Mapping distributions in the entropy-parametric space

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

Abstract. The paper discusses the possibility of approximate identification of asymmetric models in the space of information and probability signs of the sample data distributions. The paper contains a description of the features of the method for displaying asymmetric models in the entropy-parametric space of distribution signs. In particular a brief analysis of the shortcomings of the mapping of distributions in the space of parametric signs of asymmetry and kurtosis is given. For the construction a space of distribution signs, the author is proposed an entropy coefficient for asymmetric distributions that it is used as an independent interval informational estimate for observation sample data distribution. The material is illustrated by the example of a widespread family of generalized gamma distribution.

1. Introduction
In modern information-measuring systems, methods for classifying asymmetric distributions based on the space of statistical sign are used. An adequate model is selected by analysis of the probabilistic characteristics of the sample data that were obtained from experimental observations. In classical methods for determining the shape of distributions, dimensionless distribution signs such as asymmetry and kurtosis are calculated. This allows to identify approximately distribution shape [1, 2, 3, 4]. It is possible, since amidst the features of i moments of the distribution should be noted properties similar to derivatives of i orders. As for instance the work of the author [5], series were given for distributions approximation, that were constructed both by using high-order moments and by using the distribution signs of asymmetry and kurtosis.

A significant limitation of determining the distribution shape using probability signs is as follows. Distribution probabilistic signs tend to cluster on one line to a greater extent than is permissible for real detection of variability of sample observations belonging to different types of distributions. For this reason, the space of signs of asymmetry and kurtosis is used only for an approximate indication of the family of distributions it is insufficient for choosing the shape of the model [6]. The Weibull and gamma distributions are typical examples indistinguishable of model shape in the space of the statics signs of distribution so as asymmetry and kurtosis. In modeling the stochastic structure of information flows, a generalized gamma distribution [7] is used that is includes the Weibull and gamma distribution as private representatives of the subfamilies of a closed family of distributions.

2. Statistically distribution signs
A flexible family of generalized gamma distributions was introduced by Stacy E.W. [8, 9] for data analysis in the study of reliability. Distribution found distribution in studies of temporary processes. The physical basis for using distribution in statistical mechanics is [10].

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd
The generalized gamma distribution is a continuous probability distribution with three parameters: \( \alpha \) and \( \tau \) are shape parameters, and \( \lambda \) is scale parameter. All three distribution parameters are greater than zero. If the random variable \( X \) has a generalized gamma distribution, the probability density function is given by [9, 10, 11]:

\[
f_{ggd}(x, \alpha, \tau, \lambda) = \frac{\tau}{\lambda \Gamma(\alpha)} \left(\frac{x}{\gamma}\right)^{\alpha-1} e^{-\left(\frac{x}{\gamma}\right)^{\tau}}.
\]

where \( \Gamma(\alpha) \) is a gamma function.

The generalized gamma family is flexible in that it includes several well-known models as subfamilies. If both the \((\alpha, \tau)\) shape parameters are 1, then the subfamily distribution (1) is an exponential distribution. If the \(\alpha\) shape parameter is 1, then the subfamily (1) is the Weibull distribution. If the \(\tau\) shape parameter is 1, then the subfamily (1) is the gamma distribution. The distribution (1) is the chi-squared subfamily then followings values can be written for shape parameters of the distribution (1): \(\tau\) is 1, \(\alpha\) is 0.5, where \(k=1, 2, 3, \ldots\) The chi-squared subfamily distribution is converted to the half normal if \(k\) parameter is 1.

Figure 1 shows graphics of positions of asymmetry and kurtosis for subfamilies of distributions similar shape. It is illustration the imposition of graphics of distributions subfamilies in the space of distribution signs. There are graphics at numbers 1 and 2 that correspond to the positions of many shapes of the subfamilies of the Weibull and gamma distribution. The diagram also shows the shapes positions for the Pareto distributions family and for the family of the logarithmic normal distribution. There are graphics numbers 3 and 4. These distribution families overlap almost on the gamma distribution graphic in the space of asymmetry and kurtosis signs.

![Figure 1. A space of distribution signs of asymmetry Sk and kurtosis Ex.](image-url)
be seen that it is difficult to choose the shape for the asymmetric distributions of different types due
dither close position of the graphic in the space of kurtosis and asymmetry.

3. Information measure for the uncertainty interval distribution of controlled parameters

The rapid development of information technology predetermined revolutionary changes in various
fields of modern science and technology. Information carriers are observations whose outcome cannot
be predicted in advance. Since the messages are random, the amount of information is a random
variable. C. Shannon proposed a quantitative measure to ensure the informational relationship between
the objects of the system as entropy of random variable. It is convenient to use the mathematical
expectation of the amount of information to characterize the ensemble of data received from the object
of observation. The Shannon entropy is calculated using formula [13]:

$$H(x) = - \int_{-\infty}^{\infty} f(x) \ln f(x) dx.$$  (2)

Shannon entropy (2) is one of the most important metrics in information theory as it allows to
estimate amount of information when receiving, storing, transforming and transmitting in complexity
systems. At its core entropy measures the uncertainty associated with a random variable. Since the
calculation of entropy is based on the distribution probability of the observation results, the Shannon
entropy is an independent quantitative measure of the uncertainty of the distribution of the observed
value [11, 12, 14]. In modern literature, the entropy of distributions is given as an independent feature
of statistical distributions [15]. The formula for calculating of entropy of the generalized gamma
distribution is given [10, 16]

$$H_{ggd}(x, \alpha, \tau, \lambda) = \ln \left( \tau^{\alpha} \lambda \Gamma(\alpha) \right) + \left( \tau^{-1} - \alpha \right) \psi(\alpha) + \alpha.$$  (3)

where $\psi(\alpha)$ is the Psi function of the $\alpha$ shape parameter.

Formulas (3) are the entropies of the distributions of random variables these are indistinguishable
in the space of probability signs of asymmetry and kurtosis. Since the $\alpha$ scale parameter is specified in
units of the observed magnitude, it is differing for uniform distributions. For this reason in
experimental research, it is inconvenient to use the distribution entropy to specify an independent sign
of the distribution model.

As an independent coordinate, the author of this paper uses the information sign of the interval
estimation of distributions it is based on the Shannon measure. For asymmetric distributions, author
introduces the $k_H$ coefficient of the ratio of the $\Delta_H$ entropy interval of uncertainty of the asymmetric
distribution to the $m_2$ second initial moment of the distribution. For the entropy coefficient, the
formula is obtained

$$k_H = \frac{\Delta_H}{\sqrt{m_2}}.$$  (4)

Additionally, we can also write the equality between the $m_2(\lambda)$ initial moment of the second order
of the non-symmetric distribution and the $\sigma^2(Y)$ square of the standard deviation of the symmetric
distribution. Since the entropy coefficient (4) is an independent information feature of an asymmetric
distribution as the entropy coefficients were obtained for the families of asymmetric distributions
considered above. The entropy coefficient of the family of generalized gamma distribution is given by

$$k_{H_{ggd}}(\alpha, \tau, \lambda) = \frac{\left( \Gamma(\alpha) \right)^{1/5}}{\tau \sqrt{\Gamma(\alpha + 2\tau^{-1})}} \exp \left( \left[ \frac{1}{\tau} - \alpha \right] \psi(\alpha) + \alpha \right).$$  (5)

For the entropy coefficients the obtained expressions allow to display asymmetric distributions in
the space of entropy – parametric signs of asymmetry, anty-kurtosis and entropy coefficient. Entropy
coefficients has been used before in the development of a system for stochastic monitoring of the electrophysiological characteristics of the heart [17, 18]

4. Entropy – parametric signs space for asymmetry distribution

It is advisable to use in practical if we systematize the asymmetric distribution to determine analytical models’ laws by displaying them in the space of entropy-parametric signs. Since numerical estimates of the shape signs of the distribution such as asymmetry, kurtosis, anty-kurtosis and the entropy coefficient we may determine with an error of less than 5 ... 10% even with a small sample size with the number of samples less 40 [19], then the mapping of models in the attribute space shows their proximity or remoteness both regarding the results of observations, and among themselves. In doing so, we can see the distributions that it is used to approximate data or to replace complex graphic shapes with simpler models.

Special peculiar opportunities of classification for the asymmetric distribution models in the signs space are illustration then it is given in figure 2. There are used the following notation. Charts of possible shape for the Weibull family and for the gamma family are indicated by the numbers 1 and 2, respectively. A graph number 3 illustrate possible positions for a generalized gamma distribution with a \( \tau \) shape parameter of 0.6 and various \( \alpha \) parameters. For the logarithmic normal distribution the shape positions is show at the graph of 4. The exponential distributions position is the point of 5. For the position points of the chi-squared families the numbers have been set from 6 to 10. At these points the \( k \) shape parameters are equal to the integers from 1 to 5 they match of chi-squared distributions. The numbers from 6 to 10 have been set for the position points of the chi-squared families. These numbers correspond distributions that the shape parameters of \( k \) are equal to the integers respectively from 1 to 5. Chi-square distribution families with various integer degrees of freedom are located on the graph of the gamma subfamily the kind \( \alpha \) parameter is 1.5. A point 11 is the position of the half-normal distribution.

A diagram that it is show projections of models graphics onto the plane of distributions signs of the position of different distribution families is an effective tool for analyzing the shapes of models of asymmetric distributions. By analysis distributions we select the distribution family graph that located closest to the location point of the real data in three projections signs. For illustration of additional divide opportunities systematization of distributions in the entropic-parametric space the author are used only two projections signs that at figure 2 is shown. There are depicted the projection onto the

![Figure 2. Projection of asymmetry and entropy coefficient.](image_url)
plane of the asymmetry and entropy coefficient mappings at the figure 2, a and the projection onto the plane of the anty-kurtosis and entropy coefficient mappings at the figure 2, b. At the diagram anti-kurtosis is given by ratio of unit to root of distribution kurtosis quadratic.

5. Discussion and conclusions

If you look at graphics of the diagram at figure 2 it is obvious that distributions are well distinguishable in the entropic-parametric space of asymmetry, kurtosis and entropy coefficient. Before we have already seen at the figure 1 that in space of only statistical signs such as of asymmetry and of kurtosis these distributions were indistinguishable because they overlapped each other.

Now discuss some of the characteristic features it is the graphic display of distributions in the entropy-parametric space. First, note that in point 5 the graphics of shapes for the Weibull and gamma subfamilies of distributions is intersect. It is possible because in composition of these families there is a general shape that corresponds to an exponential distribution with an entropy coefficient equal to 1.922. This is explainable since the many shape for Weibull and gamma distributions were obtained by distorting the exponential distribution. These distributions belong to the family of generalized gamma distributions.

Further, we can also see point 7 that is located at the position of the Pearson distribution chi square with three degrees of freedom of independent standard normal random variables. As the chi-square distribution is a special case of the gamma distribution with shape parameters that are equal to 1.5 and 1 then the gamma distribution graph passes through this point. The entropy coefficient of distribution chi square is 2.014 at point 7 it is calculated by the formula (5). The asymmetry and anti-kurtosis of distribution chi square with three degrees of freedom are equal to 1.633 and 0.5, respectively. For the Weibull distribution with the shape parameter of \( \tau \) that is 1.139 similar entropy coefficients is 2.014 and same close values of signs of asymmetry and anty-kurtosis that are equal to 1.642 and 0.508 respectively. The Weibull distribution at the position point of 7, it is possible to use to approximate both a distribution with three degrees of freedom and gamma distributions with \( \alpha \) shape parameters that is 1.5. In the area between points 5 and 7 of the diagram at figure 2, both the Weibull and gamma distributions may be used to approximate the same data. These distributions are often used in various fields of technology, for example, at building reliability models.

At the diagram of figure 2 of models of asymmetric distributions is show also the position of two symmetric distributions: logistic distribution at point 12 and normal distribution at point 13. The entropy coefficients of symmetric logistic and normal distributions are 2.037 and 2.066, respectively. Both estimates of the entropy coefficient and of the anti-kurtosis at the symmetric distributions is obtained by analogy with the work of P. Novitsky [24]. The entropy coefficient is calculated as half the ratio of the intervals of entropy uncertainties to standard deviations of symmetric distributions. The anty-kurtosis was defined as the ratio of the square of the standard deviation to the square root of the fourth moment. From diagram 2 it follows that the symmetric logistic and normal distributions are clearly distinguishable on the projection of signs of asymmetry and entropy coefficient.

In the projection only parametric features of the distributions the positions of the log normal subfamily are superimposed on positions of the gamma subfamily this is illustrated at figure 1 there graphs 2 and 4. From the consideration of the diagram in figure 2 it can be seen that the use of an additional feature of the entropy coefficient of asymmetric distributions provides good distinguishability of graph 4 of the family of shapes of the logarithmic normal distribution.

Another example is the Pareto distribution. From the consideration of figure 1 it follows that the shapes of the Pareto distribution with different position parameters on the projection of the parametric signs of asymmetry and kurtosis are not distinguishable. Pareto Distribution Family is illustrated by graph 3 at figure 1. In addition, there the positions of the Pareto distribution shapes are near the shapes of the generalized gamma distributions. The use of the entropy coefficient of asymmetric distributions made it possible to obtain a separation for the Pareto distribution for various values of the position parameter \( x_0 \) of the left boundary of possible values. In figure 2, the numbers 14, 15, 16, and 17 correspond to the graphs of the positions of Pareto distributions with position parameters of 0.9, 1.1,
1.3, and 1.5. From the topographic diagram, it can be seen that the Pareto graphics should be used to approximate the data for the values of the entropy coefficients of asymmetric distributions of less than 1, asymmetry of more than 2.7, and anti-kurtosis less than 0.25. In this area, it is preferable to use simpler shapes of Pareto distribution, since the use of shapes of generalized gamma distribution is less effective due to the complexity of the model. Thus, the use of the entropy coefficient of asymmetric distributions as an additional informational sign of distribution uncertainty, together with the probabilistic signs of asymmetry and anti-kurtosis, allows us to create a space of entropy-parametric features for classifying and controlling asymmetric distributions by changing their shape.

Reference
[1] Cain M K, Zhang Z and Yuan K H 2017 Univariate and Multivariate Skewness and Kurtosis for Measuring Nonnormality: Prevalence, Influence and Estimation Behav. Res. 49 1716–35
[2] Vadzinsky R N 2001 Handbook of Probabilistic Distributions (St. Petersburg: St. Petersburg Printing House "Science" RAS)
[3] Woon W K 2016 Skewness and Kurtosis Ratio Tests: With Applications to Multiperiod Tail Risk Analysis Cardiff Economics (Cardiff: Cardiff Economics Working Papers)
[4] Darie M, Burian S, Csaszar T, Lupu L, Moldovan L, Colda C and Andris A 2016 Monte Carlo Simulations of Proficiency Testing for Geometric Distributed Test Results IOP Conf. Series. Earth and Environmental Science 44(4) 042041
[5] Polosin V G 2014 Way of the Statistical Analysis of Electrocardiographic Information Izvestiya Southern Federal University Engineering Sciences vol 159(10) 120–6
[6] Rose C and Murray D S 2002 Mathematical Statistics with Mathematica (New York: Springer-Verlag Cop.)
[7] Korolev In Yu, Krylov in A, and Kuzmin in Yu Stability of Finite Mixtures of Generalized Gamma Distributions with Respect to Parameter Perturbations Informatics and its Applications vol 5(1) 31–8
[8] Stacy E W 1962 A Generalization of the Gamma Distribution Annals of Mathematical Statistics vol 33(3) 1187–92
[9] Stacy E W and Mihram G A 1965 Parameter Estimation for a Generalized Gamma Distribution Technometrics 7(3) 349–58
[10] Lienhard J H and Meyer P L 1967 A Physical Basis for the Generalized Gamma Distribution Quarterly of Applied Mathematics 25(3) 330–4
[11] Rinne H 2009 The Weibull Distribution: a Handbook (London, New York: by Taylor & Francis Group LLC Chapman and Hall CRC Press)
[12] Wenzel V S 1998 Probability Theory (Moscow: Higher school)
[13] Shannon C E Weaver W 1949 The Mathematical Theory of Communication (Urbana, IL: The University of Illinois Press)
[14] Yaglom A M and Yaglom I M 1973 Probability and Information (Moskow: Science Publishing House)
[15] Lin J 1991 Divergence Measures Based on the Shannon Entropy IEEE Transactions on Information Theory 37(1) 145–51
[16] Khodabin M and Ahmadiani 2010 A Some Properties of Generalized Gamma Distribution Mathematical Sciences 4(1) 9–28
[17] Polosin V G 2017 The System of Stochastic Monitoring of the Electrophysiological Characteristics of the Heart Journal of Ney Medical Technologies (eEdition) 3
[18] Polosin V G 2019 An Entropy–Parametric Criterion for Examining Cardiovascular System Conditions News of Higher Educational Institutions. Volga region. Technical Sciences 51(3) 31–44
[19] Novitsky P V and Zograph I A 1985 Error Estimation of Measurement Results (Leningrad: Energoatomizdat)