АВТОМАТИЗИРОВАННЫЕ СИСТЕМЫ НАУЧНЫХ ИССЛЕДОВАНИЙ

A.A. Solopekina, A.A. Sytnik, I.V. Gvozduk

UNCERTAINTY INTERVAL EVALUATION

(Yuri Gagarin State Technical University of Saratov, Saratov, Russia)

As it is known, a random variable *M* characterizing the measurement process can be associated with a measurement interval and, consequently, with the quality of results, therefore the measure. We introduce the confidence level that can be attributed to the occurrence of each single event associated with the variable M in the space of all possible measurement results $S = \{m_{\min} \le M \le m_{\max}\}$.

So, it is possible to assign the highest confidence level, equal to one by convention, when we have the certainty that *M* belongs to *S*; vice versa, the confidence level is minimum, equal to zero by convention, when the values of *M* do not belong to *S*.

Considering a subinterval $[m_a, m_b]$ of *S*, it is possible to assign a probability to the confidence level associated with the occurrence of *M* in $[m_a, m_b]$.

From these assumptions, the random variable *M* is characterized by a probability distribution, that is a function of random events that represent the probability that the measurement belongs to one of the possible subintervals of *S*. The probability distribution associated with *M* is all that is known in the measurement interval.

According to the GUM [1, 2] we introduce:

$$P\{|M - E\{M\}| \le ku_M\} = P\{E\{M\} - k\, u_M \le M \le E\{M\} - k\, u_M\} = p \qquad (1)$$

Eq. (1) represents the probability that the measure *M* is between its expected value $E\{M\}$ plus or minus a quantity given by the product of the standard uncertainty u_M and the coverage factor *k*. The parameter *p*, denoted as confidence level, should tend to one to have a high value of the occurrence of an event.

The interval:

$$E\{M\} - k\,u_M \le M \le E\{M\} + k\,u_M \qquad (2)$$

represents the confidence interval and it can be interpreted as that interval able to guarantee a high probability that it contains a large number of possible values of *M*. Hence a rise of the value of *p* leads to an increase of the number of events in which *M* is within the interval.

If the probability density function $f_M(m)$ of *M* is known, it is possible to evaluate the confidence level by means of the following expression:

$$p = \int_{E\{M\}-k\,u_M}^{E\{M\}+k\,u_M} f_M(m)\,dm\,. \quad (3)$$

It is now possible to indicate, explicitly, the measurement result as “uncertainty interval” associated with a measurand with an assigned confidence level *p*.

So, if we suppose to know the probability density, its distribution function $F_M(m)$ is also known, given by its integral. Therefore, the uncertainty interval with confidence level p is defined by the equation:

$$P\{m_\alpha \le M \le m_{p+\alpha}\} = \int_{m_\alpha}^{m_{p+\varepsilon}} f_m(m)\,dm = F_M(m_{p+\alpha}) - F_M(m_\alpha) = p \quad (4)$$

where α is an appropriate value in the range [0, 1]. The extremes of the interval within which *M* is enclosed takes the name of quantiles of the distribution function F_M, and we have the following relationship:

$$F_M(m_\alpha) = P\{M \le m_\alpha\} = \alpha \quad (5)$$

As introduced in [3], taking again into account n independent successive observations $(o_1,...,o_n)$ and assuming each observation as a normally distributed random variable with expected value m_0 and standard uncertainty u_0, the chi-square distribution $(n-1)$ degrees of freedom can be represented by:

$$\chi^2_{n-1} = \frac{\sum_{i=1}^{n}(o_i - \overline{o})^2}{u_0^2} \quad (6)$$

being the mean of such variables $\overline{o} = \sum_{i=1}^{n} o_i \Big/ n$ also normally distributed with mean value m_0 and reduced variance u_0^2 / n [4].

The uncertainty interval [5] can be introduced by considering the Chi-square probability distribution with associated n degrees of freedom. With the pre-arranged confidence level p, this interval is defined as:

$$P\{\chi^2_\alpha \le \chi^2_\nu \le \chi^2_{p+\alpha}\} = F_\nu(\chi^2_{p+\alpha}) - F_\nu(\chi^2_\alpha) = p\,, \quad (7)$$

where α is a value in the range from zero to $(1-p)$; the extremes of the interval χ^2_α and $\chi^2_{p+\alpha}$ are, respectively, the $\alpha-$ and $(p+\alpha)-$ quantiles of the distribution function of χ^2_ν, whose cumulative distribution is given by:

$$F_\nu(m) = P\{\chi^2_\nu \le m\} = \int_0^m f_\nu(z)\,dz, \quad (8)$$

A $\beta-$quantile is an m_β so that $F_\nu(m) = \beta.$ Such quintiles are tabulated for different values of degrees of freedom ν corresponding to the respective β but they can be obtained more efficiently by means of specific statistic software.

Table 1 summarizes the results concerning the amplitude of the uncertainty interval with α = 0.025 ÷ 0.005 and ν = 1 ÷ 100 according to Eq. (7). Consequently, four histograms of 10^5 random generated observed values for different degrees of freedom fitted with Gaussian distribution can be obtained, as shown in Fig. 1. For each case, the mean and the standard deviation are also computed [3, 6].

Table 1. Uncertainty interval amplitude in function of ν and α

ν	$\chi^2_{0.025}$	$\chi^2_{0.975}$	$\chi^2_{0.005}$	$\chi^2_{0.995}$
1	0.000982	5.024	0.0000393	7.879
2	0.0586	7.378	0.01	10.597
5	0.831	12.832	0.412	16.750
10	3.247	20.483	2.156	25.188
20	9.951	34.170	7.434	39.997
50	32.357	71.420	27.991	79.490
100	74.222	129.561	47.328	140.169

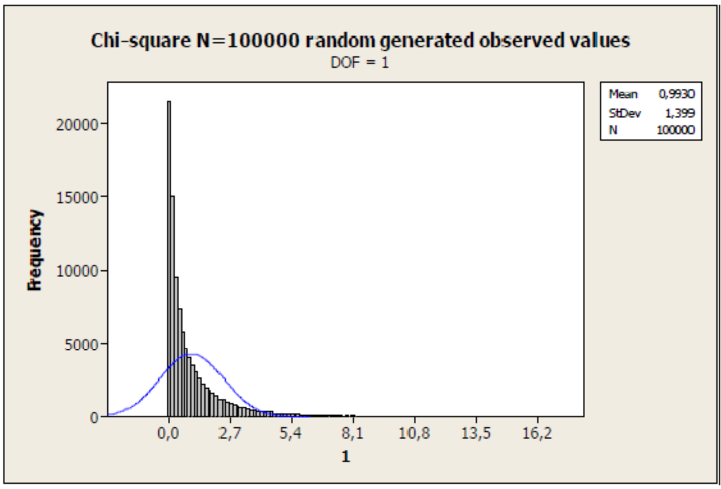
The ratio of two independent chi-square variables, each divided by its respective degrees of freedom, is a random variable F_{ν_1,ν_2} defined as follows:

$$F(v`_1, v_2) = \frac{\chi^2_{v_1}/v_1}{\chi^2_{v_2}/v_2} \tag{9}$$

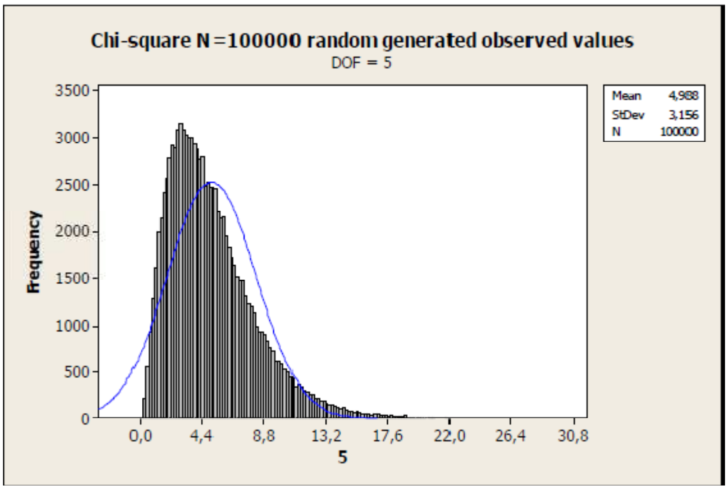
The probability density function of F_{ν_1,ν_2} can be represented by:

$$f(m; v`_1, v_2) = \frac{\Gamma[(v`_1 + v`_2)/2]}{\Gamma(v`_1/2)\Gamma(v`_2/2)} v_1^{v`_1/2} v_2^{v`_2/2} \frac{m^{(v`_1/2)-1}}{(v`_1 m + v`_2)^{(v`_1+v`_2)/2}} \tag{10}$$

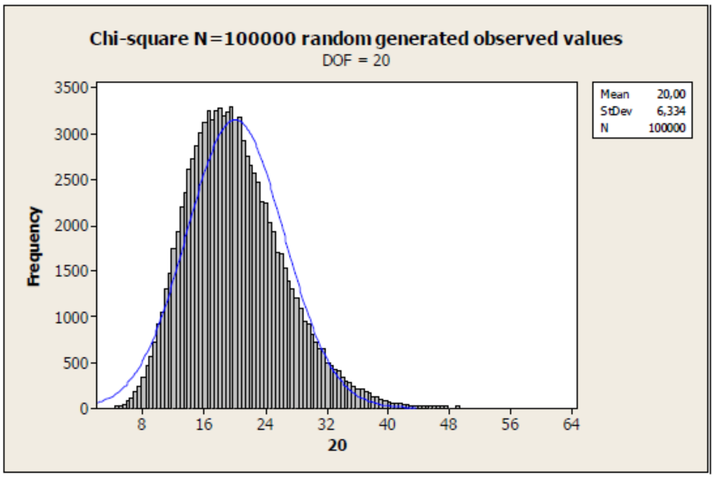
a) DOF = 1



b) DOF = 5



c) DOF = 20



d) DOF = 100

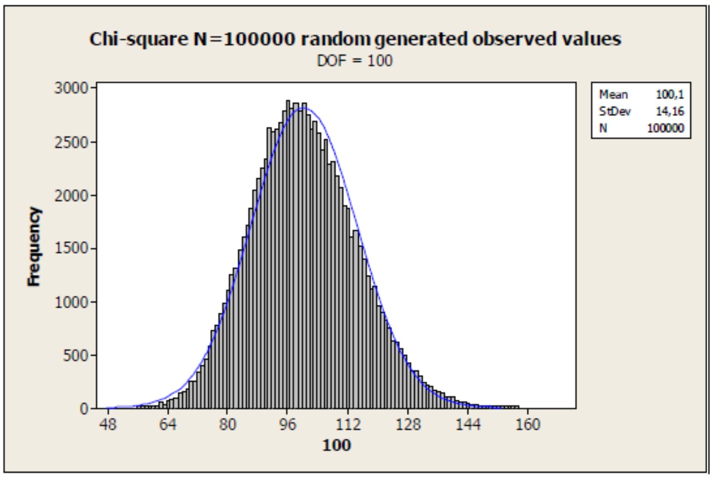


Fig. 1. Four histograms of 10^5 random generated observed values for different degrees of freedom (DOF) fitted with Gaussian distribution. Mean and standard deviation is also computed in each case.

It can be observed that the distribution is asymmetric and, in this case, the β -quintiles $m_\beta\left(_{v_1},_{v_2}\right)$ defined as:

$$P\{F(v`_1, v`_2) \le m_\beta(v`_1, v`_2)\} = \int_0^{m_\beta(v`_1, v`_2)} f(m, v`_1, v`_2) dm = \beta \quad (11)$$

This can be verified, considering that $\chi^2_{v_1+v_2} = \chi^2_{v_1} + \chi^2_{v_2}$ and $\chi^2_{v_2-v_1} = \chi^2_{v_2} - \chi^2_{v_1}$

Therefore it is possible to write the following expression:

$$m_{(1-\beta)}(v_1, v_2) = \frac{1}{m_\beta(v_1, v_2)} \quad (12)$$

The theoretical contribution to assess the uncertainty interval, with relative confidence level, in the case of n successive observations is proposed. The approach is based on the Chi-square and Fisher distributions and the validity is proved by a numerical example.

Thus, the presented work, confirms the prospects of using the method for uncertainty evaluation of different measurements.

References

1. BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, OIML. (1995). *Guide to the Expression of Uncertainty in Measurement*. International Organization for Standardization, Geneva, Switzerland.
2. BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, OIML. (2008). *Evaluation of measurement data – Supplement 1 to the "Guide to the expression of uncertainty in measurement" – Propagation of distributions using a Monte Carlo method. Joint.* Committee for Guides in Metrology, Bureau International des Poids et Mesures, JCGM 101.
3. Catelani, M. Introduction to the t and chi-square distribution for a more accurate evaluation of the measure of the Word Error Rate in Analog-to-Digital Converters. Metrol./ M. Catelani , A. Zanobini, , L. Ciani.// Meas. Syst., 15(4), 483-488.
4. Box, G.E.P., Statistics for experimenters./ G.E.P Box, W.G. Hunter, J.S. Hunter, // New York: John Wiley.
5. Solopekina, A.A. Application of the Uncertainty Method for Analysis of Multi-Port Correlator Accuracy/ A.A. Solopekina, N. Semezhev, A.A. L'vov, V.V. Komarov, M.S. Svetlov // Proceedings of the 2017 IEEE North West Russia Section Young Researchers in Electrical and Electronic Engineering Conference (2017 ElConRusNW), St.-Peterburg, Russia, 2017
6. Kendall, M. Kendall's Advanced Theory of Statistics. London: Griffin & Company Limited, 1, 2, 3.