UDC 519.2:681.2

S.V. Zabolotnii[1], V.Yu. Kucheruk[2], Z.L. Warsza[3], A. K. Khassenov[4]

[1]Cherkasy State Technological University, Ukraine;
[2]Vinnytsia National Technical University, Ukraine;
[3]Industrial Research Institute for Automation and Measurements PIAP, Poland;
[4]Ye.A. Buketov Karaganda State University, Kazakhstan
(E-mail: s.zabolotnii@chdtu.edu.ua)

Polynomial estimates of measurand parameters for data from bimodal mixtures of exponential distributions

A non-conventional approach to finding estimates of the result of multiple measurements for a random error model in the form of bimodal mixtures of exponential distributions is proposed. This approach is based on the application of the Polynomial Maximization Method (PMM) with the description of random variables by higher order statistics (moment & cumulant). The analytical expressions for finding estimates and analysis accuracy to the degree of the polynomial $r = 3$ are presented. In case when the degree of the polynomial $r = 1$ and $r = 2$ (for symmetrically distributed data) polynomial estimate equivalent can be estimated as a mean (average arithmetic). In case when the degree of the polynomial $r = 3$, the uncertainty of the polynomial estimate decreases. The reduction coefficient depends on the values of the 4th and 6th order cumulant coefficients that characterize the degree of difference while the distribution of sample data from the Gaussian model. By means of multiple statistical tests (Monte Carlo method), the properties of the normalization of polynomial estimates are investigated and a comparative analysis of their accuracy with known estimates (mean, median and center of folds) is made. Areas that depend on the depth of antimodality and sample size, in which polynomial estimates (for r = 3) are the most effective.

Keywords: bimodal distribution, measured parameter, variance of estimates, moments, cumulants, stochastic polynomial.

Introduction

In statistical processing of multiple measurement results in the presence of random errors, an important task is to obtain estimates having the smallest variance (uncertainty). Since the distribution of the errors in the measurement results is, as a rule, symmetric, the value of the estimated parameter is determined by the center of their symmetry. In other words, shift of the distribution center from the zero value (in the absence of a systematic error) determines the estimate of the measured parameter value. Although according to [1] reference value for determining the multiple measurements results is the arithmetic mean, but there are other ways of determining this parameter. For example, if we use the principle of symmetry of the probabilitydistribution function (pdf), then the obvious estimate of distribution center is its median, which is a more efficient estimate for single-modal extreme distributions (for example, the Laplace distribution).Whereas for the limited range distributions (arcsine, uniform), the most preferred is the estimate of mid-range [2].

Thus, the choice of the optimal estimation method (by the criterion of a minimum variance) depends on the type of the measurement errors distribution and requires a preliminary justification of their probabilistic model. When identifying such a model, it is recommended to consider a wider set of distribution laws, including models in the form of pdf mixtures. In this case, for any empirical distribution it is possible to construct an adequate, statistically more reasonable mathematical model [3].

Models based on mixture distributions are currently used in a variety of fields: engineering, geology, biology, medicine, economics, sociology, etc. There are various features of such models that allow describing a variety of specific real data properties, such as asymmetry, kurtosis, heterogeneity and multimodality. The importance of mixture distribution can be noted from a large number of books, for example [4-6], and from specialized publications [7, 8].

Model of bimodal symmetric mixtures of exponential distributions

The most commonly used model for bimodal data is the two-component mixture of exponential distributions belonging to the Exponential Power Family (EPF), which is considered one of the most important probability distributions in statistics [9]. The symmetric exponential power distribution with Gauss, Laplace and rectangular were mentioned for a long time in the works of many authors, for example [10-15]. In metrological problems, the two-mode distribution quite often occurs the appearance of errors in some classes of high-precision digital voltmeters, temperature errors of instruments operating in the open airand mechanical hysteresis errors of sensor elastic elements [2].

In this paper, one of the varieties of EPF is used in the form of a bimodal symmetric distribution based on exponential mixtures of the form

$$p(x) = \frac{T(\beta)}{2}\left(e^{-|x-m|^{\beta}} + e^{-|x+m|^{\beta}}\right), \tag{1}$$

where $T(\beta) = \dfrac{\beta}{2\Gamma(1/\beta)}$ — normalizing factor, depending on the exponent β, and the parameter m specifies the location of the modes on the x axis value.

The value of the parameter β significantly affects the shape of the peak (see Fig. 1). For example, when $\beta = 1$ the exponential distribution corresponds to the Laplacian peak distribution with gentle slopes, when $\beta = 2$ — the normal (Gaussian) distribution, and when $\beta \to \infty$ it is transformed into a uniform (rectangular) distribution.

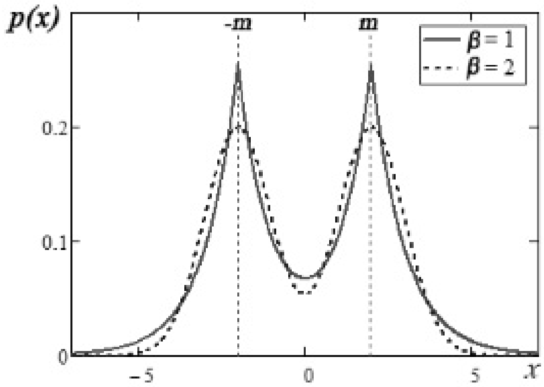


Figure 1. Bimodal symmetric distribution based on mixtures of the Laplace distribution (β=1) and the Gaussian distribution (β=2)

Model (1) is interpreted as a composition of a discrete double-digit distribution (distribution of a binary alternative at points $\pm m$) and exponential distribution with exponent β. In addition to β an important role in the properties of the model (1) is exerted by the value of the anti-modality depth parameter describing the relative content of the discrete component in the composition: $C = m/\sigma_e$, where σ_e - root-mean-square deviation of the exponential component. The value of this parameter for real distributions of measurement errors is in range from 0 to 2. It was also shown in [2; 147] that comparing classical statistics (mean and median), a more effective quantification of the coordinate of the center of bimodal distributions can be a quantile estimate in the form of a so-called center of folds, defined as half of the sum 25 % and 75 % of the quantiles corresponding to the mode vertices.

Purpose of the study

One of the alternative approaches to statistical estimation is the Polynomial Maximization Method (PMM) proposed by Kunchenko [16]. This unconventional method for finding parameter estimates is based on the apparatus of the maximization of stochastic polynomials and uses the description of random variables in the form of higher order statistics (moments and/or cumulants). In [17], a comparative analysis of the efficiency of PMM-estimations and estimates in the form of an arithmetic mean constant component for various types (arcsine, uniform, trapezoidal, triangular) non-Gaussian symmetrically distributed quantities was performed, and the properties of empirical distribution of estimates depending on the sample size were highlighted. In [18], using the trapezoidal distribution model as an example, the efficiency domains of PMM-estimations were studied in more detail in comparison with estimates based on the mean and mid-range. The obtained results show that for certain values of the distribution parameters PMM estimates can be significantly more efficient (to have a smaller variance) than classical estimates. The purpose of this paper is to study the properties and the comparative analysis of the efficiency of the a posteriori version (in the absence of a priori information on the statistical parameters of the measurement model) of PMM estimates of the coordinate of the center of symmetric bimodal distributions formed on the basis of mixtures of Gaussian and Laplace distributions.

Formulation of the problem

Let θ – can be an estimated parameter, whose value is determined on the basis of a statistical analysis of a vector $\vec{x} = \{x_1, x_2, ...x_n\}$, consisting of n independent identically distributed sampled values described by the model $\xi = \theta + \xi_0$, where ξ_0 – centered symmetrical two-modal random variable (measurement error), described by the mixture of exponential distributions.

It is necessary to investigate the properties of PMM-estimations (changing the accuracy and normalizing the empirical distribution of estimates as a function of the sample size by means of Monte Carlo Method), and also carry out a comparative analysis of the accuracy (by the variance criterion) with the classical estimates.

The determination of estimates by the Polynomial Maximization Method and properties

1. Theoretical Foundations of PMMs

Conceptually, the Polynomial Maximization Method is close to the Maximum Likelihood Method (MLM). The basic analogy is that both methods are based on the use of a certain statistical functional from sample values x_v, $v = \overrightarrow{1,n}$, which should have a maximum in the neighborhood of the true value of the estimated parameter θ. The principal difference is that the use of MLM requires a complete description of the random variables in the form of the density of probability distribution necessary for the formation of maximum likelihood statistics. The use of PMM is based on the representation of the extremum functional in the form of a stochastic polynomial $L_{sn}(\vec{x}/\theta)$ with order r:

$$L_{rn}(\vec{x}/\theta) = nk_0(\theta) + \sum_{i=1}^{r} k_i(\theta) \sum_{v=1}^{n} f_i(x_v), \tag{2}$$

$$k_0(\theta) = \int_{-\infty}^{\theta} \sum_{i=1}^{r} h_i(\theta) \Psi_i(\theta) d\theta, \; k_i(\theta) = \int_{-\infty}^{\theta} h_i(\theta) d\theta,$$

where $f_i(x)$, $i = \overrightarrow{1,r}$ — a set of ordered basic functions in a certain way, and $\Psi_i(\theta) = E\{f_i(x)\}$ — their mathematical expectations.

Thus, to find the PMM parameter estimates θ sufficiently partial description in the form of a parametric family of a sequence of mathematical expectations $\Psi_i(\theta)$.

If we use integral power-law transformations as basis functions, i.e. $f_i(x) = x^i$, then their mathematical expectations will be the initial moments $\alpha_i(\theta) = E\{x^i\}$ of the random variable ξ. Then the estimate of a parameter θ can be found from the solution of estimated parameter of stochastic power equation:

$$\frac{d}{d\theta}L_{rn}(\vec{x}/\theta)\Big|_{\theta=\hat{\theta}} = \sum_{i=1}^{r} h_i(\theta)\sum_{v=1}^{n}\left[x_v^i - \alpha_i(\theta)\right]\Big|_{\theta=\hat{\theta}} = 0. \tag{3}$$

Coefficients $h_i(\theta)$ (for $i=\overrightarrow{1,r}$) can be found by solving the system of linear algebraic equations, given by conditions of minimization of variance (with the appropriate order r) of the estimate of the parameter θ , specifically:

$$\sum_{i=1}^{r} h_i(\theta)\, F_{i,j}(\theta) = \frac{d}{d\theta}\alpha_j(\theta),\ j=\overrightarrow{1,r}, \tag{4}$$

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

In [16] it was shown that polynomial evaluations $\hat{\theta}$, which are the solutions of stochastic equations of the form (4), are consistent and asymptotically unbiased. To calculate the evaluation of uncertainty is necessary to find the volume of extracted information on the estimated parameters θ , which generally are described by the equation:

$$J_{rn(\theta)} = n\sum_{i=1}^{r} h_i(\theta)\frac{d}{d\theta}\alpha_i(\theta). \tag{5}$$

The statistical sense of function $J_{rn(\theta)}$ is similar to the classical Fisher concept of information quantity, as if $n\to\infty$ its inverse approaches to the variance of estimates:

$$\sigma^2_{(\theta)r} = \lim_{n\to\infty} J^{-1}_{rn(\theta)} . \tag{6}$$

2. *PMM-estimates of the coordinate of symmetric distributions center*

It was shown [16] that PMM is an estimate of the value of the constant component (the shift parameter, the coordinates of the distribution centers) for the polynomial degree $r=1$ is an equivalent estimate of the mean for an arbitrary distribution law of random variables. In this paper it is also shown that, with the symmetry of the distribution characterized by the equality of the cumulant coefficients of the unpaired order to zero, estimates founded when using degree polynomials $r=2$, also degenerate into linear estimates.

When using polynomial degree $r=3$ algorithm for finding PMM parameter estimates θ for the case of symmetrically distributed quantities, reduce to the necessity of solving a stochastic equation of the form:

$$h_1\sum_{v=1}^{n}(x_v-\theta)+h_2\sum_{v=1}^{n}\left[x_v^2-\left(\theta^2+\mu_2\right)\right]+h_3\sum_{v=1}^{n}\left[x_v^3-\left(\theta^3+3\theta\mu_2\right)\right]\Big|_{\theta=\hat{\theta}}=0, \tag{7}$$

where μ_2 — centralmomentof the 2-d order of the random variable ξ_0.

Weighting coefficients h_1 - h_3 (their dependence on θ is omitted for simplicity of writing), minimizing the variance of the estimated parameter, are found from the solution of a system of linear algebraic equations of the form (4) and can be described by expressions:

$$h_1=\frac{1}{\Delta_3}\left[3\theta^2\left(\mu_4-3\mu_2^2\right)+3\mu_4\mu_2-\mu_6\right],\ h_2=\frac{-3\theta}{\Delta_3}\left[\mu_4-3\mu_2\right],\ h_3=\frac{1}{\Delta_3}\left[\mu_4-3\mu_2\right], \tag{8}$$

where $\Delta_3=\kappa_2^{-2}\left(\mu_4^2-\mu_2\mu_6\right)$.

Substituting the coefficients (8) in (3), we obtain a cubic equation with respect to the parameter being estimated:

$$a\theta^3+b\theta^2+c\theta+d\,\Big|_{\theta=\hat{\theta}}=0, \tag{9}$$

where $a=1$, $b=-3\hat{\alpha}_1$, $c=3\hat{\alpha}_2-\frac{\mu_6-3\mu_4\mu_2}{\mu_4-3\mu_2^2}$, $d=\hat{\alpha}_1\frac{\mu_6-3\mu_4\mu_2}{\mu_4-3\mu_2^2}-\hat{\alpha}_3$.

Note that in (9) the statistics $\hat{\alpha}_i=\frac{1}{n}\sum_{v=1}^{n}x_v^i$, $i=\overrightarrow{1,3}$, are sample initial moments, and μ_2, μ_4 and μ_6 — theoretical central moments of the random variable ξ_0.

The required solution of the stochastic equation (9) can be analytically found by the Cardano formulas [17] or numerically on the basis of iterative procedures, for example Newton-Raphson.

3. Accuracy of PMM-estimates of the coordinate of symmetric distributions center

It is known [2] that the variance $\sigma^2_{(\theta)mean}$ of the mean estimate of does not depend on the value of the estimated parameter, but is determined only by the variance (centralmomentsof the 2d order) μ_2 and sample size n. And since PMM-estimation of the parameter θ if $r=1$ is equivalent to the estimate of the mean, then their variances coincide, i.e.;

$$\sigma^2_{(\theta)1} = \sigma^2_{(\theta)mean} = \frac{\mu_2}{n}. \tag{10}$$

Using the relation (5) and (6) describing the amount of extracted information about the estimated parameter θ, an analytical expression is obtained, the theoretical determining value of the asymptotic (for $n \to \infty$) variances $\sigma^2_{(\theta)3}$ PMM-estimates for $r=3$:

$$\sigma^2_{(\theta)3} = \frac{\mu_2\mu_6 - \mu_4^2}{n\mu_2\left(9\mu_2^3 - 6\mu_2\mu_4 + \mu_6\right)} = \frac{\mu_2}{n}\left[1 - \frac{\gamma_4^2}{6 + 9\gamma_4 + \gamma_6}\right], \tag{11}$$

where $\gamma_4 = \frac{\mu_4}{\mu_2^2} - 3$, $\gamma_6 = \frac{\mu_6}{\mu_2^3} - 15\frac{\mu_4}{\mu_2^2} + 30$ — dimensionless cumulant coefficients of the 4th and 6th orders [19].

Obviously, the value of variance ratio:

$$g_{(\theta)3} = \frac{\sigma^2_{(\theta)3}}{\sigma^2_{(\theta)1}} = 1 - \frac{\gamma_4^2}{6 + 9\gamma_4 + \gamma_6}, \tag{12}$$

whose values belong to the interval $(0;1]$, depends only on the probability properties of the distribution of measurement error, determined by cumulant coefficients γ_4 and γ_6 [17].

Statistical modeling

Based on the results obtained, the set program (for MATLAB/OCTAVE), firstly described in [17], was modernized. This set of *m*-scripts and *m*-functions allows for a comparative analysis of the accuracy of various algorithms for statistical estimation, as well as to investigate the properties of the PMM-estimates.

According to [2], among the set of classical statistics used to estimate the coordinate of the center of symmetric distributionseffective (for the considered class)can be: mean $\bar{x}$, median (50% quantile) x_{med} and the center of folds x_{cf} (half the sum of 25% and 75% of the quantiles). Thus, as the comparative efficiency criteria, we use the experimental values of the variance ratio coefficients:

$$\hat{g}_{(\theta)3} = \frac{\hat{\sigma}^2_{(\theta)3}}{\hat{\sigma}^2_{(\theta)1}} = \frac{\hat{\sigma}^2_{(\theta)3}}{\hat{\sigma}^2_{(\theta)mean}}, \quad \hat{q}_{(\theta)3} = \frac{\hat{\sigma}^2_{(\theta)3}}{\hat{\sigma}^2_{(\theta)med}}, \quad \hat{r}_{(\theta)3} = \frac{\hat{\sigma}^2_{(\theta)3}}{\hat{\sigma}^2_{(\theta)cf}}, \tag{13}$$

where $\hat{\sigma}^2_{(\theta)mean}, \hat{\sigma}^2_{(\theta)med}, \hat{\sigma}^2_{(\theta)cf}, \hat{\sigma}^2_{(\theta)3}$ - the variances of estimates θ averaged over M experiments, which are calculated on the basis of statistics of the mean, median, center of folds and PMM at $r=3$ accordingly.

Two factors influenced the validity of Monte Carlo simulation results of the statistical estimation algorithms: the total sample size n of input vector $\vec{x}$, containing the values of the estimated parameter, and the number of experiments M, conducted under the same initial conditions (the values of the model parameters that determine the probabilistic nature of the two-component Gaussian distribution).

Note that for calculating PMM-estimates, information is used not about the type of distribution, but about the values of parameters of the model (central moments or cumulants) of the measurement data distribution. In [17], these values were calculated on the basis of analytical expressions connecting the parameters of the distribution densities and their moments. However, for practical situations where information on the distribution density and the values of their parameters is a priori unavailable, the adaptive approach in this study can be used. Its essence consists in using a posteriori estimates:

$$\hat{\mu}_i = \frac{1}{n}\sum_{v=1}^{n}\left(x_v - \bar{x}\right)^i, \tag{14}$$

where $\bar{x} = \hat{\alpha}_1 = \frac{1}{n}\sum_{v=1}^{n} x_v$.

The simulation results obtained on the basis of the Monte Carlo method are presented in Table 1 and Table 2.

Table 1

The coefficients of the variance ratio of estimates for the model on the basis of Laplace distribution mixture (C_1 -anti-modality depth parameter)

C_1	$g_{(\theta)3}$	Simulation results								
		$\hat{g}_{(\theta)3}$			$\hat{q}_{(\theta)3}$			$\hat{r}_{(\theta)3}$		
		n								
		20	50	200	20	50	200	20	50	200
0.6	0.93	1.05	0.97	0.93	0.72	0.57	0.48	1.21	1.23	1.26
0.5	0.91	1.04	0.94	0.91	0.83	0.65	0.55	1.19	1.1	1.11
0.4	0.89	1.03	0.9	0.89	0.95	0.76	0.68	1.12	1.01	1.02
0.3	0.88	0.99	0.89	0.88	1.11	0.9	0.81	1.08	0.95	0.93
0.2	0.86	0.97	0.87	0.86	1.25	1.05	1	1.04	0.86	0.86

Table 2

The coefficients of the variance ratio of estimates for the model on the basis of Gaussian distribution mixture (C_2 -anti-modality depth parameter)

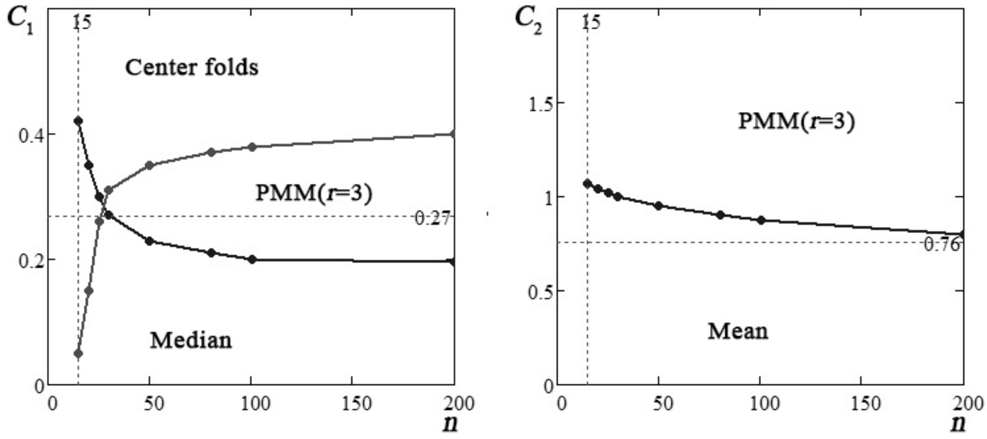
C_2	$g_{(\theta)3}$	Simulation results								
		$\hat{g}_{(\theta)3}$			$\hat{q}_{(\theta)3}$			$\hat{r}_{(\theta)3}$		
		n								
		20	50	200	20	50	200	20	50	200
2	0.39	0.54	0.42	0.39	0.13	0.06	0.04	0.63	0.62	0.62
1.5	0.66	0.78	0.71	0.67	0.27	0.2	0.16	0.74	0.71	0.71
1	0.93	1.04	0.97	0.95	0.56	0.47	0.46	0.88	0.76	0.75
0.5	0.99	1.11	1.07	1.02	0.73	0.68	0.64	0.98	0.86	0.83

Analysis of theoretical values $g_{(\theta)3}$ and experimental values $\hat{g}_{(\theta)3}$ of the ratio of variances in Table 1 and Table 2 indicates a significant correlation between the analytical calculations and the results of statistical modeling. Obviously, with an increase in volume n of the original sample $\vec{x}$ the divergence decreases (for example, if $n = 20$ the maximum discrepancy is up to 25 %, with $n = 50$ less than 8 %, and if $n = 200$ no more than 2 %). In general, these results support the asymptotic property (5), which is characteristic for the amount of information retrieved about the estimated parameter, which is used in calculating the variances of PMM estimates.

The general analysis of all results of statistical modeling presented in Table 1 and Table 2 confirms a significant dependence of the effectiveness of the application of a given statistic for finding estimates of the shift parameter from the parameters of the model distribution and the volume of the initial sample.

Figure 2 shows the areas that visualize the effectiveness of the application (by the criterion of the minimum variance) of various statistics, depending on the sample volume $n = 15...200$. These areas are obtained from the results of multiple tests with $M = 10^4$ for different values of the proportionality parameters of models of two-modal distributions.

Figure 2a shows the most effective areas of the methods for estimating the coordinates of the center of bimodal mixtures on the basis of the Laplace distribution. the threshold value of the parameter $C_1 \approx 0.27$ divides the areas of effectiveness of applying two classical assessments: the median and the center of folds. Between them is a relatively small area, which starts at $n \approx 20$ and gradually expanding (with $n \to \infty$) occupies a gap of values C_1 from 0.2 to 0.4. Inside this area, PMM estimates are the most accurate. Their variance, although not significantly (up to 10%), is the smallest relative to the variances of classical estimates.



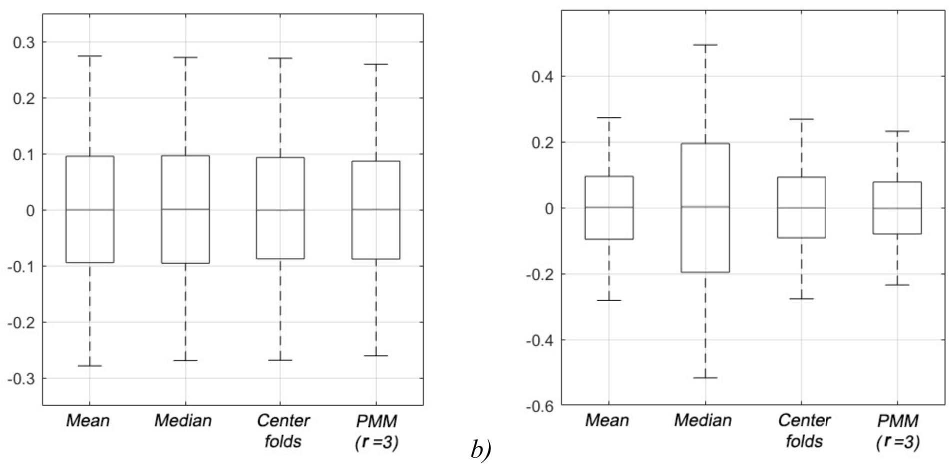
a) b)

a) on the basis of the Laplace distribution; b) on the basis of Gaussian distributions

Figure 2. Areas of effectiveness of few methods for finding estimates of coordinates of the center of two-modal mixtures

Figure 2b shows the areas that differentiate the efficiency of the application of the most accurate (mean and PMM at $r = 3$) estimates of the coordinates of the center of the two-modal mixtures based on the Gaussian distribution. For this model, the boundary for applying PMM estimates for large sample sizes (for $n \to \infty$) starts with the values of the parameter $C_2 \approx 0.76$. We note that with growth of C_2 the relative effectiveness of the application of PMM estimates significantly increases and the decrease 40-60 % of the variance can be achieved.

Examples of empirical distributions of different types of estimates obtained as a result of statistical modeling (for $M = 10^4$ experiments with a sample size $n = 50$), are shown in Figure 3. On these graphs, the main part of the boxplot contains 50% of the estimated values, and the upper and lower bounds are 2.5% and 97.5% percentiles. The results presented in Figs. 3, 4 generally correlate with the results of Tables 1 and 2.



a) *b)*

a) on the basis of the Laplace distribution ($C_1 = 0.3$); b) on the basis of Gaussian distributions ($C_2 = 1.5$)

Figure 3. Boxing-plots are empirical distribution of estimates of the coordinates of the center of two-modal mixtures

Another important result of statistical modeling is the confirmation of the theoretical assumption about the asymptotic (with growth n) normalization of the distribution of PMM-estimators [16]. In Figure 4, as one of the simulation results, an example is shown showing the approximation by a Gaussian distribution $M = 10^4$ experimental values of PMM-estimates of the coordinate of the center (for $\theta = 0$) a two-modal mixture based on Gaussian components (for $C_2 = 1.5$).

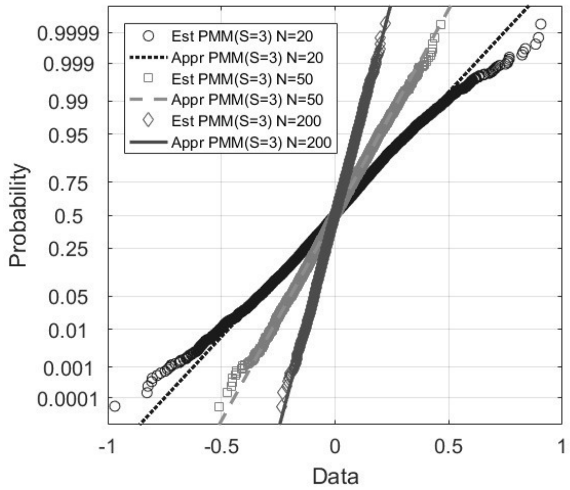


Figure 4. Probabilistic graph (Q-Q plot) of Gaussian approximation of the empirical distribution of PMM estimates

These probability charts, constructed for different volumes of the original sample n = 20, 50, 200 indicate a gradual approximation of the empirical distribution of PMM estimates to the Gaussian distribution.

The adequacy test of the hypothesis on the Gaussian distribution of PMM-estimators was also investigated using the Lillieforce test, based on the Kolmogorov-Smirnov statistics [20]. The obtained results, on the whole, correlate with the results similar to those in [17, 18]. They testify that in overwhelming majority of cases, even with the volume of sample values $n > 50$ hypothesis of a Gaussian distribution of PMM-estimates is not refuted at a significance level (p-value) equal to 0.05.

Conclusions

The presented investigations make it possible to draw a general conclusion about the possibility of applying the polynomial maximization method PMM to find estimates of center coordinate of symmetric bimodal mixtures of exponential distributions.

An analysis of the aggregate of the results of statistical modeling shows that the relative effectiveness of PMM estimates essentially depends on the probabilistic nature of the exponential components and the depth of antimodality. Thus, for models based on the Laplace distribution, the PMM estimates have better accuracy (decrease in variance) in comparison with the classical estimates, which does not exceed 10% and is observed only in a rather narrow range of antimodality depths. For mixtures based on a Gaussian distribution, with an increase in the antimodality depth $C > 1$ the relative accuracy of PMM estimates increases and leads to a significant (more than 2-time) decrease in the variance compared with known estimates.

The obtained analytical expressions describing the accuracy properties of PMM-estimates, under the conditions of normalization associated with sufficient volume of samples $n > 50$, allow to calculate the expanded uncertainty of measurement results and build confidence intervals for parameter estimates.

References

1 BIPM, I., IFCC, I., ISO, I., & IUPAP, O. The International Vocabulary of Metrology, Basic and General Concepts and Associated Terms—VIM, JCGM 200: 2008 [ISO/IEC Guide 99].

2 Новицкий П.В. Оценка погрешностей результатов измерений / П.В. Новицкий, И.А. Зограф. — Л.: Энергоатомиздат, 1991.

3 Лемешко Б.Ю. О задаче идентификации закона распределения случайной составляющей погрешности измерений / Б.Ю. Лемешко // Метрология. — 2004. — 7. — С. 8–17.

4 Lindsay B.G. Mixture models: theory geometry and applications, vol 51. In: NSF-CBMS regional conference series in probability and statistics / B.G. Lindsay // Institute of Mathematical Statistics. — Hayward, 1995.

5 McLachlan G.J. Finite mixture models / G.J. McLachlan, D. Peel. — New York: Wiley, — 2000.

6 Mengersen, K. Mixtures: estimation and applications / K.Mengersen, C.P. Robert, D.M. Titterington. — New York: Wiley, 2011.

7 Böhning D. Editorial: Advances in mixture models / D.Böhning, W.Seidel, M.Alfó, B.Garel, V.Patilea, G.Walther // Comput Stat Data Anal. — 2007. — 51:5205–5210.

8 Böhning, D. Editorial: The 2nd special issue on advances in mixture models / D.Böhning, C.Hennig, G.J.McLachlan, P.D.McNicholas // Comput Stat Data Anal. — 2014. — 71:1–2.

9 Hassan M.Y., Hijazi R.H. Pak. j. statist. — 2010. — Vol. 26(2). — 379–396 a bimodal exponential power distribution. Pak. J. Statist, 26(2), 379–396.

10 Subbotin M.T. On the law of frequency of error / M.T. Subbotin // Matematicheskii Sbornik. — 1923. — 31. — 296–301.

11 Turner, M.E. (1960). On Heuristic Estimation Methods. Biometrics, 16(2), 299-302.

12 Robertson C.A. Some descriptive properties of normal mixtures / C.A. Robertson, J.G. Fryer // Scandinavian Actuarial Journal. — 1969. — (3–4). — 137–146. — doi:10.1080/03461238.1969.10404590.

13 Rao K.S. A bimodal distribution / K.S. Rao, J.L. Narayana, and V.Papayya Sastry // Bulletin of the Calcutta Mathematical Society. — 1988. — 80(4). — 238–240.

14 Sarma P.V.S. Prabhakara On a family of bimodal distributions / P.V.S. Sarma, Rao, K.S. Srinivasa and Rao, R. Prabhakara // Sankhya, Series B. — 1990. — 52. — 287–292.

15 Box G. Bayesian inference in statistical analysis / G.Box, G.Tiao. John. — New York: John Wiley and Sons, 992.

16 Kunchenko Y. Polynomial Parameter Estimations of Close to Gaussian Random variables / Y. Kunchenko. — Germany, Aachen: Shaker Verlag, 2002.

17 Warsza Z.L., Zabolotnii S.W. A Polynomial Estimation of Measurand Parameters for Samples of Non-Gaussian Symmetrically Distributed Data. Szewczyk, R., Zieliński, C., and Kaliczyńska, M. (eds): Automation 2017: Innovations in Automation, Robotics and Measurement Techniques, serie: Advances in Inteligent Systems and Computing. vol. 550, Springer International Publishing AG, 2017. — P. 468–480. — doi: 10.1007/978-3-319-54042-9_45.

18 Warsza Z.L., Zabolotnii S.W. Ocena niepewności pomiarów o rozkładzie trapezowym metodą maksymalizacji wielomianu. Przemysł Chemiczny. —2017. — 1(12). — 68–71. — doi:10.15199/62.2017.12.6.

19 Cramér H. Mathematical Methods of Statistics / H. Cramér. PMS v.9 Princeton University Press, 2016.

20 Lilliefors H.W. On the Kolmogorov-Smirnov test for normality with mean and variance unknown / H.W. Lilliefors // Journal of the American Statistical Association. — 1967. — 62(318). — 399–402.

С.В. Заболотный, В.Ю. Кучерук, З.Л. Варша, А.К. Хасенов

Экспоненциалды таралулардың бимодалды қоспаларының берілгендері үшін өлшенетін параметрлердің полиноминалды бағалары

Экспоненциалды таралулардың екімодалды қоспалары түрінде кездейсоқ қателіктер моделі үшін қайталама өлшеулер нәтижелерінің бағасын табуға болатын стандарты емес тәсіл ұсынылды. Берілген тәсіл полиномды барынша көбейту әдісін қолданумен және кездейсоқ шамаларды жоғарғы ретті статистикалармен (моменттер және кумулянттармен) сипаттаумен негізделген. Полиномның $r = 3$ дәрежесіне дейінгі дәлдікпен сараптау және бағаларды табу үшін аналитикалық өрнектер ұсынылды. Полиномның дәрежесі $r = 1$ және $r = 2$ болған кезде (симметриялы таралған мәліметтер үшін) полиноминалды бағалар орташа арифметикалық бағаларға сай. Полиномның дәрежесі $r = 3$ болғанда полиноминалды бағаның белгісіздігі кемиді. Кему коэффициенті іріктелген мәліметтер таралуының Гаусс моделінен ерекшелігінің деңгейін сипаттайтын 4 және 6 ретті кумулянтты коэффициенттер мәндеріне байланысты. Көптеген статистикалық сынақтар жолымен (Монте-Карло әдісі) полиноминалды бағалардың қалыптандыру қасиеттері зерттеліп, олардың дәлдігін белгілі бағалармен (орташа, медианды және бүгілу центрі) салыстыру сараптамалары жүргізілді. Полиноминалды бағалар ($r = 3$ кезінде) едәуір тиімді болатын антимодалды тереңдігі мен сұрыптау көлеміне тәуелді облыстар құрастырылған.

Кілт сөздер: екімодалды таралу, өлшенетін параметр, бағалар дисперсиясы, моменттер, кумулянттар, стохастикалық полином.

С.В. Заболотный, В.Ю. Кучерук, З.Л. Варша, А.К. Хасенов

Полиномиальные оценки измеряемых параметров для данных из бимодальных смесей экспоненциальных распределений

Предложен нестандартный подход к нахождению оценок результата многократных измерений для модели случайных ошибок в виде двумодальных смесей экспоненциальных распределений. Данный подход основан на применении метода максимизации полинома и описании случайных величин статистиками высших порядков (моментами и кумулянтами). Представлены аналитические выражения для нахождения оценок и анализа их точности до степени полинома $r = 3$. При степени полинома $r = 1$ и $r = 2$ (для симметрично-распределенных данных) полиномиальные оценки эквивалентны оценкам в виде среднего арифметического. При степени полинома $r = 3$ неопределенность полиномиальных оценок уменьшается. Коэффициент уменьшения зависит от значений кумулянтных коэффициентов

4-го и 6-го порядка, которые характеризуют степень отличия распределения выборочных данных от гауссовской модели. Путем многоразовых статистических испытаний (методом Монте-Карло) исследованы свойства нормализации полиномиальных оценок и проведен сравнительный анализ их точности с известными оценками (средним, медианой и центром сгиба). Построены области, зависящие от глубины антимодальности и объёма выборки, в которых полиномиальные оценки (при $r = 3$) являются наиболее эффективными.

Ключевые слова: двумодальное распределение, измеряемый параметр, дисперсия оценок, моменты, кумулянты, стохастический полином.

References

1 BIPM, I., IFCC, I., ISO, I., & IUPAP, O. The International Vocabulary of Metrology, Basic and General Concepts and Associated Terms—VIM, JCGM 200: 2008 [ISO/IEC Guide 99].

2 Novickij, P.V., Zograf, I.A., (1991). Otsenka pohreshnostei rezultatov izmerenii [Estimation of the measurement result errors]. Leningrad: Enerhoatomizdat [in Russian].

3 Lemeshko, B.Y. (2004). O zadache identifikatsii zakona raspredeleniia sluchainoi sostavliaiushchei pohreshnosti izmerenii [On the problem of identification of the distribution law of the random component of the measurement error]. *Metrolohiia – Metrology, 7,* 8–17 [in Russian].

4 Lindsay, B.G. (1995). Mixture models: theory geometry and applications, Vol 51. In: NSF-CBMS regional conference series in probability and statistics, Institute of Mathematical Statistics, Hayward.

5 McLachlan, G.J., & Peel, D. (2000). Finite mixture models. New York: Wiley.

6 Mengersen, K., Robert, C.P., & Titterington, D.M. (2011). Mixtures: estimation and applications. New York: Wiley.

7 Böhning, D., Seidel, W., Alfó, M., Garel, B., Patilea, V., Walther G. (2007). Editorial: Advances in mixture models. Comput Stat Data Anal 51:5205–5210.

8 Böhning, D., Hennig, C., McLachlan. G.J., McNicholas, P.D. (2014). Editorial: The 2nd special issue on advances in mixture models. Comput Stat Data Anal 71:1–2.

9 Hassan, M.Y., & Hijazi, R.H. (2010). Pak. j. statist. 2010 Vol. 26 (2), 379-396 a bimodal exponential power distribution. Pak. J. Statist, 26(2), 379–396.

10 Subbotin, M.T. (1923). On the law of frequency of error. Matematicheskii Sbornik, 31, 296–301.

11 Turner, M.E. (1960). On Heuristic Estimation Methods. Biometrics, 16(2), 299–302.

12 Robertson, C.A., & Fryer, J.G. (1969). Some descriptive properties of normal mixtures. *Scandinavian Actuarial Journal, 1969(3–4),* 137–146. doi:10.1080/03461238.1969.10404590.

13 Rao, K.S., Narayana, J.L. & Papayya Sastry, V. (1988). A bimodal distribution. *Bulletin of the Calcutta Mathematical Society, 80(4),* 238–240.

14 Sarma, P.V.S., Rao, K.S. Srinivasa, & Rao, R. Prabhakara (1990). On a family of bimodal distributions. Sankhya, Series B, 52, 287–292.

15 Box, G., & Tiao, G. (1992). Bayesian inference in statistical analysis. New York: John Wiley and Sons.

16 Kunchenko, Y. (2002). Polynomial Parameter Estimations of Close to Gaussian Random variables. Germany, Aachen: Shaker Verlag.

17 Warsza, Z.L., Zabolotnii, S.W. (2017). A Polynomial Estimation of Measurand Parameters for Samples of Non-Gaussian Symmetrically Distributed Data. Szewczyk, R., Zieliński, C., and Kaliczyńska, M. (Eds): Automation 2017: Innovations in Automation, Robotics and Measurement Techniques, serie: Advances in Inteligent Systems and Computing. Vol. 550, Springer International Publishing AG, pp. 468–480. doi: 10.1007/978-3-319-54042-9_45.

18 Warsza, Z.L., & Zabolotnii, S.W. (2017). Uncertainty of measuring data with trapeze distribution evaluated by the polynomial maximization method. Przemys Chemiczny, 1(12), 68–71. doi:10.15199/62.2017.12.6 (in Polish).

19 Cramér, H. (2016). Mathematical Methods of Statistics. PMS v.9 Princeton University Press.

20 Lilliefors, H.W. (1967). On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association, 62(318),* 399–402.