SUPPLEMENT I
Estimation of the admissible interval for an unknown constant value under conditions of uncertainty

Example 1. Consistent homogeneous sample

Estimation of the constant value is carried out. The value is a reference weight. The true value is $G^* = 0.25$ gr (in grams). The model of the error contains only the absolute component. The probability distribution law or any other statistical characteristics of the error are absent. The sample is homogeneous, *i.e.*, the maximal value of the error is the same for all measurements. The constraint on its value is $\varepsilon_{\max} = 0.1$ gr.

The following sample of the length $N = 12$ is given (the errors are obtained by the standard random number generator from the interval $[-\varepsilon_{\max}, \varepsilon_{\max}]$):

$\{x_i, i = 1,...,12\} =$
= {0.2910, 0.2570, 0.2660, 0.2280, 0.2110, 0.3050,
0.1550, 0.3020, 0.3130, 0.2870, 0.3390, 0.2330} gr.

The bounds of uncertainty set $H_i = [h_{\mathrm{lo}i}, h_{\mathrm{up}i}]$, i=1,12 for each measurement are calculated and take the values

[0.1910, 0.3910], [0,1570, 0,3570], [0.1660, 0.3660], [0.1280, 0.3280],
[0.1110, 0.3110], [0.2050, 0.4050], [0.0550, 0.2550], [0.2020, 0.4020],
[0.2130, 0.4130], [0.1870, 0.3870], [0.2390, 0.4390], [0.1330, 0.3330] gr.

The auxiliary values of boundaries of the admissible interval are

$h_{\min} = \max_i \{h_{\mathrm{lo}i}\} = 0.2390$ gr, by the measurement No.11,

$h_{\max} = \min_i \{h_{\mathrm{up}i}\} = 0.2550$ gr, by the measurement No.7.

Since evidently $h_{\min} < h_{\max}$, the sample is consistent and its informational set is not empty. The bounds of the segment $I(G)$ of admissible values of the constant directly coincide with the auxiliary values

$I = [h_{\min}, h_{\max}] =$ [0.2390, 0.2550] gr.

Estimation of the central value is

$x_c = (h_{\min} + h_{\min})/2 =$ 0.2470 gr.

The maximal deviation is

$\Delta x = (h_{\max} - h_{\min})/2 = 0.0080$ gr.

Deviations of measurements relatively the central estimation are

0.0440, 0.0100, 0.0190, –0.0190, –0.0360, 0.0580,
–0.0920, 0,0550, 0.0660, 0.0400, 0.0920, –0.0140 gr.

Using the maximal measurement No.11, $x_{\max} =$ 0.339 gr and minimal measurement No.7, $x_{\min} =$ 0.155 gr, the minimal limit of the error in the given sample is

$\Delta^*_{\max} =$ (0.339 – 0.155)/2 = 0.092 gr,

and the limit value of estimation of the unknown constant is

$x^* =$ (0.339 + 0.155)/2 = 0.247 gr.

The sample and results of calculations are shown in Fig.SI.1. The horizontal line marks the true level of the constant to be evaluated. Crosses show the measurements. The vertical segments presents the uncertainty sets of measurements. The uncertainty sets of the minimal and maximal

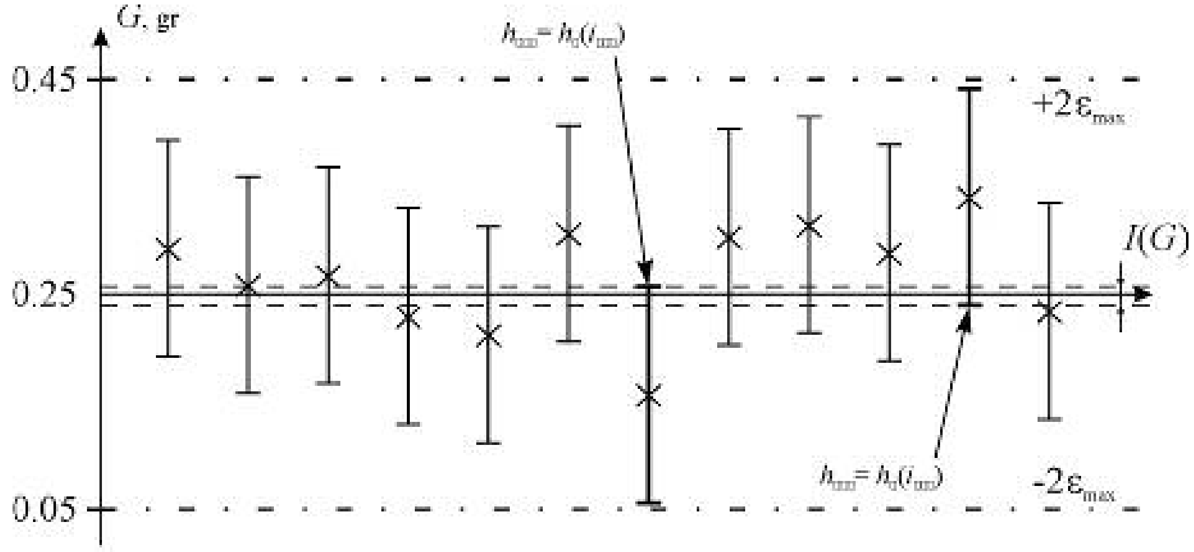

Figure SI.1: Consistent sample of measurements of a constant value

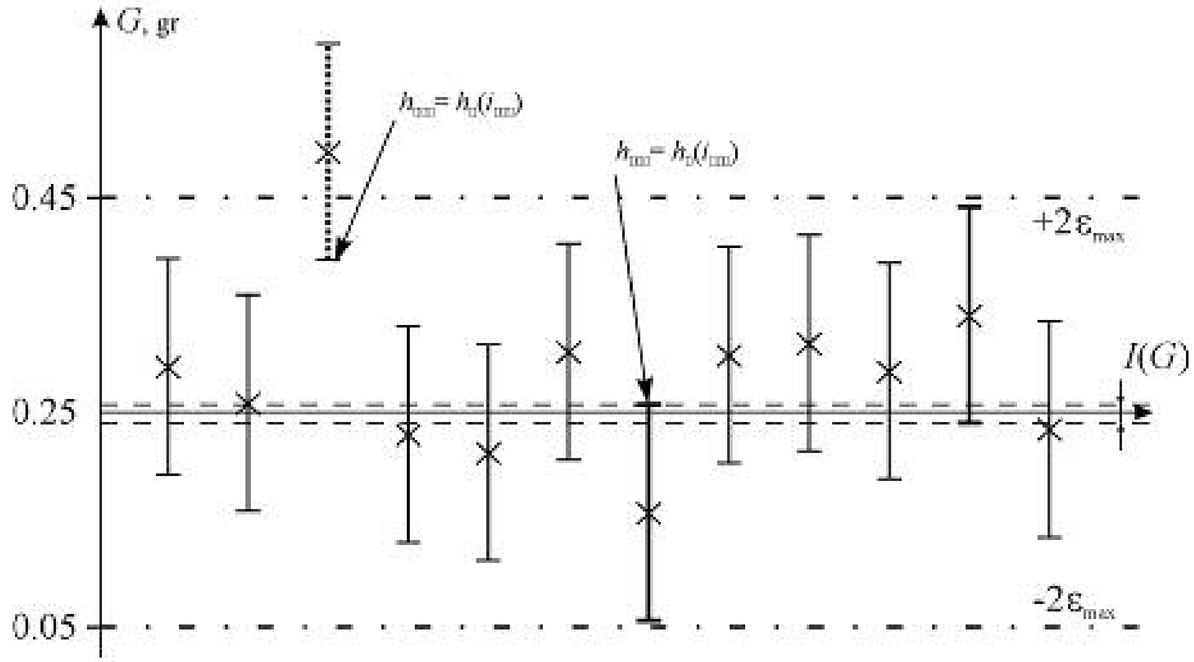

Figure SI. 2: Inconsistent sample of measurements of a constant value

measurements are marked with thick vertical segments, the arrows show their bounds which determine the bounds of the uncertainty set, *i.e.*, the segment $I(G)$ of the admissible values of the constant. The lines in dashes (the bounds of the *IS*) allow to compare the original uncertainty sets with the span of the *IS* $I(G)$. The dotted-dashed lines mark the demonstrational bounds of the segment $\pm 2\varepsilon_{max}$ of two maximal values of the error around its true value.

For comparison application of the standard relations of the corresponding GOST to processing the given sample is implemented. It can be only realized in a formal way, because there no any probability information about the error and the sample is rathert short.

The mean is calculated

$$\bar{x} = (\sum_{i=1,N} x_i)/N = 0.2655 \text{ gr},$$

the square mean is

$$\sigma = \sqrt{(\sum_{i=1,N}(x_i - \bar{x})^2)/(N-1)} = 0.0516 \text{ gr}.$$

Deviations $\{\gamma_i = x_i - \bar{x},\ i = 1, N\}$ of measurements relatively the mean are

0.0254, –0.0085, –0.0004, –0.0375, –0.0545, 0.0394,
–0.1105, 0,0364, 0.0474, 0.0214, 0.0734, –0.0325 gr.

Note that the estimation $\bar{x} = 0.2655$ gr, determined by the GOST is not consistent with the exact estimation given by the informational set [0.2390, 0.2550] gr, so as $\bar{x} \notin I(G)$.

It can be seen also that the square mean exceeds in three times the maximal deviation $\Delta x = 0.0080$ gr.

Using formally the restriction $2\sigma = 0.1033$ gr, the standard "confidence segment" is

$$[\bar{x} - 2\sigma, \bar{x} + 2\sigma] = [-0.1622, 0.3688] \text{ gr}.$$

It significantly exceeds the span of the informational set.

Moreover, formal application of the GOST "two σ"-rule to detection of the gross errors leads to improper elimination of the authentic measurement No.7, because its deviation exceeds the value 2σ. As a result, the improperly truncated sample (without the measurement No.7) becomes additionally shifted to the improper direction. The mean in this case is equal to $\bar{x} = 0.2756$ gr, and the square mean is $\sigma = 0.0394$ gr.

Example 2. Inconsistent homogeneous sample

Similarly to Example 1, estimation of the constant value is carried out. The value is a reference weight. The true value is $G^* = 0.25$ gr (in grams). The model of the error contains only the absolute component. The probability distribution law or any other statistical characteristics of the error are absent. The sample is homogeneous, *i.e.*, the maximal value of the error is the same for all measurements. The constraint on its value is $\varepsilon_{max} = 0.1$ gr.

The following sample of the length $N = 12$ is given (the errors are obtained by the standard random number generator from the interval $[-\varepsilon_{max},\ \varepsilon_{max}]$):

$\{x_i,\ i = 1,...,12\} =$

={0.2910, 0.2570, **0.4950**, 0.2280, 0.2110, 0.3050,
0.1550, 0.3020, 0.3130, 0.2870, 0.3390, 0.2330} gr.

But in this sample, the measurement No.3 (outlined in thick and its uncertainty set is marked by the thick dashed vertical line) was artificially made with the error to be "suspicious" (Fig.SI.2).

The bounds of uncertainty set $H_i = [h_{\text{lo}i},\ h_{\text{up}i}]$, i=1,12 for each measurement are calculated

[0.1910, 0.3910], [0,1570, 0,3570], **[0.3950, 0.5950]**, [0.1280, 0.3280],
[0.1110, 0.3110], [0.2050, 0.4050], [0.0550, 0.2550], [0.2020, 0.4020],
[0.2130, 0.4130], [0.1870, 0.3870], [0.2390, 0.4390], [0.1330, 0.3330] gr.

The auxiliary values are

$h_{\min} = \max_i \{h_{\mathrm{lo}i}\} =$ **0.3950** gr, **by the false measurement No.3,**

$h_{\max} = \min_i \{h_{\mathrm{up}i}\} = 0.2550$ gr, by the measurement No.7.

Since $h_{\min} > h_{\max}$, the input sample is inconsistent. The additional analysis is necessary. By the corresponding procedure, the table of pair-incidence flags is calculated (Table S1.1).

The analysis of the Table rows shows that the whole zero rows are absent. Thus, the input sample does not contain single (separate) outliers.

Appling the program of partition the sample into a collection of consistent sub-samples, two of them are obtained: the consistent sub-sample that is composed of the measurement numbers {1,3,6,8,9,11} and the consistent sub-sample of the maximal length consisting the measurements of numbers {1,2,4,5,6,7,8,9,10,11,12}.

Table SI.1. The table of pair-incidence flags

i/*j*	1	2	3	4	5	6	7	8	9	10	11	12
1		1	1	1	1	1	1	1	1	1	1	1
2	1		0	1	1	1	1	1	1	1	1	1
3	1	0		0	0	1	0	1	1	0	1	0
4	1	1	0		1	1	1	1	1	1	1	1
5	1	1	0	1		1	1	1	1	1	1	1
6	1	1	1	1	1		1	1	1	1	1	1
7	1	1	0	1	1	1		1	1	1	1	1
8	1	1	1	1	1	1	1		1	1	1	1
9	1	1	1	1	1	1	1	1		1	1	1
10	1	1	0	1	1	1	1	1	1		1	1
11	1	1	1	1	1	1	1	1	1	1		1
12	1	1	0	1	1	1	1	1	1	1	1	

Remark: The trivial cells, which do not participate in the analysis, are shadowed in gray.

The truncated consistent sub-sample of the maximal length with number {1,2,4,5,6,7,8,9,10,11,12} is processed and all output data are calculated like in Example 1:

– the informational set is evidently the same $I(G)$= [0.2390, 0.2550] gr;

– the central estimation is evidently the same $x_c = 0.2470$ gr;

– the maximal deviation is evidently the same $\Delta x = 0.0080$ gr.

Apply now the standard relations from the GOST to the consistent sub-sample of the maximal length. Elimination of the measurement No.3 and a formal processing the sample of measurements {1,2,4,5,6,7,8,9,10,11,12} gives the mean $\bar{x} = 0.2655$ gr. Note that again this estimation is not consistent with the informational set [0.2390, 0.2550] gr since it is out of the *IS*.

For the truncated sample the square mean is $\sigma = 0.0516$ gr, and its double value is $2\sigma = 0.1033$ gr. Deviations from the mean $\bar{x}$ are

0.0254, –0.0085, (Nos.4 – 6): –0.0375, –0. 0545, 0.0394,
(Nos.7–12): –0.1105, 0.0364, 0.0474, 0.2145, 0.0734, –0.0325 gr.
It is seen that that the formal application of the standard rules leads to improper elimination of the authentic measurement No.7.

Thus, Examples 1 and 2 show that for the sample of a short length and with a high level of the errors really appeared in the sample, the approach on the basis of the information sets allows to implement more sophisticated analysis and gives the better results (in the accuracy) than the standard procedures on the basis of the classic statistical methods.