Shape measures of generalized beta distributions

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. This paper presents shape measures for generalized beta distributions that unite many subfamilies of distributions. For the study of complex systems, the information entropy of the whole family of the generalized beta distribution is obtained. The paper uses the interval of entropy uncertainty as an estimate of the entropy uncertainty for probable models, which are given in units of an observable random variable. The entropy uncertainty interval was used to construct the entropy coefficient of unbiased subfamilies of the generalized beta distribution. Particular entropy coefficients are given for frequently used subfamilies of beta distribution, that greatly facilitates the use of coefficients as independent information measures in determining the shape of models. The paper contains the most general formulas for probabilistic measures of the distributions shape also.

1. Introduction
Probabilistic models provide a framework for analyzing empirical data and constructing statistical procedures. With the development of complex systems, there is an increased interest in universal distribution models. A complex system consists of many interacting subsystems, as a result of which it acquires chaotic properties. Typical examples of complex systems are climatic systems, living organisms, medical systems, transport or communication systems, etc. Since such systems are characterized by chaotic behavior then probabilistic models are the most useful for analyzing data in the study of complex systems. Probabilistic models are often used for a simplified description of chaotic properties [1, 2, 3].

A huge number of models cover the classification of unbiased subfamilies of distributions that are based on the generalization of the beta distribution. A huge number of models cover the classification of unbiased subfamilies of distributions that are based on the generalization of the beta distribution as it is discussed in paper [4]. The five-parametric unbiased beta distribution can be specified by combining the generalized distributions of the first and second kind. This is given as

$$f_{gb}(y; a, c, u, v) = \frac{1}{B(u, v)} \frac{a}{\theta} y^{u-1} \left(1-(1-c)\left(\frac{y}{\theta}\right)^w\right)^{v-1} \left(1+c\left(\frac{y}{\theta}\right)^w\right)^{-(u+v)} . \quad (1)$$

Where $J$ is scaled parameter, $a, c, u$ and $v$ are the parameters of the distribution shape. For generalized beta distributions of the first and second types the $c$ parameter equal to zero or one, respectively. In these cases, the $y$ random variable is specified on the interval
It follows from equation (2) that the random variable $Y$ is specified in the interval $[0, J]$, if the value of the $c$ parameter is zero, and in the interval from zero to infinity if the value of the $c$ parameter is one.

Various subfamilies of distributions are included in generalized beta distributions. If you specify the characteristic values of the parameters, it is allowing you to select subfamilies of distributions. If you setting the $c$ parameter with equal to zero and the $a$ and $u$ shape parameters equal to one ensures the selection of such unbiased subfamilies as Beta of the first kind and Kumaraswamy, respectively. If the $\vartheta$ scale parameter is equal to 1, these unbiased subfamilies are specified on the interval of a random variable from 0 to 1. If subfamilies are included in the generalized beta of the second kind, then the $c$ parameter is 1. In this case, the scale parameter equal to 1 selects the normalized distribution subfamilies specified in the interval from 0 to infinity. Additional conditions imposed on some of the more important subfamilies of distributions are given in the table 1.

**Table 1.** Subfamily parameters for the generalized beta distribution.

| Distribution name                              | $\vartheta$ | $c$ | $a$ | $u$ | $v$ |
|-----------------------------------------------|-------------|-----|-----|-----|-----|
| **Generalized beta of the first kind**        | -           | 0   | -   | -   | -   |
| Beta of the first kind                        | 1           | 0   | 1   | -   | -   |
| Kumaraswamy                                   | 1           | 0   | -   | 1   | -   |
| **Generalized beta of the second kind**       | -           | 1   | -   | -   | -   |
| Pareto IV                                     | -           | 1   | -   | 1   | -   |
| Pareto II                                     | -           | 1   | 1   | 1   | -   |
| Pareto III                                    | -           | 1   | $\gamma^1$ | 1   | 1   |
| Beta of the second kind                       | 1           | 1   | -   | -   | -   |
| Burr type XII                                 | 1           | 1   | -   | 1   | -   |
| Burr type III                                 | 1           | 1   | -   | -   | 1   |
| Loglogistic                                   | -           | 1   | -   | 1   | -   |
| Paralogistic                                  | 1           | 1   | -   | 1   | $a$ |
| Lomax                                         | -           | 1   | 1   | 1   | -   |
| Inverse Lomax                                 | -           | 1   | 1   | -   | 1   |

The use of parameters makes it possible to unify the measures of distributions when determining and studying its shape. This is convenient since it is possible to use generalized expressions at construction formulas for subfamilies of distributions, which are also true for special cases.

**2. Probability measures of the shapes for generalized beta distributions**

The shapes variety of the generalized beta distribution is determined by the presence of six parameters in equation (1), of which two parameters are responsible for the bias and the scale of the distribution, at the same time 4 parameters determine the shape of the distribution. Obviously, a way is needed to pre-select possible implementations. For these purposes, various methods of moments are often used, that is contains in the paper [3, 5, 6, 7].

The method of moments was introduced by Pafnuty Chebyshev in 1887 at the proof of the central limit theorem [3, 8, 9]. Distribution moments allow obtaining measures of distribution shape, such as asymmetry and kurtosis. [5, 6]. For shape measures of asymmetry and kurtosis the most general formulas are given by
Expressions for calculating the initial moments of generalized beta of the first and second types can be found in paper [4]. The formula for determining the moments is

\[
m_s = \frac{\mathcal{B}(u + sa^{-1}, v - sa^{-1}(1-c))}{B(u, v)}
\]

where \(c\) is equal to 0 or 1 for a generalized beta distribution of type 1 or 2, respectively.

Diagrams of standardized moments are often used to pre-select the shape of the distribution. You can be found in [10, 11, 12, 13] moment ratio diagrams. This diagram is a useful theoretical tool for understanding the relationships between continuous univariate distributions. In [14], moment ratio diagrams are illustrated there, that includes the most complete families of beta distributions.

Diagrams based on their second, third and fourth points serve such purposes as to quantify the closeness between different univariate distributions and to create a short list of potential probabilistic models from a dataset [14]. Moment diagrams allow you to illustrate different well-known families of distributions and to establish constraining relationships between these families.

By plotting the sampled moments in the moment ratio diagram, you can select the candidate distributions for data modeling. This can be used as to organize simulation studies and as to model the behavior of complex systems.

If methods for estimating the shape of distribution is based on moments or on moment ratio diagrams than the main disadvantage is that statistical moments are often clustered so closely that they do not allow us distinguishing distributions even for one family. This is shown for example in [15, 16]. Moreover, for many families there are no analytical expressions for the standardized moments of asymmetry and kurtosis. For this reason, the diagrams of moment ratios contain only the boundaries for specifying possible distributions and it is do not allow a preliminary estimate of the parameters of the distribution shape. In these cases, additional information about possible realizations can be obtained on the basis of a study of the entropy intervals of the uncertainty of the distribution of random variables.

3. Information measures of shapes

The uncertainty of the state of a complex system is due to the randomness of the processes controlled by its output values. The state of the system is characterized by a sample array of recorded values. The change in the shape of the distribution of the array of values can be judged by the change in the information measure. Then, for a preliminary selection of the shape of the approximating beta model, it is possible to use the Shannon information entropy. For a continuous random variable with a probability density \(f(x)\), the Shannon entropy is given as

\[
H(Y) = -\int_{-\infty}^{\infty} f(x) \ln f(x) \, dx
\]

Shannon's entropy is a mathematical measure that estimates the reduction in the uncertainty of the resulting random variable. In modern literature, Shannon's entropy is used as an independent property of data distribution that is characterizing the information content of the obtained sample values.

This is for calculating the Shannon entropy of the entire family of generalized beta distribution, the formula was obtained from expression (1) that is given as
Expression (6) for the generalized entropy can be used to study the properties of both subfamilies of generalized beta distributions of the first and second kind.

The measure of Shannon’s entropy is not convenient as a characteristic of the shape of the beta distribution, since it depends on the scale of the distribution and on the choice of physical units of random variables. The operation of potentiating Shannon’s entropy allows one to represent the information measure in the form of an interval of entropic uncertainty. Then, the interval of entropy uncertainty is given as

\[ \Delta H_{GB}(Y, \theta, a, c, u, v) = \exp\left( H_{GB}(Y, \theta, a, c, u, v) \right) \]

(7)

The disadvantage of the entropy uncertainty interval is that the interval is proportional to the scaling parameter. The proportionality of the interval to the distribution scale is the basis of the information interval model. The units of measure for the interval of entropy uncertainty coincide with the units of measure of the investigated random variable. Similar properties characterize the probabilistic Euclidean intervals for the distributions of random variables in the space of elementary events. For this reason, the author uses the ratio of information and probability intervals to obtain new distribution measures that depend only on the shape parameters. The author of [15, 17] proposed the entropy coefficients for biased and unbiased non-symmetric distributions. Since the entropy coefficients depend only on the shape parameters, this can be used for preliminary analysis of distributions together with the known probability measures of shape, such as skewness and kurtosis. In modern literature, this is often used to analyze the shape of distributions.

4. Entropy coefficients for generalized beta distribution

It’s to eliminate the scaling parameter, the ratios of the entropy uncertainty interval to the statistical uncertainty intervals are used. It possible to obtain effective coefficients for the study of distribution shapes that do not depend on its scale.

It was shown in [15] that for unbiased non-symmetric distributions, the entropy coefficient of unbiased distributions should be used as an informational characteristic of the distribution shape, that it is defined as the ratio of the \( \Delta \mu(Y) \) entropy uncertainty interval to the mean square value of a random variable, which is equal to the square root of the second \( m_2(Y) \) initial moment of the same random value. This is given as

\[ K_{\mu} = \frac{\Delta \mu(Y)}{\sqrt{m_2(Y)}}. \]

(8)

If we substitute the entropy uncertainty interval of (7) and the 2-nd initial moment from (4) into expression (8), then we obtain the most general formula for the entropy coefficient for unbiased distributions of the generalized beta family. The formula is given

\[ K_{HnGB} = \frac{(B(u, v))^{1.5}}{\sqrt{B(u + 2a^{-1}, v - 2a^{-1}(1 - c))}} \cdot \frac{\exp\left[ a^{-1} u \left( \psi(u) - \psi(v + u - cu) + (v + c(u + 1) - 1)(\psi(u) - \psi(v)) \right) \right]}{\prod \left( \Gamma(a) \Gamma(c) \right)^{1/2}}. \]

(9)

The resulting expression is the most general form of the entropy coefficient for unbiased distributions of the generalized beta family. It should be noted that it is inconvenient to use the ratio of beta functions at performing calculations. In this reason, if particular problems are solving, it is more convenient to use the entropy coefficients for individual subfamilies of distributions. Since the initial statistical moments for many subfamilies of distributions do not contain beta functions, this makes it possible to reduce the computational load and increase the accuracy of the calculations. In addition, if mathematical analysis
is conducting, it is convenient to use simplified expressions. Therefore, in Table 2, the entropy coefficients of non-shifted distributions for a number of subfamilies of the generalized beta distributions.

Since the Transformed beta distribution corresponds to the unbiased family of generalized beta distributions, the table gives the entropy coefficient of the unbiased generalized beta distribution. The entropy coefficient of the unbiased Pareto Feller distribution can also be easily obtained from the general beta of the second kind by replacing the parameter a with the inverse of the parameter $\gamma$.

### Table 2. Entropy coefficients of subfamilies of the generalized beta distribution

| Distribution name                        | Entropy coefficients for non-shifted distributions |
|------------------------------------------|----------------------------------------------------|
| Generalized beta of the first kind       | $K_{\text{Hn GB I}}(a,u,v) = \left( B(u,v) \right)^{1.5} \frac{e^{\left[a^{-1}-u\right]\left[\psi(u) - \psi(v+u)\right] + (v-1)\left[\psi(u+v) - \psi(v)\right]}}{a\sqrt{B(u+2a^{-1}, v-2a^{-1})}}$ |
| Beta of the first kind                    | $K_{\text{Hn B I}}(u,v) = \left( B(u,v) \right)^{1.5} \frac{e^{\left[-1-u\right]\left[\psi(u) - \psi(v+u)\right] + (v-1)\left[\psi(u+v) - \psi(v)\right]}}{\sqrt{B(u+2, v-2)}}$ |
| Kumaraswamy                               | $K_{\text{Hn Ku}}(a,v) = \frac{\exp(1-v^{-1}) \left[1-v^{-1} + \psi(v) - \psi(1)\right]}{v^{1.5}a\sqrt{B(u+2a^{-1}, v-2a^{-1})}}$ |
| Generalized beta of the second kind       | $K_{\text{Hn GB II}}(a,u,v) = \left( B(u,v) \right)^{1.5} \frac{e^{\left[a^{-1}-u\right]\left[\psi(u) - \psi(v+u)\right] + (v-1)\left[\psi(u+v) - \psi(v)\right]}}{a\sqrt{B(u+2a^{-1}, v-2a^{-1})}}$ |
| Pareto IV                                 | $K_{\text{Hn Pareto IV}}(\gamma, v) = \frac{\gamma}{\sqrt{v}} \cdot \frac{1}{\sqrt{B(1+2\gamma, v)}} \exp\left(\gamma - 1\right)\left[\psi(1) - \psi(v)\right] + 1 + \frac{1}{v}$ |
| Pareto II                                 | $K_{\text{Hn Pareto II}}(v) = \frac{1}{\sqrt{v} \cdot B(3, v)} \exp\left(\frac{1+1}{v}\right)$ |
| Pareto III                                | $K_{\text{Hn Pareto III}}(\gamma) = e^{\gamma\sqrt{1+2\gamma}}$ |
| Burr type XII                             | $K_{\text{Hn Burr XII}}(a,v) = \frac{1}{av\sqrt{v \cdot B(1+2a^{-1}, v)}} \exp\left(-\frac{a-1}{a}\left[\psi(1) - \psi(v)\right] + \frac{v+1}{v}\right)$ |
| Burr type III                             | $K_{\text{Hn Burr III}}(a,u) = \frac{1}{au} \cdot \sqrt{\frac{au+2}{a}} \exp\left(\frac{1-au}{a} \left[\psi(u) - \psi(1)\right] + \frac{1+u}{u}\right)$ |
| Beta of the second kind                   | $K_{\text{Hn B II}}(u,v) = \left( B(u,v) \right)^{1.5} \exp\left(1-u\right)\psi(u) - (1+v)\psi(u) + (v+u)\psi(u+v)$ |
| Loglogistic                               | $K_{\text{Hn loglog}}(a) = \frac{e^{\frac{2}{1}}}{\sqrt{a}}$ |
| Paralogistic                              | $K_{\text{Hn Paralog}}(a) = \frac{1}{a^{2.5}} \sqrt{B(1+2a^{-1}, a)} \exp\left(\frac{1-a}{a} \left[\psi(1) - \psi(v)\right] + \frac{a+1}{a}\right)$ |
| Lomax                                     | $K_{\text{Hn Lomax}}(v) = \frac{1}{v\sqrt{B(3, v)}} \cdot \frac{1}{\sqrt{v}} \exp\left(-\frac{v+1}{v}\right)$ |
| Inverse Lomax                             | $K_{\text{Hn InvLomax}}(a) = \frac{1}{u} \left\{\sqrt{1+\frac{2}{u}} \cdot \exp\left(1-u\right)\left[\psi(u) - \psi(1)\right] + \frac{v+1}{v}\right\}$ |
As you can see from the consideration of table 2, the entropy coefficients depend on three parameters only for the generalized beta distributions of the first and second kind. The entropy coefficients are determined by two shape parameters for the most common subfamilies such as beta distributions of the first and second types, Kumaraswamy, Pareto IV, Burr XII. The entropy coefficients of simple forms are set by changing one parameter.

5. Conclusion
Thus, the coefficients of the entropies of unbiased distributions can be used to independently estimate the shape parameters. A prerequisite for using the formulas in table 2 is that all distributions correspond to unbiased distributions. This means that formulas can be used if the distribution position parameter is zero.

References
[1] Lisienko V G, Trofimova O G, Trofimov S P, Druzhinina N G and Dyugay P A 2011 Modeling of complex probabilistic systems (Yekaterinburg: Publishing house Ural Federal University)
[2] Ye Y, Oluyede B O and Pararai M 2012 Weighted generalized beta distribution of the second kind and related distributions Journal of Statistical and Econometric Methods 1(1) 13-31
[3] Prokhorov S A 2002 Modelling and analysis of random processes Laboratory practice (Samara: Samara Scientific Canter of the Russian Academy of Sciences)
[4] McDonald J B and Xu Y J 1995 A generalization of the beta distribution with applications Journal of Econometrics 66(1-2) 133-5
[5] Vazdinsky R N 2001 Handbook of probability distributions (St. Petersburg: Nauka Publishing House of the Russian Academy of Sciences)
[6] Hall A R 2005 Generalized method of moments Advanced texts in econometrics (Oxford: University Press)
[7] Kobzar A I 2006 Applied mathematical statistics For engineers and scientists (Moskow: Publishing House of Physical, mathematical and technical literature)
[8] Rybnikov K A 1974 History of mathematics 2nd ed (Moscow: Publishing House of Moscow State University)
[9] Chebyshev P L 1955 Selected works Series classics of science (Moscow: Publishing House of the USSR Academy of Sciences)
[10] Craig C C 1936 A new exposition and chart for the Pearson system of frequency curves Annals of mathematical statistics 7 16-28
[11] Johnson N L, Kotz S and Balakrishnan N 1994 Continuous Univariate Distributions (New York, NY: John Wiley & Sons) I
[12] Meeker W Q and Escobar L A 1998 Statistical Methods for Reliability Data (New York: Wiley-Intercedence)
[13] Tadikamalla P R 1980 A look at the Burr and related distributions Int. Stat. Rev. 48(3) 337-44
[14] Vargo E, Pasupathy R and Leemis L 2010 Moment-Ratio Diagrams for Univariate Distributions Journal of Quality Technology 42(3) 1-10
[15] Polosin V G 2020 Mapping distributions in the entropy-parametric space J. Phys.: Conf. Ser. 1515 032044
[16] Polosin V G 2020 Space of entropy-parametric characteristics for control of the non-symmetric distributions form University proceedings Volga region 3(55) 44-57
[17] Polosin V G 2020 Topographic space for distributions of the Amoroso family J. Phys.: Conf. Ser. 1679 022010