

# On the Borders of Applicability of the Weighted Mean in Processing Measurement Results

E. Eremin

Division of development of normative documentation, ZAO "Centr MO" JSC "Transneft",  
Dobrolyubova street, no. 1, body 16, Moscow, 127254, Russia  
Email: eremin@centermo.ru

**A theorem on the law of large numbers for the estimator of the mean weighted first (initial) moment of the joint distribution of independent random variables is proved. The conditions for statistical stability of this moment and consistency of its mean weighted estimator are established.**

**Keywords:** estimated (measured) physical quantity, mean weighted, random variables, consistency.

## 1. INTRODUCTION

IN BOTH physical and metrological investigations it is necessary to utilize all the available information about the estimated (measured) physical quantity. This leads to the problem of combining the obtained (available) results of measurements  $X_1, X_2, \dots, X_N$  for  $N$  groups (series) of observations. A quite substantial literature is devoted to the general methods and practical examples of solving this problem; we cite just [1, 2] as the most typical ones.

Thus in [1 and other sources] based on the results of  $X_1, X_2, \dots, X_N$  (usually the arithmetic means of the obtained measurement results in individual groups of corresponding sizes), a construction of a more exact estimator  $\tilde{X}_N$  is suggested to be implemented, based on the linear function of the form

$$\tilde{X}_N = a_1 X_1 + a_2 X_2 + \dots + a_i X_i + \dots + a_N X_N = \sum_{i=1}^N a_i X_i \quad (1)$$

where  $a_i, i=1, \dots, N$  are the coefficients whose values are determined (established) from the condition of attaining the minimum of the variance of the estimator  $\tilde{X}_N$  itself. Moreover, proceeding from the condition of unbiasedness of the estimator (1) in the case of *equal mathematical expectations* of the random variables  $X_i$ , we obtain a condition (restriction) for the sum of coefficients

$$\sum_{i=1}^N a_i X_i = 1.$$

Estimations of the form (1) are of the mean weighted type; the conditions under which the mean weighted first (initial) moment itself of the joint distribution of independent random variables  $X_1, X_2, \dots, X_N$  will be statistically stable (exist) as well as the condition under which the mean weighted estimator of this moment of form (1) will be consistent in [3] are certain.

## 2. SUBJECT AND METHODS

It is known that the estimator  $\tilde{X}_N$  of the form (1) is consistent if, as  $N \rightarrow \infty$ , it converges in probability to the parameter  $E\{\tilde{X}_N\}$  - the mathematical expectation of the variable  $\tilde{X}_N$ , i.e., if for all  $\varepsilon > 0$  the condition

$$Pr\left\{\tilde{X}_N - E\{\tilde{X}_N\} < \varepsilon\right\} \rightarrow 1, \quad N \rightarrow \infty \quad (2)$$

is fulfilled.

Here the primary and the necessary condition is the *existence* of the  $E\{\tilde{X}_N\}$  as  $N \rightarrow \infty$ , i.e., of the finite limit

$$\lim_{N \rightarrow \infty} E\left\{\sum_{i=1}^N a_i X_i\right\} = \lim_{N \rightarrow \infty} \sum_{i=1}^N a_i E\{X_i\} = R, \quad R < \infty, \quad (3)$$

which is actually the condition for the existence of the mean weighted first (initial) moment of the joint distribution of independent random variables  $X_1, X_2, \dots, X_N$ .

If the random variables  $X_i$ , appearing in the estimator of the form (1), possess finite variances  $D\{X_i\}$ ,  $0 < c_i \leq D\{X_i\} \leq C_i$ , then they also possess bounded mathematical expectations  $|E\{X_i\}| \leq A_i$ .

Set  $A = \max_{i \in \{1, \dots, N\}} \{A_i\}$ , then  $|E\{X_i\}| \leq A_i$  and condition (3) for

the existence of the finite value of  $E\{\tilde{X}_N\}$  as  $N \rightarrow \infty$  becomes

$$\begin{aligned} \lim_{N \rightarrow \infty} \sum_{i=1}^N a_i E\{X_i\} &\leq \lim_{N \rightarrow \infty} \left| \sum_{i=1}^N a_i E\{X_i\} \right| \leq \lim_{N \rightarrow \infty} \sum_{i=1}^N |a_i E\{X_i\}| = \\ &= \lim_{N \rightarrow \infty} \sum_{i=1}^N |a_i| \cdot |E\{X_i\}| \leq A \cdot \lim_{N \rightarrow \infty} \sum_{i=1}^N |a_i| = R', \quad R' < \infty. \end{aligned} \quad (4)$$

The simplest and the basic notions of the theory of limits yield

the necessity of admitting the condition

$$\sum_{i=1}^N |a_i| = k = Const \quad (5)$$

for the fulfilment of (4).

If  $X_1, X_2, \dots, X_N$  are independent, having finite variances

$$D\{X_i\} = E\{[X_i - E\{X_i\}]^2\}, \quad 0 < D\{X_i\} \leq C_i,$$

then for the estimators of the form (1) and the values of the weights  $|a_i| = |a_j| = k/N$  for any  $\varepsilon > 0$ , we have

$$\lim_{N \rightarrow \infty} Pr\{ |(a_1 X_1 + \dots + a_N X_N) - (a_1 E\{X_1\} + \dots + a_N E\{X_N\})| < \varepsilon \} = 1$$

(6)

provided  $\sum_{i=1}^N |a_i| = k = Const$  is fulfilled.

Hence to prove (6), we use the Bienayme-Chebyshev inequality which connects the probability of deviation of a random variable  $\xi$  from its mathematical expectation  $E\{\xi\}$  by the amount  $x > 0$  and the variance  $D\{\xi\} = E\{\xi - E\{\xi\}\}^2$  as follows

$$Pr\{ |\xi - E\{\xi\}| < x \} \geq 1 - \frac{D\{\xi\}}{x^2}. \quad (7)$$

Set  $C = \max_{i \in [1, \dots, N]} \{C_i\}$ , then  $D\{\tilde{X}_N\} = \sum_{i=1}^N D\{X_i\} \leq NC$ .

For  $|a_i| = |a_j| = k/N$  and taking into account that  $a^2 = |a|^2$ , we obtain from (7) that for any  $\varepsilon > 0$  we shall have

$$1 \geq Pr\{ |\tilde{X}_N - E\{\tilde{X}_N\}| < \varepsilon \} \geq 1 - \frac{k^2 D\{\tilde{X}_N\}}{\varepsilon^2 N^2} \geq 1 - \frac{k^2 C}{\varepsilon^2 N}. \quad (8)$$

From (8) for  $k = Const$  and  $N \rightarrow \infty$  we arrive at (6).

### 3. RESULTS

Consider now consistency of the estimators of the form (1) under general assumptions on the values of weight coefficients. Let us take advantage of an inequality (7), which takes

$$D\{\tilde{X}\} = \sum_{i=1}^N a_i^2 D\{X_i\} \leq C \sum_{i=1}^N a_i^2$$

into account, where

$C = \max_{i \in [1, \dots, N]} \{C_i\}$ , becomes

$$1 \geq Pr\{ |\tilde{X}_N - E\{\tilde{X}_N\}| < x \} \geq 1 - C \sum_{i=1}^N a_i^2 / x^2. \quad (9)$$

We shall utilize the connection between the absolute values of the weights and represent the absolute value of an individual, for example the  $a_N$ th, weight in the form  $|a_N| = p \cdot k/N$ , where  $0 < p < N$  and the values of  $|a_i|$  for the remaining  $i = 1, \dots, (N-1)$  are assumed to be equal. Since

$$\sum_{i=1}^{N-1} |a_i| = k - p \cdot k/N \text{ we obtain } |a_i| = k(N-p)/N(N-1).$$

Next, after some straightforward transformations, we have

$$\sum_{i=1}^N a^2 = \sum_{i=1}^N |a_i|^2 = |a_N|^2 + \sum_{i=1}^{N-1} \left[ \frac{k(N-1)}{N(N-1)} \right]^2 = k^2 \left[ \frac{1}{N} + \frac{(p-1)^2}{N(N-1)} \right]$$

and for (9) we obtain

$$1 \geq Pr\{ |\tilde{X}_N - E\{\tilde{X}_N\}| < x \} \geq 1 - \frac{Ck^2}{x^2} \left[ \frac{1}{N} + \frac{(p-1)^2}{N(N-1)} \right].$$

In order that (2) be fulfilled, it is necessary that

$$\lim_{N \rightarrow \infty} (p-1)^2 [N(N-1)]^{-1} \rightarrow 0,$$

which, in turn, requires that  $(p-1)^2 = o\{N(N-1)\}$  be satisfied [3]. The last condition is fulfilled if, for example,

$$(p-1)^2 \leq [N(N-1)]_r^{-1}, \quad r > 1. \quad (10)$$

Since the inequality (10) should be fulfilled for the differences of the weights values, we obtain

$$(|a_N| - |a_i|)^2 = \frac{k^2 (p-1)^2}{(N-1)^2} \leq k^2 [N(N-1)]_r^{-1-2r}, \quad r > 1,$$

whence to assure consistency of the estimator of the form (1) in the general case follows the requirement concerning the maximal spread of the values of the weight coefficients in the

$$\text{form } \max_{i \in [1, \dots, N]} \{|a_i|\} - \min_{i \in [1, \dots, N]} \{|a_i|\} \leq k \cdot [N(N-1)]_r^{-1-2r}, \quad r > 1.$$

4 CONCLUSION

General properties of consistency of statistical estimators of the form (1) and thus conditions for its applicability as such for the processing of, for example, unequally spaced (along random as well as systematic errors) results of measurements are established: if random variables  $X_i$  appearing in the estimator of the form (1) possess finite variances  $D(X_i) = E\{X_i - E\{X_i\}\}^2$ , then provided the conditions for nonrandom weight coefficients

$$\left. \begin{aligned} \max_{i \in [1, \dots, N]} \{a_i\} - \min_{i \in [1, \dots, N]} \{a_i\} &\leq k \cdot [N(N-1)^{1-2r}]^{\frac{1}{2r}}, \quad r > 1 \\ \sum_{i=1}^N |a_i| &= k = Const \end{aligned} \right\}$$

We note the following generalization. The consistency property is required for any estimation rule; it is, however, an asymptotic property and it is not connected with the properties of estimators for a fixed sample size (unlike unbiasedness and optimality). In the direct and literal sense, it is thus

impossible to provide specific practical recommendations to "assure" consistency of estimators of the form (1) for any fixed sample sizes (specific values of  $N$ ) since consistency is an asymptotic property and the limit defining this property is of a probabilistic nature. However, the results obtained serve as a kind of "warning" against excessive "optimism" in utilizing estimators of the type (1) to increase the precision of processing measurement results: to assure its statistical stability (consistency), the spread of the values of the weight coefficients should increase no faster than  $\sim \sqrt{N}$ .

REFERENCES

- [1] Vatutin, V., Televinova, T., Chistyakov, V. (1985). Probabilistic Methods in Physical Investigations. (in Russian). Nauka, Moscow.
- [2] Tyurin N. (1982). Introduction to Metrology. (in Russian). Izdatelstvo Standartov, Moscow.
- [3] Eremin E. (1999). Law of large numbers for the weighted mean. *Measurement Techniques*, Consultant Bureau, New York. 42(7), 635-642.