

WOJCIECH BATKO\*, BARTOSZ PRZYSUCHA\*\*

## IDENTIFICATION OF PROBABILITY DISTRIBUTION FORM FOR RESULTS OF SOUND LEVEL MEASUREMENTS

### SUMMARY

The authors focus their attention on the identification of the probability distribution density function of the sound level, which constitutes the basis for the proper statistical inferences and uncertainty assessments in the environment acoustical hazard control. Their functional form is a metric for the analysis of acoustical measurement results burdened with random errors. Its proper selection conditions the rightness of statistical inferences in relation to the analysed noise effect. The problem of identification of the noise level probability distribution form was presented on the grounds of the sound  $L_A$  level monitored at one of the main streets in Lublin. The analysis of differences and references to the normal distribution form, commonly applied to statistical analysis of the acoustical measurement results, was carried out. It is the aim of the authors, that the presented results should become the basis of a broader discussion concerning new estimation procedures of the controlled noise indicators and their uncertainty assessment. Also new verification procedures of the rightness of model acoustical formalisms, assumed in numerous environment acoustic investigations, are required.

**Keywords:** acoustical measurements, statistical analysis of obtained results, estimation of the distribution, uncertainty

### SPOSÓB ESTYMACJI FUNKCJI GĘSTOŚCI ROZKŁADU PRAWDOPODOBIENSTWA DLA LOSOWEJ PRÓBY WYNIKÓW POMIARU POZIOMÓW DŹWIĘKU

W artykule przedstawiono sposób estymacji funkcji gęstości rozkładu prawdopodobieństwa dla losowej próby wyników pomiaru poziomów dźwięku  $L_A$  przesuniętym rozkładem gamma. Sformułowano potrzebę jej znajomości w kontekście środowiskowych akustycznych badań kontrolnych i związanych z nimi ocen ich niepewności. Dokonano analizy różnic i odniesień do postaci rozkładu normalnego, powszechnie wykorzystywanego w statystycznych analizach wyników pomiarów akustycznych. W zamierzeniu autorów zaproponowane rozwiązanie winno stać się podstawą szerszej dyskusji nad poszukiwaniami nowych algorytmów estymacji kontrolowanych wskaźników hałasu oraz ocen ich niepewności, a także procedur identyfikacji i weryfikacji akustycznych formalizmów modelowych przyjmowanych w badaniach.

**Słowa kluczowe:** pomiary akustyczne, statystyczna analiza wyników pomiarów, estymacja rozkładów prawdopodobieństw, niepewność

### 1. INTRODUCTION

Control processes of the environment acoustical hazards are of an undeniable importance for the environment protection politics. At a macro level, related to the development of the environment protection programs as well as at a micro level concerning the given object of excessive noise emissions. The proper control of the environment acoustical state is crucial for undertaking the proper administrative decisions leading to limiting the noise hazards. They are also inseparably related to the verification processes of the strategic noise maps and environment acoustical protection programs corresponding to them.

The measurement procedure of the control system of the environment acoustical hazard is divided into two stages. In the first stage the results of the sound  $L_A$  level are collected and recorded, while in the second stage they are processed in order to assess values of the controlled noise indicator. In this approach traditional statistical inferences – determined

in the international document (International Organization for Standardization 1995), issued by seven most important metrological organizations – are applied. Their concept assumes the equivalence of measurement assessments treated in a form of random variables, which are further undergoing processing in accordance with the calculus of probability. It assumes the axiomatic that the most probable value of the unknown observation is the arithmetical mean of the data collection obtained under the same measurement conditions. Such procedure is hedged with the assumption that at a sufficiently large number of measuring results it is possible to assume that the obtained values distribution is in a form of the density probability distribution. Such conditions are commonly used in the current control methods of noise indicators describing the environment acoustical climate state. In general, the correctness of such approach, especially the likelihood of the assumption of the normal probability distribution of the random test results of sound level measurements, is not discussed.

\* AGH University of Science and Technology, Faculty of Mechanical Engineering and Robotics, Department of Mechanics and Vibroacoustics, al. Mickiewicza 30, 30-059 Krakow, Poland; batko@uci.agh.edu.pl

\*\* Lublin University of Technology, Faculty of Management, Department of Quantitative Methods in Management of Economy, 20-618 Lublin, Nadbystrzycka, 38, Poland

The correctness problem of this assumption can be the source of errors of the controlled noise indicators estimation. This problem is recently more and more often indicated (Batko and Przysucha 2011; Batko and Stępień 2009, 2010, 2011; Wszolek 2006), and also questioned by the results of the environment acoustical monitoring.

Thus, investigations directed towards the determination of the proper form of the probability density distribution of sound  $L_A$  levels are justified. Solution of this problem constitutes the contents of the paper. The authors directed their attention to the possibility of attributing the gamma distribution to the random test results of the sound  $L_A$  level measurements. Its properties were analysed in relation to their approximation by the normal distribution and by the trimmed-normal distribution.

## 2. ANALYSIS OF THE PROBABILISTIC VARIABILITY OF RESULTS OF SOUND $L_A$ LEVEL MEASUREMENTS IN A RANDOM TEST

The practice of measuring sound levels and carrying out related to it statistical analysis of results variability in random test indicate the necessity of introducing different characteristics of the probability distribution than the normal one. The main difference between the assumed normal distribution of the sound  $L_A$  level measurements and its empirical representation, is generally observable in acoustical measurements of the empirical distribution asymmetry.

In statistical investigations of the variability sound  $L_A$  level value usually the left hand side asymmetry is observed (the mean situated on the left from the modal value), less often the right hand side (the means situated on the left from the modal value). The coefficient of skewness of the probability density distribution is often different from zero, in contrast to the normal distribution symmetry properties.

Considered in the paper analysis of sound level probability distribution was carried out for the results of the sound  $L_A$  level measurements, performed at one of the main arteries in Lublin. It can serve as the illustration of the mentioned above observations.

The empirical histogram, describing the density probability distribution for the obtained results of the sound level control, was determined for the measurements illustrating variability of the control test results (Fig. 1). This is described by the distributive sequence of parameters presented in table 1.

**Table 1**

Parameters of the distributive sequence for data obtained in measurements

Maximum value	$x_{\max}$	101.3 dB
Minimum value	$x_{\min}$	39.9 dB
Class number	$k$	62
Data range	$R$	61.4 dB
Class length	$B$	1 dB

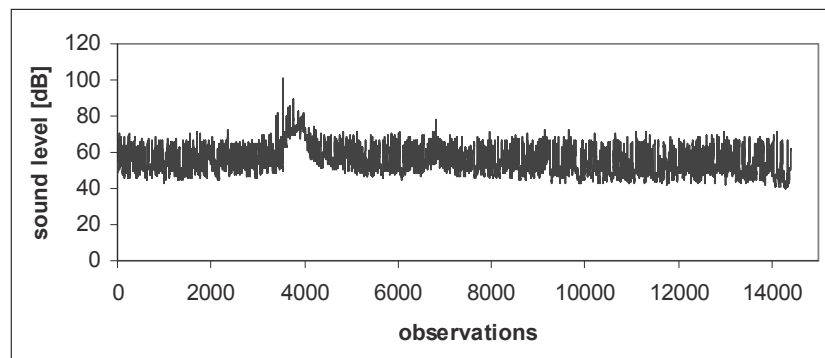
While creating the frequency histogram, the class number was assumed as 62, at the class length being 1 dB. The frequency histogram representing the measurement results of the noise emission is presented in figure 2.

Squares in the diagram illustrate minimal and maximal values of the noise level, while the triangle marks the mean value representing sound level variability. Statistical parameters: mean value, standard deviation, skewness coefficient or kurtosis are given in table 2.

**Table 2**

Statistical parameters of the distributive sequence for the measurement data

Statistical parameters	Symbol	Value	Error
Mean	$\mu$	55.2 dB	0.1 dB
Standard deviation	$\sigma$	7.1 dB	0.1 dB
Skewness coefficient	$\rho$	0.525	0.0204
Kurtosis	$K$	-0.06 (2.94)	0.0408



**Fig. 1.** Sound level measurement results

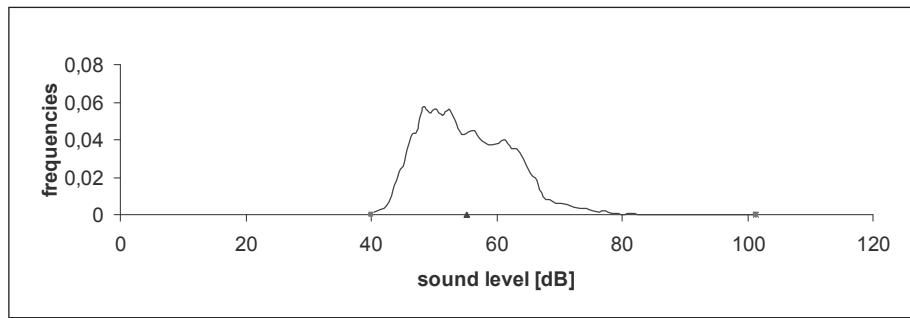


Fig. 2. Empirical frequency histogram for data obtained in measurements

As is seen from table 2, for the empirical data considered, the skewness coefficient  $\rho = 0.525$  with deviation 0.0204. Such a result is statistically different from the normal distribution, for which  $\rho_N = 0$ . This can indicate that the hypothesis of the normal distribution of the sound level results is not fully reliable.

This problem constituted also the basis of the verification by means of the Kolmogorov test, with the assumed allowable significance level of error:  $\alpha = 0.1$ . The approximation of the density probability distribution function for the sound level results by the normal distribution, necessary in this test, was realized by the method of moments (MM), the non-linear least squares method (NLMS) and the minimal likelihood method (MLM). The set of coefficients obtained in these procedures is presented in table 3.

Table 3

List of test coefficients for the hypothesis of the estimation errors normal distribution

Estimation method	MM	NLSM	MLM
Coefficient in the Kolmogorov test, at the significance level: $\alpha = 0.1$	0.56	0.33	0.32
Critical statistics	0.153	0.153	0.153

Results presented in the table 3 revealed that the assumption of the normality of estimation errors is not met. In case of applying the mentioned estimation procedure, approximation of the probabilistic variability of the acoustical measurements by the normal distribution led to the rejection of the hypothesis of its permissibility at the significance level:  $\alpha = 0.1$ .

This effect can result from the observed certain dual-modality of the sound level results distribution, which is more visible when the empirical distribution is presented in 5-second intervals (Fig. 3).

The properties of the random acoustical investigations discussed above pose a problem of looking for the proper probability distribution representing the measured acoustical results. Its knowledge is necessary for proper and efficient programming of the control investigations of the environment acoustical hazard state.

In the analysis of the selection of the proper function form, describing density of the probability distribution for the obtained results of the sound level control, we limited ourselves to their representation by the normal distribution, trimmed-normal distribution and the shifted-gamma distribution.

Let us mark the function of the normal distribution density  $N(\mu, \sigma)$  by  $f_N(x)$ :

$$f_N(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad (1)$$

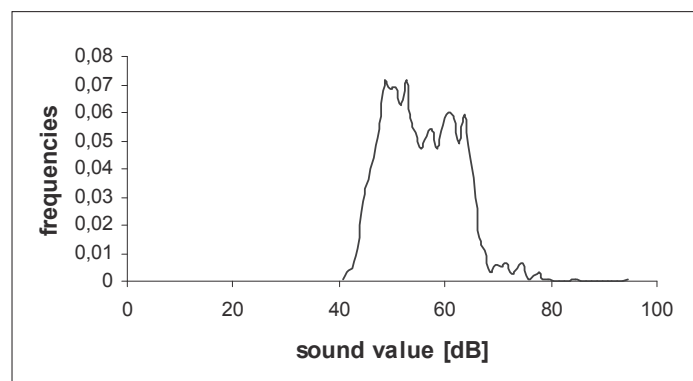


Fig. 3. Frequency histogram of 5-second empirical measurement data of sounds  $L_A$  level

Let us mark the function of the trimmed normal distribution, being the representation of the sound level measurement  $L_A$  in the range:  $x_{\max} = L_{A\max}$  and  $x_{\min} = L_{A\min}$  by  $f_U(x)$

$$f_U(x) = \frac{1}{F_N(x_{\max}) - F_N(x_{\min})} f_N(x) \quad (2)$$

where:

$$F_N(x) = \int_{-\infty}^x f_N(y) dy \quad (3)$$

However, by  $g(x)$  let us mark the density function of the shifted gamma distribution:

$$g(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} (x - x_{\min})^{\alpha-1} e^{-\frac{x-x_{\min}}{\beta}}, \quad x \geq x_{\min} \quad (4)$$

where:

$$\Gamma(z) = \int_0^{+\infty} t^{z-1} e^{-t} dt \quad (5)$$

is the gamma Euler function:  $z \in C, \alpha > 0, \beta > 0$ .

The essential problem in approximating the results of control tests by gamma distribution is the problem of selecting estimators for parameters determining its form. They can be determined by the methods: MM, NLSM and MLM. The Method of Moments was applied in this work. However, this method usually does not provide optimal estimators, for its selection persuades the easiness of calculating the parameters looked for. Moreover, the estimators of the distribution parameters assure the same characteristics as the characteristics calculated from the data.

Estimators of the looked for parameters determined by the method of moments are given by equations:

$$\begin{cases} \bar{\alpha} = \frac{(x_{\min} - \mu)^2}{\sigma^2} \\ \bar{\beta} = \frac{\sigma^2}{\mu - x_{\min}} \end{cases} \quad (6)$$

However, the estimation performed by this method has certain limitations being the result of the estimators properties. But it is very convenient, since the calculation of two first moments of distribution is sufficient.

Alternatively, estimation of parameters of the assumed density probability distribution function for the analysed results of the sound level control can be carried out by means of the NLSM (non-linear least squares method) or MLM (minimal likelihood method). The solution based on the NLSM consists in minimizing the sum given by the equation:

$$\sum_{i=1}^k (f_e(x_i) - f(x))^2 \rightarrow \min \quad (7)$$

In case of the MLM (minimal likelihood method) application, the idea of the estimation process consists in the maximisation of the likelihood function:  $\max_{(\alpha, \beta) \in \Theta} \ln \prod_{i=1}^n f(x_i; \alpha; \beta)$ ,

which brings the estimation process to solving the following set of equations:

$$\begin{cases} \ln \beta + \frac{d \ln \Gamma(\alpha)}{\alpha} = \frac{1}{n} \sum_{i=1}^n \ln x_i \\ \alpha = \frac{1}{n\beta} \sum_{i=1}^n x_i \end{cases} \quad (8)$$

Equations (8), leading to the determination of parameters of the density probability function, in a similar fashion as the majority of the MLM estimations, can be only solved by numerical methods. Identical situation is in case of looking for the estimator determined by the NLSM from equation (7). In case of models in which estimation errors have normal distribution and are not correlated with each other, the NLSM and MLM estimators should coincide and have the desirable properties of compatibility, effectiveness and be unbiased (Rydlewski 2009).

Estimations by the NLSM (non-linear least squares method) and MLM (minimal likelihood method) provide theoretically (at fulfilling certain assumptions in their application) much better estimators than the method of moments (MM). However, on account of a dual-modality property, observed in the histogram of the sound level measurement, results calculated from 5-seconds samples (Fig. 3), these estimative solutions were given up for the method of moments (MM).

### 3. APPROXIMATION OF THE PROBABILITY DENSITY OF RANDOM SOUND LEVEL MEASUREMENT RESULTS BY GAMMA DISTRIBUTION

In numerous environmental investigations and assessments of acoustical hazards, the basic measurement parameter is the sound A level value, determined as the logarithm of the ratio of the acoustic pressure square  $p$  to the constant value  $p_0 = 2 \cdot 10^{-5}$  Pa:  $L_A = 10 \log \frac{1}{t_0} \int_{t_i}^{t_{i+1}} \left( \frac{p(t)}{p_0} \right)^2 dt$ , where  $t_0$  – is length of time interval, which the integral is determined (from  $t_i$  to  $t_{i+1}$   $i = 1, 2, \dots, n$ ).

The sequence of its measurement results:  $L_{A_i}; i = 1, 2, \dots, n$  determines the random test from the bases of possible measurement assessments, on the grounds of which the indicators of environment hazard, applied universally in the European Union countries, are estimated (Directive 2002/49/WE). Knowledge of the probability density function of the sound A level measurement results is necessary for their likelihood measuring estimation.

After analysing the random sequence of sound level measurements results, it was decided that the gamma distribution form, defined by equations (4) and (5), will be subjected to analysis of its suitability for the density probability distribution description. Its references to the – generally assumed in investigations – normal and trimmed-normal distribution were also carried out.

Parameters of the singled out gamma distribution are as follows:

$$\begin{cases} EX = \alpha\beta + x_{\min} \\ VarX = \alpha\beta^2 \\ \rho = \frac{2}{\sqrt{\alpha}} \\ K = \frac{6}{\alpha} \end{cases} \quad (9)$$

The root-mean square criterion of differences between the estimated sequence value and densities of the matched functions was the basis for the verification of the correctness of the approximation model selection:

$$r(f_e(x), f(x)) = \sum_{i=1}^k (f_e(x_i) - f(x))^2 \quad (10)$$

It provides information in what way the postulated probability distribution differs from the empirical distribution, determined on the basis of the random test from the distributive sequence.

The estimators of the analysed distribution parameters determined by the NLSM gave the best estimation with respect to this measure. The obtained results are given in table 4.

As it results from the waveforms and shapes of the analysed distributions presented in figure 4, as well as from the related calculations shown in table 2, the sound  $L_A$  level probability distribution differs from the normal distribution. The skewness of this distribution is essentially different than in case of the Gaussian distribution. The data are more concentrated below the mean value (right-hand skeweness). The gamma distribution form assigned to the results of the sound  $L_A$  level random test, better characterizes properties of the empirical distribution.

It gives smaller value of the metric of root-mean-squares between the analysed sequence and the histogram of observation (Tab. 4).

**Table 4**

Distance coefficients given by equation (10) between the empirical probability distribution and estimated forms of probability distributions

Probability function	$r$
Normal distribution $f_N(x)$	0.00251
Trimmed-normal distribution $f_U(x)$	0.00254
Gamma distribution LSM $g(x)$	0.00084
Gamma distribution MM $g_1(x)$	0.00137
Gamma distribution MLM $g_2(x)$	0.00103

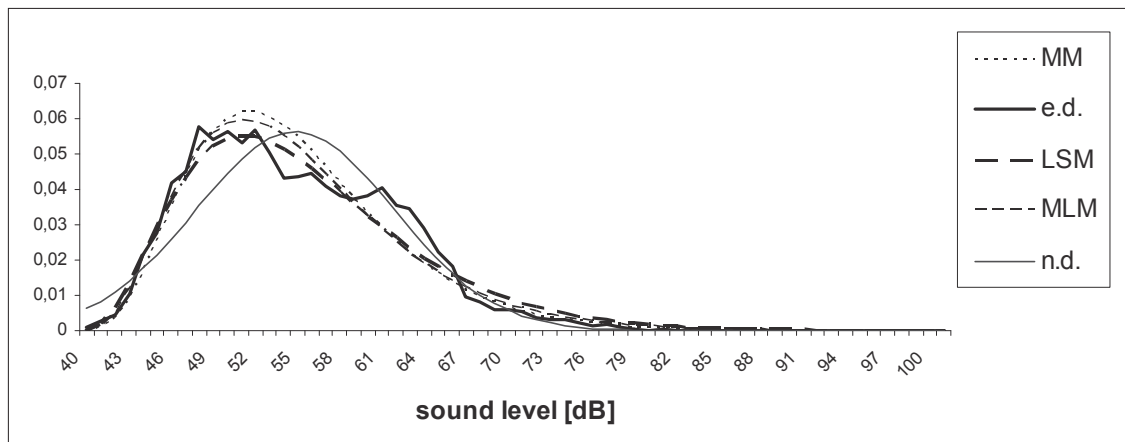
Since the assumption of the normality of errors generated by the applied approximation is related to the estimation of parameters performed by the NLSM and MLM methods, such solutions can not be always efficient. Thus, the behaviour of their characteristics calculated by various methods was analysed.

**Table 5**

Estimators of the gamma distribution coefficients at various methods and characteristics calculated from these distributions

Parameter	MLM	NLSM	MM
Alfa	4.19	3.76	4.65
Beta	3.65	4.22	3.29
Characteristics			
$EX$	55.2 dB	55.8 dB	55.2 dB
$sX$	7.5 dB	8.2 dB	7.1 dB

The results presented in table 5 provide information on differences between characteristics calculated from the density function parameters estimated by various methods. The MM estimation gives the best results with respect to compatibility



**Fig. 4.** Probability distribution curves obtained from the estimation of the distribution function  $g(x)$  by : LSM,  $g_1(x)$ , MM,  $g_2(x)$ , MLM compared with the normal distribution (n.d.)  $f_N(x)$  with parameters  $\mu, \sigma$

of characteristics from the theoretical distribution with the ones calculated from the test. At the MLM estimation  $EX = 55.2$  dB, which is compatible with the value calculated from the test, while the standard deviation:  $sX = 7.5$  dB and the standard deviation calculated from the test:  $sX_{MM} = 7.1$  dB. The worst effects, despite the best matching in relation to  $R$  measure, give characteristics calculated by means of parameters obtained at the estimation by the NLSM,  $EX = 55.8$  dB, while  $sX_{NLSM} = 8.2$  dB. The confidence interval for the mean value, at our number of observations, equals: (55.1 : 55.3). Thus, it is seen that the expected value, determined from parameters of the distribution estimated by the NLSM method, does not fall even into the confidence interval for the mean value.

Such effect can be caused by the lack of normal distributions of the rests in the model, or by their correlation. It is worth to emphasise, that the NLSM estimation is performed on the basis of the empirical distribution determined from the distributive sequence, while the MM estimation is done on the basis of the moments calculation and in the case of the MLM on the bases of all values from the test.

Thus, it is worth to draw the attention to the fact, that in case of estimation by the NLSM method, on the basis of the empirical distribution obtained from the distributive sequence, we can have for the disposal various distributive sequences of various centres and class length. As an example, parameters of another distributive sequence – assumed for the analysed data – are illustrated in table 6.

**Table 6**

Parameters of the 2<sup>nd</sup> distributive sequence

Maximum value	$x_{\max}$	101.3 dB
Minimum value	$x_{\min}$	39.9dB
Number of classes	$k$	120
Data range	$R$	61.4 dB
Class length	$B$	0.52 dB

Characteristics obtained from this sequence coincide with the characteristics of the distributive sequence from table 6. Whereas parameters obtained by the NLSM estimation for this 2<sup>nd</sup> distributive sequence are as follows.

**Table 6**

Parameters and characteristics calculated by the NLSM for the distributive sequence from table 5

	NLSM
Alfa	1.74
Beta	14.83
Characteristics	
$EX$	65.7 dB
$sX$	19.55 dB
$r$	0.007

In the case of this 2<sup>nd</sup> sequence the obtained characteristics are different than the ones obtained from the test.

The analysis of the realised process of selecting the approximation for the density probability distribution function for the results of the sound level measurements indicated that choosing the method of moments for the estimation of its parameters is justified despite of the lack of the possibility of attributing to it the highest effectiveness value. However, it assures the compatibility of the obtained by it characteristics with the ones from the test. Good effects can be expected in the considered case at the estimation of the probability density function by the minimal likelihood method. In case of testing the characteristic of the sound levels probability distribution, in consideration of its tendency to bimodal values, we can only speak about its approximation by the gamma distribution.

#### 4. CONCLUSIONS

This paper can be regarded as a basis for the statistical analyses of environmental acoustical hazards, both based on simulation and on empirical investigations. It's main achievement has been the description of the probability density distribution of sound levels  $L_A$  by the gamma distribution. This distribution is capable of proper approximation of the measurement results of the sound level  $L_A$ , having properties of the right-handed asymmetry of the random test.

This paper presents the discussion of divergence – with generally applied in acoustic investigations – the Gauss function proper for the normal distribution. The results are documented by analyses of monitoring of sound level  $L_A$  changes, carried out at one of the main arteries in Lublin.

The distinctive form of the gamma distribution, for the statistical description of measurement results  $L_{A_i}$ ,  $i = 1, 2, \dots, n$ , constitutes a starting point for the estimation of the uncertainty of results of the environment acoustical control and related to it analysis of the probability distributions of long-term noise indicators. The procedures of estimating such indicators are given in other papers of the authors (Batko and Przysucha 2010, 2011).

The argumentation, presented in the paper, can be a recommendation and encouragement for avoiding routine approaches in uncertainty assessments of the controlled noise indicators, and for assuming deliberate (more cautious) approaches. This is a tool for estimating the control results likelihood and related to it error intervals, within which the right control value will occur with the given probability. It allows for a better understanding of the control process and conditions of the uncertainty budget construction. It provides the possibility of obtaining more reliable analytical results of the realized vibroacoustic experiments. In practice, this can have an essential meaning in case of acoustical measurements at the highest accuracy level as well as in case of environmental acoustical investigations.

## Acknowledgments

*The study was performed within the framework of the project: N R03 0030 06/2009.*

## References

- Batko W., Stępień B. 2009, *Non-parametric methods of estimation of type A uncertainty of the environmental noise hazard indices*. Archives of Acoustics, 34, No. 3, pp. 295–303.
- Batko W., Stępień B. 2010, *Application of the bootstrap estimator for uncertainty analysis of the long-term noise indicators and their uncertainty*. Acta Physica Polonica A, 118, pp. 11–16.
- Batko W., Stępień B. 2011, *Application of the Bayesian inference for estimation of the long-term noise indicators and their uncertainty*. Acta Physica Polonica A, 11, pp. 916–920.
- Batko W., Przysucha B. 2010, *Determination of the Probability Distribution of the Mean Sound Level*. Archives of Acoustics, vol. 35, No. 4, pp. 543–550.
- Batko W., Przysucha B. 2011, *Random Distribution of Long-Term Indicators of Variable Emission Conditions*. Acta Physica Polonica A, 119, pp. 1086–1090.
- Directive 2002/49/WE of the European Parliament and of the Council of 25 June 2002, relating to the assessment and management of environmental noise, Official Journal of the European Communities 18.07.2002.
- International Organization for Standardization 1993, 1995 (corrected and reprinted), *Guide to Expression of Uncertainty in Measurement*.
- Rydlewski J. 2009, *Estymatory największej wiarygodności w uogólnionych modelach regresji nieliniowej*. [The Minimal likelihood estimators in the generalized nonlinear regression models.] PhD thesis written at the Jagiellonian University Faculty of Mathematics and Computer Science at the Institute of Mathematics, Krakow.
- Wszolek T. 2006, *Uncertainty analysis of the environmental noise research*. Proc. of the conference XXXIV Winter School of Vibroacoustical Hazards Suppression, pp. 205–216.