GEODESY AND CARTOGRAPHY
ISSN 2029-6991 / eISSN 2029-7009
2015 VOLUME 41(3): 131–136
http://dx.doi.org/10.3846/20296991.2015.1086127

UDK 528.94:551.4

EVALUATION OF EXTREMELY ASYMMETRIC TYPES OF DISTRIBUTIONS OF GEOFEATURES VALUES

Azimkhan KURMANKOZHAYEV[1], Guldana SYZDYKOVA[2], Erzhan SARYBAEV[3]

Department Of Mine Survey And Geodesy, Kazakh National Technical University after K. I. Satpayev, 050300 Almaty, 22 Satpayev street, Kazakhstan
E-mails: [1]kurmankozhaev_a@mail.ru; [2]guldanasd@mail.ru (corresponding author); [3]edilait@mail.ru

Received 08 April 2015; accepted 15 August 2015

Abstract. Statistical and geometric patterns were disclosed that are inherent to extremely asymmetric types of formation of stochastic frequencies of geofeatures values with the help of the exponential, probabilistic, structural, and Zipf distribution laws, which are most often used to describe them. A sophisticated pattern with exponential and hyperbolic nature of the probabilistic frequencies formation development was found. Based on these features, a leading concept of using major structure-forming characteristics of the distribution of variable parameters as desired theoretical distribution function was used for parameterization of an extremely asymmetric distribution model.

A median and modal frequency of geofeatures distribution were defined as the theoretical parameters of the desired distribution model. A structure was constructed and formulas were derived for determination of the statistical characteristics of the recommended distribution model. Approbation of the distribution model was conducted on an example of morphometric signs empirical distributions among the localities of varying complexity and qualitative indicators of a number of gold and rare metal deposits; inhesion of its approximating power and flexibility of use was revealed.

Keywords: model, extremely-asymmetric distribution, evaluation, parameter, georesourse.

Introduction

Substantiation of a rational distribution model was carried out by generalizing the properties of extremely asymmetric types of geofeatures' distribution among the objects of georesources of varying complexity.

According to the analysis of extremely asymmetric types of empirical distributions assessment practices it was found that the distribution of maximum relief and heights of evenly spaced points in the upland and foothill areas, as well as the distribution of quality indicators of gold and rare metal deposits often correspond to these types of distributions; that the theoretical distributions favorably describing these types of distributions are scarce due to the diversity of their frequencies formation. There are highly sophisticated types of extremely asymmetric types of georesources distribution signs with totally unknown theoretical distribution convergence.

The statistical geometric regularities inherent to extremely asymmetric distributions of geofeatures were studied. For this purpose, the most common indicative, structural and probabilistic and Zipf's distribution laws were used. It was found that the extreme types of asymmetric distribution are characterized by rather complex pattern including exponential and hyperbolic nature of probability frequencies formation development. Based on these features, a leading concept used with regard to the extremely asymmetric distribution types was that of providing their structural empirical parameters as theoretical parameters of the desired distribution function.

The main structural parameters of the extremely asymmetric distribution – median and modal relative frequency were involved as theoretical parameters at parameterizing of the recommended distribution model. The structure of the distribution function was constructed with the help of new theoretical parameters, formulas determining the statistical parameters were derived, testing of the distribution model was conducted on the example of a number of geofeature sempirical distributions. Density of the distribution function varies depending on its parameters and thus its curve may take the form ranging from extremely to moderately asymmetric type.

Use of sustainable median unbiasedness property and its consistency in the assessment allowed to increase the approximation power and flexibility of the recommended distribution, which were strengthened by parametric conversion of the feature in the median lobes, thereby allowing us to obtain more reliable results and reduce the risk in the case of small values and uncertainties of statistical sampling that is particularly important in the assessment of the extremely asymmetric types of empirical distributions.

1. Parametrization of the distribution model

Diversity of the types of extremely asymmetric empirical distributions of geological and morphometric signs specifies the need in a reasonable distribution model that would most reliably take into account a wide range of frequencies fluctuation typical for these types of distributions. To parameterize the structure of the desired distribution model the principle of the use of positive properties of the basic probability distributions was adopted, often describing these types of geofeature sempirical distributions. As is obvious from the analysis of evaluation practices of the extremely asymmetric geofeatures distributions, the exponential, probabilistic and structural distribution and Zipf's law can be referred to these types of distributions. In this connection, various statistical and geometrical properties inherent in these probability distributions were studied. Exponential distribution to date is most frequently used to describe the extremely asymmetric types of empirical distributions. Basic theoretical stochastic distributions describing the types of moderately asymmetric distributions for specific values of its parameters converge with exponential distribution since exponential distribution coincide with gamma distribution (when $p = 1$), the Weibull's distribution (when $m = 1$, $V = 100\%$), Pearson distribution (when $b = b_2 = 0$, $\mu_3 = 0$). Equally important is the stochastic structural c distribution, which for certain values converge with exponential, gamma and Weibull's distributions.

Basic regularities of probabilistic frequencies development inherent in the extreme types of asymmetric geofeatures frequencies formation were identified on the basis of assessment of the approximation properties of these theoretical distributions:

- development of probabilistic frequencies formation at extremely asymmetric types of distributions are often described with the help of exponential hyperbolic regularities;
- development of frequencies formation at extremely asymmetric – type distributions is sometimes described only by hyperbolic regularity.

In the above mentioned known theoretical distributions the patterns of exponential and hyperbolic development of probabilistic frequencies are used separately and thus the important properties arising from a complex combination of these regularities remain unused.

Parameterization of the developed distribution model was implemented by way of generalizing the features of the extreme types of asymmetric distributions signs among the georesources objects of varying complexity; it was taken into account that the theoretical descriptions for these types of distributions are few in number due to the diversity of formation of the frequencies, distribution of quality indicators of the deposits of gold and rare metals, maximum elevations and heights of evenly spaced points in the upland areas, as well as the distribution of heights and minimum altitudes common to these areas often correspond to the extreme asymmetric types and variety of different geometry of curves.

The patterns inherent to extremely asymmetric – empirical distributions of the signs on georesources objects were assessed involving exponential, probabilistic and structural and Zipf's distribution. According to the results of the statistical-geometric evaluation of the properties of these distributions, patterns of exponential and hyperbolic character are inherent to the development of probabilistic frequencies formation for the extremely asymmetric types of distribution. These laws are often reflected in the form of $e^{-\lambda x}$ under the exponential distribution, sometimes $e^{-m(x-x_0)}$ under – probabilistic structural distribution, and rarely in the form of $\frac{A}{x}$ – under the Zipf's distribution.

Developmental pattern of probabilistic frequencies in the form of $e^{-\lambda x}$, according to the exponential distribution, depends on only one theoretical parameter λ. This creates an approximation limit with regard to the statistical range of frequency formation distribution, and whatever the value of the parameter λ, this limitation is not eliminated. Pattern of probabilistic frequencies development in the form of $e^{-m(x-x_0)}$ is a two-parameter and are rigidly connected to the property modal value. However, the property modal values in highly asymmetric types of frequencies formation are usually close to zero, and it loses its useful properties. Developmental pattern of probabilistic frequencies in the form of $\frac{A}{x}$ according to the Zipf's distribution law has a significant disadvantage due tousageofonly one parameter A and the hyperbolic function that lacks high elasticity as the exponential function.

The theory of analytical structure of the distribution model is based on the concept of using the main structure forming parameters of the extremely asymmetric types of distributions as theoretical parameters of the desired distribution function (Kurmankozhaev 1990; Kurmankozhaev, Zhumanov 2012). The median and modal relative frequency with a multiplier were chosen as theoretical parameters of a recommended distribution model.

The median was chosen as adistribution model parameter thanks to its positive properties which were much more valuable in comparison with properties of other indicators- parameters. The median as the main structural and statistical characteristic of distribution replaces a role an arithmetic-mean deviation value and dispersion as the main well-founded assessment. It is known that the median is that value of a random variable which divides a distribution to two equal parts, and is on the probability curve of probabilities distribution at $P_m = 0.5$; respectively, the median is closely connected with a form of a distribution curve.

The median has significant advantages in comparison with the arithmetic mean (Smyslov *et al.* 1979):

– the median value remains unchanged under its assessment both by a definite value and according to its functions; this property simplifies the computational procedures and facilitates the random variables statistic s evaluation;
– the median value is more stable in the statistical samples with greater excess than that of a normal distribution law; as far as the median divides the area of the probability distribution into two equal parts, its value remains stable for different distribution curve changes caused by changes in the magnitude of excess;
– the median value is slightly affected by errors compared with the arithmetic mean value, i.e. with systematic errors and feature values close to the threshold of sensitivity; this property is important for the evaluation distribution laws of geofeatureswith their specific natural complexities;
– the most important property of the median is that its value can be calculated with sufficient reliability in the presence of a statistical population of uncertain or approximate numerical values of the quantities, and in the presence of various abnormal (hurricane, weak etc.) values.

It follows from the analysis that the median of random variables can be used instead of other features in order to obtain consistent results when the source data have significant analytical error.

The second parameter of the distribution - modal relative frequency of feature distribution of the studied object as the most informative value is of theoretical and practical importance and serves as the pivotal frame for frequencies formation of a variable *y*; it plays a crucial role in justifying the forecast estimates and probabilistic approaches. The combination of the median and modal properties of the relative frequency distribution is aimed at improving the reliability of estimates of spatial and statistical regularities of geofeatures with their distinguishing natural specific features.

2. Analitical basis of a model of the extremely asymmetric type of geofeatures distribution

Density of static distribution function for the recommended new "median" model looks as follows:

$$\begin{cases} f(x) = f_0 e^{-c\left(\frac{x}{Me}\right)}, \\ x > 0 \end{cases} \quad (1)$$

where f_0 is a relative modal distribution frequency of the feature under consideration; c is a statistical parameter; $\frac{x}{Me}$ – value of feature (x) in the median shares (Me).

The distribution function is shown in Figure 1:

$$F(x) = \frac{f_0}{c}\left[1 - e^{-c\left(\frac{x}{Me}\right)}\right], \; 0 < x < \infty. \quad (2)$$

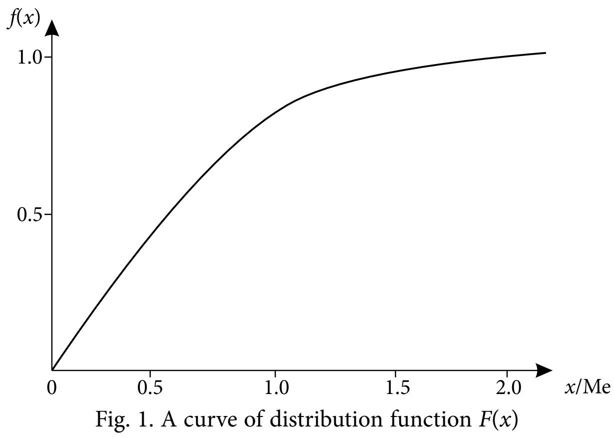

Fig. 1. A curve of distribution function $F(x)$

The density of distribution function depends on the values of a statistical parameter, thereby takes different forms (Fig. 2).

Basic statistical characteristics of the recommended distribution

Mathematical expectation:

$$Mo = f_0 \frac{1}{C^2}. \quad (3)$$

Average value:

$$x_{cp} = f_0 \frac{1}{C}. \quad (4)$$

Dispersion:

$$\begin{cases} D = f_0 \frac{1}{C^2}\left(3\overline{x} - 2\right) \\ D = Mo\left(3\overline{x} - 2\right) \end{cases}. \quad (5)$$

The standard:

$$\begin{cases} \sigma = \frac{1}{C}\sqrt{f_0\left(3\overline{x} - 2\right)}, \\ \sigma = \sqrt{\left(3\overline{x} - 2\right)Mo} \end{cases}. \quad (6)$$

The median:

$$\begin{cases} Me = \frac{2}{3}ac \\ Me = \frac{x_0 + 2\overline{x}}{3} \end{cases}. \quad (7)$$

The formula off inding the statistical exponent c was derived from the main property of the probabilities sum and is as follows:

$$c = f_0 Me, \text{ with } 0 < x < \infty;$$

$$c \approx \frac{3Me}{2a} \text{ with } 0 < x < a. \quad (8)$$

It was found that the recommended distribution model for specific values of the feature median converges with the exponential and probabilistic and structural distributions, and is associated with the distribution laws by Pareto, Erlang and geometric distribution.

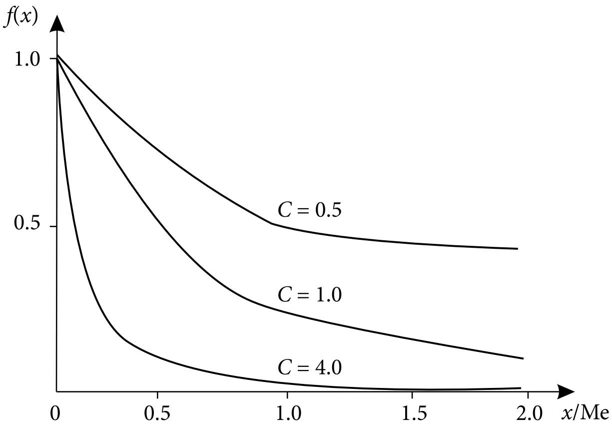

Fig. 2. The character of changes in the density of distribution function depending on parameter c

Calculation formulas for determining the values of the median for the extreme asymmetric and asymmetrical types of the main probability distributions are given in Table 1.

3. Comparative analysis of acceptability of the recommended model of extremely asymmetric type of distribution

A comparative evaluation of the acceptability of the distribution model was conducted with a wide variety of highly-asymmetric types of empirical distributions of geofeatures on actual statistical aggregate qualitative indicators of gold, rare metal, as well as complex morphometric features (heights, slopes, areas) of different localities in the regions of Kazakhstan. The

Table 1. Analytical assessment of the median calculation for the main highly asymmetric and asymmetric types of distributions

Highly asymmetric types of distributions	The median calculation formula	Asymmetric types of distribution	The median calculation formula
Exponential distribution	$Me = \frac{1}{\lambda}\ln 2$	Weibull's distribution	$Me = \frac{1}{k}\overline{x}_{cp}$ $\kappa = \varphi(\nu, p)$
Structural probability distribution	$Me = \frac{\Phi_0}{m}e^{mx_0}$	Gamma distribution	$Me = \beta \cdot \ln 2$ $\beta_x = \frac{\overline{x}}{a+1}$
Zipf's distribution	$Me = \frac{1}{2}\ln\frac{x_{max}}{x_{min}}$	Logarithmic normal distribution	$Me = 10^{\ln \overline{x}}$
Recommended median distribution	$Me = \frac{2}{3}ac$ $a = x_{max}$	Pearson's Distribution (type VII)	$Me = X_{mo} = a$

convergence analysis of these highly asymmetric types of empirical distributions with known the oretical distribution laws was carried out involving the exponential and structural probability distribution, Zipf's law and the recommended median distribution. Values of average ($\overline{x}_{cp}$), and standard deviation (σ), amplitude variability (*d*) and the Pearson criterion were calculated.

The comparative evaluation showed that in the shares of the considered amount of these sets of empirical distributions of geofeatures, more than 70% of their types are favourably and more accurately described by the recommended model, 20% – by the exponential distribution, and the rest 10% – by the structural probability distribution and Zipf's law (Table 2). This confirms the validity of using the median and modal frequency values of the geofeatures as theoretical parameters of the new distribution model, which provides the approximation power and flexibility of description of various kinds of highly-asymmetric distributions of geofeatures.

Convergence of this model with different kinds of extremely asymmetric types of empirical distributions of the geofeatures shows the feasibility of its use instead of the recognised exponential structural probability and Zipf's laws often used to evaluate the highly asymmetric distributions.

Table 2. Results of the natural-experimental convergence analysis of the main distributions acceptable for the evaluation of highly asymmetric distribution of the geofeatures

Natural-experimental varieties of comparative evaluation	Results of comparative evaluation of distributions			
	Exponential distribution $f(x)=\lambda\, e^{-\lambda x}$	Stochastic structural distribution $(x)=\Phi_0 e^{m(x-x_0)}$	Zipf's distribution law $f(x)=\frac{A}{x}$	Recommended median distribution $f(x)=f_0 e^{-c\frac{x}{Me}}$
Distribution of the heights (topographic plan M:1:5000) N = 1600; σ = 4.1; d = 4.7; $\overline{x}=4.033$	$\overline{X}=4.033$ λ = 0.10 $\chi^2=20.1>\chi^2_{\partial on}$	$x_0=1.5$ $\Phi_0=138$ m = 0.0122 N = 630 $\chi^2=11.4>\chi^2_{\partial on}=11$	$\chi^2>\chi^2_{\partial on}$	Me = 2.4 f_0 = 0.401 c = 1.25 $\chi^2=14.6>\chi^2_{\partial on}$
Distribution of the elementary areas of terrestrial sites (topographic plan M:1:1000, Zhambyl region) N = 1445; σ = 4.4; d = 5.0; $\overline{x}=18.9$	$\overline{X}=1.514$ λ = 0.661 $\chi^2=20.1>\chi^2_{\partial on}$	$x_0=6.0$ m = 0.0021 $\Phi_0=360$ N = 930 $\chi^2=11.9>\chi^2_{\partial on}=120$	A = 0.25 $\overline{X}=1.2$ $\chi^2=48.1>\chi^2_{\partial on}$	Me = 4.3 f_0 = 0.501 c = 2.88 $\chi^2=11.3>\chi^2_{\partial on}$
Distribution of underlay values (topographic plan M:1:10000, Almaty region) N = 445; σ = 7.8; d = 4.5; $\overline{x}=20.0$	$\overline{X}=12.2$ λ = 0.09 $\chi^2=19.5>\chi^2_{\partial on}$	$x_0=0.5$ $\Phi_0=360$ m = 0.155 $\chi^2=15.5>\chi^2_{\partial on}=15.65$	A = 0.527 $\overline{X}=12.2$ $\chi^2=40.7>\chi^2_{\partial on}$	Me = 0.529 f_0 = 0.584 c = 2.40 $\chi^2=1.23>\chi^2_{\partial on}$
Distribution of gold in Maikaiyng old ore deposit N = 767; σ = 2.12; $\overline{x}$ = 2.4	$\overline{X}=2.4$ λ = 0.171 $\chi^2>\chi^2_{\partial on}$	$\Phi_0=447$ m = 0.49 $\chi^2=28>\chi^2_{\partial on}$	A = 0.495 $\chi^2=31>\chi^2_{\partial on}$	Me = 1.6 f_0 = 0.59 c = 2.1 $\chi^2=8.9<\chi^2_{\partial on}$
Distribution of gold in Bakyrchik gold ore deposit N = 532; σ = 1.434; $\overline{x}$ = 1.9	$\overline{X}=1.7$ λ = 0.105 $\chi^2=9.5<\chi^2_{\partial on}$	$\Phi_0=189$ m = 0.620 $\chi^2=14.7>\chi^2_{\partial on}$	$\chi^2>\chi^2_{\partial on}$	Me = 1.43 f_0 = 0.36 $\chi^2=9.0<\chi^2_{\partial on}$
Distribution of silver in Maikaiyn gold ore deposit N = 339; σ = 1.97; $\overline{x}$ = 29.5	$\overline{X}=30.1$ λ = 0.020 $\chi^2=17>\chi^2_{\partial on}$	$\Phi_0=70$ m = 0.035 $\chi^2=16.5>\chi^2_{\partial on}$	$\chi^2>\chi^2_{\partial on}$	Me = 28.9 f_0 = 0.23 $\chi^2=9.8<\chi^2_{\partial on}$
Lead distribution in the Karagalinsky polymetallic field N = 293; σ = 0.17; $\overline{x}$ = 0.17	$\overline{X}=0.169$ λ = 0.022 $\chi^2=9.2<\chi^2_{\partial on}$	$\Phi_0=168$ m = 3.72 $\chi^2=7.6<\chi^2_{\partial on}$	$\chi^2>\chi^2_{\partial on}$	Me = 0.25 f_0 = 0.53 $\chi^2=5.8<\chi^2_{\partial on}$

Conclusions

Distributions used to describe the extremely asymmetric distributions of quality indicators of the gold and rare metal mineral deposits, and morphometric features of complex relief areas are scarce, due to the variety of their forms, the exponential and other distributions often used to describe them do not cover a wide range of highly-varying types of asymmetric distributions of the geofeatures. Such empirical extremely asymmetric distributions can be come across which require a special approach for finding their convergence with any known theoretical distribution.

The validity of the concept is confirmed of using structure-forming parameters of the empirical distribution of a variable as theoretical parameters of the distribution function for the effective evaluation and use of the geofeatures distribution with sufficient reliability. This conceptual approach seems to be scientifically sound, especially for the extreme conditions of complex extremely asymmetric types of the geofeatures distributions for which to date any proper theoretical justification has been developed.

According to the results of a comparative evaluation, it is confirmed that the new distribution model has an approximate power and flexibility, which provide its convergence with highly-sophisticated asymmetric types of empirical distributions of the geofeatures. Its use in highly variable values of the geofeature sallows to obtain more reliable results and reduce the risk allowed for small values and considerable uncertainty of statistical sampling.

References

Kurmankojaev, A. K. Stochastic-structural model of mineral resources indication distribution XXII, in *International Symposium APCOM,* 17–21 September 1990, Berlin, West-Germany, 1–6.

Kurmankojaev, A. K.; Jumanov, B. M. 2012. The structural and empirical to evaluate the distribution of quality indicators of minerals, in *34th International Geological Congress (IGC),* 5–10 August 2012, Brisbane, Australia.

Smyslov, A. A.; Rudnik, V. A.; Dinkova, N. M.; Panayotova, A. I. 1979. *The principles and method of geochemical studies under the ore fields forecasting.* Leningrad: Nedra. 129–145.

Azimkhan KURMANKOZHAYEV, Professor, Doctor of Technical Sciences, works at Kazakh National Technical University after K. I. Satpaeva. Author of more than 400 publications.
Research interests: mining, ecology, mine surveying.

Guldana SYZDYKOVA has a Master degree in Geodesy and remote sensing with skills profiles in Geodesy and Space Geodesy and Navigation. Author of more than 15 publications, works at National Company Kazakhstan Gharysh Sapary.
Research interests: creation of the ground based infrastructure of high-accuracy satellite navigation system in Kazakhstan.

Erzhan SARYBAEV has a Master degree at Kazakh National Technical University after K. I. Satpaeva. Co-author of 3 publications.
Research interests: mining, mine surveying.