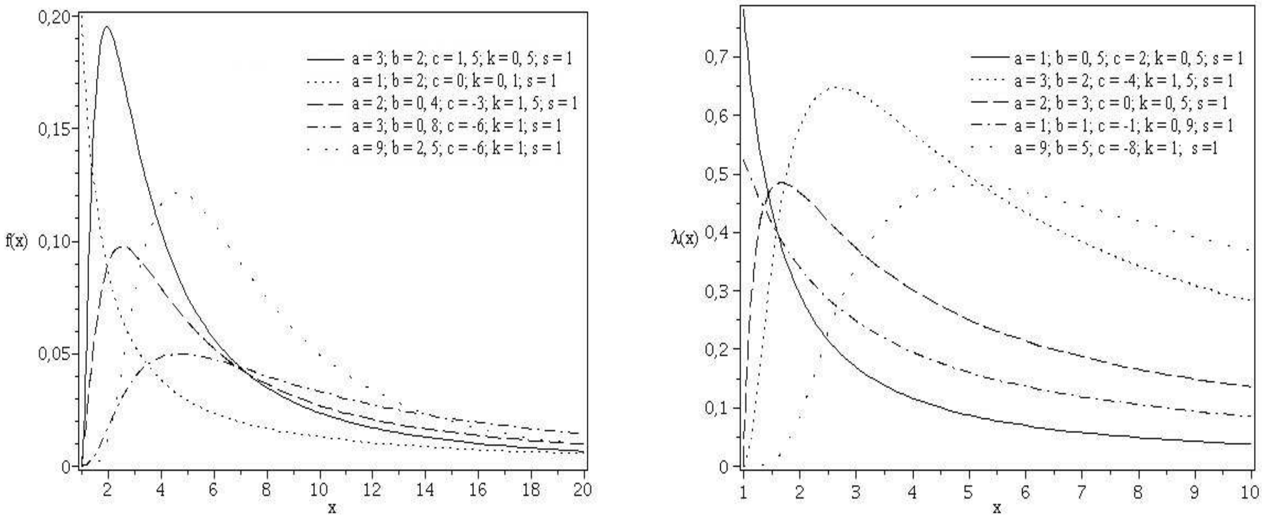

FIGURA 1: Plots of the PCH density and hazard rate function for selected parameter values.

Proof:

$$E\left(X^n\right)=\int_s^{\infty}\frac{ks^{bk}x^{n-(bk+1)}\left[1-(x/s)^{-k}\right]^{a-1}\exp\left[c\,(x/s)^{-k}\right]}{\exp(c)B(a,b)\ {}_1F_1\left(a;a+b;-c\right)}dx \tag{14}$$

substituting $t=(x/s)^{-k}$, the integral (14) can be rewritten as

$$E\left(X^n\right)=\frac{s^n}{\exp(c)B(a,b)\ {}_1F_1\left(a;a+b;-c\right)}\int_s^{\infty}t^{b-1-\frac{n}{k}}\left(1-t\right)^{a-1}\exp\left(ct\right)dt \tag{15}$$

If $b>n/k$, the direct application of Lemma 2 shows that (15) can be rewritten as

$$E\left(X^n\right)=\frac{s^nB\left(a,b-\dfrac{n}{k}\right)\ {}_1F_1\left(b-\dfrac{n}{k};a+b-\dfrac{n}{k};c\right)}{\exp\left(c\right)B\left(a,b\right)\ {}_1F_1\left(a;a+b;-c\right)} \tag{16}$$

The result of the theorem follows by direct application of Kummer's relation

$${}_1F_1\left(c-a;c;z\right)=\exp\left(c\right)\ {}_1F_1\left(a;c;-z\right) \tag{17}$$

■

In particular,

$$E\left(X\right)=\frac{sB\left(a,b-\dfrac{1}{k}\right)\ {}_1F_1\left(a;a+b-\dfrac{1}{k};-c\right)}{B\left(a,b\right)\ {}_1F_1\left(a;a+b;-c\right)} \tag{18}$$

$$E\left(X^2\right)=\frac{s^2B\left(a,b-\dfrac{2}{k}\right)\ {}_1F_1\left(a;a+b-\dfrac{2}{k};-c\right)}{B\left(a,b\right)\ {}_1F_1\left(a;a+b;-c\right)} \tag{19}$$

$$E\left(X^3\right)=\frac{s^3B\left(a,b-\dfrac{3}{k}\right)\ {}_1F_1\left(a;a+b-\dfrac{3}{k};-c\right)}{B\left(a,b\right)\ {}_1F_1\left(a;a+b;-c\right)} \tag{20}$$

$$E\left(X^4\right) = \frac{s^4 B\left(a, b - \frac{4}{k}\right)\ {}_1F_1\left(a; a + b - \frac{4}{k}; -c\right)}{B\left(a, b\right)\ {}_1F_1\left(a; a + b; -c\right)} \tag{21}$$

The variance, skewness, and kurtosis measures can now be calculated using the relations

$$Var\left(X\right) = E\left(X^2\right) - E^2\left(X\right) \tag{22}$$

$$Skewness\left(X\right) = \frac{E\left(X^3\right) - 3E\left(X\right)E^2\left(X\right) + 2E^3\left(X\right)}{Var^{3/2}\left(X\right)} \tag{23}$$

$$Kurtosis\left(X\right) = \frac{E\left(X^4\right) - 4E\left(X\right)E^3\left(X\right) + 6E\left(X^2\right)E^2\left(X\right) - 3E^4\left(X\right)}{Var^2\left(X\right)} \tag{24}$$

4 Entropy

An entropy of a random variable X is a measure of variation of the uncertainty. Rényi entropy is defined by

$$H_\alpha\left(X\right) = \frac{1}{1-\alpha}\log\left[\int f^\alpha(x)dx\right] \tag{25}$$

for $\alpha > 0$ and $\alpha \neq 1$.

Theorem 2: *If X has the PCH pdf (9), then its Rényi entropy is given by*

$$H_\alpha\left(X\right) = \log s - \log k + \frac{c\alpha}{\alpha-1} + \frac{\alpha}{\alpha-1}\log B(a,b) + \frac{1}{1-\alpha}\log B\left(\alpha a - \alpha + 1, \alpha b + \frac{\alpha}{k} - \frac{1}{k}\right)$$
$$+\frac{\alpha}{\alpha-1}\log\ {}_1F_1\left(a; a+b; -c\right) + \frac{1}{\alpha-1}\log\ {}_1F_1\left(\alpha b + \frac{\alpha}{k} - \frac{1}{k}; \alpha a + \alpha b - \alpha + 1 + \frac{\alpha}{k} - \frac{1}{k}; c\alpha\right) \tag{26}$$

for $\alpha\left(bk+1\right) > 1$ and $\alpha\left(a-1\right) > -1$.

Proof: One can express

$$\int f^\alpha(x)dx = \int_s^\infty \frac{k^\alpha s^{bk\alpha} x^{-\alpha(bk+1)}\left[1 - (x/s)^{-k}\right]^{\alpha(a-1)}\exp\left[c\alpha\left(x/s\right)^{-k}\right]}{\exp(c\alpha)B^\alpha(a,b)\ {}_1F_1^\alpha\left(a; a+b; -c\right)}dx \tag{27}$$

substituting $t = (x/s)^{-k}$, the integral (27) can be rewritten as

$$\int f^\alpha(x)dx = \frac{k^{\alpha-1}s^{1-\alpha}\exp(-c\alpha)}{B^\alpha(a,b)_1F_1^\alpha\left(a; a+b; -c\right)}\int_s^\infty t^{b\alpha+\frac{\alpha}{k}-\frac{1}{k}-1}(1-t)^{\alpha(a-1)}\exp(c\alpha t)dt \tag{28}$$

If $\alpha\left(bk+1\right) > 1$ and $\alpha\left(a-1\right) > -1$, the direct application of Lemma 2 shows that (28) can be rewritten as

$$\int f^\alpha(x)dx = \frac{k^{\alpha-1}s^{1-\alpha}B\left(\alpha a - \alpha + 1, \alpha b + \frac{\alpha}{k} - \frac{1}{k}\right)}{\exp\left(c\alpha\right)B^\alpha\left(a,b\right)\ {}_1F_1^\alpha\left(a; a+b; -c\right)}\ {}_1F_1\left(\alpha b + \frac{\alpha}{k} - \frac{1}{k}; \alpha a + \alpha b - \alpha + 1 + \frac{\alpha}{k} - \frac{1}{k}; c\alpha\right) \tag{29}$$

The result of the theorem follows by substituting equations (29) into (25).

■

5 ESTIMATION

In this section, we consider estimation of the five parameters by method of moments and the maximum likelihood of the PCH distribution. Let $x_1, \ldots, x_n$ be a random sample of size n from the PCH distribution given by (9). Under the method of moments, equating $E\left(X^j\right)$ with the corresponding sample moment,

$$M_j = \frac{1}{n}\sum_{i=1}^{n} x_i^j, \quad j = 1, \ldots, 5 \tag{30}$$

respectively, one obtains the system of equations

$$\frac{s^j B\left(a, b - \frac{j}{k}\right)\ {}_1F_1\left(a; a+b-\frac{j}{k}; -c\right)}{B\left(a, b\right)\ {}_1F_1\left(a; a+b; -c\right)} = M_j, \quad j = 1, \ldots, 5 \tag{31}$$

which can be solved simultaneously to give estimates for a, b, c, k and s.

The log-likelihood for a random sample $x_1, \ldots, x_n$ from the PCH distribution given by (9) is:

$$\begin{aligned}\log L\left(a,b,c,k,s\right) &= n\log k + nbk\log s - nc - n\log B(a,b) - n\log\ {}_1F_1\left(a; a+b; -c\right) + c\sum_{i=0}^{n}\left(\frac{x_i}{s}\right)^{-k} \\ &-(bk+1)\sum_{i=0}^{n}\log x_i + (a-1)\sum_{i=0}^{n}\log\left[1-\left(\frac{x_i}{s}\right)^{-k}\right]\end{aligned} \tag{32}$$

The derivatives of this log-likelihood with respect to a, b, c, k, and s are:

$$\frac{\partial\log L}{\partial a} = n\psi(a+b) - n\psi(a) + \sum_{i=0}^{n}\log\left[1-\left(\frac{x_i}{s}\right)^{-k}\right] - \frac{n}{{}_1F_1\left(a;a+b;-c\right)}\frac{\partial\ {}_1F_1\left(a;a+b;-c\right)}{\partial a} \tag{33}$$

$$\frac{\partial\log L}{\partial b} = nk\log s + n\psi(a+b) - n\psi(b) - k\sum_{i=0}^{n}\log x_i - \frac{n}{{}_1F_1\left(a;a+b;-c\right)}\frac{\partial\ {}_1F_1\left(a;a+b;-c\right)}{\partial b} \tag{34}$$

$$\frac{\partial\log L}{\partial c} = s^k\sum_{i=0}^{n}x_i^{-k} - \frac{n}{{}_1F_1\left(a;a+b;-c\right)}\frac{\partial\ {}_1F_1\left(a;a+b;-c\right)}{\partial c} - n \tag{35}$$

$$\begin{aligned}\frac{\partial\log L}{\partial k} &= nb\log s + cs^k\log s\sum_{i=0}^{n}x_i^{-k} - cs^k\sum_{i=0}^{n}x^{-k}\log x_i - b\sum_{i=0}^{n}\log x_i + s^k(1-a)\log s\sum_{i=0}^{n}\frac{1}{x^k - s^k} \\ &+\frac{n}{k} + s^k(a-1)\sum_{i=0}^{n}\frac{\log x_i}{x^k - s^k}\end{aligned} \tag{36}$$

$$\frac{\partial\log L}{\partial s} = \frac{nbk}{s} + cks^{k-1}\sum_{i=0}^{n}x_i^{-k} + ks^{k-1}(1-a)\sum_{i=0}^{n}\frac{1}{x^k - s^k} \tag{37}$$

where $\psi(x) = d\log\Gamma(x)/dx$ is the digamma function. Setting these expressions to zero and solving them simultaneously yields the maximum-likelihood estimates of the five parameters.

6 ACKNOWLEDGMENTS

The authors would like to thank the referees and the editor for carefully reading the paper and for their great help in improving the paper.

REFERÊNCIAS

[1] J. de A. Rodrigues, L. M. Chaves e F. Castellares: *Uma nova classe de distribuições generalizadas*. Ten. Mat. Apl. Comput., 13(2):167–178, 2012.

[2] A. Akinsete, F. Famoye e C. Lee: *The beta-Pareto distribution*. Statistics, 42:547–563, 2008.

[3] M. M. Ali, M. Pal e J. Woo: *Some exponentiated distributions*. Korean Comm. Stat., 14:93–109, 2007.

[4] G. M. Cordeiro e M. Castro: *A new family of generalized distributions*. J. Stat. Comp. Simulation, 81:883–898, 2011.

[5] N. Eugene, C. Lee e F. Famoye: *Beta-normal distribution and its application*. Commun. Stat. - Theory and Methods, 31:497–512, 2002.

[6] M. B. Gordy: *Computationally Convenient Distributional Assumptions for Common-Value Auctions*. Comp. Econom., 12(1):61–78, 1998.

[7] I. S. Gradshteyn e I. M. Ryzhik: *Table of Integrals, Series, and Products*. Academic Press, 2000.

[8] M. S. Khan: *The beta inverse Weibull distribution*. Int. Trans. Math. Sciences and Comp., 3(1):113–119, 2010.

[9] C. Lee, F. Famoye e O. Olumolade: *Beta-Weibull distribution*. J. Mod. Appl. Stat. Meth., 6(1):173–186, 2007.

[10] G. S. Mudholkar, D. K. Srivastava e M. Friemer: *The beta-Pareto distribution*. Technometrics, 37:436–445, 1995.

[11] S. Nadarajah e S. Kotz: *The beta Gumbel distribution*. Math. Prob. Eng., 10:323–332, 2004.

[12] S. Nadarajah e S. Kotz: *The beta exponential distribution*. Reliab. Eng. and Syst. Safety, 10:689–697, 2006.

[13] K. Oldham, J. Myland e J. Spanier: *An atlas of functions: with equator, the atlas function calculator*. Springer, 2009.

[14] P. F. Paranaíba, E. M. M. Ortega, G. M. Cordeiro e R. R. Pescim: *The beta Burr XII distribution with application to lifetime data*. Commun. Stat. - Theory and Methods, 55(2):1118–1136, 2011.

[15] M. A. R. de Pascoa, E. M. M. Ortega e G. M. Cordeiro: *The Kumaraswamy generalized gamma distribution with application in survival analysis*. Statistical Methodology, 8(5):411–433, 2011.

[16] G. O. Silva, E. M. M. Ortega e G. M. Cordeiro: *The beta modified Weibull distribution*. Math. and Stat., 16(3):409–430, 2010.