UDC 519.2:004.94

POLYNOMIAL ESTIMATION OF DATA MODEL PARAMETERS WITH NEGATIVE KURTOSIS

Chepynoha V. V. – Postgraduate student of the Department of Computer Science and System Analysis, Cherkasy State Technological University, Cherkasy, Ukraine.

Chepynoha A. V. – PhD, Associate Professor, Dean of the Faculty of Information Technologies and Systems, Cherkasy State Technological University, Cherkasy, Ukraine.

Palahin V. V. – Dr. Sc., Professor, Head of the Department of Robotics and Telecommunication Systems and Cyber Security, Cherkasy State Technological University, Cherkasy, Ukraine.

ABSTRACT

Context. The paper focuses on the problem of estimating the center of distribution of the random component of experimental data for density models with a negative kurtosis.

Objective. The goal of this research is to develop methods to improve the efficiency of polynomial estimation of parameters of experimental data with a negative kurtosis coefficient.

Method. The study applies a relatively new approach to obtaining estimates for the center of the probability distribution from the results of experimental data with a stochastic component. This approach is based on polynomial estimation methods that rely on the mathematical apparatus of Kunchenko's stochastic polynomials and the description of random variables by higher-order statistics (moments or cumulants). A number of probability density distributions with a negative kurtosis coefficient are used as models of the random component.

As a measure of efficiency, the ratio of variance of the estimates for the center of the distribution found using polynomial and classical methods based on the parameter of amount of information obtained is used.

The relative accuracy of polynomial estimates in comparison with the estimates of the mean, median and quantile estimates (center of curvature) is researched using the Monte Carlo method for multiple tests.

Results. Polynomial methods for estimating the distribution center parameter for data models of probability distribution density with a negative kurtosis coefficient have been constructed.

Conclusions. The research carried out in this paper confirms the potentially high efficiency of polynomial estimates of the coordinates of the center of the experimental data, which are adequately described by model distributions with a negative kurtosis. Statistical modeling has confirmed the effectiveness of the obtained estimates in comparison with the known non-parametric estimates based on the statistics of the mean, median, and quantile, even with small sample sizes.

KEYWORDS: data sampling, estimation, stochastic polynomial, cumulants, negative kurtosis.

ABBREVIATIONS

PMM – polynomial maximization method;

NOMENCLATURE

$a_i(\theta) = E\{x_i\}$ are mathematical expectations that depend on the parameter being estimated;

$g_{(\theta)mean}$ is a coefficient of variance reduction relative to the method of moments;

$h_1 - h_3$ are weighting coefficients;

$J_{rn}(\theta)$ is an amount of obtained information;

N is a number of experiments to obtain a predetermined accuracy;

n is a sample volume;

r is a predefined order of the polynomial;

x is a sample value;

$\vec{x}$ is a sample of equally distributed random variables;

$\hat{\alpha}_i$ are sample statistical initial moments;

γ_3 and γ_6 are cumulant coefficients;

θ is an informative parameter;

$\hat{\theta}$ is an estimation of informative parameter;

ξ is a random component of sample data;

$\sigma^2{}_{(\theta)r}$ is a theoretical variance of the parameter estimated by the PMM;

$\hat{\sigma}^2{}_{(\theta)mean}$ is an empirical value of the variance of the mean estimates;

$\hat{\sigma}^2{}_{(\theta)med}$ is an empirical value of the variance of the median estimates;

$\hat{\sigma}^2{}_{(\theta)qvant}$ is an empirical value of the variance of the quantile estimates;

$\hat{\sigma}^2{}_{(\theta)PMM3}$ is an empirical value of the variance of the PMM estimates ($r = 3$);

χ_2 is a second-order cumulant;

INTRODUCTION

Statistical estimation methods are the usually core mathematical tool for many tasks where the analysis of experimental data is required, such as interpretation of scientific experiments, stochastic signal processing, product quality assurance, non-destructive testing of structures, network intrusion detection systems, theoretical fundamental metrology etc. The need to use them is predetermined by the impact of various noises and interferences, random errors or errors of measuring instruments on the experimental data. In such cases, a typical ap-

DOI 10.15588/1607-3274-2023-3-7

proach is to present the data model as an additive interaction of informative parameter and a stochastic component [1, 2].

There are several tasks in which a stochastic component of the experimental data symmetrical in nature, and under the following condition, the center of the probability distribution of this stochastic component can be determined as center of symmetry, and its coordinate will be the actual value of the informative parameter [3].

Quite often, in practice, simple averaging is used to calculate an estimate of the coordinate the center of the stochastic experimental data distribution. However, the application of arithmetic mean to estimate the parameters of random series is actually optimal only for models with a Gaussian probability distribution density. But in fact, it is impossible to represent all the variety of real probability distributions of the random component with this law alone. Therefore, some tasks require the use of other distribution laws, sometimes significantly different from the normal one [2–5].

The object of study is the process of estimating the parameters of random variable models with the moment-cumulant description.

The subject of study is models of experimental data with negative kurtosis and methods of polynomial estimation of their parameters.

The purpose of the work is to increase the efficiency of polynomial estimation of parameters of experimental data with a negative kurtosis coefficient when creating methods and tools for mathematical and computer modelling based on the application of the polynomial maximization method and moment-cumulant description of random variables to improve the quality of the estimates obtained by increasing their accuracy in the functioning of data processing systems of the corresponding class.

1 PROBLEM STATEMENT

It is necessary to estimate the value of the informative parameter θ based on statistical processing of the vector $\vec{x} = \{x_1, x_2, ... x_n\}$ of sample values $x = \theta + \xi$, which are statistically independent and identically distributed and obtained by conducting multiple experiments.

It is assumed that the mathematical model of the random component of the data ξ is adequately described by the distribution laws shown in Table 1.

The research is supposed to obtain estimates of the center of the distribution, analytical expressions describing the variances of the PMM estimates, and, using Monte Carlo statistical modelling, to check the actual efficiency of the estimates compared to the ones obtained by other methods, depending on the amount of sample data and the distribution with a negative kurtosis.

2 REVIEW OF THE LITERATURE

In a number of tasks of processing experimental data of various origins, it turns out that the random component in them has a distribution with a negative kurtosis, and the type of distribution itself is concave (bimodal). Models of such distributions are used in information and measurement systems [2, 7], in determining electromagnetic compatibility [8–10], in sociology [11], evolutionary biology [12], hydrology [13], and other fields.

Among such probability density models, according to which the random component of the experimental data is distributed, the following can be distinguished: U-quadratic distribution, arcsine distribution, V-shaped distribution and symmetrical Kumaraswamy distribution. These distributions are symmetric, have a negative kurtosis, and have another property: they are defined on a limited interval. A graphical representation of their probability density is shown in Figure 1. Table 1 shows the mathematical expressions for the probability density function of the corresponding laws:

Table 1 – Mathematical representation of probability distribution

№	Type of distribution	Distribution parameters	
a)	$p(x) = \alpha(x-\beta)^2$	$\alpha = \dfrac{12}{(b-a)^3}$	$\beta = \dfrac{b+a}{2}$
b)	$p(x) = \dfrac{1}{\pi\sqrt{w^2-(x-h)^2}}$	$w = \dfrac{b-a}{2}$	$w = \dfrac{b+a}{2}$
c)	$p(x) = \dfrac{\lvert x-a\rvert}{b^2}$	$a = 0$	$b = 1$
d)	$p(x) = a \cdot b \cdot x^{a-1}\left(1-x^a\right)^{b-1}$	$a = 0.5$	$b = 0.5$

DOI 10.15588/1607-3274-2023-3-7

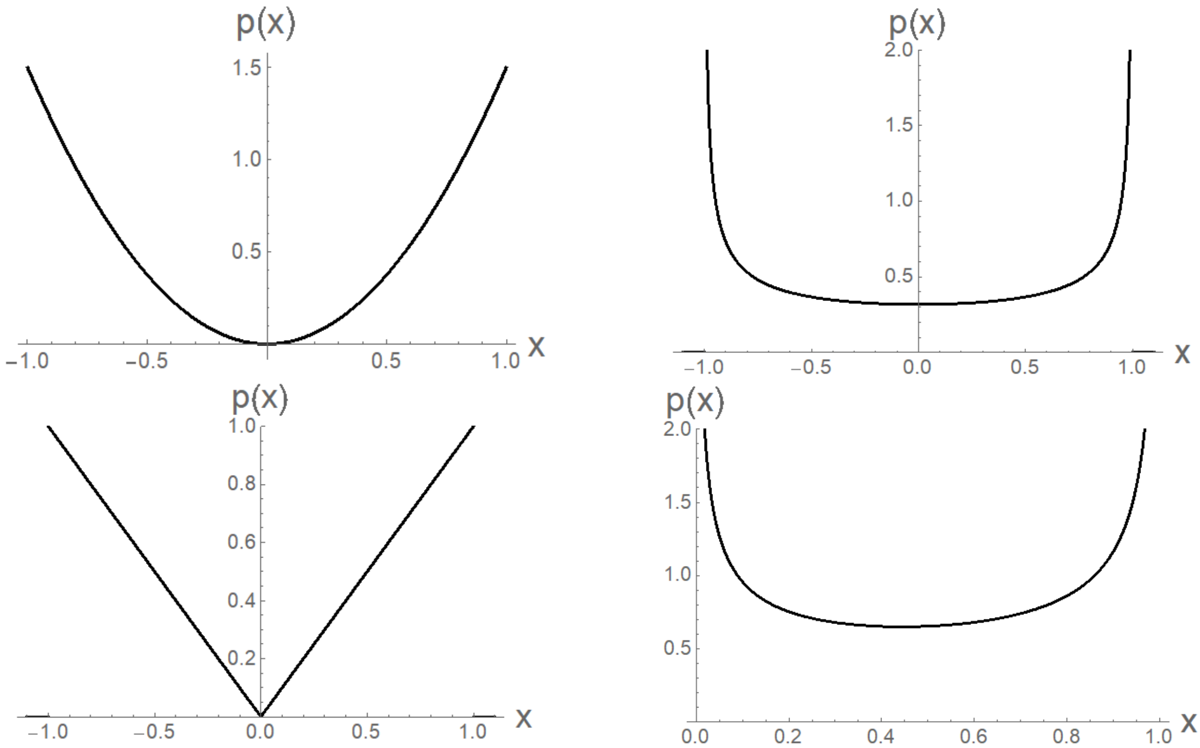

Figure 1 – Symmetric distributions with negative kurtosis: a – U-quadratic distribution, b – arcsine distribution, c – V-shaped distribution, d – symmetrical Kumaraswamy distribution

It is well known that the application of mathematical models of the random component of experimental data with any probability distribution requires the use of certain mathematical procedures for statistical processing. The existing fundamental approaches to finding estimates of the parameters of such models are based on the method of moments and the method of maximum likelihood or its modifications, respectively. The first one has a relatively low accuracy, but is notable for its simple calculation. The other requires the implementation of rather complex algorithms and the availability of as large a sample size as possible, which characterizes it as asymptotically efficient [13–15]. In addition to these methods, other statistics can also be used to obtain an estimate: median, quantile (center of curvature) or rank estimates [5, 16].

3 MATERIALS AND METHODS

In general, finding estimates of an unknown scalar parameter θ by the polynomial maximization method over a sample of equally distributed random variables $\vec{x} = \{x_1, x_2, ... x_n\}$ involves solving a polynomial equation of the form:

$$\sum_{i=1}^{r} h_i(\theta)\left[\frac{1}{n}\sum_{v=1}^{n}\left(x_v^i - \alpha_i(\theta)\right)\right]\Bigg|_{\theta=\hat{\theta}} = 0, \qquad (1)$$

where r – predefined order of the polynomial, $a_i(\theta) = E\{x_i\}$ – mathematical expectations that depend on the parameter being estimated. The weighting coefficients $h_i(\theta)$ of equation (1) can be found under the condition of ensuring the minimum variance of the estimates of the parameter θ when using a polynomial of power r [17]. These coefficients are found from solving a system of linear algebraic equations of the form:

$$\sum_{i=1}^{r} h_i(\theta) F_{i,j}(\theta) = \frac{d}{d\theta}\alpha_j(\theta), \; j = \overline{1,r}, \qquad (2)$$

where the centered correlants $F_{i,j}(\theta) = \alpha_{i+j}(\theta) - \alpha_i(\theta)\alpha_j(\theta)$, $i, j = \overline{1,r}$. When forming a stochastic polynomial of power r for equation (2), it is necessary to have a partial probability description in the form of a set of moments up to the $2r$-th order.

It is known from [17–19] that when using power transforms as basis functions, polynomial estimates obtained with the powers of the polynomial $r = 1,2$ are equivalent to estimates of the mean for any symmetric distribution law of random variables. Therefore, in order to obtain estimates of the PMM that can prevail in accuracy, it is necessary to use a stochastic polynomial of power $r = 3$. To form the polynomial equation (1) with the unknown parameter θ, the relations for the first 6 initial moments were calculated. They will depend on the desired parameter θ, the second-order cumulant χ_2, and the cumulant coefficients γ_4 and γ_6:

DOI 10.15588/1607-3274-2023-3-7

$$\alpha_1(\theta)=\theta,\ \alpha_2(\theta)=\theta^2+\chi_2,\ \alpha_3(\theta)=\theta^3+3\theta\chi_2,$$
$$\alpha_4(\theta)=\theta^4+6\theta^2\chi_2+(3+\gamma_4)\chi_2^2,$$
$$\alpha_5(\theta)=\theta^5+10\theta^3\chi_2+5\theta(3+\gamma_4)\chi_2^2, \qquad (3)$$
$$\alpha_6(\theta)=\theta^6+15\theta^4\chi_2+15\theta^2(3+\gamma_4)\chi_2^2+$$
$$+(15(1+\gamma_4)+\gamma_6)\chi_2^3.$$

To find the weighting coefficients that minimize the variance of the PMM estimates of the distribution center parameter, the necessary expressions for the derivatives of the first 3 moments are as follows:

$$\frac{d}{d\theta}\alpha_1(\theta)=1, \frac{d}{d\theta}\alpha_2(\theta)=2\theta, \frac{d}{d\theta}\alpha_3(\theta)=3(\theta^2+\chi_2). \qquad (4)$$

If we use a polynomial of power $r=3$, then the equation for finding the PMM estimates $\hat{\theta}$ for the case of a symmetrically distributed random component of the experimental data can be written as follows:

$$h_1\sum_{v=1}^{n}\left[x_v-\theta\right]+h_2\sum_{v=1}^{n}\left[x_v^2-(\theta^2+\chi_2)\right]+ +h_3\sum_{v=1}^{n}\left[x_v^3-(\theta^3+3\theta\chi_2)\right]\Bigg|_{\theta=\hat{\theta}} \qquad (5)$$

where h_1-h_3 – are the weighting coefficients obtained from the analytical solution of the system of equations (2) by the Kramer method, and taking into account expressions (3) and (4), and are as follows:

$$h_1=\frac{3\theta^2\gamma_4-(6+12\gamma_4+\gamma_6)\chi_2}{((\gamma_4-9)\gamma_4-6-\gamma_6)\chi_2^2},$$
$$h_2=\frac{-3\theta\gamma_4}{((\gamma_4-9)\gamma_4-6-\gamma_6)\chi_2^2}, \qquad (6)$$
$$h_3=\frac{\gamma_4}{((\gamma_4-9)\gamma_4-6-\gamma_6)\chi_2^2}.$$

If the found coefficients (6) are substituted into the into the PMM equation (5), then a polynomial equation of the third power with respect to the parameter θ is obtained:

$$P_1\theta^3+P_2\theta^2+P_3\theta+P_4\Big|_{\theta=\hat{\theta}}=0, \qquad (7)$$

where the coefficients P_1-P_4 depend on the second-order cumulant χ_2 and the cumulant coefficients γ_4 and γ_6. They can be found by using sample statistical initial moments $\hat{\alpha}_i=\frac{1}{n}\sum_{v=1}^{n}x_v^i$, $i=\overline{1,6}$.

It is known [16, 19] that the solution of the cubic equation of the form (7) is obtained analytically on the basis of Cardano's formulas. However, there are alternative methods for obtaining an analytical solution, for example, the general cubic formula, Viete`s substitution, trigonometric and hyperbolic solutions, and the Lagrange method [20]. The effectiveness of these methods depends on the type of roots of the cubic equation. It is known that the correct solution of a PMM is a real root, which in general can be more than one, which requires additional verification by the criterion of minimum variance.

In [18], it was proposed to use numerical iterative algorithms for solving nonlinear equations. As an initial approximation for finding the root of equation (7), it is proposed to use a linear estimate of the desired parameter in the form of a sample mean $\hat{\alpha}_1$. The condition for stopping iterations is set based on the standard deviation of the estimates.

It is known that the Newton-Raphson method is most commonly used to solve nonlinear equations and their systems, but the method of simple iteration, the method of fastest descent, etc. can also be applied.

So, in fact, the question of using mathematical methods to solve a polynomial maximization equation requires a more thorough study, comparing analytical and numerical methods, the speed of algorithms, and the types of roots obtained.

It was proved in [17] that the estimates obtained as a result of solving the PMM equation are valid and asymptotically unbiased. It is also important that there is an analytical solution for the variance of such estimates. It is based on the concept of the amount of information obtained about the estimated parameter θ and is calculated using a well-known formula:

$$J_{rn(\theta)}=n\sum_{i=1}^{r}h_i(\theta)\frac{d}{d\theta}\alpha_i(\theta). \qquad (8)$$

The previously found values of the moments and weighting coefficients are used here. The inverse of the amount of obtained information in the asymptotic case tends to the variance of the estimated parameter, namely:

$$\sigma^2_{(\theta)r}=\lim_{n\to\infty}J^{-1}_{rn(\theta)}. \qquad (9)$$

Analogous to [16–17], for a comparative analysis of the relative accuracy of estimates, it is advisable to introduce the concept of the coefficient of variance reduction:

$$g_{(\theta)mean}=\frac{\sigma^2_{(\theta)PMM3}}{\sigma^2_{(\theta)mean}}. \qquad (10)$$

DOI 10.15588/1607-3274-2023-3-7

This coefficient indicates the ratio of the variance of the PMM estimates of the parameter θ obtained from the r-order polynomial equation to the variance of the PMM estimates when $r=1$ or to the variance of the method of moments estimates.

The variance $\sigma^2_{(\theta)mean}$ of the estimates of the parameter θ, found by the method of moments, does not depend on the value of this parameter. It is uniquely determined by the ratio of the second-order cumulant χ_2 to the volume of the sample n [17]:

$$\sigma^2_{(\theta)mean} = \frac{\chi_2}{n}. \tag{11}$$

If substituting the values of the obtained weighting coefficients (6) and the derivatives of the moments (4) into expression (8), then in the asymptotic case (at $n \to \infty$), it is possible to determine the variance of the PMM estimates of the order of the polynomial $r=3$ through the value of the cumulant coefficients:

$$\sigma^2_{(\theta)PMM3} = \frac{\chi_2}{n}\left[1 - \frac{\gamma_4^2}{6+9\gamma_4+\gamma_6}\right]. \tag{12}$$

In this case, the coefficient of variance reduction (10) will depend on the cumulant coefficients characterizing the kurtosis of the random variable and has the form:

$$g_{(\theta)mean} = 1 - \frac{\gamma_4^2}{6+9\gamma_4+\gamma_6}. \tag{13}$$

As a result of the research, theoretical coefficients for reducing the variance of estimates of the center of distributions with a negative kurtosis coefficient were calculated, as shown in Table 2. Since the selected distribution models have constant cumulant coefficients characterizing the kurtosis, the values obtained are numerical:

Table 2 – Theoretical coefficients of variance reduction

№	Вид розподілу	$g_{(\theta)mean}$	γ_4	γ_6
a)	U-quadratic distribution	0,037	–1,81	13,69
b)	Arcsine distribution	0,1	–1,5	10
c)	V-shaped distribution	0,074	–1,67	12
d)	Symmetrical Kumaraswamy distribution	0,058	–1,46	9,42

As shown in Table 2, all distributions are non-Gaussian, but it is the coefficient γ_4 that is negative. The results obtained do not contradict the domain of permissible values of cumulant coefficients shown in [17], and they cannot take on arbitrary values.

Based on the results of the research on the use of the PMM to estimate the parameter θ, it is possible to say that such estimates are superior in accuracy to the classical average method for distributions with a negative kurtosis. Thus, when the order of the stochastic polynomial is $r=3$, the efficiency advantage of the PMM estimates of the distribution center is more than 10 times. Although, of course, in solving certain practical problems, the efficiency advantage of such estimates using the PMM may be somewhat less.

4 EXPERIMENTS

Usually, it is quite problematic to check and prove the effectiveness of both methods and algorithms in practice due to the limited amount of data obtained through a real-world experiment. A way out of this situation may be the use of simulation modeling methods, among which the most common is the Monte Carlo method, which is based on obtaining a large number of random data realizations that have stochastic properties of the real process being reproduced.

In this work, a software implementation of a series of repeated experiments was developed in Mathematica to compare the accuracy of the PMM estimates of the polynomial order $r=3$ with other classical estimates. For this purpose, the experiment was supplemented with the calculation of median and quantile estimates (center of curve), which are also used to determine the center of the distribution.

As shown above, expression (10) can serve as a precise criterion of efficiency, but for experimental data it will contain empirical values of the variance reduction coefficients of the corresponding estimates:

$$\hat{g}_{(\theta)mean} = \frac{\hat{\sigma}^2_{(\theta)PMM3}}{\hat{\sigma}^2_{(\theta)mean}}, \quad \hat{g}_{(\theta)med} = \frac{\hat{\sigma}^2_{(\theta)PMM3}}{\hat{\sigma}^2_{(\theta)med}}, \quad \hat{g}_{(\theta)qvan} = \frac{\hat{\sigma}^2_{(\theta)PMM3}}{\hat{\sigma}^2_{(\theta)qvan}}, \tag{14}$$

where $\hat{\sigma}^2_{(\theta)mean}$, $\hat{\sigma}^2_{(\theta)med}$, $\hat{\sigma}^2_{(\theta)qvan}$, $\hat{\sigma}^2_{(\theta)PMM3}$ – average values of the variance of estimates for N experiments, obtained using the mean, median, quantile estimate, and PMM ($r=3$), respectively.

In the implementation of the experiment, the estimates of the sample initial moments for substitution into the stochastic polynomial equation were obtained using the known formula:

$$\hat{\alpha}_i = \frac{1}{n}\sum_{v=1}^{n} x_v^i, \quad i = \overline{1,6}. \tag{15}$$

DOI 10.15588/1607-3274-2023-3-7

When analyzing expressions (11) and (12), we can talk about the asymptotic efficiency of the methods, i.e., to obtain valid results, the sample of one experiment should be as large as possible. As for averaging the variance of the estimates, it is also necessary to have a certain number of experiments to ensure a predetermined accuracy and statistical stability of the modeling results. This calculation is performed using confidence intervals [21].

If a parameter $\hat{\theta}$ with mathematical expectation θ is estimated from a sample and has a mean $M(X)$, then there exists a value ε such that $|\theta - M(X)| < \varepsilon$, which will be the accuracy of the estimate. The probability of fulfillment of this inequality is its reliability $P(|\theta - M(X)| < \varepsilon) = E$.

Since the scheme of independent statistical tests with a relative frequency of occurrence of the required event m/N is considered, for sufficiently large N, a distribution close to normal will be obtained. Therefore, for each value of reliability E, the value of the function $\Phi(t) = E$ can be selected from the probability integral tables for such a value t_E that the accuracy ε is equal:

$$\varepsilon = t_E \sqrt{D(m/N)} = t_E \sqrt{D(X)/N}\,. \tag{16}$$

Thus, the number of experiments to obtain a given accuracy will be found as follows:

$$N = t_E^2 D(X)/\varepsilon^2\,. \tag{17}$$

If a number of random variables are evaluated during one experiment, then $\max D_i(X)$ is usually chosen to calculate the number of experiments N.

Typically, the reliability is set at $E = 0.95$, so the table value found is $t_E = 1.96$. The accuracy of calculating the variance and the coefficient of variance reduction will be adequate when the number has two decimal places, i.e. $\varepsilon = 0.01$. Calculating the variance values for the given probability distribution models with constant parameters, it is obtained $\max D_i(X) = 0.6$. Thus, the required number of experiments to get a given accuracy is as follows:

$$N = \frac{t_E^2 D(X)}{\varepsilon^2} = 23000\,. \tag{18}$$

5 RESULTS

The results of statistical modeling by the Monte Carlo method for the various distributions considered in the research and the volume of experimental sample values $n = 20 \div 200$ obtained at the number of experiments $N = 23 \cdot 10^3$ are presented in Table 3.

For a more visual representation and comparison, the coefficients of variance reduction obtained as a result of Monte Carlo simulation relative to the method of moments and their theoretical values were plotted, since they were calculated analytically. Figure 2 shows them depending on the volume of the sample n.

6 DISCUSSION

Usually, methods that perform nonlinear processing of any data, especially stochastic data, lead to the complication of mathematical and computational algorithms. The polynomial estimation of the parameters of the experimental data proposed in this study is no exception. Moreover, it is known that increasing the order of the stochastic polynomial r does not lead to a linear dependence of the increase in the accuracy of the obtained estimates. This fact somewhat reduces the application of the method in practice and forces a compromise between increasing accuracy or complexity. That is why in this study it was decided to choose the order $r = 3$ of the stochastic polynomial.

The analysis of the results of theoretical (Table 2) and experimental (Table 3) studies of variance reduction coefficients shows their significant correlation. This is especially evident when the sample size n increases, and also indicates the correctness of the statistical modeling. Such results fully confirm the asymptotic efficiency of polynomial estimation based on the property of the amount of obtained information (8).

Table 3 – Coefficients of variance reduction of estimates

Вид розподілу	$g_{(\theta)3}$	Результати статистичного моделювання											
		$\hat{g}_{(\theta)mean}$				$\hat{g}_{(\theta)med}$				$\hat{g}_{(\theta)qvan}$			
		n											
		20	50	100	200	20	50	100	200	20	50	100	200
U-quadratic distribution	0,037	0.16	0.044	0.038	0.035	0.018	0.002	0.001	0.001	0.284	0.161	0.151	0.153
Arcsine distribution	0,1	0.305	0.138	0.114	0.102	0.077	0.031	0.024	0.021	0.236	0.108	0.088	0.082
V-shaped distribution	0,074	0.25	0.097	0.08	0.071	0.039	0.009	0.005	0.003	0.319	0.17	0.154	0.14
Symmetrical Kumaraswamy distribution	0,058	0.196	0.093	0.068	0,061	0.109	0.044	0.031	0.026	0.284	0.138	0.108	0.094

DOI 10.15588/1607-3274-2023-3-7

p-ISSN 1607-3274 Радіоелектроніка, інформатика, управління. 2023. № 3
e-ISSN 2313-688X Radio Electronics, Computer Science, Control. 2023. № 3

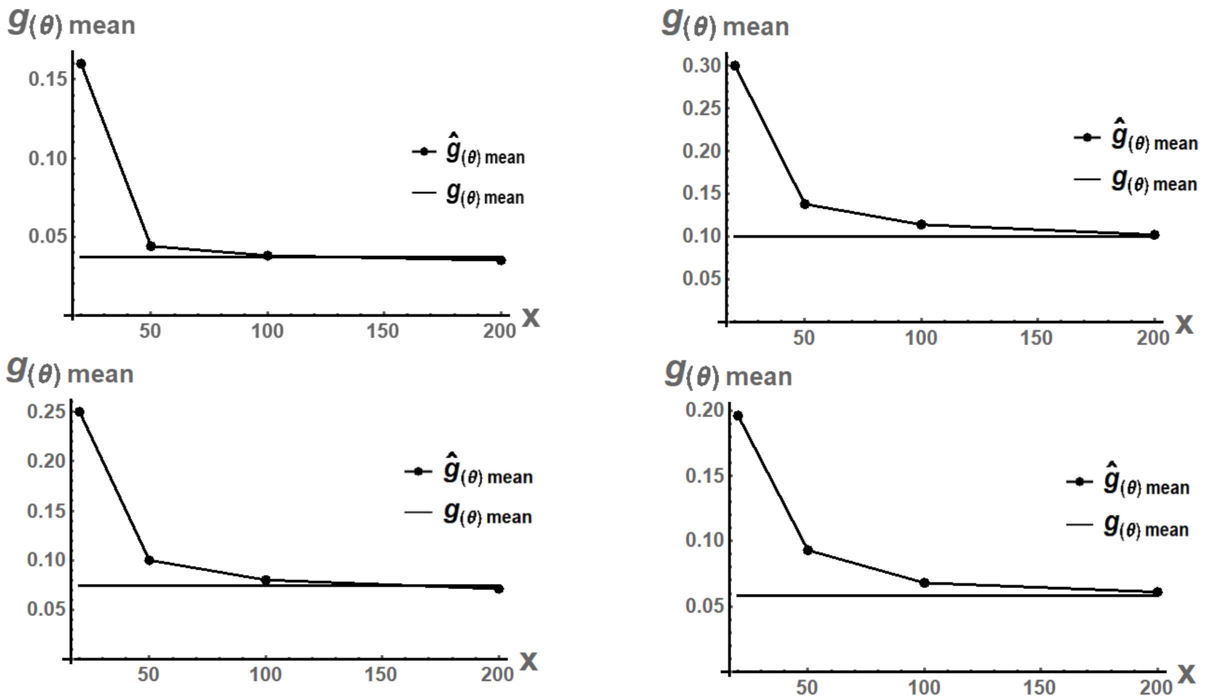

Figure 2 – Comparison of the efficiency of methods for finding center coordinate estimates depending on the volume of the sample: a – U-quadratic distribution, b – arcsine distribution,
c – V-shaped distribution, d – symmetrical Kumaraswamy distribution

A general analysis of the results of comparing different estimation methods with the PMM shows its high efficiency. It is demonstrated by the fact that the PMM has an average 3-5 times higher accuracy of estimates than classical methods, as shown in Figure 2. This is also true for small (20 values) sample volumes, which is important for limited experiments, although the smallest reduction in the variance of the estimate in such cases is observed compared to the method of quantile estimates. As for the median estimate, the PMM has the greatest advantage over it in terms of accuracy for models with a gentler curve.

As the sample volume increases, the coefficient of variance reduction tends to its theoretical value, as it should be, which is confirmed by the results of statistical modeling. This trend can be observed both in the numerical results (Table 3) and in the graphical interpretation (Figure 2), which shows a comparison of the experimental value and the theoretical one. Although the coefficient of variance reduction increases with decreasing non-Gaussianity of the experimental data (the advantage in accuracy decreases), the advantage of polynomial estimates remains significant.

CONCLUSIONS

The research conducted in this paper confirms the potentially high efficiency of the polynomial maximization method in determining estimates of the coordinates of the center of distribution of experimental data that are adequately described by models of probability distributions with a negative kurtosis. An important and essential feature of this approach is the fact that it does not require any prior information about the type and parameters of the distributions. However, for the purpose of demonstration, the necessary well-known models were used, which are adequate to the task at hand. The paper shows that the algorithm for finding the PMM estimates of an informative parameter using the polynomial order $r = 3$ can be reduced to solving a cubic equation. For this purpose, both explicit and numerical methods can be used. The coefficients of this equation are obtained from the calculation of a posteriori estimates of the initial moments up to the 6-th order of the stochastic component of the experimental data.

The basis for finding the coefficient of variance reduction in this paper is the concept of the amount of information extracted, which is identical to Fisher's information for the Maximum Likelihood Estimation. However, the amount of information extracted was calculated based on sample statistical characteristics, such as moments and cumulants. The result is an analytical expression that, for each model distribution, when substituting the numerical values of the interval on which it is defined, gives a number in the final case.

Statistical modeling has confirmed the effectiveness of the PMM estimates compared to known nonparametric estimates based on the mean, median, and quantile statistics, even with small sample sizes. The simulation results show that the accuracy of the proposed approach is sometimes significantly (more than 3 times) higher than classical nonparametric estimates.

ACKNOWLEDGEMENTS

The work is supported by the registered research project of Cherkasy State Technological University «Models, methods and tools of joint signal detection and parameter

DOI 10.15588/1607-3274-2023-3-7

estimation in non-Gaussian noise» (state registration number 0122U201835).

REFERENCES

1. International Vocabulary of Metrology Fourth edition – Committee Draft (VIM4 CD). [Effective from 2021-01-11]. JCGM, 2021, 55 p.
2. Dorozhovets M. Opratsiuvannia rezultativ vymiriuvan: Navch. posibnyk. Lviv, Vydavnytstvo Natsionalnoho universytetu «Lvivska politekhnika», 2007, 624 p.
3. Dikhtiievskyi O. V., Kvasnikov V. P., Vozniakovskyi A. O. Otsinka systematychnoi ta vypadkovoi skladovykh pokhybky vymiriuvannia tsylindrychnykh zubchastykh kolis, *Perspektyvni tekhnolohii ta prylady*, 2019, № 14, pp. 62–66. DOI: 10.36910/6775-2313-5352-2019-14-10
4. Nevyznachenist vymiriuvan. Chastyna 3. Nastanova shchodo podannia nevyznachenosti u vymiriuvanni (GUM:1995): DSTU ISO/IEC Guide 98-3:2018 (ISO/IEC Guide 98-3:2008, IDT). [Chynnyi vid 2018-12-18]. Kyiv, DP «UkrNDNTs», 2019, 272 p. (Natsionalnyi standart Ukrainy)
5. Zakharov Y. P., Shtefan N. V. Opredelenye эffektyvnыkh otsenok tsentra raspredelenyia pry statystycheskoi obrabotke rezultatov nabliudenyi, *Radioe'lektronika i informatika,,* 2002, Vyp. 3, S. 97–99
6. Dodge Y. The Concise Encyclopedia of Statistics. NY,: Springer New York, 2008, 612 p. DOI: 10.1007/978-0-387-32833-1
7. Ribeiro A. S., Sousa J. A., Castro M. P. Some remarks on the use of U-shape probability distribution functions in Monte Carlo simulation, *Advances of Measurement Science 2004: 10th IMEKO TC7 Symposium, Saint-Petersburg, 30 June – 2 July 2004: proceedings*. NY, IMEKO-International Measurement Federation Secretariat, 2004, pp. 124–130.
8. Ceekala M. Stochastic ADC with random U-quadratic distributed reference voltages to uniformly distribute comparators trip point : thesis … Master of Applied Science. Nova Scotia, Dalhousie University, 2013, 91 p.
9. Carpenter D. A demystification of the U-shaped probability distribution, *Electromagnetic Compatibility: 2003 IEEE Symposium, Boston, 18–22 August 2003: proceedings*. Los Alamitos, IEEE, 2003, Vol. 2, pp. 521–525, DOI: 10.1109/ISEMC.2003.1236656.
10. Harris I. A., Warner F. L. Re-examination of mismatch uncertainty when measuring microwave power and attenuation *IEE Proceedings H: Microwaves Optics and Antennas*, 1981, Vol. 128, Issue 1, pp. 35–41, DOI: 10.1049/ip-h-1.1981.0007
11. Badircea R. M., Florea N. M., Manta A. G. et al. Comparison between Romania and Sweden based on three dimensions: Environmental performance, green taxation and economic growth, *Sustainability,* 2020, Vol. 12, Issue 9, pp. 3817–3834, DOI: 10.3390/su12093817
12. Sella G., Hirsh A. E. The application of statistical physics to evolutionary biology, *Proceedings of the National Academy of Sciences of the United States of America,* 2005, Vol. 102, Issue 27, pp. 9541–9546. DOI: 10.1073/pnas.0501865102
13. Javanshir Z., Habibi Rad A., Arghami N. R. Exp-Kumaraswamy distributions: some properties and applications, *Journal of Sciences, Islamic Republic of Iran,* 2015, Vol. 26, Issue 1, pp. 57–69.
14. Emam W., Tashkandy Y. The arcsine Kumaraswamy-generalized family: Bayesian and classical estimates and application, *Symmetry*, 2022, Vol. 14, Issue 11, pp. 2311–2330. DOI: 10.3390/sym14112311
15. Muhammad M., Rano S., Sani R. et al. Parameter estimation of exponentiated U-quadratic distribution: alternative maximum likelihood and percentile methods, *Asian Research Journal of Mathematics,* 2018, Vol. 9, Issue 2, pp. 1–10., DOI: 10.9734/ARJOM/2018/39446
16. Zabolotnii S. V., Kucheruk V. Yu., Warsza Z. L. et al. Polynomial estimates of measurand parameters for data bimodal mixtures of exponential distributions, *Bulletin of the Karaganda University*, 2018, Vol. 2, Issue 90, pp. 71–80.
17. Kunchenko Y. P. Polynomial parameter estimation of close to Gaussian random variables. Aachen, Shaker Verlag, 2002, 396 p.
18. Zabolotnii S. V., Chepynoha A. V., Bondarenko Yu. Yu. et al. Polynomial parameter estimation of exponential power distribution data, *Visnyk NTUU KPI Seriia – Radiotekhnika Radioaparatobuduvannia*, 2018, Vol. 75, pp. 40–47. DOI: 10.20535/RADAP.2018.75.40-47
19. Warsza Z. L., Zabolotnii S. W. A Polynomial Estimation of Measurand Parameters for Samples of Non-Gaussian Symmetrically Distributed Data, *Advances in Intelligent Systems and Computing*, 2017, Vol. 550, pp. 468–480, DOI: 10.1007/978-3-319-54042-9_45
20. Cooke R. Classical Algebra: Its Nature, Origins, and Uses. New Jersey, John Wiley & Sons, 2008, 224 p. DOI:10.1002/9780470277980
21. Tomashevskyi V. M. Modeliuvannia system. Kyiv, Vydavnycha hrupa BHV, 2005, 352 p.

Received 12.06.2023.
Accepted 25.08.2023.

УДК 519.2:004.94

ПОЛІНОМІАЛЬНЕ ОЦІНЮВАННЯ ПАРАМЕТРІВ МОДЕЛЕЙ ДАНИХ З ВІД'ЄМНИМ КОЕФІЦІЄНТОМ ЕКСЦЕСУ

Чепинога В. В. – аспірант кафедри комп'ютерних наук та системного аналізу, Черкаський державний технологічний університет, Черкаси, Україна.

Чепинога А. В. – канд. техн. наук, доцент, декан факультету інформаційних технологій та систем, Черкаський державний технологічний університет, Черкаси, Україна.

Палагін В. В. – д-р техн. наук, професор, завідувач кафедри робототехнічних і телекомунікаційних систем та кібербезпеки, Черкаський державний технологічний університет, Черкаси, Україна.

АНОТАЦІЯ

Актуальність. В роботі розглянуто задачу оцінювання центру розподілу випадкової складової експериментальних даних для моделей щільності з від'ємним коефіцієнтом ексцесу.

Мета. Метою роботи є отримання методів підвищенні ефективності поліноміального оцінювання параметрів експериментальних даних з від'ємним коефіцієнтом ексцесу.

Метод. В дослідженні застосовано відносно новий підхід для отримання оцінок центру розподілу імовірності з результатів експериментальних даних, що мають стохастичну складову. Цей підхід засновано на поліноміальних методах оціню-

DOI 10.15588/1607-3274-2023-3-7

вання, котрі спираються на математичний апарат стохастичних поліномів Кунченка та опис випадкових величин статистиками вищих порядків (моментами чи кумулянтами). В якості моделей випадкової складової в роботі використано ряд розподілів щільності імовірності з від'ємним коефіцієнтом ексцесу.

В якості міри ефективності оцінок було використано відношення дисперсії оцінки центру розподілу, знайденої з використанням поліноміальних та класичних методів, виходячи із параметра кількості добутої інформації.

Досліджено, із застосуванням методу Монте-Карло для багаторазових випробувань, відносну точність поліноміальних оцінок у порівнянні з оцінками середнього, медіани та квантильних оцінок (центру перегину).

Результати. Побудовано поліноміальні методи оцінювання параметра центру розподілу для моделей даних щільності розподілу імовірності з від'ємним коефіцієнтом ексцесу.

Висновки. Дослідження, що були проведені в даній роботі, підтверджують потенційно високу ефективність поліноміальних оцінок координати центру експериментальних даних, що адекватно описуються модельними розподілами з від'ємним коефіцієнтом ексцесу. Статистичне моделювання підтвердило ефективність отриманих оцінок в порівнянні із відомими непараметричними оцінками, на основі статистик середнього, медіани і квантильної оцінки, причому навіть при малих об'ємах вибірки.

КЛЮЧОВІ СЛОВА: вибірка даних, оцінювання, стохастичний поліном, кумулянти, від'ємний ексцес.

REFERENCES

1. International Vocabulary of Metrology Fourth edition – Committee Draft (VIM4 CD). – [Effective from 2021-01-11]. – JCGM, 2021. – 55 p.
2. Дороховець М. Опрацювання результатів вимірювань: Навч. посібник / М. Дороховець. – Львів : Видавництво Національного університету «Львівська політехніка», 2007. – 624 с.
3. Діхтієвський О. В. Оцінка систематичної та випадкової складових похибки вимірювання циліндричних зубчастих коліс / О. В. Діхтієвський, В. П. Квасніков, А. О. Возняковський // Перспективні технології та прилади. – 2019. – № 14. – С. 62-66 DOI: 10.36910/6775-2313-5352-2019-14-10
4. Невизначеність вимірювань. Частина 3. Настанова щодо подання невизначеності у вимірюванні (GUM:1995): ДСТУ ISO/IEC Guide 98-3:2018 (ISO/IEC Guide 98-3:2008, IDT). – [Чинний від 2018-12-18]. – К.: ДП «УкрНДНЦ», 2019. – 272 с. – (Національний стандарт України)
5. Захаров И. П. Определение эффективных оценок центра распределения при статистической обработке результатов наблюдений / И. П. Захаров, Н. В. Штефан //, Радиоэлектроника и информатика. – 2002. – Вып. 3. – С. 97–99
6. Dodge Y. The Concise Encyclopedia of Statistics / Y. Dodge – NY: Springer New York, 2008. – 612 p., DOI: 10.1007/978-0-387-32833-1
7. Ribeiro A. S. Some remarks on the use of U-shape probability distribution functions in Monte Carlo simulation / A. S. Ribeiro, J. A. Sousa, M. P. Castro // Advances of Measurement Science 2004: 10th IMEKO TC7 Symposium, Saint-Petersburg, 30 June – 2 July 2004: proceedings. – NY: IMEKO-International Measurement Federation Secretariat, 2004. – P. 124–130
8. Ceekala M. Stochastic ADC with random U-quadratic distributed reference voltages to uniformly distribute comparators trip point : thesis … Master of Applied Science / Mithun Ceekala. – Nova Scotia: Dalhousie University, 2013. – 91 p.
9. Carpenter D. A demystification of the U-shaped probability distribution / D. Carpenter // Electromagnetic Compatibility: 2003 IEEE Symposium, Boston, 18–22 August 2003: proceedings. – Los Alamitos: IEEE, 2003. – Vol. 2 – P. 521–525. DOI: 10.1109/ISEMC.2003.1236656.
10. Harris I. A. Re-examination of mismatch uncertainty when measuring microwave power and attenuation / I. A. Harris, F. L. Warner // IEE Proceedings H: Microwaves Optics and Antennas. – 1981. – Vol. 128, Issue 1. – P. 35–41. DOI: 10.1049/ip-h-1.1981.0007
11. Comparison between Romania and Sweden based on three dimensions: Environmental performance, green taxation and economic growth / [R. M. Badircea, N. M. Florea, A. G. Manta et al.] // Sustainability. – 2020. – Vol. 12, Issue 9. – P. 3817–3834, DOI: 10.3390/su12093817
12. Sella G. The application of statistical physics to evolutionary biology / G. Sella, A. E. Hirsh // Proceedings of the National Academy of Sciences of the United States of America. – 2005. – Vol. 102, Issue 27. – P. 9541–9546., DOI: 10.1073/pnas.0501865102
13. Javanshir Z. Exp-Kumaraswamy distributions: some properties and applications / Z. Javanshiri, A. Habibi Rad, N.R. Arghami // Journal of Sciences, Islamic Republic of Iran. – 2015. – Vol. 26, Issue 1. – P. 57–69.
14. Emam W. The arcsine Kumaraswamy-generalized family: Bayesian and classical estimates and application / W. Emam, Y. Tashkandy // Symmetry. – 2022. – Vol. 14, Issue 11. – P. 2311–2330., DOI: 10.3390/sym14112311
15. Parameter estimation of exponentiated U-quadratic distribution: alternative maximum likelihood and percentile methods / [M. Muhammad, S. Rano, R. Sani et al.] // Asian Research Journal of Mathematics. – 2018. – Vol. 9, Issue 2. – P. 1–10., DOI: 10.9734/ARJOM/2018/39446
16. Polynomial estimates of measurand parameters for data bimodal mixtures of exponential distributions / [S. V. Zabolotnii, V. Yu. Kucheruk, Z. L. Warsza et al.] // Bulletin of the Karaganda University. – 2018. – Vol. 2, Issue 90. – P. 71–80.
17. Kunchenko Y. P. Polynomial parameter estimation of close to Gaussian random variables / Y. P. Kunchenko. – Aachen: Shaker Verlag, 2002. – 396 p.
18. Поліноміальні оцінки параметрів для даних з експоненційним степеневим розподілом / [С. В. Заболотній, А. В. Чепинога, Ю. Ю. Бондаренко та ін.] // Вісник НТУУ «КПІ». Серія: Радіотехніка. Радіоапаратобудування. – 2018. – Вип. 75. – С. 40–47. DOI: 10.20535/RADAP.2018.75.40-47
19. Warsza Z. L. A Polynomial Estimation of Measurand Parameters for Samples of Non-Gaussian Symmetrically Distributed Data / Z. L. Warsza, S. W. Zabolotnii // Advances in Intelligent Systems and Computing, – 2017. – Vol. 550. – P. 468–480, DOI: 10.1007/978-3-319-54042-9_45
20. Cooke R. Classical Algebra: Its Nature, Origins, and Uses / R. Cooke. – New Jersey: John Wiley & Sons, 2008. – 224 p. DOI:10.1002/9780470277980
21. Томашевський В.М. Моделювання систем / В. М. Томашевський – К. : Видавнича група BHV, 2005. – 352 с.

DOI 10.15588/1607-3274-2023-3-7