METHOD FOR ASSESSING RELIABILITY OF EXPERT MEASUREMENTS

A. I. Buravlev

An approach to assessment of reliability of expert estimation procedures based on information-statistical measurement theory is discussed.

Expert estimates continue to receive wide application in various areas of practical activity. They are especially often used in those sciences where no instrumental methods of measuring are available (such as sociology, psychology, economics, etc.) or when it is impossible to utilize these methods. In these cases, an expert — namely a competent person (a specialist) possessing knowledge, experience and professional intuition — serves as an information-measuring system of a kind. [1, 2].

Unlike instrumental measurements, expert measurements do not have a corresponding metrological basis which give rise to problems related with determination of accuracy and reliability of these measurements.

The majority of expert procedures [1] utilize statistical methods of processing information allowing us to estimate the random error of observations and reliability of statistical inference of the experts. However, these estimates are provisional since the first one characterizes only the scatter of expert opinions and the second only their consistency. Both estimators depend heavily on the empirical scale of observations selected by the experts, the methods of transforming it into a numerical scale and the adopted criterion of measurements precision.

Below we shall consider an approach to estimating reliability of expert measurements based on the information-statistical theory of measurements [2].

**Statement of the Problem.** Let realizations of a random variable $U \in [0, 1]$ with the probability density $f(u)$, which cannot be experimentally measured utilizing experimental devices, be subject to an expert estimation. To estimate the variable $U$ on the interval $[0, 1]$ a scale of measurements $M$ with the readings $0 = u_0 < u_1 < \ldots < u_n = 1$ and the required precision of observations

$$\delta = \max_{1 \leq i < n} |u_i - u_{i-1}|,$$

is formed. An expert’s task is then to determine an interval $\Delta_i = (u_{i-1}, u_i)$, in which the location of the values of the random variable $U$ is feasible in accordance with his/her subjective impression.

A result of such an expert assessment involves, unavoidably, a measurement error $\varepsilon$ due to an incorrect classification of the variable $U$ with respect to the given scale of measurements $M$.

We shall assume that the error $\varepsilon$ is additive with respect to the results of measurements. In this case, an expert estimate $U^*$ is related to the true value $U$ as follows:

$$U^* = U + \varepsilon,$$

where $\varepsilon$ is a random error of measurements with the conditional distribution $\varphi(\varepsilon/\mu_i)$, $(i = 1, n)$ relative to the readings of the scale $M$.

The distribution $\varphi(\varepsilon/\mu)$ is a subjective measure of the expert’s preference probability relative to the measured quantity $U$. It characterizes the level of his/her competence and the possibility of a reliable estimation of the measured quantity.
Reliability of expert’s assessment can be characterized by the probability of correct classification of the quantity $U$ in the given scale of measurements. If $N$ independent experts are used, the probability of correct classification of the measured quantity is:

$$D = \frac{1}{N} \sum_{k=1}^{N} \sum_{i=1}^{n} d_{ik},$$

where

$$d_{ik} = D(U^*_k \in \Delta_i, U \in \Delta_i) = \int_{\Delta_i} \int f(u) q_k(u-u) du du^*$$

is the probability of a correct assessment of the quantity $U$ over the range of measurements $\Delta_i = (u_{i-1}, u_i)$, $(i = 1, n)$ by the $k$-th expert.

Depending on the level of competence of experts, two limiting cases of expert measurements are possible — coarse and accurate.

**Coarse measurements** correspond to a low competence level of an expert when the error of measurements is distributed uniformly (or close to it) on the whole interval of measurements $[0, 1]$:

$$q_k(\varepsilon/u) = \begin{cases} 1, & \varepsilon \in [0, 1]; \\ 0, & \varepsilon \not\in [0, 1]. \end{cases}$$

In this case, reliability of expert measurements in accordance with (1) equals

$$D = \sum_{i=1}^{n} |\Delta_i| f_i,$$

where $|\Delta_i| = u_i - u_{i-1}$ is the width of the $i$-th range of measurements; $f_i = 1/(|\Delta_i|) \int_{u_{i-1}}^{u_i} f(u) du$ is the mean value of the probability density $f(u)$ over the $i$-th range of measurements.

In the case of a high competence level and coordination between the experts, the distribution of the error of measurements is close to the normal law with the zero mean:

$$q_k(\varepsilon/u) = \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\frac{(u-u)^2}{2\sigma_k^2}}, \ v\in(-\infty, \infty),$$

which corresponds to the definition of **accurate measurements** [2].

In this case, reliability of expert measurements is approximately equal to:

$$D \approx \sum_{i=1}^{n} |\Delta_i|^2 \gamma_i f_i,$$

where $\alpha_i = (2/|\Delta_i|) \Phi_0 \left( |\Delta_i|/2\sigma_k \right)$ is the weight coefficient which characterizes the reliability of expert's estimation over the $i$-th range of observations.

It follows from the analysis of the expressions obtained that as the scale factor increases ($n \to \infty$), i.e. the precision of measurements increases ($\delta \to 0$), the reliability of expert measurements decreases ($D \to 0$). This phenomenon is explained by the fact that as the number readings increases, it becomes harder to classify correctly the measured quantity with respect to the readings of the scale.

In this connection, the problem of determining an optimal scale which will assure maximal reliability of expert measurements arises.