УДК 519.2:681.2

Comparative Analysis of Polynomial Maximization and Maximum Likelihood Estimates for Data with Exponential Power Distribution

Zabolotnii S. V.[1], Chepynoha A. V.[2], Chorniy A. M.[2], Honcharov A. V.[2]

[1]Cherkasy State Business College, Ukraine
[2]Cherkasy State Technological University, Ukraine

E-mail: *a.chepynoha@chdtu.edu.ua*

The work is devoted to the estimate accuracy comparative analysis of the experimental data parameters with exponential power distribution (EPD) using the classical Maximum Likelihood Estimation (MLE) and the original Polynomial Maximization Method (PMM). In contrast to the parametric approach of MLE, which uses the description in the form of probability density distribution, PMM is based on a partial description in the of higher-order statistics form and the mathematical apparatus of Kunchenko's stochastic polynomials. An algorithm for finding PMM estimates using 3rd order stochastic polynomials is presented. Analytical expressions allowing to determine the variance of PMM-estimates of the asymptotic case parameters and EPD parameters with a priori information are obtained. It is shown that the relative theoretical estimates accuracy of different methods significantly depends on the EPD shape parameter and matches only for a separate case of Gaussian distribution. The effectiveness of different approaches (including valuation of mean values estimates) both with and without a priori information on EPD properties was investigated by repeated statistical tests (through Monte Carlo Method). The greatest efficiency areas for each of methods depending on EPD shape parameter and sample data volume are constructed.

Key words: exponential power distribution; stochastic polynomials; high-order statistics; parameter estimation

Intro

Current information and measurement technologies trends are focused on the use and software systems development that will provide a solution to the problem of statistical processing the experimental results in any form of recorded measurements and necessarily includes modern statistical analysis methods. They may solve a wide range of measurement problems, but also should be used properly when applied for researching, like in case of using a model that includes the additive interaction of the measured parameter and random error [1].

It is well known that random measuring instruments errors comply with the normal (Gaussian) law seldom [2]. Moreover, the basis of measuring instruments and systems are different physical principles, different measurement methods and different measuring signals transformations. And in fact, measurement errors are the result of many factors that influence, both random and non-random, acting constantly or periodically. Therefore, it is clear that

only if certain preconditions (theoretical and technical) have place – measurement errors are enough described by the normal distribution law [3].

EPD is offered as an alternative to the Gaussian distribution law assumption for measurement error probability while the model of exponential power distribution is applied [4]. And since the distribution of experimental data is usually symmetric according to its center, the exponential power distribution can be claimed as a model of error measurement distribution law.

A typical view of the EPD model is described by a species function:

$$\mathrm{w}(x) = \frac{1}{2\sigma p^{1/p}\,\Gamma\left(1+1/p\right)}\exp\left(-\frac{|x|^p}{p\sigma^p}\right), \quad (1)$$

where σ – scale parameter, p – shape parameter.

The EPD usage is substantiated as a model for error measurement probability distribution [5,6], to describe the probabilistic characteristics of random noise while video and audio signals processing [7], biomedical research statistics [8], risk management [9], etc. The

great advantage of this model is that it can describe random variables that have both positive (flat-peak distributions) and negative (acute-peak distributions) kurtosis coefficient [6, 10]. And such models are in demand in many fields of science and technology. In addition, EPD model modifications for asymmetrically distributed random variables have already been developed [11–14].

1   Purpose of research

As a detailed analysis of works [1, 2, 15, 16] shows, due to the simplicity of their calculation simple statistical characteristics are used for getting the center coordinate of symmetrically distributed experimental data estimates. Incidentally, with certain predetermined requirements for the accuracy of obtaining estimates, such simple estimates may not be sufficiently accurate. Turning to the description of random variables using the probability density distribution, we actually have at present the use of the most accurate method for obtaining estimates – the method of maximum likelihood. But this approach produces problems associated with nonlinearity of processing [9, 11, 12]. The most complicated thing here is related to the practical application of numerical methods to solve nonlinear equations [13, 14, 17].

Thus, the researcher is often faced with the task of choosing either a simpler method, which may not satisfy the accuracy, or cumbersome calculations that require the implementation of complex computational procedures.

To solve this dilemma this paper proposes to apply an original approach to statistical parameter estimation, which is based on the polynomial maximization method (PMM) by describing random variables in the form of a finite moments or cumulants number [18].

In [19] was developed an algorithm for finding estimates for multiple measurements results on random errors background that are described by the EPD model; besides comparative analysis of the PMM estimates accuracy with estimates based on three nonparametric statistics: median, mean, mid-range. The purpose of this study is a comparative analysis of PMM estimates accuracy and parametric maximum likelihood estimates (MLE).

The analysis procedure involves obtaining both: theoretical expressionsfor parameters estimates variance ratio and the comparison of estimates variances empirical values using statistical modeling by the Monte Carlo method.

2   Mathematical problem statement

Supposing θ − is informative parameter, the value of which must be estimated on basis of set of values $\vec{x} = \{x_1, x_2, \ldots, x_n\}$ analysis. This vector contains independent and equally distributed sample values of the measurement model $x = \theta + \xi$, where ξ − random error, which is adequately described by the exponential power distribution of the form (1).

It is necessary to determine the accuracy (parameter estimates variance) of different estimation methods, as well as to assess the relative accuracy impact of the sample size and whether there is a priori information about the EPD model parameters.

3   Finding estimates using the polynomial maximization method

As shown in [19–22], the application of the polynomial maximization method is based on finding the functional extremum in the form of a stochastic functional polynomial of a certain r order. When using basic power transformations to form such a polynomial, a description in the form of a finite sequence of moments or cumulants up to the $2r$ order must be used. This polynomial is formed in a way that provides an extremum (maximum) in the true value vicinity of various evaluated parameter θ. Thus, finding the desired parameter estimates is reduced to finding the root of the function $l_{rn}(\theta)$, that is formed as a derivative of the corresponding parameter and depends on the sample values of $\vec{x} = \{x_1, x_2, \ldots, x_n\}$:

$$l_{rn}(\theta) = \sum_{i=1}^{r} h_i(\theta) \left[\frac{1}{n} \sum_{v=1}^{n} x_v^i - \alpha_i(\theta) \right] \Bigg|_{\theta=\hat{\theta}} = 0, \quad (2)$$

where $\alpha_i(\theta) = E\{x^i\}$ − mean values which are the initial points depending on the estimated parameter.

In fact, the function $l_{rn}(\theta)$ is also the sum of differences r of theoretical and sample moments, considered by certain coefficients $h_i(\theta)$. These coefficients can be found as the solution of a linear algebraic set of equations shown below. This set is based on the minimum variance ensuring criterion of the desired parameter θ estimates when the r degree of polynomial is used:

$$\sum_{i=1}^{r} h_i(\theta) \ F_{i,j}(\theta) = \frac{d}{d\theta} \alpha_j(\theta), \ \ j = \overrightarrow{1, r}, \quad (3)$$

where the centered correlants $F_{i,j}(\theta) = \alpha_{i+j}(\theta) - \alpha_i(\theta)\alpha_j(\theta); i, j = \overrightarrow{1, r}$.

In [19] it was shown that finding the θ parameter estimates using a symmetric error distribution model requires polynomial degree $r = 3$ to be the least as possible.

The correlations for the first 6 initial moments required for the formation of the set of equations (3)

have the form:

$$\alpha_1 = \theta, \ \ \alpha_2 = \theta^2 + \mu_2, \ \alpha_3 = \theta^3 + 3\theta\mu_2,$$
$$\alpha_4 = \theta^4 + 6\theta^2\mu_2 + 3\mu_2^2 + \mu_4,$$
$$\alpha_5 = \theta^5 + 10\theta^3\mu_2 + 15\theta\mu_2^2 + 5\theta\mu_4, \tag{4}$$
$$\alpha_6 = \theta^6 + 15\theta^4\mu_2 + 45\theta^2\mu_2^2 + 15\mu_2^3 + 15\theta^2\mu_4 +$$
$$+ 15\mu_2\mu_4 + \mu_6,$$

where μ_i – the central moments of the distribution (1), depending only on the parameters of scale σ and shape p [17]:

$$\mu_i = \Gamma\left(\frac{i+1}{p}\right)\Gamma\left(\frac{1}{p}\right)^{\frac{i-2}{2}}\Gamma\left(\frac{3}{p}\right)^{-\frac{i}{2}}\sigma^i. \tag{5}$$

$$h_1 = -\frac{\Gamma(3p^{-1})\left(9\theta^2\Gamma(3p^{-1})^3 - 3(\theta^2+\sigma^2)\Gamma(p^{-1})\Gamma(3p^{-1})\Gamma(5p^{-1}) + \sigma^2\Gamma(p^{-1})^2\Gamma(7p^{-1})\right)}{\mu_2^2\Gamma(p^{-1})^2\left(\Gamma(5p^{-1})^2 - \Gamma(3p^{-1})\Gamma(7p^{-1})\right)},$$

$$h_2 = -\frac{3\theta\Gamma(3p^{-1})^2\left(-3\Gamma(3p^{-1})^2 + \Gamma(p^{-1})\Gamma(5p^{-1})\right)}{\mu_2^2\Gamma(p^{-1})^2\left(\Gamma(5p^{-1})^2 - \Gamma(3p^{-1})\Gamma(7p^{-1})\right)}, \tag{7}$$

$$h_3 = \frac{\Gamma(3p^{-1})^2\left(-3\Gamma(3p^{-1})^2 + \Gamma(p^{-1})\Gamma(5p^{-1})\right)}{\mu_2^2\Gamma(p^{-1})^2\left(\Gamma(5p^{-1})^2 - \Gamma(3p^{-1})\Gamma(7p^{-1})\right)}.$$

The equation of polynomial degree $r = 3$ maximization for finding PMM estimates $\hat{\theta}$ for the case of symmetrically distributed experimental data can be represented as:

$$h_1\sum_{v=1}^{n}(x_v - \theta) + h_2\sum_{v=1}^{n}\left[x_v^2 - (\theta^2 + \mu_2)\right] +$$
$$+ h_3\sum_{v=1}^{n}\left[x_v^3 - (\theta^3 + 3\theta\mu_2)\right]\Bigg|_{\theta=\hat{\theta}} = 0. \tag{8}$$

Substituting the values of the found optimal coefficients (7) in equation (8), we obtain directly the expression for estimating the desired parameter of the distribution center θ [19].

4 Polynomial maximization method estimates accuracy

It is known that the main statistical estimation methods accuracy criterion is the variance value of the resulting parameter estimates. One of the ways to obtain analytical expressions describing the method of maximum likelihood estimates variance is to use the apparatus of quantitative information according to Fisher

$$I_\theta = -E\left[\frac{\partial^2\ln[l(\theta,\sigma,p)]}{\partial\theta^2}\right], \tag{9}$$

where $l(\theta,\sigma,p)$ – logarithm of the likelihood function, θ – desired parameter of the distribution center, σ – scale parameter , p – shape parameter.

To find the optimal coefficients that minimize the variance of PMM estimates, expressions for derivatives from the first 3 moments should be calculated:

$$\frac{\partial}{\partial\theta}\alpha_1 = 1,$$
$$\frac{\partial}{\partial\theta}\alpha_2 = 2\theta + \mu_2, \tag{6}$$
$$\frac{\partial}{\partial\theta}\alpha_3 = 3\theta^2 + 3\mu_2.$$

Via analytical solution of the set of equations (3) by the Cramer method (taking into account expressions (4) and (6)), the optimal coefficients expressions for the power stochastic polynomial are obtained:

When using a probabilistic model of the form (1), as shown in [17], the logarithm of the likelihood function is described by the function:

$$l(\theta,\sigma,p) = \ln L(\theta,\sigma,p) =$$
$$= -n\ln\left[2\sigma p^{1/p}\Gamma(1+1/p)\right]^{-n} - \frac{\sum_{i=1}^{n}|x_i - \theta|}{p\sigma^p}. \tag{10}$$

The variance estimates for the asymptotic case (at $n \to \infty$) can be found as the inverse of the Fisher information. Thus, using (9) and (10) allows to obtain an analytical expression:

$$\operatorname*{var}_{n\to\infty}\left(\widehat{\theta}\right) = \sigma^2_{(\theta)\ MLE} = \frac{\Gamma(1+1/p)^2}{\Gamma(2-1/p)\cdot\Gamma(3/p)}\frac{\sigma^2}{n}. \tag{11}$$

In [18] it was shown that for the polynomial maximization method a certain analogue of the amount of Fisher information is the so-called amount of obtained information about the parameter estimated using the stochastic polynomial degree r, which is generally described by the expression:

$$J_{rn(\theta)} = E\left[\frac{d}{d\theta}l_{rn}(\theta)\right]^2 = -E\frac{d^2}{d\theta^2}l_{rn}(\theta). \tag{12}$$

In addition, it is proved that in the asymptotic case (at $r \to \infty$) the amount of extracted information goes to the amount of Fisher information.

After certain basic power transformations, the expression for the amount of information obtained

about the parameter θ can be represented as:

$$J_{rn(\theta)} = -E\frac{d^2}{d\theta^2}l_{rn}(\theta) =$$

$$= n\sum_{i=1}^{r}\sum_{j=1}^{r}h_i(\theta)h_j(\theta)F_{i,j}(\theta) = n\sum_{i=1}^{r}h_i(\theta)\frac{d}{d\theta}\alpha_i(\theta). \tag{13}$$

This quantitative characteristic has statistical properties similar to Fisher's information. That is, in the asymptotic case (at $n \to \infty$) the variance of PMM-estimates is defined as the inverse of the amount obtained at the r degree of information, i.e.:

$$\sigma_{(\theta)r}^2 = \lim_{n\to\infty} J_{rn(\theta)}^{-1}. \tag{14}$$

For a comparative analysis of the estimates relative accuracy, let us use the concept of the variance coefficient [19–22]:

$$g_{(\theta)\,r} = \frac{\sigma_{(\theta)\,PMM}^2}{\sigma_{(\theta)\,MLE}^2}. \tag{15}$$

The obtained coefficient is the ratio of the PMM parameter estimates variance θ when the polynomial r-th degree is used and the variance of the PMM estimates.

Using form (13), it is easy to determine an analytical expression for the amount of information obtained through the EPD shape parameter p
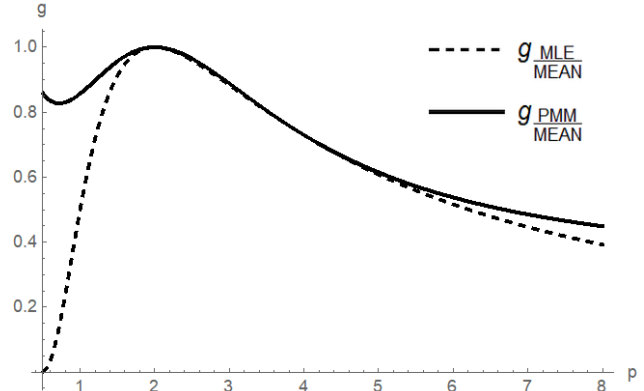
$$J = -\frac{n\Gamma\left[3p^{-1}\right]\left(9\Gamma\left[3p^{-1}\right]^3 - 6\Gamma\left[p^{-1}\right]\Gamma\left[3p^{-1}\right]\Gamma\left[5p^{-1}\right] + \Gamma\left[p^{-1}\right]^2\Gamma\left[7p^{-1}\right]\right)}{\sigma^2\Gamma[p^{-1}]^2\left(\Gamma[5p^{-1}]^2 - \Gamma\left[3p^{-1}\right]\Gamma\left[7p^{-1}\right]\right)}. \tag{16}$$

Thus, for the asymptotic case (at $n \to \infty$) expression (16) allows to theoretically determine the PMM estimates variance at the $r = 3$ polynomial degree:

$$\sigma_{PMM}^2 = -\frac{\sigma^2\Gamma\left[p^{-1}\right]^2\left(\Gamma\left[5p^{-1}\right]^2 - \Gamma\left[3p^{-1}\right]\Gamma\left[7p^{-1}\right]\right)}{n\Gamma\left[3p^{-1}\right]\left(9\Gamma[3p^{-1}]^3 - 6\Gamma\left[p^{-1}\right]\Gamma\left[3p^{-1}\right]\Gamma\left[5p^{-1}\right] + \Gamma[p^{-1}]^2\Gamma\left[7p^{-1}\right]\right)}. \tag{17}$$

It is also known [3] that the parameter estimates variance in the form of the arithmetic mean depends only on the random component variance and the sample values volume:

$$\sigma_{MM}^2 = \frac{\sigma^2}{n}. \tag{18}$$

Using expressions (11) and (18), it isn't hard to obtain the variance reduction coefficient of the PMM-estimates compared to the arithmetic mean estimates:

$$g_{\frac{MLE}{MM}} = \frac{\Gamma\left[1 + p^{-1}\right]^2}{\Gamma\left[2 - p^{-1}\right]\Gamma\left[3p^{-1}\right]}. \tag{19}$$

Fig. 1. Dependency of variance reduction coefficients $g_{PMM/MM}$ and $g_{MLE/MM}$ with EPD shape parameter p.

Using expressions (17–19), graphs can be presented (see Fig. 1) showing the dependence of the MLE relative accuracy and PMM estimates compared with the arithmetic mean estimates of the EPD shape parameter.

Using expressions (17) and (19) it is also easy to obtain a function (the graph of which is presented in Fig. 2), which makes it possible to directly compare the accuracy (magnitude of asymptotic variances) of MLE and PMM-estimates:

$$g_{\frac{MLE}{PMM}} = -\frac{\Gamma\left[1 + p^{-1}\right]^2\left(9\Gamma\left[3p^{-1}\right]^3 - 6\Gamma\left[p^{-1}\right]\Gamma\left[3p^{-1}\right]\Gamma\left[5p^{-1}\right] + \Gamma\left[p^{-1}\right]^2\Gamma\left[7p^{-1}\right]\right)}{\Gamma\left[2 - p^{-1}\right]\Gamma[p^{-1}]^2\left(\Gamma[5p^{-1}]^2 - \Gamma\left[3p^{-1}\right]\Gamma\left[7p^{-1}\right]\right)}. \tag{20}$$

Obviously, the magnitude of different methods relative efficiency depends solely on the EPD shape parameter p.
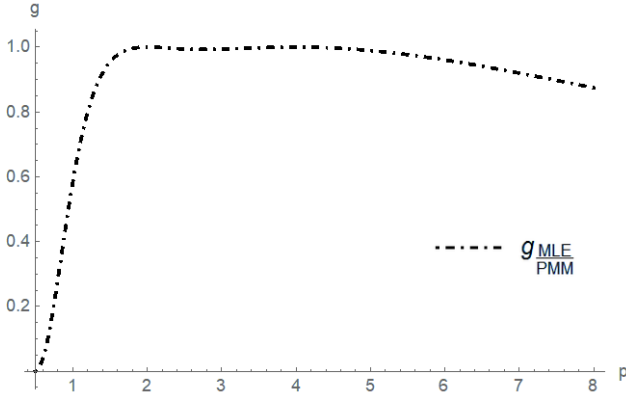
Fig. 2. Dependency of variance reduction coefficient $g_{MLE/PMM}$ with EPD shape parameter p.

Analyzing the dependencies shown in Fig. 1 and Fig. 2, it should be noted that for a sufficiently wide range of shape parameter values $p \in (2; 5)$ the theoretical efficiency of MLE and PMM estimates actually matches. For sharp-topped distributions at $p < 2$ the MLE efficiency can be essentially higher. Also at increase $p > 5$, i.e. at essentially flat-topped (in asymptotics uniform) distribution the related efficiency of PMM-estimations increases too.

At the end of this analysis it is necessary to emphasize once again that it was obtained the theoretical expressions of estimates variances ratio (19–20) and built on Fig. 1 and Fig. 2 corresponding graphs depending on the presence of a priori information about the probabilities of measurement errors, i.e. EPD parameter valuestype (1).

5　Statistical modeling

To experimentally verify the obtained theoretical results of the different evaluation methods effectiveness analysis, statistical modeling by the Monte Carlo Method was implemented. For the statistical modeling implementation programming language R is used. This choice was made due its free distribution, as well as the presence of a large number of libraries focused on the tasks of statistical data analysis. Among them is the software module *normalp* package, which contains a

set of functions for the generating statistical evaluation by maximum likelihood method and visualization of random EPD variables [17].

Empirical values of the coefficients of the parameter estimates variance were used as quantitative criteria characterizing the different methods efficiency:

$$\hat{g}_{(\theta)3} = \frac{\hat{\sigma}^2_{(\theta)PMM3}}{\hat{\sigma}^2_{(\theta)MEAN}}, \quad \hat{q}_{(\theta)3} = \frac{\hat{\sigma}^2_{(\theta)MLE}}{\hat{\sigma}^2_{(\theta)MEAN}},$$

$$\hat{r}_{(\theta)3} = \frac{\hat{\sigma}^2_{(\theta)PMM3}}{\hat{\sigma}^2_{(\theta)MLE}}, \tag{21}$$

where $\hat{\sigma}^2_{(\theta)MEAN}$, $\hat{\sigma}^2_{(\theta)MLE}$, $\hat{\sigma}^2_{(\theta)PMM3}$ – based on m experiments obtained using the arithmetic mean, MLE and PMM (at $r = 3$) estimates variance value, respectively.

The experiment was carried out depending on presence or absence of a priori information about measurement errors probability. For the second case, instead of a priori information posteriori model parameters estimates values (1) (required for the maximum likelihood method) and estimates of central moments of 2,4 and 6-th degrees (required for the polynomial maximization method) were used. It is obvious that the reliability of the simulation results was also significantly influenced by the amount of sample data n and the number of experiments m.

The set of results of statistical modeling for different values of the EPD shape parameter ($p = 1 \div 10$) and the sample values volumes ($n = 20 \div 200$), obtained by repeated $m = 10^4$ experiments are presented in Tab. 1 (with presence of a priori information) and Tab. 2 (with absence of a priori information).

The theoretical and experimental values analysis of the estimates variance coefficients ratio shows that there is a certain correlation between analytical calculations and results obtained by statistical modeling. The emergence of the difference (as noted earlier) is due to the fact that the expressions describing the variances of the MLE and PMM-estimates obtained for the asymptotic case (at $n \to \infty$). Tab. 1 data confirms that with increasing sample size, the difference between theoretical and experimental data decreases.

Tab. 1 Estimates variance coefficients with presence of a priori information

p	$\hat{g}^a_{(\theta)3}$				$g_{(\theta)3}$	$\hat{q}^a_{(\theta)3}$				$q_{(\theta)3}$	$\hat{r}^a_{(\theta)3}$				$r_{(\theta)3}$
	n					n					n				
	20	50	100	200		20	50	100	200		20	50	100	200	
10	0.89	0.42	0.41	0.41	0.4	0.45	0.35	0.32	0.32	0.31	2	1.2	1.26	1.27	1.28
5	0.69	0.64	0.61	0.62	0.61	0.68	0.64	0.62	0.62	0.61	1.01	1	1	1	1.02
3	0.9	0.9	0.89	0.89	0.89	0.9	0.9	0.89	0.89	0.88	1	1	1	1	1.01
2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	0.98	0.86	0.86	0.86	0.85	0.66	0.61	0.57	0.56	0.5	1.49	1.43	1.5	1.53	1.71

Tab. 2 Estimates variance coefficients with absence of a priori information

p	$\hat{g}_{(\theta)3}$				$g_{(\theta)3}$	$\hat{q}_{(\theta)3}$				$q_{(\theta)3}$	$\hat{r}_{(\theta)3}$				$r_{(\theta)3}$
	n					n					n				
	20	50	100	200		20	50	100	200		20	50	100	200	
10	0.65	0.50	0.44	0.42	0.4	0.91	0.57	0.49	0.48	0.31	0.71	0.88	0.91	0.88	1.28
5	0.82	0.71	0.67	0.63	0.61	1.09	0.85	0.80	0.74	0.61	0.74	0.88	0.83	0.86	1.02
3	1.02	0.97	0.93	0.91	0.89	1.26	1.11	0.99	0.94	0.88	0.81	0.87	0.93	0.97	1.01
2	1.13	1,07	1.05	1.02	1	1.23	1,14	1.07	1.03	1	0.91	0.93	0.98	0.97	1
1	0.96	0,83	0.80	0.83	0.85	0.77	0,63	0.58	1.03	0.5	1.24	1.32	1.34	0.99	1.71

The experimental results also confirm the previously stated thesis that the accuracy of obtaining MLE and PMM estimates is significantly affected by the factor of presence / absence of a priori information on the error model properties (exponential distribution parameters for MLE and even central moments up to 6-th degree for PMM). Tab. 2 data reflects the important fact that in the absence of a priori information on EPD properties for flat-top distributions ($p > 2$) the efficiency of PMM estimates is mostly higher than the MLE estimates. This can be explained by the fact that upon these conditions the influence of the uncertainty of the EPD shape parameter value on the MLE is more significant than the influence of the uncertainty of the central moments of the PMM. The relative decrease in the variance of the MLE estimates (compared to the PMM) depends on both the shape parameter and the sample values volume. And it is especially significant (up to 30%) with least samples and significant flatness.

Such results are especially important from a practical point of view, because the vast majority of real situations lacks a priori information about the true values of random component model parameters (measurement errors). Therefore, for such a situation Fig. 3 shows the boundaries delimiting the areas of greatest efficiency (based on the minimum variance criterion) on different estimation methods background: arithmetic mean, MLE and PMM. These areas were also obtained by statistical modeling via Monte Carlo method (for $m = 10^4$ experiments) at different values of EPD shape parameter p and sample size n.

Based on the totality of the above results the following conclusions can be drawn:

- in case of the normal (Gaussian) errors distribution, the most effective (both in terms of accuracy and ease of implementation) is the use of arithmetic mean estimates;

- for sharp-topped distributions, the maximum likelihood method is generally more effective, the application of which requires large amount of sample data (the boundary separating the efficiency areas depends significantly on the sample values volume and is parabolic);

- for flat-topped distributions, the polynomial maximization method is more effective (the boundary separating the areas, although to a lesser extent, also depends on the sample values volume).
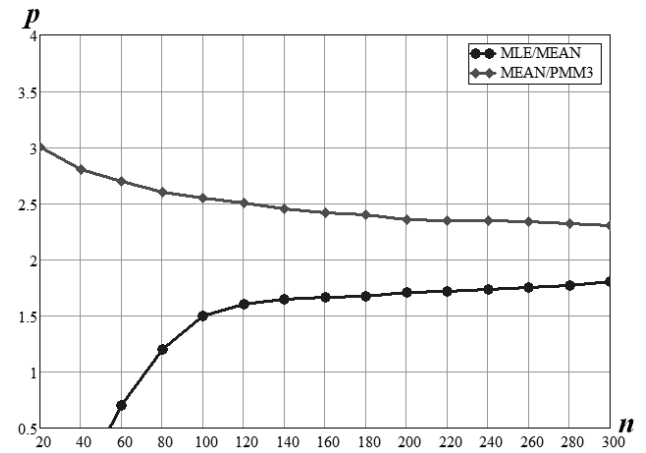

Fig. 3. Efficiency areas of methods for finding the center coordinates of the symmetric exponential distributions family

Conclusion

The obtained results are the second part that completes the research initiated in [19] on the validity of applying the polynomial maximization method for finding the experimental data parameters estimates with exponential power distribution.

The set of obtained results allows to position the proposed approach as a compromise in terms of the balance between simple estimates (mean, median and mid-range) and potentially more accurate estimates of maximum likelihood, the use of which requires additional a priori information. In this context, the advantages of using the polynomial maximization method are:

- providing additional opportunities (compared to simple non-parametric methods) to increase the informative parameter estimates accuracy by taking into account the measurement errors properties (a posteriori values of the central moments);

- reducing the impact (compared to the parametric maximum likelihood method) of the lack of a priori information on the results of the informative parameter estimates accuracy.

The obtained efficiency areas of different methods allow us to recommend their choice depending on the probabilistic error model properties (the value of the EPD shape parameter) and the experimental data sample values amount.

References

[1] Bezuglov D.A., Andrjushhenko I.V., Shvidchenko S.A. Informacionnaja tehnologija identifikacii zakona raspredelenija na baze kumuljantnogo metoda analiza rezul'tatov izmerenij [Information technology for identification of the distribution law based on the cumulant method of results measurement analysis]. *Informacionnye sistemy i tehnologii. Teorija i praktika: cbornik nauchnyh trudov* [Information systems and technologies. Theory and Practice: Collection of Scientific Papers], Shahty, 2011, pp. 186-194.

[2] Krasilnikov A. (2019) Family of Subbotin Distributions and its Classification. *Èlektronnoe modelirovanie*, Vol. 41, Iss. 3, pp. 15-32. DOI: 10.15407/emodel.41.03.015

[3] Novitskii P. V., Zograf I. A. *Otsenka pogreshnostei rezul'tatov izmerenii* [Estimation of errors of measurement results]. Moscow, Energoatomizdat Publ., 1991, 304 p.

[4] Subbotin M. T. (1923) On the law of frequency of error. *Mat. Sb.*, vol. 31, no. 2, pp. 296-301.

[5] Gui W. (2013) Statistical Inferences and Applications of the Half Exponential Power Distribution. *Journal of Quality and Reliability Engineering*, pp. 1-9. DOI: 10.1155/2013/219473

[6] Hassan M. Y., Hijazi R. H. (2010) A bimodal exponential power distribution. *Pak. J. Statist*, vol. 26, no. 2, pp. 379–396.

[7] Maugey T., Gauthier J., Pesquet-Popescu B. and Guillemot C. (2010) Using an exponential power model forwyner ziv video coding. *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 2338-2341. DOI: 10.1109/icassp.2010.5496065

[8] Saatci E. and Akan A. (2010) Respiratory parameter estimation in non-invasive ventilation based on generalized Gaussian noise models. *Signal Processing*, Vol. 90, Iss. 2, pp. 480-489. DOI: 10.1016/j.sigpro.2009.07.015

[9] Chan J., Choy B., Walker S. On the Estimation of the Shape Parameter of a Symmetric Distribution. *Journal of Data Science*, 2020, vol. 18, no. 1, pp. 78-100.doi: 10.6339/JDS.202001_18(1).0004

[10] Komunjer I. (2007) Asymmetric power distribution: Theory and applications to risk measurement. *Journal of Applied Econometrics*, Vol. 22, Iss. 5, pp. 891-921. DOI: 10.1002/jae.961

[11] Olosunde A. A., Soyinka A. T. (2019) Interval Estimation for Symmetric and Asymmetric Exponential Power Distribution Parameters. *Journal of the Iranian Statistical Society*, Vol. 18, Iss. 1, pp. 237-252. DOI: 10.29252/jirss.18.1.237

[12] Dominguez-Molina J. A., González-Farías G., Rodríguez-Dagnino R. M. (2003) A practical procedure to estimate the shape parameter in the generalized Gaussian distribution. *ResearchGate*.

[13] Olosunde A. A. (2013) On Exponential Power Distribution and Poultry Feeds Data: a Case Study. *Journal of The Iranian Statistical Society*, vol. 12, no. 2, pp. 253-270.

[14] Duda J. (2020) Adaptive exponential power distribution with moving estimator for nonstationary time series. *ResearchGate*, 6 p. doi: 10.13140/RG.2.2.26797.64483

[15] Zakharov I. P., Shtefan N. V. (2002) Opredelen-iej effektivnyh ocenok centra raspredelenija pri statisticheskoj obrabotke rezul'tatov nabljudenij [Definition of effective distribution center value at statistical processing of measurement observations]. *Radioelektronika ta informatyka - Radioelectronics and informatics*, vol. 3, no. 20, pp. 97-99.

[16] Warsza Z. L., Galovska M. (2009) About the best measurand estimators of trapezoidal probability distributions. *Przeglad Elektrotechniczny*, vol. 85, no. 5, pp.86–91.

[17] Mineo A. M., Ruggieri M. (2005) A Software Tool for the Exponential Power Distribution: The normalp Package. *Journal of Statistical Software*, vol. 12, iss. 4, pp. 1-24. doi: 10.18637/jss.v012.i04

[18] Kunchenko Yu. P. (2002) *Polynomial parameter estimations of close to Gaussian random variables*. Aachen, Shaker Verlag, 396 p.

[19] Zabolotnii S. V., Chepynoha A. V., Bondarenko Y. Y. and Rud M. P. (2018) Polynomial parameter estimation of exponential power distribution data. *Visnyk NTUU KPI Seriia - Radiotekhnika Radioaparatobuduvannia*, Iss. 75, pp. 40-47. DOI: 10.20535/radap.2018.75.40-47

[20] Warsza Z. L. and Zabolotnii S. W. (2017) A Polynomial Estimation of Measurand Parameters for Samples of Non-Gaussian Symmetrically Distributed Data. *Automation 2017*, pp. 468-480. DOI: 10.1007/978-3-319-54042-9_45

[21] Warsza Z. L., Zabolotnii S. W. (2017) Uncertainty of measuring data with trapeze distribution evaluated by the polynomial maximization method. *Przemysl Chemiczny*, no. 12, pp. 68–71. doi: 10.15199/62.2017.12.6

[22] Zabolotnii S. V., Kucheruk V. Yu., Warsza Z. L. (2018) Polynomial Estimates of Measurand Parameters for Data from Bimodal Mixtures of Exponential Distributions. *Bulletin of the Karaganda University. Physics Series.*, vol. 2, no. 90, pp. 71-80.

Порівняльний аналіз оцінок методів максимізації поліному та максимальної правдоподібності для даних з експоненційним степеневим розподілом

Заболотній С. В., Чепинога А. В., Чорній А. М., Гончаров А. В.

Робота присвячена порівняльному аналізу точності оцінок параметрів експериментальних даних із експоненціальним степеневим розподілом (ЕСР), що знаходяться із застосуванням класичного методу максимальної правдоподібності (ММП) і оригінального методу максимізації поліномів (ММПл). На відміну від параметричного підходу ММП, що використовує опис у вигляді щільності розподілу ймовірностей, ММПл базується на частковому описі у вигляді статистик вищих порядків і математичному апараті стохастичних поліномів Кунченка.

Наведено алгоритм для знаходження ММПл-оцінок із застосуванням стохастичних поліномів 3-го порядку. Отримані аналітичні вирази, що дозволяють визначати дисперсію ММПл-оцінок параметрів для асимптотичного випадку та при наявності апріорної інформації

про параметри ЕСР. Показано, що відносна теоретична точність оцінок різних методів суттєво залежить від параметра форми ЕСР і співпадає лише для окремого випадку гаусового розподілу.

Шляхом багаторазових статистичних випробувань (методом Монте-Карло) досліджено ефективність (за критерієм величини дисперсій оцінок) різних підходів (у тому числі оцінок середнього значення) при наявності та відсутності апріорної інформації щодо властивостей ЕСР. Побудовано області найбільшої ефективності для кожного із методів в залежності від параметра форми ЕСР і обсягу вибіркових даних.

Ключові слова: експоненційний степеневий розподіл; стохастичні поліноми; статистики вищих порядків; оцінка параметра

Сравнительный анализ оценок методов максимизации полинома и максимального правдоподобия для данных с экспоненциальным степенным распределением

Заболотний С. В., Чепинога А. В., Чорний А. М., Гончаров А. В.

Работа посвящена сравнительному анализу точности оценок параметров экспериментальных данных с экспоненциальным степенным распределением (ЭСР), которые находятся с применением классического метода максимального правдоподобия (ММП) и оригинального метода максимизации полинома (ММПл). В отличие от параметрического подхода ММП, который использует описание в виде плотности распределения вероятностей, ММПл базируется на частичном описании в виде статистик высших порядков и математическом аппарате стохастических полиномов Кунченко.

Приведен алгоритм для нахождения ММПл-оценок с применением стохастических полиномов 3-го порядка. Получены аналитические выражения, позволяющие определять дисперсию ММПл-оценок параметров для асимптотического случая и при наличии априорной информации о параметрах ЭСР. Показано, что относительная теоретическая точность оценок различных методов существенно зависит от параметра формы ЭСР и совпадает только для частного случая гауссова распределения.

Путем многократных статистических испытаний (методом Монте-Карло) исследована эффективность (по критерию величины дисперсий оценок) различных подходов (в том числе оценок среднего значения) при наличии и отсутствии априорной информации о свойствах ЭСР. Построены области наибольшей эффективности для каждого из методов в зависимости от параметра формы ЭСР и объема выборочных данных.

Ключевые слова: экспоненциальное степенное распределение; стохастические полиномы; статистики высших порядков; оценка параметра