METROLOGY AND MEASUREMENT SYSTEMS
VOL. XV, NUMBER 1 (2008)

IGOR P. ZAKHAROV, SERGEY V. VODOTYKA

Kharkov National University Of Radioelectronics
Ukraine
e-mail: newzip@ukr.net

APPLICATION OF MONTE CARLO SIMULATION FOR THE EVALUATION OF MEASUREMENTS UNCERTAINTY

The Monte Carlo procedure for evaluation of uncertainty in measurements is considered. Algorithms of formation correlation and non correlation data files of the input quantity estimated on types A and B are developed.

Keywords: uncertainty in measurement, Monte Carlo simulation, correlation, standard uncertainty

1. INTRODUCTION

The common approach to uncertainty evaluation described in the "Guide to the Expression of Uncertainty in Measurement" GUM [1] has a series of limitations. First, the basis of that method consists of the law of uncertainty propagation that rested on first-order Taylor series of nonlinear function of the model. Application of that approach lead to the biased estimator of the measurement result and to the unjustified estimator of the combined standard uncertainty u_c. Secondly, when one estimates the expanded uncertainty according to the Central Limit Theorem it assumes a normal distribution of the output value and the coverage factor approximated by the coverage factor of Student's distribution with particular degrees of freedom obtained from the Welch-Satterthwaite formula. The first assumption is justified when the model function is linear, there is a big amount of input values and their laws of distribution are symmetrical. Besides, the expression of the coverage factor by way of the coverage factor of Student's distribution is not always justified. In this way the minimal value of the coverage factor of Student's distribution for a level of confidence 0.95 is 1.96 with infinite degrees of freedom. However, in case of a single measurement or when the contributions of type B evaluation of uncertainty is dominant, the value 1.96 of the coverage factor is maximum. Then the minimum value of the coverage factor can be equal to 1.65 for rectangular law of the distribution and 1.45 for arcsine law of the distribution. In those cases the estimation of the coverage factor according to the GUM gives a too high estimator of the expanded uncertainty correspondingly for 20% and 40% [2].

Complication of applying the Welch-Satterthwaite formula exist even in the cases when there are no contributions of type B evaluation of uncertainty [3]. That formula was obtained by the approximate analytical methods in the thirties and forties of the previous century and its correctness was not checked by methods of computational modeling. Besides, the Welch-Satterthwaite formula is fundamentally not designed to deal with correlated input values and in this case its application can lead to an unjustified estimator of the expanded uncertainty [6].

Applying the calculus of approximations for evaluation of uncertainty of the measurement can eliminate all drawbacks like that described above.

Statistical modeling (Monte Carlo) based on the law of propagation of the distributions is the most universal. This method is the foundation of Supplement 1 to the GUM “Numerical methods of the propagation of distributions” developed by Working Group 1 (WG1) of the Joint Committee for Guides in Metrology headed by the director of the BIMP [7]. Note that the Supplement 1 (currently under development) does not consider the combined laws of Student’s distribution in case when there is type A evaluation of uncertainty and the observed correlation, and it also does not consider the combined laws of rectangular distribution in case when there is type B evaluation of uncertainty and logical correlation.

The present article has the statement of the evaluation of measurement uncertainty by Monte Carlo simulation, taking into consideration the observed and logical correlation.

2. MONTE CARLO SIMULATION AND ITS REALIZATION

In the Monte Carlo simulation the input values $X_1, X_2, \ldots, X_m$ represented as random values with densities of the probability distributions $g_1, g_2, \ldots, g_m$. Expectation and standard deviation of these probability distributions is assigned to be equal to the estimators of the input values and their standard deviations respectively.

In this application the Monte Carlo simulation encompasses the realization of the next operations (Fig. 1):

1. Generation of m arrays of the random numbers x_j, $j = 1, 2, \ldots, m$ of the given size $n(n = 10^5 \ldots 10^6)$, appropriate for the required laws of distribution.
2. Obtaining the array y of the output value estimator. The size of that array equal to n.
3. Counting the parameter estimators of the obtained distribution: expectation $\hat{M}(y)$, combined standard uncertainty $\hat{u}_c(y)$, coverage factor k and expanded uncertainty $\hat{U}_p$ for the given level of confidence p.
4. Repeating l times ($l = 50 \ldots 10$) the steps 1-3 to obtain the average values of parameter estimators from step 3 and counting the estimator of their standard deviation to obtain their reliability.

An advantage of the Monte Carlo simulation is the practical possibility of unlimited increase of accuracy by increasing the size $n \times l$ of the random numbers arrays.

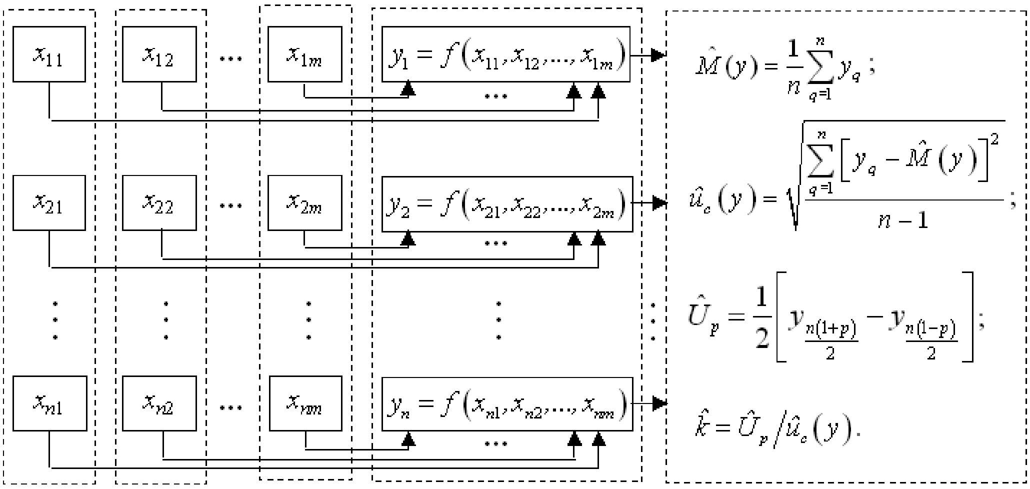

Fig. 1. Realization of the Monte Carlo simulation.

3. PRODUCING THE ARRAYS OF THE INPUT VALUES THAT ARE EVALUATED BY TYPE B

When one evaluates the uncertainty of type B one needs to obtain data arrays with normal, rectangular, triangular, arcsine and other distributions applied in that case. Generally, integrated generators of random numbers in mathematical and statistical software provide generation of numbers with normal and rectangular distributions. Other required laws of distribution could be obtained by the inverse function method.

For realization of the inverse function method one needs to generate random numbers that have the rectangular law of distribution in the interval [0; 1]. These numbers are reduced according to equation:

$$x^*_{ji} = G^{-1}\left(x_{ji}\right),$$

where x_{ji} - the initial i^{th} random number of j^{th} input value that has the rectangular law of distribution; x^*_{ji} – the unknown i^{th} random number of j^{th} input value that has the given law of distribution; G^{-1} – the inverse integral function of the given law of distribution.

For obtaining G^{-1} one needs to write the analytic equation for the integral function of distribution in the form of $x = G(x^*)$, then express from this the next equation: $x = G^{-1}(x^*)$.

Thus, for obtaining the array of the random numbers that have the arcsine distribution with given expectation and standard deviation we realize inverse transformation of the integral function of the arcsine distribution:

$$x = G\left(x^*\right) = \frac{1}{2} + \frac{1}{\pi} arcsin \frac{x^* - \mu}{\sigma\sqrt{2}},$$

from which we obtain:

$$x^{*} = G^{-1}(x) = \mu + \sigma\sqrt{2} sin\pi(x - 0{,}5) \cdot$$

We can obtain some laws of distribution by composition of random numbers that have the known law of the distribution. Thus, triangular distribution obtained by composition of two arrays of random numbers that have the rectangular distribution with equal standard deviation, trapezoidal distribution is the result of composition of two arrays of random numbers that have the rectangular distribution with different standard deviations σ_1 and σ_2. That standard deviation of the result distribution is defined according to the rules of the variance composition: $\sigma_\Sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$.

When one evaluates type B uncertainty there is the possibility of the existence of several mutually correlated input values. Correlation of these values (so-called logical correlation) is conditioned upon the use of the same measuring instrument, physical reference or the same reference data that have significant uncertainty during the measurement.

In that case for realization of the Monte Carlo simulation, the generation of the combined (two-dimensional) law of distribution correlated input values is needed. That task has an easy solution for normal distributions [7]. However for other distributions that procedure does not succeed, but there is a critical need of it, because generally contributions of the uncertainties evaluated by type B have the law of distributions not normal (Fig. 2).

The developed method of modeling of the combined distribution of the two correlated values that have arbitrary laws of distribution includes the next steps:

1. Generation of two sequences of uncorrelated random values ξ_1 and ξ_2 that have normal distribution, their expectation equal to zero and their standard deviation equal to one.
2. Generation of the third sequence from those sequences $\xi_3 = r_{1,2}\xi_1 + \sqrt{1 - r_{1,2}^2}\xi_2$.
3. Realization of the transformation of ξ_1 and ξ_3 in the form of an integral function of normalized normal distribution $\upsilon = F_H(\xi)$ with obtaining of the sequences of the correlated random numbers υ_1 and υ_2 in the interval [0; 1] that have rectangular distributions and correlation coefficient $r_{1,2}$ close to the initial coefficient r.
4. Obtaining normalized random numbers υ_{1H} and υ_{2H} with rectangular distributions, with expectations equal to zero and standard deviation equal to one is realized according to $\upsilon_{1,2H} = (2\upsilon_{1,2} - 1)\sqrt{3}$.
5. Realization of the inverse function method for obtaining the two sequences of the correlated random values ψ_1 and ψ_2 that have the given laws of distribution $\psi_{1,2} = G^{-1}(\upsilon_{1,2})$ where G – the integral function of the given law of distribution.

The proposed approach allows the realization of the generation of the mutually correlated random numbers that have a given correlation coefficient and any law of distribution, including different laws. The research has shown [8] that the methodical bias of reproduction of the correlation coefficient due to nonlinear transformation of initial laws of distribution in the case of simulation does not exceed -0.018 in the interval $-1 \le r \le 1$ when the laws of distribution are rectangular and it achieves a maximum at the point $r \approx \pm 0.6$.

Realization of the inverse function method to obtain the correlated input values that have arcsine distribution increases the maximum mean value of the added bias of the correlation coefficient reproduction to -0.043. Taking into account the corresponding correction for the mean value of the correlation coefficient allows to compensate the mentioned bias.

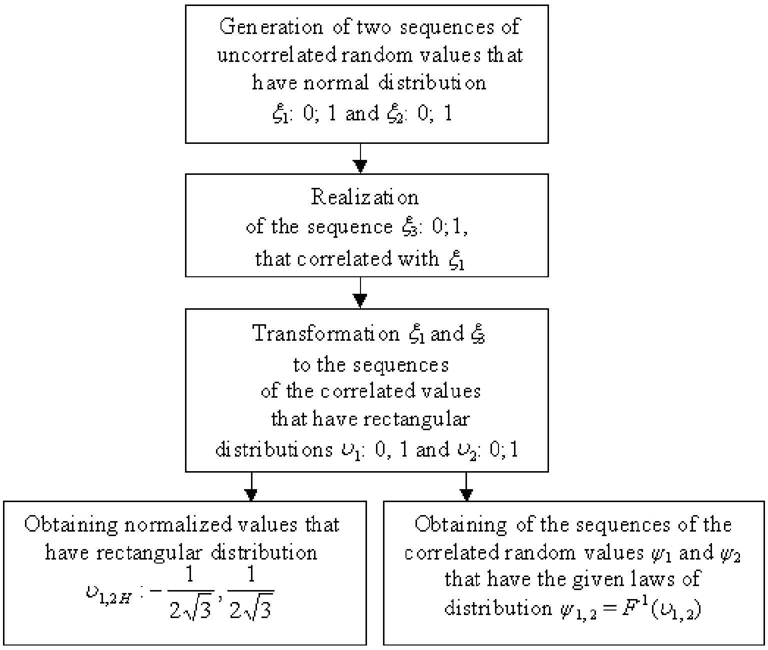

Fig. 2. Modeling method of the combined distribution of the two correlated values that have arbitrary laws of distribution.

4. PRODUCING THE ARRAYS OF THE INPUT VALUES THAT ARE EVALUATED BY TYPE A

When one evaluates type A uncertainty one needs to obtain the data arrays that have Student's distribution. The basis for its obtainment is the known equation for parameter T:

$$T = \frac{\overline{x} - M_x}{s(\overline{x})} . \tag{1}$$

To obtain the array of random numbers that have Student's distribution with given degrees of freedom v one needs to perform the following operations:

1. Generation of the amount $N = v + 1$ of the random numbers x_i, which have normal distribution and the following parameters: $M_x = 0$ and $s(x) = 1$.
2. Calculation of the average mean value $\overline{x}$, the estimator of its standard deviation $s(\overline{x})$ by known equation and calculation of the mean value of the first realization of the parameter T by using Eq. (1).
3. Subsequent calculation of the data array of the realization of the parameter T of the given size n.

During evaluation of type A uncertainty we have to face the case when the input values are mutually correlated. In that case the cause of the (so-called) observed correlation is the measurement of the two or more input values in the same conditions simultaneously. Mostly we can meet the mutual observed correlation in indirect multiple measurements.

For the realization of Monte Carlo simulation one needs the generation of combined (two-dimensional) Student's distribution of the correlated input values. That task has the following solution:

1. Generation of the two series of random numbers x_i of the size N that have normal distribution with parameters $M_1 = 0$, $M_2 = 0$, standard deviation equal to one and with values mutually correlated with the correlation coefficient $r_{1,2}$.
2. Calculation of the average means $\bar{x}_1$ and $\bar{x}_2$, and estimators of their standard deviations $s(\bar{x}_1)$ and $s(\bar{x}_2)$ of every series and calculation of the mean values of the first realization of the parameters T_1 and T_2 according to Eq. (1).
3. Subsequent calculation of the arrays of realization of the parameters T_1 and T_2 of the given size $n(n = 10^5 \ldots 10^6)$.

5. OBTAINING THE MEAN VALUES OF THE EXPANDED UNCERTAINTY AND OF THE COVERAGE FACTOR BY USING THE STATISTICAL DATA OF THE INPUT VALUES

After producing the data arrays of all the input values $X_1, X_2, \ldots, X_m$ obtainment of the data array of the output value is realized according to the equation:

$$Y = f(X_1, \ldots, X_m) \ .$$

After obtaining the data array of the output value one calculates:

– the measurement result estimator according to the equation:

$$\hat{y} = \frac{1}{n}\sum_{q=1}^{n} f\left(x_{1q}, x_{2q}, \ldots, x_{mq}\right) = \frac{1}{n}\sum_{q=1}^{n} y_q \, ,$$

– the standard combined uncertainty estimator of the measurement result:

$$u_c(y) = \sqrt{\frac{1}{(n-1)}\sum_{q=1}^{n}\left(y_q - \hat{y}\right)^2} \ .$$

The resulting data array of the output value needs to be sorted for estimation of the expanded uncertainty. Also we need to calculate the interquartile interval according to the equation:

$$\hat{U} = (Y_{n(1+p)/2} - Y_{n(1-p)/2})/2 \ ,$$

where $Y_{n(1+p)/2}$ and $Y_{n(1-p)/2}$, are the $n(1 + p)/2$ and $n(1 - p)/2$ terms of the sorting data array of the output value respectively.

Thus, for $p = 0.95$ and $n = 10^5$ quartiles of the output value, distribution is estimated for the 97500[th] and 2500[th] terms of the sorting array.

The coverage factor estimator is obtained by dividing the expanded uncertainty estimator by the standard combined uncertainty:

$$\hat{k} = \hat{U} / \hat{u}_c(y).$$

It is necessary to repeat the mentioned operations l times (l = 50 … 100) for specification obtained uncertainty estimators and coverage factor and for evaluation of the standard deviation of these estimators.

6. AUTOMATION OF THE EVALUATING MEASUREMENT UNCERTAINTY

The authors have developed the software for evaluating measurement uncertainty which realized Monte Carlo simulation and which saves the operator from routine work and speeds up the process of evaluation [9].

7. CONCLUSIONS

It's described Monte Carlo simulation for estimating measurement result and its uncertainty that have arbitrary model function. Application of this method allows eliminate the shortcomings of the uncertainty distribution law in the presence of the essential nonlinearity of the model function, correlation between input and output values that have non-normal law of distribution. The present approach allows generation correlated random numbers in pairs with given correlation factor and any distribution laws including not the same laws.

REFERENCES

1. *Guide to the Expression of Uncertainty in Measurement*, ISO, Geneva, 1st Ed. 1995 – 101 p. Translation from English. S. Peterburg: VNIIM D. I. Mendeleeva, 1999 – 126 p. (in Russian)
2. Zakharov I. P.: "Calculating coverage factor for rectangular and normal distributions of the uncertainty contributions", *System of information processing*, 2005, no. 6, pp. 52–57. (in Russian)
3. Zakharov I. P.: "Composition of Student's laws of distributions", *System of information processing*, 2005, no. 8, pp. 28–35. (in Russian)
4. Welch B. L.: "The significance of the differences between two means when the population variances are unequal", *Biometrika*, 1938, no. 29, pp. 350–362.
5. Satterthwaite F. E.: "An approximate distribution of estimates of variance components", *Biometrics Bulletin*, 1946, no. 2, pp. 110–114.
6. Zakharov I. P.: "Calculation of correlation during estimation of multiple measurements", *System of information processing*, 2005, no. 9, pp. 43–45. (in Russian)
7. Kocks M., Harris P., Zibert B. R.-L.: "Measurement uncertainty estimation based on transformation of distributions by Monte Carlo simulation", *Measurement technique*, 2003, no. 9, pp. 9–14. (in Russian)
8. Zakharov I. P.: "Correlated data simulation during processing results of measurements", *Simulations and informational technologies*, 2005, no. 33, pp. 35–40. (in Russian)
9. Zakharov I. P., Vodotyka S. V.: "Software for measurements uncertainty evaluating", *System of information processing*, 2007, no. 6, pp. 41–43. (in Russian)