Metric Identification for Random Value Density Distribution

E.I. Tsvetkov and E.S. Suloeva
Department of Information and Measurement Systems and Technology, Saint Petersburg Electrotechnical University “LETI”, Prof. Popov str. 5,197376 Saint Petersburg, Russian Federation
E-mail: suloeva@list.ru

Abstract. The problem of metric identification of probability density distribution for a random value is considered. It is shown that the accuracy is determined both by the sample size and by the number of intervals into which the domain of its determining is divided.

1. Introduction
Metric identification of dependencies [1] is a rather new trend in metrology. The objective of metric identification is to determine the dependence type, which allows specifying its value at a fixed number of the argument. Moreover, accuracy characteristics are determined for the obtained results according to the principles of metrology [2]. Metric identification of dependencies is found in between the value measurement and objects and relations identification. Metric dependency identification is the basis for metrological properties analysis of random values described by functional probabilistic characteristics determined upon a set of its possible values \( \Lambda \). In other words, for a random value of \( \lambda \), the membership relation \( \lambda \in \Lambda \) is true.

The numeric probabilistic characteristic \( \Phi[\lambda] \) is determined by the following correlation:

\[
\Phi[\lambda] = \lim_{i \to \infty} \frac{1}{I} \sum_{i=1}^{I} g_{\Phi}[\lambda_i] / I,
\]

where \( \lambda_i \) is the \( i \)-th count of \( \lambda \); \( g_{\Phi}[\lambda] \) is the transformation forming the basis of the definition \( \Phi[\lambda] \).

The functional probabilistic characteristic \( \Phi[\alpha/\lambda] \) is determined by the following correlation:

\[
\Phi[\alpha/\lambda] = \lim_{i \to \infty} \frac{1}{I} \sum_{i=1}^{I} g_{\Phi}[\alpha/\lambda_i] / I,
\]

Correspondingly, the estimates of the numeric \( \Phi[\lambda] \) and functional probabilistic characteristics \( \Phi[\alpha/\lambda] \) obtained during the \( j \)-th identification experiment are represented by the following expressions:

\[
\Phi^*_j[\lambda] = \sum_{i=1}^{I} g_{\Phi}[\lambda_{ij}] / I,
\]

\[
\Phi^*_j[\alpha/\lambda] = \sum_{i=1}^{I} g_{\Phi}[\alpha/\lambda_{ij}] / I.
\]
Metric identification of a probabilistic characteristic is aimed at determining the value of $\Phi[\alpha / \lambda]$ at the fixed value of $\alpha$.

A particular feature of the metric identification of random function probabilistic characteristics is that their determining supposes performance of averaging upon the aggregate of counts. Trustworthiness assessment [3] of the probabilistic characteristics of the error in identification of the functional probabilistic characteristic obtained while performing the metrological analysis requires additional averaging. It is just the error of the metric identification results of the probabilistic characteristic that is represented by the following difference:

$$
\delta \Phi_j^*[\alpha / \lambda] = \frac{\sum_{i=1}^{I} g_\Phi[\alpha / \lambda]_i / I}{1} - \Phi[\alpha / \lambda] = \lim_{i \to \infty} \sum_{i=1}^{I} g_\Phi[\alpha / \lambda] / I,
$$

$$
\Theta[\delta \Phi_j^*] = \lim_{N \to \infty} \sum_{j=1}^{N} g_{\delta \Phi}[\alpha / \delta \Phi] / N.
$$

The abovementioned particularities of metrological analysis of the procedures used for assessment of the probabilistic characteristics have still been paid little attention, whereas the practical demand for use of the relevant information increases. Particularities of metric identification of the probability density distribution for random functions are considered below.

2. **Metrological Procedure Analysis for Probabilistic Characteristics Evaluation**

Generally, identification of $w_j(\lambda)$ assumes forming up the array $\{\lambda_{ij}\}_{i=1}^{I}$ with further determining the frequency (probability assessment) of counts hitting into the established intervals of the domain of determining $\Lambda$ of the random value $\lambda$. The correspondent equation has the following representation:

$$
w^*_j(\lambda) = \left\{ \left\{ \frac{n_s}{\Delta_s} I_s \in (\lambda_{S-1}, \lambda_s) \right\}_{i=1}^{I} \right\}_{s=1}^{S}.
$$

Here, the following definitions are introduced: $I$ is the number of used counts in the $j$-th identification experiment; $S$ is the number of intervals; $\Delta_s = \lambda_s - \lambda_{S-1}$ is the length of the $i$-th interval; $I_s$ is the number of counts, for which $\lambda_{ij} \in (\lambda_{S-1}, \lambda_s)$ is true; $n_s$ is the number of counts hitting into the $s$-th interval.

The obtained results allow determining the identification error for the dependence $w(\lambda)$:

$$
\Delta w^*_j(\lambda) = w^*_j(\lambda) - w(\lambda)
$$

and studying its dependence upon the parameters $I, S, \{\Delta_s\}_{s=1}^{S}$.

$$
\Theta[\Delta w^*_j(\lambda)] = \left[ \frac{\int_{\lambda_{S-1}}^{\lambda_s} w(\lambda) d\lambda}{\lambda_s - \lambda_{S-1}} \right]
$$

where $N$ is the number of estimates used at determining $\Theta[\Delta w^*_j(\lambda)]$, $N_s = N \cdot \int_{\lambda_{S-1}}^{\lambda_s} w(\lambda) d\lambda$.

Analytical study of the dependence of $\Theta[\Delta w^*_j(\lambda)]$ upon the parameters $I, S, N, \{\Delta_s\}_{s=1}^{S}$ is related to the use of approximations and assumptions, which substantially decreases the trustworthiness of the results assessment. Furthermore, the parameters $I, S, N, \{\Delta_s\}_{s=1}^{S}$ are established while planning metric
identification of probability density distribution for the random value of $w(\lambda)$. Generally, their optimal values match the equation system solution:

$$
\begin{align*}
\frac{d(\Theta[\Delta w_j^*(\lambda)])}{dI} &= 0 \\
\frac{d(\Theta[\Delta w_j^*(\lambda)])}{dS} &= 0 \\
\frac{d(\Theta[\Delta w_j^*(\lambda)])}{dN} &= 0
\end{align*}
$$

The provided assumptions lead us to the performance selection of the machine experiment metrological analysis. The correspondent mapping sequence has the following representation:

$$
\begin{align*}
\{\lambda^*_j\}_{j=1}^{S} \rightarrow w_j^*(\lambda) = \{n_j \Delta_s^{-1} I_s^{-1} / (\lambda^*_j \in \lambda_{s-1}, \lambda_s)\}_{j=1}^{S} \\
\Delta w_j^*(\lambda) = \{n_j / I_s \Delta_s - w(\lambda)\}_{j=1}^{S} \rightarrow \\
\Theta[\Delta w_j^*(\lambda)] = N^{-1} \sum_{j=1}^{N} N_s^{-1} \sum_{s=1}^{S} g_\Theta(\Delta w_j^*)
\end{align*}
$$

The distribution identification possibility based on the Simpson’s distribution, with the sample size amounting to 100 counts, is provided as an example.

**Fig. 1.** Comparison of the experimental and theoretical values of probability density distribution with error indication (the difference between these values) at $S=10$.

The interval used for determining the domain of values $[-2;2]$ is subdivided into ten intervals ($S=10$). Graphical representation of the data obtained at imitational modeling (the Experiment), the theoretical part obtained via approximation of the values (the Theory), as well as the error in determining the probability density distribution (the Error) are provided below.

Assessment of the systematic error for the identification results amounted to 0.011, which can be explained by finiteness of the sample size, as well as by the number of used intervals. It should be taken into consideration that the number increasing of counts $I$ provides for a correspondent decreasing of error. Nevertheless, increasing of accuracy with the increase of the number of intervals $S$ is not apparent, because in this case relative dispersion is increasing.

The abovementioned can be confirmed by the results of the machine experiment performed at similar initial data, except for the value of $S$ ($S=30$).
3. Conclusion

It is noteworthy that the obtained results do not consider the approximation errors of the identified dependence, i.e. the errors are correlated with the middle of the intervals.

The provided preliminary results show that while planning metric identification of the probability density distribution one should be based not on a monotonous dependence of accuracy upon the accepted number of intervals, but on the specific dependence of errors upon the controlled parameters and accepted approximation.

References

[1] Tsvetkov E I 2016 Metrology. Models. Metrological analysis. Metrological synthesis. Additional chapters (St. Petersburg: St. Petersburg State University ETU) 144 (in Russian)

[2] Avdeyev B Y 2014 Metrology, standardization and certification: a textbook for students of higher education institutions (under the editorship of V V Alekseyev) (Moscow: Academy) 384 (in Russian)

[3] Tsvetkov E.I and Suloeva E S 2018 Analysis of the parameters that determine the reliability of the results of a verification of measuring instruments Measurement techniques 61 (9) 872–877