Modeling of Expert Estimation*

Antonina V. Ganicheva 1[[0000-0002-0224-8945] and Aleksey V.Ganichev 2[0000-0003-3389-7582]

[1]Tver State Agriculture Academy, 7 Marshala Vasilevskogo Street (Sakharovo), Tver, 170026, Russia
[2]Tver State Technical University, 22 Af. Nikitina Embankment, Tver, 170026, Russia,
alexej.ganichev@yandex.ru

Abstract. The problem of collective decision-making by a group of experts is a crucial one in the theory and practice of decision-making. To obtain a high-quality collective estimation of an object or process, individual expert opinions should be coordinated. In this article, we use the absolute error of estimation (deviation of individual expert estimates from their arithmetic mean) and the relative error (absolute error divided by the arithmetic mean of estimates) as indicators of individual estimates consistency within a collective decision. We consider errors as random variables that fall under the normal probability distribution law. For the selected indicators, their variances, probability distribution densities, and confidence intervals (for averages and variances) are obtained. The preference criterion of methods for constructing confidence intervals is obtained for variances based on the number of experiments. An algorithm is given for finding the required number of experiments when the variance of estimation errors is not specified. A method is developed for obtaining the distribution density of the relative estimation error, its average value, the spread, and the probability of falling within a certain range of values, with a given degree of accuracy. We consider the determination of the required number of experiments (the number of questions or test tasks) for obtaining reasonable estimates by students based on the test results, as the practical implementation of the proposed methods of expert estimation.

Keywords: Sampling, Estimation Error, Confidence Interval.

1 Introduction

The expert estimation method is used to solve difficult to formalize problems for which classical optimization methods cannot be applied. To choose the best solution, in this case, the methods of voting (simple, weighted) and detachment of committees and coalitions are used. In the case of collective estimation, the task of coordinating the individual opinions of experts emerges. For random estimates, the minimum sum of the coefficients of variation or the minimum sum of centered estimates (we call it the estimation error) may be used as an indicator of decision consistency. Not only specialists in a certain field of knowledge may be considered experts, but testing systems, intelligent agents, algorithms, diagnostic systems, measuring instruments, and neural networks as well. The use of expert estimates in the case of distance learning is of special interest. Testing of each trainee may be considered a collective estimation of their knowledge concerning different domains (topics, sections, lectures, practical training, and laboratory classes) of a certain discipline. In this case, the collective body of experts represents the questions and tasks of a test that are considered experiments.

The reference literature pays much attention to the methods for determining the optimal combination of the confidence interval and the confidence probability [4, 12], the use of Chebyshev inequality for making the confidence intervals of expert estimates [5], making the robust (stable and independent of the type of distribution law) estimates [9], and the determination of the degree of various factors influence upon the result based on expert estimation [7]. A recent trend is the use of artificial intelligence methods in expert estimation: the methods of subjective probabilities [8] and fuzzy intervals [10]. The expert estimation models are used to determine the quality of the education process [11].

The objective of the article is to develop mathematical models for consistent expert estimates and estimates of collective decision quality.

2 Setting the Task of Expert Estimation

Let us assume that *n* independent experts estimate some object *A*. Let $x_i \left(i = \overline{i,n} \right)$ denote the estimate given by the *i*th expert and let $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ denote the average estimate given by all experts. Data on the expertise conducted by those *n* experts may be considered as the results of *n* experiments. All possible estimates of object *A* form a general population of values for some random variable *X*.

Before experimenting, we will consider the sample units $X_1, X_2, ..., X_n$ to be pairwise independent random variables that have the same distribution law as *X* has. Let us assume that *X* has a normal distribution, m_x and σ_x is its mathematical expectation and mean square deviation, accordingly. The variable $\bar{x}$ is random, since it is determined by the sample [1], has a normal distribution (as the sum of normally distributed variables), and also $M\left[\bar{x}\right] = m_x$, $D\left[\bar{x}\right] = \frac{\sigma_x^2}{n}$.

Let us assume that *x* is an arbitrary value of a random variable *X*. Then the variable $y = x - \bar{x}$ characterizes the degree of consistency of the experts' opinions. This variable will be called an absolute error of estimation.

We assume that the interval $\left(x_{\min}, x_{\max}\right)$ (here, $x_{\min}$ and $x_{\max}$ are correspondingly the minimum and maximum values of the random variable *X*) coincides with the interval $\left(u_1, u_2\right) = \left(m_x - 3\sigma_x, m_x + 3\sigma_x\right)$ and $m_x - 3\sigma_x \geq 0$. If the first interval is less than the second one, we stretch it by the appropriate number of times, moving to the new $x_{\min}$ and $x_{\max}$.

Now the distribution density of the random variable *Y* may be found.

3 The absolute error of estimation

Consider a random variable $Y = X - \bar{x}$. Using the composition of two normally distributed random variables, we may obtain the distribution law *Y*. When subtracting random variables, their mathematical expectations are subtracted, and the variance is calculated using the formula:

$$D[Y] = D\left[X - \bar{x}\right] = D[X] + D\left[\bar{x}\right] - 2K_{X,\bar{x}} = \begin{cases} \dfrac{n+1}{n} D_x, \text{ если } x \notin \{x_1, ..., x_n\}, \\ \dfrac{n-1}{n} D_x, \text{ если } x \in \{x_1, ..., x_n\}. \end{cases}$$

The distribution density of the random variable *Y* is equal to:

$$f(y) = f\left(x - \bar{x}\right) = \begin{cases} \dfrac{\sqrt{n}}{\sqrt{2\pi} \cdot \sqrt{n+1}\, \sigma_x} e^{-\frac{\left(x-\bar{x}\right)^2 \cdot n}{2(n+1)\sigma_x^2}}, \text{ если } x \notin \{x_1, ..., x_n\} \\ \dfrac{\sqrt{n}}{\sqrt{2\pi} \cdot \sqrt{n-1}\, \sigma_x} e^{-\frac{\left(x-\bar{x}\right)^2 \cdot n}{2(n-1)\sigma_x^2}}, \text{ если } x \in \{x_1, ..., x_n\} \end{cases}. \tag{1}$$

Note that $\sqrt{n} \approx \sqrt{n+1}$ with an accuracy of 3.3 % if $n \geq n_0 = 40$.

In this formula, $m_y = 0$ and 99.73 % of *Y* values fall within the interval $[u_1, u_2]$.

It is inconvenient to make calculations using the formula (1) because σ_x^2 is unknown. Consider replacing σ_x^2 with $S_x^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$. With such a replacement, there will be a calculation error.

Let us find the number *n*, at which the indicated replacement will take place with the specified accuracy ε and probability α. The problem is reduced to constructing a confidence interval for the variance.

Without violating generality, we compare two methods (described in [1] and [5]) for constructing a confidence interval for the variance.

[1] discusses the approximate method for constructing a confidence interval for the variance σ_x^2 when the number of observations $n \geq 30$. It is the interval $\left(S_x^2 - \varepsilon - t_\beta\sqrt{\dfrac{2}{n-1}}\left(S_x^2 - \varepsilon\right), S_x^2 + \varepsilon + t_\beta\sqrt{\dfrac{2}{n-1}}\left(S_x^2 + \varepsilon\right)\right)$, where ε is the accuracy of the estimate and t_β is the argument of the Laplace function for the confidence probability β. This suggests

$$n \geq \frac{2t_\beta^2\left(S_x^2 + \varepsilon\right)^2}{\varepsilon^2} + 1. \tag{2}$$

[3] discusses the method for constructing a confidence interval for the variance when $n > 30$. The confidence interval looks as follows:

$$\left(\frac{(n-1)S_x^2}{(n-1)+t_\beta\sqrt{2(n-1)}},\ \frac{(n-1)S_x^2}{(n-1)-t_\beta\sqrt{2(n-1)}} \right).$$

Then

$$\left|S_x^2-\sigma_x^2\right| \le \frac{2(n-1)S_x^2 t_\beta\sqrt{2(n-1)}}{(n-1)^2-2t_\beta^2(n-1)} = \frac{2S_x^2 t_\beta\sqrt{2(n-1)}}{n-1-2t_\beta^2} < \varepsilon.$$

In this case,

$$n > \frac{8t_\beta^2 S_x^4}{\varepsilon^2} + 4\cdot t_\beta^2 - 0{,}52\cdot t_\beta^4 + 1. \qquad (3)$$

We obtain a sufficient condition for the second method to prevail over the first one relative to the number of experts.

Let us assume that $\varepsilon = \varepsilon_1 S_x^2$, i.e. ε_1 is the fraction of ε in S_x^2. Then (2) and (3) will correspondingly look as follows:

$$n > \frac{2t_\beta^2 + 4t_\beta^2\varepsilon_1 + 2t_\beta^2\varepsilon_1^2}{\varepsilon_1^2} + 1,$$

$$n > \frac{8t_\beta^2 + 4t_\beta^2\varepsilon_1^2 - 0{,}52\cdot t_\beta^4\varepsilon_1^2}{\varepsilon_1^2} + 1.$$

Now we find the difference Δ between the right-hand members of the latter inequalities (subtracting the first one from the second one). The result is:

$$\Delta = \frac{6t_\beta^2 - 0{,}52\cdot t_\beta^4\varepsilon_1^2 - 2t_\beta^2\varepsilon_1^2}{\varepsilon_1^2}.$$

Hence it is not difficult to demonstrate that $\Delta > 0$ if and only if the following inequality is correct:

$$11\varepsilon_1^2 + 4\varepsilon_1 - 6 < 0.$$

Maximum value $t_\beta^2 = 25$. Then $\Delta > 0$ if and only if

$$\varepsilon_1 < 0{,}579.$$

This means that $\left|S_x^2-\sigma_x^2\right| < 0{,}775 S_x^2$.

Thus, if $\left|S_x^2-\sigma_x^2\right| \ge 0{,}579 S_x^2$ then the number of experts, according to the second method, will be less than those under the first method.

This is a criterion for the second method prevailing over the first one in terms of the number of experts.

We will use the first method.

We assume that $\varepsilon_2' S_x^2 \le \varepsilon \le \varepsilon_2 S_x^2$ and $0 < \varepsilon_2' < \varepsilon_2 < 1$, then the condition (2) is converted into the following:

$$n \ge n_1 = \left\lfloor \frac{2t_\beta^2(1+\varepsilon_2')^2}{\varepsilon_2^2} \right\rfloor + 1. \qquad (4)$$

So, for $\alpha = 0{,}1;\ \varepsilon_2 = 0{,}15;\ \varepsilon_2' = 0{,}09$ we find: $n \geq 292$.

When calculating the relative error of σ_x^2 replacement by S_x^2 we have:

$$\frac{\left|\sigma_x^2 - S_x^2\right|}{\sigma_x^2} \leq \frac{\varepsilon_1}{S_x^2 - \varepsilon_1} = \frac{\varepsilon_2 \cdot S_x^2}{S_x^2 - \varepsilon_2 S_x^2} = \frac{\varepsilon_2}{1-\varepsilon_2}.$$

Thus,

$$\frac{\left|\sigma_x^2 - S_x^2\right|}{\sigma_x^2} \leq \frac{\varepsilon_2}{1-\varepsilon_2}. \tag{5}$$

Inequality (5) is good under $n \geq n_1$ probability $1-\alpha$.

Here is an algorithm for finding *n* when S_x is not specified; iterations are made to determine the value S_x so that the inequality (2) is valid. Let us assume that $n_2 = \max\{n_0, n_1\}$.

Steps of the algorithm (we will call it Algorithm 1).

1. Given: β, ε, n_2 , $n_{\max}$ is the maximum possible number of experts in this situation.
2. Find the right member of the inequality (2) when $S_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$ and $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. For this purpose, estimates of n_2 independent experts are considered.
3. If the right member of (2) obtained in step 2 is not less than n_2 and not more than $n_{\max}$, we get an estimate for the number of experts *n* so that the inequality: $n_2 \leq n \leq n_{\max}$ is valid. Go to the End.
4. If the right member of (2) is less than n_2, then we assume $n_2 = n_2 + 1$. At that, if $n_2 \leq n_{\max}$, move to step 2.
5. Otherwise, it is necessary to decrease the value ε or increase the value β.

If $n \geq n_2 \max\{n_0, n_1\}$ the formula (1) is converted to the formula

$$f_1(y) = f_1\left(x - \bar{x}\right) = \frac{1}{\sqrt{2\pi} \cdot S_x} e^{-\frac{\left(x-\bar{x}\right)^2}{2S_x^2}}, \tag{6}$$

since for these values of *n* $\sigma_x^2 \approx S_x^2$ with probability $1-\alpha$ and with accuracy ε .

It may be demonstrated that the relative error of formula (5) does not exceed

$$\varepsilon_3 = \frac{1}{2\sigma_x}\left|-1 + \frac{(x-\bar{x})^2}{\sigma_x \cdot (n+1)}\right| \cdot \frac{\varepsilon_2}{1-\varepsilon_2}. \tag{7}$$

with probability β .

4 The Relative Error of Estimation

When estimating the errors made by experts, an important role is assigned to the relative error of estimation; such an error may be defined either as $\delta = \frac{\left|x-\bar{x}\right|}{\bar{x}}$ or by $\delta_1 = \frac{x-\bar{x}}{\bar{x}}$ (the case $\frac{\bar{x}-x}{\bar{x}}$ is symmetric to δ_1).

Let us assume that $\delta = \frac{\left|x-\bar{x}\right|}{\bar{x}}$ is a relative error. If $\frac{\left|x-\bar{x}\right|}{\bar{x}} > 1$, then the estimate x is not consistent with $\bar{x}$; if $\frac{\left|x-\bar{x}\right|}{\bar{x}} \le 1$, then x is consistent with $\bar{x}$ and the better the smaller the relative error is.

If δ is no more than 3 %, then the estimate x has increased accuracy; if δ falls within the range from 3 % to 10 %, then this is the usual accuracy; δ from 10 % to 20 % results in an approximate estimate [2].

Consider $\delta_1 = \frac{x-\bar{x}}{\bar{x}}$ as a relative error. This is the value of a random variable Δ_1, represented by the ratio of variables $Y = X - \bar{x}$ and $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, which have a normal distribution, moreover, $M[Y] = 0$, $M\left[\bar{x}\right] = m_x$, if $n \ge n_2$ with probability $1-\alpha$ and with accuracy $\frac{\varepsilon}{n} D\left[\bar{x}\right] \approx \frac{S_x^2}{n}$.

One may replace an unknown value m_x with an exact estimate $\bar{x}$ that meets the condition:

$$\bar{x} - \frac{S_x}{\sqrt{n}} t_{\beta,n-1} < m_x < \bar{x} + \frac{S_x}{\sqrt{n}} t_{\beta,n-1}, \tag{8}$$

where $t_{\beta,n-1}$ is the argument of the Student function $\theta(t,k)$, which is such that $\theta\left(t_{\beta,n-1}, n-1\right) = \beta = 1-\alpha$ and

$$n \ge n_3 = \max\left\{ n_2, \frac{S_x^2 t_{\beta,n_2-1}^2}{\varepsilon_4^2} \right\}. \tag{9}$$

The relative error of estimate m_x:

$$\varepsilon_6 = \left|\frac{\bar{x} - m_x}{m_x}\right| \le \frac{\varepsilon_4}{m_x} = \frac{\varepsilon_4}{\bar{x} - \varepsilon_4},$$

where $\varepsilon_4 = \bar{x} \cdot \varepsilon_5$ and $0 < \varepsilon_5 < 1$, i.e.

$$\varepsilon_6 \le \frac{\bar{x}\varepsilon_5}{\bar{x} - \bar{x}\varepsilon_5} = \frac{\varepsilon_5}{1-\varepsilon_5}. \tag{10}$$

[6] proved that δ has a normal distribution if $n \ge n_3$ and $\varepsilon_2' S_x^2 \le \varepsilon \le \max\left\{\varepsilon_2 S_x^2, \bar{x}\varepsilon_5\right\}$,

while

$$f_2(\delta_1) = \frac{1}{\sqrt{2\pi} \cdot S_x \cdot c} e^{-\frac{(\delta_1)^2}{2S_x^2 c^2}}, \tag{11}$$

where

$$c = \frac{1}{m_{\bar{x}}} + \frac{1}{(m_{\bar{x}})^3} \cdot \frac{S_x^2}{n}. \tag{12}$$

Then

$$P(-d < \delta_1 < d) = \Phi\left(\frac{d}{S_x \cdot c}\right). \tag{13}$$

So, formulas (11) to (13) give a fairly accurate value of the distribution density of the relative estimation error, its average value, the spread, and the probability of falling within the range $(-d, d)$.

Here is the algorithm for solving these problems.

1. First, Algorithm 1 is applied. If it ends with Go to the End, move to step 2.
2. n_3 is calculated.
3. If n_2 is equal to n_3, then $n \geq n_2$. Move to 4.

Otherwise, we assume $n_2 = n_2 + 1$ and move to step 2 of Algorithm 1.

4. Calculation of the spread of the relative estimation error.

4.1. First, we calculate "c" using the formula (12).

4.2. Use the right member of (12) for $n = n_1$.

5. Calculation of probability of the relative error falling within the interval $(-d, d)$ where d is specified.

5.1. "*c*" is calculated using the formula (12).

5.2. The value of the Laplace function at a point $\frac{d}{S_x c}$ is calculated.

5 Practical Implementation

When we tested students in the probability theory and mathematical statistics at the Tver State Agriculture Academy, the maximum score was 10 points, taking into account the complexity of the test. We used prompts (no more than 5); each of the prompts reduces the score by 0.7 *k* points, where *k* is the number of useful prompts. The maximum number of prompts, which is equal to 5, was determined based on these trial tests that allowed consulting. The decrease in the score occurs in arithmetic progression. The parameter 0.7 is selected from the condition of the maximum "penalty" for a prompt because in this case the score is reduced for five prompts to the maximum level. If the answer to this task is incorrect, two approaches are possible. The first (standard) approach: a score of 0 points is graded for the task regardless of the number of prompts the student used. The second approach is to use estimates of ontologies, various fragments of this task, and methods of fuzzy control in an adequate system for estimating the quality of teaching.

When selecting the volume of n=300 test tasks, on average, the group of students got: $\overline{x} = 5{,}1$; $S_x^2 = 3{,}06$ and $n \geq 292$ if $\varepsilon = 0{,}27$.

When statistical data were approximated by the normal distribution law, the significance level was 0.1.

Note that the results of this article are valid for traditional testing without prompts as well. At the same time, well-defined prompts greatly contribute to the use of the test not only for monitoring but also for training, i.e. they increase the teaching potential of test tasks.

6 Conclusion

Creating reliable and high-quality methods of making collective decisions by experts is a significant top-of-the-agenda topic of modern research in the field of complex systems modeling. This issue is crucial for remote testing of learners.

The developed method enables to obtain quantitative estimates of the required number of experts to make a joint decision of a given quality.

This method may be used not only for testing learners, but also in diagnostic systems, product quality control, and other areas as well.

References

1. Ventcel', E.S.: Probability Theory: Textbook. JUSTICE (2018).
2. Ganicheva, A.V.: Mathematical Models and Methods for Evaluating Events, Situations, and Processes: Textbook. Saint Petersburg: «Lan'» (2017).
3. Mihin, M.N.: Mathematical Statistics: Textbook. Moscow: MIREA (2016). (2016).
4. Simankov, V.S., Buchackaya, V.V., Teplouhov, S.V.: Determination of the Optimal Combination of the Confidence Interval and Confidence Probability. The Bulletin of the Adyghe State University. Series 4: Natural-Mathematical and Technical Sciences. 3 (246), 69-74 (2019). (In Russian)
5. Suvorova, A.V., Pashchenko, A.E., Tulup'eva, T.V., Tulup'ev, A.L.: Construction of Confidence Intervals for Assessing the Intensity of Risky Behavior Based on Chebyshev's Inequality. SPIIRAS Proceedings. 10, 96-109 (2009). DOI: 10.15622/sp.10.6.
6. Ganicheva, A.V.: Test Technologies in Training. Tver: TGSHA (2011).
7. Halikova, K.S., Ryzhkova S.K.: Assessment of the Influence of Factors Based on Cognitive Modeling and Expert Assessment. Humanities Research. 2 (54), 300-303. (2016).
8. Boiko, Ye.: Methods of Forming an Expert Assessment of the Criteria of an Information System for Managing Projects and Programs Technology Transfer. Fundamental Principles and Innovative Technical Solutions. 2, 9-11 (2018). DOI: http://dx.doi.org/10.21303/2585-6847.2018.00766.
9. Bonett, D.G.: Robust Confidence Interval for a Ratio of Standard Deviations. Applied Psychological Measurement. 30(5), 432-439 (2006). DOI: 10.1177/0146621605279551.

10. Chakraborty, D.: Structural Quantization of Vagueness in Linguistic Expert Opinions in an Evaluation Program. Fuzzy Sets and Systems. 119 (1), 171-186 (2001) DOI: 10.3745/KIPSTB.2005.12B.2.203.
11. Safargaliev, E.R., Eremina, I.I., Savitsky, S.K., Camelina, V.A.: Mathematical Model and Qualimetric Assessment of Graduate Education Quality in Environment Saturated with Information and Communication Technologies. International Education Studies. 8(2), 78-83 (2015). DOI: 10.5539/ies.v8n2p78.
12. Thulin, M.: On Split Sample and Randomized Confidence Intervals for Binomial Proportions. Statistics & Probability Letters. 92, 65-71 (2014). DOI: 10.1016/j.spl.2014.05.005.