İSTATİSTİK: JOURNAL OF THE TURKISH STATISTICAL ASSOCIATION

İSTATİSTİK

Vol. 6, No. 2, July 2013, pp. 66–72

ISSN 1300-4077 | 13 | 2 | 66 | 72

NUMERICAL CHARACTERISTICS OF PARETO DISTRIBUTION UNDER UNCERTAINTY

Turkan ERBAY DALKILIC

Karadeniz Technical University, Faculty of Sciences
Department of Statistics and Computer Sciences
61080, Trabzon/TURKEY

Tulay KESEMEN

Karadeniz Technical University, Faculty of Sciences
Department of Mathematics
61080, Trabzon/TURKEY

Fatih TANK*

Ankara University, Faculty of Sciences
Department of Statistics
06100 Tandogan-Ankara / TURKEY

Abstract: In this study, we consider the Pareto distribution where the value of k parameter is uncertain. Some statistical quantities such as expected value, variance, skewness and kurtosis of the pareto distribution are derived by using fuzzy numbers. In addition to the these numerical characteristics, a numerical study is provided. Furthermore, large claims data from Society of Actuaries (SOA) Group Medical insurance is examined and k parameter of Pareto distribution is obtained by using the maximum likelihood estimation method and the numerical characteristics of this distribution are calculated under uncertainty.

Key words: Pareto distribution, uncertain probability, medical insurance.

History: Submitted: 5 November 2012; Revised: 1 April 2013; Accepted: 3 April 2013

1. Introduction

Pareto distribution, which is introduced by Italian-born Swiss professor Vilfredo Pareto (1848-1923), is a heavy tailed distribution [10]. Heavy-tailed distributions are crucial for modeling extreme loss in insurance applications such as medical and catastrophic insurance. On the other hand, in financial applications, heavy-tailed distributions gives opportunity to interpreting of financial status such as ruin. Detailed information about Pareto distribution is summarized by Kleiber and Kotz (2003) [11].

It's known that fuzzy logic is used in numerous scientific areas, for instance, electronic, inventory theory, reliability theory, queuing theory, biology, medical, decision theory, mathematics, statistics, and specially in insurance theory. Note that, insurance theory has numerous areas with applications for fuzzy logic. These include classification, fuzzy future and present values, pricing, asset location and cash flows, and investment ([1, 2, 4, 7, 13, 15, 16, 18]). As discussed by Shapiro (2004), measures of fuzziness can be used to classify life insurance risks. Buckley (1987) investigated fuzzy time-value-of-money aspects of actuarial pricing, when he investigated the fuzzy future and present values of fuzzy cash amounts, using fuzzy interest rates, and both crisp and fuzzy periods. Lemaire (1990) discussed the computation of the fuzzy premium for a pure endowment policy using fuzzy arithmetic by a model which is given by Buckley (1987). Young (1996) described how fuzzy logic could be used to make pricing decision in group health insurance. Carretero and Viejo (2000) investigate

* Corresponding author. E-mail address: tank@ankara.edu.tr (F. Tank)

to use of fuzzy mathematical programming for insurance pricing decisions. Pareto distribution can be used for modelling of probabilities in extreme losses. Due to the uncertainty, investigating of fuzzy numerical characteristics of Pareto distribution have vital importance for insurance industry. In accordance with Buckley and Eslami' method [3], we investigated the Pareto distribution and its numerical characteristics where k parameters of the probability values are uncertain. After proving that the Pareto distribution fits to medical insurance claims data, k parameter of Pareto distribution is obtained by using the maximum likelihood estimation method then the numerical characteristics of this distribution are recalculated under uncertainty.

We might use the standard notation $X \sim Par(m,k)$. Distribution function (c.d.f) and density function (p.d.f) of the Pareto distribution is given by

$$F(x) = 1 - \left(\frac{m}{x}\right)^k \ , \ 0 < m \le x < \infty, \tag{1.1}$$

$$f(x;k,m) = k\frac{m^k}{x^{k+1}} \ , \ 0 < m \le x < \infty, \tag{1.2}$$

respectively. In here, k is called the Pareto index [5, 14, 21], $k > 0$ is a shape parameter (also measuring the heaviness of the right tail) and m is a scale parameter.

2. Literature Review

After Zadeh's well konown study [19], fuzzy logic had gained recognition and was intensively used in applied mathematics and computer sciences ([6],[8]-[12],[20]).

Definition 1. Let X is a classical set of objects, called the universe, whose generic elements are denoted by X. The membership in a crisp subset of X is often viewed as characteristic function μ_A from X to $\{0,1\}$ such that:

$$\mu_A(x) = \begin{cases} 1 & , \ \text{if } x \in A \\ 0 & , \ \text{if } x \notin A \end{cases} \tag{2.1}$$

where $\{0,1\}$ is called a valuation set. If the valuation set is allowed to be the real interval [0, 1], A is called *fuzzy set*. $\mu_A(x)$is the degree of membership of X in A ([17],[20]).

Definition 2. If A is a fuzzy set, then $\tilde{A}(x) \in [0,1]$ is the membership function for $\tilde{A}$ evaluated a real number x. An α-cut of $\tilde{A}$ written $\tilde{A}_\alpha$, is defined as $\left\{x \,\middle|\, \tilde{A}(x) \ge \alpha\right\}$, for $0 < \alpha \le 1$.

Definition 3. Let A be a fuzzy set in $\mathbb{R}$. A is a called a *fuzzy number* if: A is normal, A is convex, A has a bounded support, and all α-cuts of A are closed intervals in $\mathbb{R}$. Two special classes of fuzzy numbers often are used in practice; they are triangular and trapezoidal fuzzy numbers. Since the triangular fuzzy number will be used in application in this work, we will be defined this kind of fuzzy number. A triangular fuzzy number A is defined by three numbers $a < b < c$. In this regard a triangular membership function is given by

$$\mu_A(x) = \begin{cases} \frac{x-a}{b-a} & , \ a \le x < b \\ 1 & , \ x = b \\ \frac{c-x}{c-b} & , \ b < x \le c \\ 0 & , \ x > c \text{ or } x < a \end{cases} \tag{2.2}$$

where fuzzy set A in $\mathbb{R}$

It is called *triangular fuzzy number*. This triangular number is based on the interval $[a,c]$ and $\mu_A(b) = 1$ [17].

Let $[a_1,c_1]$ and $[a_2,c_2]$ be two closed, bounded, intervals of real numbers. The basic operations from interval arithmetics are,

$$[a_1,c_1] + [a_2,c_2] = [a_1 + a_2, c_1 + c_2] \tag{2.3}$$

$$[a_1, c_1] - [a_2, c_2] = [a_1 - c_2, c_1 - a_2] \tag{2.4}$$

$$[a_1, c_1] \,/\, [a_2, c_2] = [a_1, c_1] \left[\frac{1}{c_2}, \frac{1}{a_2}\right] \tag{2.5}$$

and

$$[a_1, c_1] \cdot [a_2, c_2] = [\gamma, \beta] \tag{2.6}$$

where

$$\gamma = \min\{a_1a_2, a_1c_2, c_1a_2, c_1c_2\}, \tag{2.7}$$

$$\beta = \max\{a_1a_2, a_1c_2, c_1a_2, c_1c_2\}. \tag{2.8}$$

Multiplication and division may be further simplified if $a_1 > 0$ and $c_2 < 0$, or $c_1 > 0$ and $c_2 < 0$, are known. For instance, if $a_1 \geq 0$ and $a_2 \geq 0$, then

$$[a_1, c_1] \cdot [a_2, c_2] = [a_1a_2, c_1c_2] \tag{2.9}$$

and if $c_1 < 0$ but $a_2 \geq 0$, it can be seen that

$$[a_1, c_1] \cdot [a_2, c_2] = [a_1c_2, a_2c_1]. \tag{2.10}$$

Also, assuming $c_1 < 0$ and $c_2 < 0$ we get

$$[a_1, c_1] \cdot [a_2, c_2] = [c_1c_2, a_1a_2]\ , \tag{2.11}$$

but $a_1 \geq 0$, $c_2 < 0$ produces

$$[a_1, c_1] \cdot [a_2, c_2] = [a_2c_1, c_2a_1]\ . \tag{2.12}$$

In division, it is assumed that zero does not belong to $[a_2, c_2]$ [1].

3. Pareto Distribution under Uncertainty

Fuzzy probability can be calculated in $[c, d]$ interval by

$$\widetilde{P}_\alpha[c, d] = \left\{ \int_c^d k\frac{m^k}{x^{k+1}}dx \,\middle|\, k \in \widetilde{k}[\alpha] \quad \forall\alpha \right\} \tag{3.1}$$

For $0 \leq \alpha \leq 1$ then $\widetilde{P}_\alpha[c, d] = [p_1(\alpha), p_2(\alpha)]$ where

$$p_1(\alpha) = \min\left\{ \int_c^d k\frac{m^k}{x^{k+1}}dx \,\middle|\, k \in \widetilde{k}[\alpha] \right\} \tag{3.2}$$

and

$$p_2(\alpha) = \max\left\{ \int_c^d k\frac{m^k}{x^{k+1}}dx \,\middle|\, k \in \widetilde{k}[\alpha] \right\} \tag{3.3}$$

Let

$$h(k) = \int_c^d k\frac{m^k}{x^{k+1}}dx, k \in \tilde{k}_\alpha = [k_1(\alpha), k_2(\alpha)] \tag{3.4}$$

then

$$p_1(\alpha) = h(k_2(\alpha)) \ \text{ and } p_2(\alpha) = h(k_1(\alpha)) \tag{3.5}$$

where $h(.)$ is increase function with respect to parameter k.

EXAMPLE 1. Let $m = 1$, $k = 5$. Since k is uncertain, it is assumed that $\widetilde{k} = (4.9/5.0/5.1)$ for k. So fuzzy probability, $\widetilde{P}_\alpha[2,3]$, can be calculated by using equations given by (3.2-3.5) with $\widetilde{k}_\alpha = [k_1(\alpha), k_2(\alpha)] = [4.9+0.1\alpha, 5.1-0.1\alpha]$.

Fuzzy probability, $\widetilde{P}_\alpha$, and other fuzzy numerical characteristics are calculated according to the different values of α for k in the interval $\widetilde{k}_\alpha$ by a MATLAB code and given in Table 1.

α	$\widetilde{P}_\alpha[2,3]$	$\widetilde{\mu}_\alpha$	$\widetilde{\sigma}^2_\alpha$	$\widetilde{\gamma}_{1,\alpha}$	$\widetilde{\gamma}_{2,\alpha}$
0.0	[0.0255, 0.0289]	[1.1951, 1.3077]	[0.0940, 0.1156]	[4.2372, 5.1073]	[55.5432, 91.1285]
0.1	[0.0256, 0.0287]	[1.2005, 1.3018]	[0.0950, 0.1144]	[4.2372, 5.0589]	[56.8869, 88.8113]
0.2	[0.0258, 0.0285]	[1.2059, 1.2959]	[0.0960, 0.1132]	[4.2372, 5.0110]	[58.2674, 86.5636]
0.3	[0.0260, 0.0284]	[1.2113, 1.2901]	[0.0969, 0.1120]	[4.2372, 4.9637]	[59.6862, 84.3829]
0.4	[0.0261, 0.0282]	[1.2167, 1.2843]	[0.0979, 0.1109]	[4.2372, 4.9170]	[61.1443, 82.2667]
0.5	[0.0263, 0.0280]	[1.2222, 1.2785]	[0.0989, 0.1097]	[4.2372, 4.8708]	[62.6432, 80.2127]
0.6	[0.0265, 0.0278]	[1.2277, 1.2727]	[0.1000, 0.1086]	[4.2372, 4.8251]	[64.1842, 78.2186]
0.7	[0.0266, 0.0277]	[1.2333, 1.2670]	[0.1010, 0.1075]	[4.2372, 4.7800]	[65.7689, 76.2823]
0.8	[0.0268, 0.0275]	[1.2388, 1.2613]	[0.1020, 0.1063]	[4.2372, 4.7353]	[67.3987, 74.4018]
0.9	[0.0270, 0.0273]	[1.2444, 1.2556]	[0.1031, 0.1052]	[4.2372, 4.6912]	[69.0751, 72.5749]
1.0	[0.0271, 0.0271]	[1.2500, 1.2500]	[0.1042, 0.1042]	[4.6476, 4.6476]	[70.8000, 70.8000]

Table 1. $\widetilde{P}_\alpha$, $\widetilde{\mu}_\alpha$, $\widetilde{\sigma}^2_\alpha$, $\widetilde{\gamma}_{1,\alpha}$(skewness) and $\widetilde{\gamma}_{2,\alpha}$(kurtosis) values for α-cut

The fuzzy probability shown in Figure 1 for Example 1.

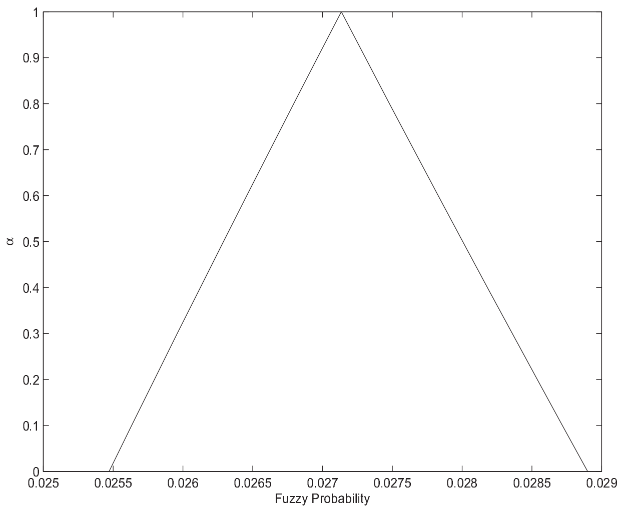

Figure 1. Fuzzy Probabilities for Pareto Distribution ($k = 5$, $m = 1$)

We now determinate the fuzzy expected value and fuzzy variance of fuzzy Pareto probability density function.

$$\tilde{\mu}_\alpha = \left\{ \int_m^\infty k\left(\frac{m}{x}\right)^k dx \,\middle|\, k \in \tilde{k}[\alpha] \right\} \tag{3.6}$$

Since, each integral in the equation (3.6) equals $k\frac{m}{k-1}$, $k > 1$. Hence, $\widetilde{\mu}_\alpha = \widetilde{k}\frac{m}{\widetilde{k}-1}$. Similarly, $\tilde{\sigma}^2_\alpha = \tilde{k}\frac{m^2}{(\tilde{k}-2)(\tilde{k}-1)^2}, \tilde{k} > 2$.

Now let us calculate the fuzzy expected value and fuzzy variance of fuzzy Pareto probability density function for values which are used in Example 1. And they are given in Table 1 for some α-cuts. The graph of the fuzzy expected value and fuzzy variance are shown in Figure 2 and Figure 3 respectively.

Now let us calculate the fuzzy skewness and fuzzy kurtosis of fuzzy Pareto probability density function for values which are used in Example 1. And they are given in Table 1 for some α-cuts, respectively.

$$\widetilde{\gamma}_{1,\alpha} = \frac{\widetilde{\mu}_{3,\alpha}}{\widetilde{\sigma}_{\alpha}^{3}} \tag{3.7}$$

$$\widetilde{\gamma}_{2,\alpha} = \frac{\widetilde{\mu}_{4,\alpha}}{\widetilde{\sigma}_{\alpha}^{4}} - 3 \tag{3.8}$$

where

$$\widetilde{\mu}_{i,\alpha},\ i = 3,4 \tag{3.9}$$

this called as fuzzy central moment.

The graph of the the fuzzy skewness and fuzzy kurtosis are shown in Figure 4 and Figure 5 respectively.

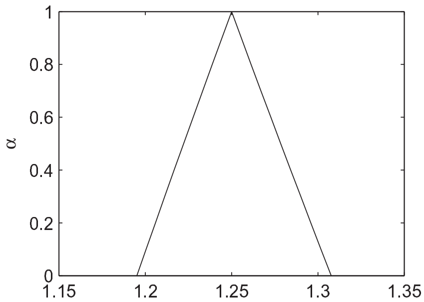

Figure 2. Fuzzy expected value

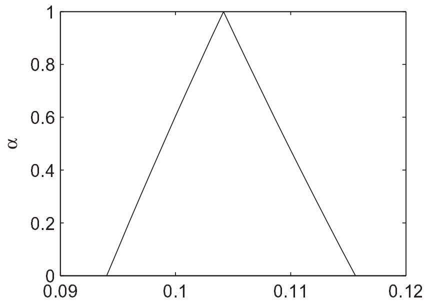

Figure 3. Fuzzy variance

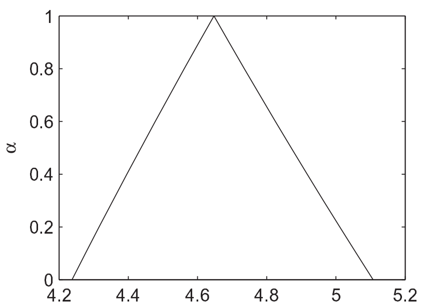

Figure 4. Fuzzy skewness

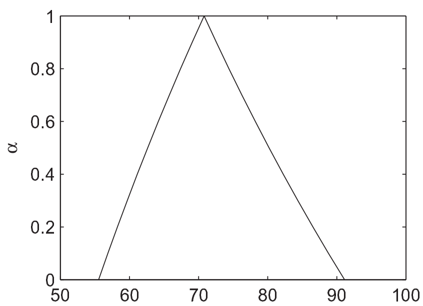

Figure 5. Fuzzy kurtosis

4. Application to the SOA Group Medical Insurance Database

We handled the SOA Group Medical insurance large claims database, which is available at http://www.soa.org, the period 1991. For calculating, a sample drawn by systematic sampling. Systematic Sampling is a method of selecting sample members from a larger population according to a random starting point and a fixed periodic interval. Typically, every $n-th$ member is selected from the total population for inclusion in the sample population. Since we select every $100-th$ observation. This sample has 758 observations. According to the Kolmogrov-Smirnov goodness-of-fit test, this observation has pareto distribution with ($\widehat{k} = 1.6577$, $m = 25007$) parameters. The calculated values and the table values of the Kolmogorov-Smirnov test are shown in Table 2.

Threshold	n	$\widehat{k}$	K-S (calculated value)	K-S (table)
25007	758	1.6577	0.0374	< 0.0592

Table 2: Goodness-of-fit to Pareto Distribution

The observed and theoretical frequencies plots are shown in Figure 6. From the figure 6, we can also see that the variables are fit to pareto distribution.

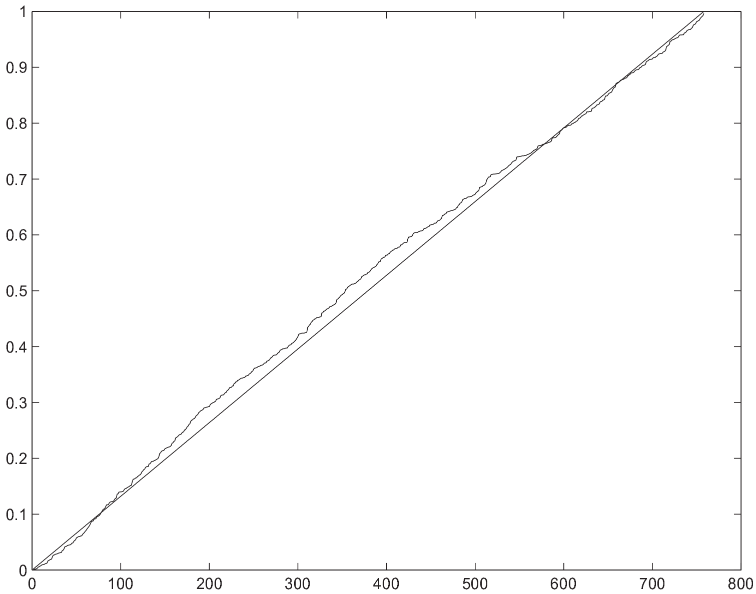

Figure 6. Pareto Fit Whole Data Set

Now we calculate the expectation value for the population by using estimated $\widehat{k}$, under uncertainty. Let $m = 25007$, $\widehat{k} = 1.6577$. Since $\widehat{k}$ is uncertain we use $\widetilde{\widehat{k}} = \left[\widehat{k}_1(\alpha), \widehat{k}_2(\alpha)\right] = [1.5577 + 0.1\alpha\ ,\ 1.7577 - 0.1\alpha]$ for $\widehat{k}$. We will calculate the fuzzy expected value $\widetilde{\widehat{\mu}}$ from Equation (3.6):

α	$\widetilde{\widehat{\mu}}$
0.0	[51410 , 78814]
0.1	[52432 , 76986]
0.2	[53482 , 75220]
0.3	[54560 , 73515]
0.4	[55669 , 71866]
0.5	[56809 , 70272]
0.6	[57982 , 68730]
0.7	[59188 , 67236]
0.8	[60431 , 65790]
0.9	[61710 , 64388]
1.0	[63029 , 63029]

Table 3: $\widetilde{\widehat{\mu}}$ values for α-cut

5. Conclusions

It's known that mean value of population is 67632. However, the expected value $\widehat{\mu}$ is calculated as 63029 from sample by using $\widehat{k}$. On the other hand, it can be reached to $\widetilde{\widehat{\mu}} = [51410, 78815]$ interval by using $\widetilde{\widehat{k}}$. This is shown that this interval includes the mean value. So, it can be noted that $\widetilde{\widehat{\mu}}$ gives more useful estimation than $\widehat{\mu}$.

References

[1] Buckley, J.J., 2006, Fuzzy Probability and Statistics, Springer-Verlag, Berlin Heidelberg.

[2] Buckley, J.J., 1987, The Fuzzy Mathematic of Finance, Fuzzy Sets and Syst., 21, 257-273.

[3] Buckley, J.J. and Eslami, E., 2004, Uncertain Probabilities II: The Continuous Case, Soft Computing, 8, 193-199.

[4] Carretero, R.C., Viejo, A.S., 2000, A Bonus-Malus System in The Fuzzy Set Theory [insurance pricing decisions]. IEEE Int.-Conf. Fuzzy Syst. 2, 1036-1036.

[5] Cramer, H., 1963, Mathematical Methods of Statistics, Princeton University Press.

[6] Dubois, D. and Prade, H., 1980, Fuzzy Sets and Systems: Theory and Application, Academic, New York.

[7] Ebanks, B., Kanvowski, W., Ostaszewski, K.M., 1992, Application of Measures of Fuzziness to Risk Classification in Insurance, In: Computing and Information. IEEE Computer Society Press, Los Alamitos, CA, 290-291.

[8] Feng, Y., Hu, L., Shu, H., 2001, The Variance and Covariance of Fuzzy Random Variables and Their Applications, 120, 487-497.

[9] Garcia, J.N., Kutalik, Z., Cho, K.H. and Wolkenhauer, O., 2003, Level Sets and Minimum Volume Sets of Probability Density Functions, International Journal of Approximate-Reasoning 34, 25-47.

[10] Johnson, N.L., Kotz, S. and Balakrishnan, N., 1995. Continuous Univariate Distributions (2nd Edition, Vol. 1 ed.),, Wiley, New York.

[11] Kleiber, C., Kotz, S., 2003, Statistical Size Distributions in Economics and Actuarial Sciences, John Wiley & Sons, Hoboken NJ.

[12] Körner, R., 1997, On the Variance of Fuzzy Random Variables, Fuzzy Set and Systems, 92, 83-93.

[13] Lemaire, J. 1990, Fuzzy Insurance, ASTIN Bull. 20(1), 33-55.

[14] Meyer, P.L.,1970, Introductory Probability and Statistical Applications, Second Edition, Addison-Wesley Publishing Company, USA.

[15] Muzzioli, S., Reyna, H., 2008, American Option Pricing With Imprecise Risc-Neutral Probabilities, International Journal of Approximate-Reasoning 49, 140-147.

[16] Shapiro, A.F., 2004, Fuzzy Logic in Insurance, Insurance: Mathematics and Economics, 35, 399-424.

[17] Wang, L.X., 1997, A Course in Fuzzy Systems and Control, Prentice Hall PRT, USA.

[18] Young, V.R., 1996, Insurance Rate Changing: a Fuzzy Logic Approach. J. Risk Insurance, 63, 461-483.

[19] Zadeh, L.A.,1965, Fuzzy Set. Information Control, 8, 338-353.

[20] Zimmermann, H., 1996, Fuzzy Set Theory and Its Applications, Third ed. Kluwer Academic Publishers, Boston, USA.

[21] Zisheng O., Chi, X., 2006, Generalized Pareto Distribution Fit to Medical Insurance Claim Data, Appl. Math. J. Chinese Univ. Ser. B. 21(1), 21-29.

1 306020 660729 >