Topographic space for distributions of the Amoroso family

V G Polosin
Medical Cybernetics and Computer Science Department, Penza State University, 40, Krasnay Str., Penza 440026, Russia
E-mail: polosin-vitallij@yandex.ru

Abstract. The paper contains material about the possibility of systematization and analysis of shifted non-symmetric models of the Amoroso family in the topographic entropy-parametric space of distribution shape. To form the space of distribution forms, it is proposed to use the entropy coefficient of shifted asymmetric models as an independent information measure of the shape of the family distributions. The material is illustrated with an example of displaying curves in the space of information and statistical coefficients. Mapping the positions of distributions in the shape space of the entropy coefficient combined with the statistical coefficients of skewness and kurtosis improves the distinguish ability of the distributions.

1. Introduction

Statistical distributions are widely used in applied problems of modeling physical and information processes. Statistics are applied throughout the life cycle of a process control scheme for evaluated economic benefits, developing output properties, processing data, defining dynamic models, monitoring performance and troubleshooting [1].

Lately the Amoroso distribution has been used for data analysis, which should be considered as the most general form of many distributions common in practice, such as gamma, Weibula-Gnedenko, CHI-square, Poisson, Nakagami, Rayleigh, and many others. The Amoroso distribution is a four-parametric continuous one-dimensional probability density with the possibility of distributing a random variable on an infinite number semiaxis [2, 3]. The Amaroso distribution is given by

$$Amoroso(x | \alpha, \beta, \sigma, \xi) = \frac{1}{\Gamma(\alpha)} \frac{\beta}{\sigma} \left( \frac{x-a}{\sigma} \right)^{\alpha-1} \exp \left( - \left( \frac{x-a}{\sigma} \right)^{\beta} \right).$$  \hspace{1cm} (1)

Where $\Gamma(\alpha)$ is a gamma function, $\alpha, \beta, \sigma, \xi$ are the shifted and scaled parameters, respectively; $\alpha, \beta$ are the shapes parameters, at that $\alpha>0$.

Although that the Amoroso distribution was obtained to generate interpolation curves of income series in economic problems [2], this distribution includes, as particular or as limiting forms, many known probability distributions, that is makes possible to use the Amoroso distributions to systematize known particular models [4]. For example, the well-known generalized gamma distribution is a special case of the Amoroso distribution if the bias $\alpha$ parameter is set to zero [1]. When solving problems of the theory of reliability at various stages of the product life cycle, a set of known probabilistic and stochastic models is used. These models were obtained on the basis of the standard Weibull distribution type VII [5]. One-dimensional and two-dimensional models are also included in the Amoroso distribution family.
The Weibul distribution of type I and II are called the Gumbel distribution and the Frechette distribution, respectively [6, 7].

It was discussed in [4] that four special cases are possible for the Amoroso distribution. The first case is a mixed one. This case includes a subfamily of the three-parameter generalized gamma distribution, known as the Stacy distribution.

The second case is positive integers for the $\beta$ shape parameter. This case includes such subfamilies of distributions as the Gamma, Pearson type III, Nakagami, Rayleigh and Maxwell, Pearson's chi distribution, Pearson's chi squared distribution and other distributions.

The third case is negative integers of the form parameter $\beta$. This case includes inverse distributions of shapes such as Pearson type V subfamilies, inverse gamma, scaled inverse chi-square, scaled inverse chi; inverse Rayleigh distribution; Levy distribution and others.

The fourth case is an extreme-order statistic that includes the subfamily of the generalized Weibull distribution, the generalized subfamily of Fresthet, the extreme distributions of maximum and minimum values and others.

Since special cases of the family of Amoroso distributions include a lot of known probability distributions or these are the limiting form of various more general distributions, recently there has been an increase in interest in using the family of Amoroso distributions for constructing systematization schemes. The purpose of this paper is to construct an entropy-parametric space for classifying distributions and performing statistical analysis.

2. Signs of the distributions shape

Statistical distribution models are used for high data variability. Since distribution models characterize the causes of origin, then for an adequate description of the results it is necessary to choose a distribution with an adequate shape. When looking for a suitable model, the concept of the shape of a probability distribution is arises that we can used to approximate a sample data. In statistics, quantitative signs of the shape, such as skewness and kurtosis, are traditionally used to digitally assess the shape of distribution.

It is shown in [8] that the use of the coordinate space of statistical coefficients of asymmetry and kurtosis does not allow choosing the parameters of the shape of the generalized distribution gamma due to the close location of the curves corresponding to different subfamilies distribution. Since probabilistic quantitative signs tend to cluster to a greater extent than is acceptable for real detection of variability of data belonging to different distributions [9], the space of probabilistic coefficients of asymmetry and kurtosis can be used to establish one of the shape parameters with apriori known distribution subfamily.

For more information about the shape of the distribution, you can get using the information measure of the distribution. Claude Shannon used information entropy as an independent measure of the information contained in the distribution of the received data [10, 11]. The modern literature on statistics contains information on the differential information entropy of most known distributions. The formula for calculating the entropy of the Amoroso family is given as [4]

$$H(X, \underline{\theta}, \alpha, \beta) = \ln \left( \frac{\theta}{\beta} \frac{\Gamma(\alpha)}{\Gamma(\alpha + 1)} \right) + \alpha \ln \left( \frac{1}{\beta} - \alpha \right) + \psi(\alpha).$$

(2)

From expression (2) it follows that the entropy of the Amoroso distribution is determined by the shape parameters and, by the distribution scale parameter and does not depend on the displacement parameter $\alpha$, this makes it possible to construct an independent feature of the distribution shape.

As an independent indicator of the shape of shifted distributions, the author uses the entropy coefficient of shifted non-symmetric distributions $k_{Hs}$, that is equal to the ratio of the intervals of entropy and statistical uncertainty. The formula for calculating the entropy coefficient of displaced non-symmetric distributions is given as

$$k_{Hs} = \frac{\Lambda_{H}}{2\sigma}.$$  

(3)
Statistical coefficients of asymmetry and kurtosis together with the entropy coefficient are three independent measures of the three-dimensional entropy parametric space of features of the shape, that we may use to analyze of the Amoroso distributions models. The expression for calculation of the informational sign of the shape of distributions from the Amarezo family is given as

\[ k_{Hs} = \frac{\Gamma(\alpha)}{\beta} \frac{\exp\left\{\alpha + \left(\frac{1}{\beta} - \alpha\right)\psi(\alpha)\right\}}{\sqrt{\frac{\Gamma(\alpha + 2 \cdot \beta^{-1})}{\Gamma(\alpha)} - \left(\frac{\Gamma(\alpha + \beta^{-1})}{\Gamma(\alpha)}\right)^2}} \]  

(4)

A distinctive feature of the entropy coefficient of displaced asymmetric distributions is that in the structure of the information measure of the distribution form, uncertainty intervals is used that characterize different distribution properties. The interval of statistical uncertainty of the standard deviation of a random variable is given symmetrically with respect to its mathematical expectation and characterizes the uncertainty of the values provided that an estimate of the distribution center is obtained. The model of the entropy uncertainty interval is set relative to the beginning of the semi-axis of the possible positions of the random variable \( X \). This model characterizes the spread uncertainty relative to the boundary of the allowed values of the random variable moreover uncertainty boundary is determined by the position point with coordinates equal to \( a \) shifted parameter.

3. Entropy-parametric coordinate space of the distribution’s shapes for the Amoroso family

It is possible analytical models of non-symmetric distributions are systematized that are part of the Amoroso family, if in entropy-parametric space the distribution of shape features is displayed. The joint display of curves in the shape feature space is aimed at obtaining additional information about the shape of the data approximation models. As the axes of of the entropy-parametric space of the distribution’s shapes, both asymmetry and peakedness statistical signs were used, that they are the coefficients of skewness and kurtosis of the distribution. As the third axis, an information feature is applied, that is specified using the entropy coefficient of shifted asymmetric distributions.
Figure 1. Position diagram of distribution models of the Amoroso family

Diagrams of models of distributions of the Amoroso family in the entropy-parametric space of the shape of skewness and the coefficient of entropy of shifted asymmetric distributions are given in figure 1, where the following designations were used. The curves graphs for the possibles shapes of the Weibull-Gnedenko and gamma distributions subfamilies are numbered 1 and 2. For the Weibull-Gnedenko subfamily, the \( \alpha \) shape parameter is 1. For the gamma distribution subfamily, the \( \beta \) shape parameter is 1. The graphs of 1 and 2 curves cross at point 7 that is an exponential distribution with shape parameters \( \alpha \) and \( \beta \), if both parameters is equal to 1. It is the fact that the Weibull-Gnedenko and gamma distributions subfamilies were obtained by distorting the exponential distribution, the shape of this distribution is contained in the both subfamilies.

Number 3 is the curves graphs for the possibles shapes of the subfamily of inverse gamma distributions. For this subfamily the \( \beta \) shape parameter is -1.

Number 4 is the curves graphs for the possibles shapes of the subfamily of Frechet distributions, if the \( \alpha \) shape parameter is 1 and the \( \beta \) shape parameter is any number less than zero. Formally, the Frechet distribution corresponds to the inverse Weibull distribution. Curves Frechet distribution and inverse gamma distribution are crossed an point 12 that corresponds to the distribution of the maximum value.

Number 5 is the curves graphs for the possibles shapes of the Nakagami subfamily, if the \( \alpha \) shape parameters is 2. Points 9, 10, and 11 on curve 5 for the shape of the Nakagami distributions correspond to the Maxwell, Rayleigh and semi-normal distributions, if the shape parameters of this distributions are 1.5, 1, and 0.5, respectively. It should be noted that the Rayleigh distribution shape is contained in the subfamilies of the Nakagami and Weibull distribution. For this reason the graphs of 5 and 2 curves cross at point 10.

Number 6 is the curve graph of the possibles shapes of the logarithmic normal distribution that is the limiting form of the gamma family under the limiting transition \( \lim (\beta \to 0) \). The formula for the entropy coefficient of a three-parameter lognormal distribution to plot the graph of 6 curve was obtained.

The Pearson subfamily for different numbers of degrees of freedom is shown in figure 1, a, b as a set of points located on curves. Points 8 correspond to the Pearson’s chi-distribution if the \( \nu \) scale parameter is 1.1414 and the \( \alpha \) and \( \beta \) shapes parameters are 0.5\( \cdot n \) and 2, respectively, where \( n \) is a positive integer equal to the number of degrees of freedom. It should be noted that for Pearson’s chi distributions, the signs of skewness and kurtosis tends to normal distribution properties if the number of degrees of freedom \( n \) is more than 1. For this reason, the Pearson’s chi distribution tends to cluster around an entropy coefficient of 2. Asymmetry and kurtosis coefficients also slightly differ from zero.
Points 13 illustrate the position of the Pearson’s chi-square distribution if the scale parameter is 2 and the $\alpha$ and $\beta$ shape parameters are $0.5n$ and 1, respectively. These points are located on the curve of 2. If the number of degrees of freedom of $n$ is 2, then the position of the chi-squared distribution coincides with the position of the exponential distribution.

4. Discussion and conclusions

An advantage of the considered method of classification of distributions of the Amoroso family is the analyzing possibility of shifted distributions for finite values of the shifted parameter. The ratio of uncertainty parameters made it possible to obtain an entropy coefficient of shifted non-symmetric distributions independent of the scaled and shifted parameter, as it represents an informational sign of the shape of skewness shifted distributions.

It is possible an entropy-parametric space for systematizing the distributions of the Amoroso family is possible to construct due to the known statistical coefficients of the skewness and kurtosis shape with the entropy coefficient shifted skewness distributions was united.

Diagrams of models of distribution of the Amoroso family in the space of entropy-parametric signs of asymmetry, kurtosis and entropy coefficient of displaced asymmetric distributions made possible to strengthen the separation of subfamilies distributions.

It is follows from the diagram of the models in figure 1a, that in the space the skewness $Sk$ and the entropy coefficient $kHs$, the distributions of the Amoroso family are well distinguishable when the coefficient of skewness is greater than 1.

From the diagrams in figure 1b of the positions of the models in the space of kurtosis $Ex$ and the entropy coefficient $kHs$, it is following that the distinguish-ability of distributions is achieved with excess, the values of which are greater than 2 ... 5. Thus, this method of graphical analysis and systematization of the distributions of the Amoroso family we can used for large values of the coefficients of skewness and kurtosis. With asymmetry values less than 0.5 and kurtosis less than 1, the positions of various distributions shapes tend to cluster to a greater extent than is acceptable for real detection of data variability. For this reason, with small values of skewness and kurtosis, other methods of analyzing the distribution shape we should use, for example, that is considered in the article [8].

Insufficient distinguish-ability of the distributions of the Amoroso family in the space of shapes features of asymmetry, kurtosis and entropy coefficient of displaced asymmetric distributions is explained by the fact that the information feature of the shape is built on the basis of uncertainty intervals that is characterizing various properties of the distributions spread.

On the one hand, the entropy of the distribution is determined under the condition that the boundary of the values of the random variable on the numerical semiaxis with the origin of the given shift parameter is known. On the other hand, the standard deviation is defined relative to the mean.

Due to the fact that at low values of asymmetry and kurtosis, the forms of distributions tend to a normal distribution, the differences in the features of the forms are not sufficient for the analysis. According to [12], the error in estimating shape parameters of asymmetry, kurtosis, and entropy coefficient is more than 5% with a sample size of 40.

Thus, the entropy-parametric space of skewness, kurtosis and entropy coefficient of shifted non-symmetric distributions is an effective tool for systematizing and analyzing distributions of the Amoroso family at large values of skewness and kurtosis.

References

[1] King M 2017 Statistics: A Practical Approach for Process Control: A Practical Approach (Wiley: John Wiley & Sons Ltd)

[2] Amoroso L 1925 Richerche intorno alla curve die redditi Ann. Mat. Pura Appl 21 123-159
[3] Johnson N L, Kotz S and Balakrishnan N 1994 *Continuous univariate distributions* (New York, Wiley: A Wiley-Interscience Publication)

[4] Crooks G E 2019 *Field Guide to Continuous Probability Distributions* (Berkeley: Berkeley Institute for Theoretical Sciences BITS)

[5] Murthy P D N, Xie M and Jiang R 2004 *Weibull models* (Published by John Wiley & Sons, New Jersey)

[6] Horst R 2009 *The Weibull distribution: a handbook* (Boca Raton: Published by Taylor & Francis Group, LLC)

[7] Hamdy M. Salem, Mahmoud Ali Selim 2014 The Generalized Weibull-Exponential Distribution: Properties and Applications *International Journal of Statistics and Applications* 4(2) 102-112

[8] Polosin V G 2020 Mapping distributions in the entropy-parametric space *Journal of Physics: Conf. Ser.* 1515 032044

[9] Rose C and Smith M D 2002 *Mathematical Statistics with Mathematica* (New York: Springer-Verlag)

[10] Shannon C E and Weaver W 1963 *The Mathematical Theory of Communication* (Urbana: The University of Illinois Press)

[11] Shannon C E 1963 *Works on information theory and cybernetics* (Moscow: Foreign Literature Publishing House)

[12] Novitskiy PV and Zograf IA 1985 *Estimation of errors in measurement results* (Leningrad: Energoatomizdat)