Metrol. Meas. Syst., Vol. XX (2013), No. 2, pp. 249–262.

METROLOGY AND MEASUREMENT SYSTEMS

Index 330930, ISSN 0860-8229

www.metrology.pg.gda.pl

ESTIMATION OF RANDOM VARIABLE DISTRIBUTION PARAMETERS BY THE MONTE CARLO METHOD

Sergiusz Sienkowski

University of Zielona Góra, Faculty of Electrical Engineering, Computer Science and Telecommunication, Institute of Electrical Metrology, Podgórna 50, 65-246 Zielona Góra, Poland (s.sienkowski@ime.uz.zgora.pl)

Abstract

The paper is concerned with issues of the estimation of random variable distribution parameters by the Monte Carlo method. Such quantities can correspond to statistical parameters computed based on the data obtained in typical measurement situations. The subject of the research is the mean, the mean square and the variance of random variables with uniform, Gaussian, Student, Simpson, trapezoidal, exponential, gamma and arcsine distributions.

Keywords: Monte Carlo method, mean, mean square, variance.

1. Introduction

Measurement accuracy is a basic characteristic of both measurement tools and results. Accuracy is characterized indirectly by giving an opposite quantity in the form of uncertainty (inaccuracy) or measurement error. In the metrological regulations, the description of measurement uncertainty is based on the Guide [1], as well as on the Supplements [2] and [3] to the Guide. From the moment that the Joint Committee for Guides in Metrology announced the commencement of its work on designing the Supplements, and also after their publication, a steady growth has been observed in interest in the Monte Carlo method and its application to the analysis of measurement uncertainty [4-10]. It takes place when particular measurement situations are defined by complicated measurement models. Then the use of the Monte Carlo method makes it possible to avoid complex mathematical apparatus and to take into account the influence of all the parameters characterizing a specific measurement situation.

The growth in interest in the Monte Carlo method voiced in prestigious scientific journals has induced the author to study the application of such a method to the estimation of random variable distribution parameters, which are determined based on the probability density function. Such quantities can correspond to statistical parameters computed based on data obtained in typical measurement situations [11, 12]. In the literature there are many references to the issues discussed in the paper [13-25]. Against this background, the author's original achievement consists in the formalization of the selection principles for the method's input parameter values for a fixed probability distribution. The author has focused on the distributions mentioned in the Supplements [2, 3] and commonly employed in the measurement uncertainty analysis. In the author's view, the obtained results can be used to automate the estimation of density distribution parameters by the Monte Carlo method. The obtained results can also be employed to compare the evaluation of distribution parameters with the estimators computed based on measurement results possessing a known probability density. It is particularly important when the number of measurements is not large, there is no possibility of carrying out measurements in reproducible measurement conditions, the sets of

Article history: received on Dec. 05, 2012; accepted on Mar. 21, 2013; available online on Jun. 03, 2013; DOI: 10.2478/mms-2013-0022.

measurement results are incomplete, or measurement results are not sufficiently accurate. In practice, the obtained results can be applied in virtual measuring devices as an implemented module for distribution parameter estimation. The author made available a computer program and web page illustrating the functioning of such a module [26].

The subject of the research is the mean, the mean square and the variance of a random variable with uniform, Gaussian, Student, Simpson, trapezoidal, exponential, gamma and arcsine distributions.

2. Random variable distribution parameters

In the analysis of the results of measurements by probabilistic methods, it is assumed that they can be modeled with a random variable. The probability density function is the best description of a random variable. Let X be a random variable with density $f_X(x)$, $x\in\mathbf{R}$. The parameters of the random variable X include the mean, the mean square, the variance, the standard deviation, the median, and the modal value [27]. The subject of the research is the mean:

$$\bar{X} = E\left[X\right] = \int_{-\infty}^{+\infty} x f_X\left(x\right)\mathrm{d}x, \tag{1}$$

the mean square:

$$\overline{X^2} = E\left[X^2\right] = \int_{-\infty}^{+\infty} x^2 f_X\left(x\right)\mathrm{d}x, \tag{2}$$

and the variance:

$$Var\left[X\right] = \int_{-\infty}^{+\infty} \left(x - E\left[X\right]\right)^2 f_X\left(x\right)\mathrm{d}x = E\left[X^2\right] - E^2\left[X\right]. \tag{3}$$

Quantities (1) and (2) are ordinary first- and second-order moments, respectively, quantity (3) is the second central moment [27]:

We will assume that parameters (1)-(3) are computed over the interval $[a, b]$, $-\infty<a\leq b<\infty$, a, $b\in\mathbf{R}$, and that $f_X(x)$ has a non-zero value for any $x\in\mathbf{R}$. Then the error reduction:

$$\delta_{\bar{X}_{[a,b]}} = \frac{\left|\bar{X}_{[a,b]} - \bar{X}\right|}{\left|\bar{X}\right|}100\%, \tag{4}$$

$$\delta_{\overline{X^2}_{[a,b]}} = \frac{\left|\overline{X^2}_{[a,b]} - \overline{X^2}\right|}{\overline{X^2}}100\%, \tag{5}$$

$$\delta_{Var[X]_{[a,b]}} = \frac{\left|Var\left[X\right]_{[a,b]} - Var\left[X\right]\right|}{Var\left[X\right]}100\%, \tag{6}$$

leads to an increase in the accuracy of the computation of parameters (1)-(3).

3. Estimation of random variable distribution parameters by the Monte Carlo method

Let there be given a continuous function $g(x)$, $x\in\mathbf{R}$, integrated over the interval $[a, b]$. Let us assume that the integral of the function $g(x)$ has the form of the formula:

Metrol. Meas. Syst., Vol. XX (2013), No. 2, pp. 249–262.

$$\theta = \int_a^b g(x)\,\mathrm{d}x. \tag{7}$$

One of the most popular methods of estimation of the integral θ is the hit-or-miss Monte Carlo method [13-15, 17-19, 24]. The popularity of this method is owed to the fact that it has a geometric character and is usually described with uncomplicated mathematical apparatus. Additionally, in comparison with other Monte Carlo methods, the presented method ensures greater uniqueness of the generated data on which the performed computations are based. It is particularly important because of the finite lengths of the cycles of pseudo-random number generators used for data generation. In practice, the accuracy of the method is dependent on the number of data and the quality of the pseudo-random number generator [28]-[30]. The article presents results obtained with the use of LabVIEW. LabVIEW uses a triple-seeded very-long-cycle linear congruential generation (LCG) algorithm to generate the uniform pseudorandom numbers [31].

The hit-or-miss Monte Carlo method makes it possible to numerically integrate the function $g(x)$. Let us assume that the values of the function $g(x)$ are positive and negative, and are situated within the area:

$$\Omega = \{(x, y) \in \mathrm{R} : a \le x \le b,\, c \le y \le d\}, \tag{8}$$

where $-\infty < c \le 0$, $0 \le d < \infty$, $c, d \in \mathbf{R}$.

The estimation of the integral θ is carried out based on N points $p_i=(x_i, y_i)$, $i=0,1,\ldots,N-1$, with the coordinates x_i and y_i drawn from uniform distribution over the intervals $[a, b]$ and $[c, d]$, respectively (Fig. 1). The result of the integral estimation is the quantity:

$$\tilde{\theta} = \frac{k_N}{N}(b-a)(d-c), \tag{9}$$

where k_N is the power of the set $\{i\in\mathbf{N}: 0<y_i\le g(x_i)\}$ reduced by the power of the set $\{i\in\mathbf{N}: g(x_i)\le y_i<0\}$. If $y_i=0$, then k_N does not change its value.

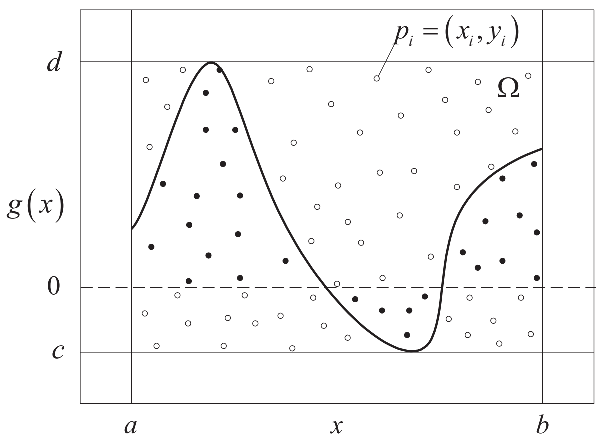

Fig. 1. Illustration of numerical integration by the hit-or-miss Monte Carlo method.

In the case when the values of the function $g(x)$ are only nonnegative or nonpositive, and there exist $c>0$ and $c\le g(x)$ or $d<0$ and $d\ge g(x)$, then:

$$\tilde{\theta} = \frac{k_N}{N}(b-a)(d-c) + \begin{cases} (b-a)c, & c>0, \\ (b-a)d, & d<0, \\ 0, & c \le 0 \wedge d \ge 0. \end{cases} \tag{10}$$

From (10) it follows that in the result $\tilde{\theta}$ of the estimated integral θ, the area of a rectangle with the sides *b-a* and *c*, or *b-a* and *d* is taken into account. Simultaneously, for the same number *N* of points p_i, the way of estimating the integral θ described by formula (10) makes it possible to increase the accuracy of the estimation of this quantity.

The estimation error of the integral θ can be described with the formula:

$$\delta_{\tilde{\theta}} = \frac{|\tilde{\theta}-\theta|}{|\theta|} 100\%. \tag{11}$$

Because in (1)-(3) an integration operation appears, then for fixed values of *a* and *b*, as well as for:

$$g_1(x) = x f_X(x), \quad g_2(x) = x^2 f_X(x), \quad g_3(x) = (x - E[X])^2 f_X(x), \tag{12}$$

the hit-or-miss Monte Carlo method can be adapted for the estimation of parameters (1)-(3). The result of the estimation of parameters (1)-(3) are the quantities: $\tilde{\bar{X}} = \tilde{\theta}_1$, $\overline{X^2} = \tilde{\theta}_2$, $\tilde{V}ar[X] = \tilde{\theta}_3$. In practice, the values of these quantities are often determined based on measurement results.

A basic problem during the estimation of parameters (1)-(3) is to establish the values of the ends of the interval [*c*, *d*]. According to the Weierstrass theorem, a function continuous on the interval [*a*, *b*] attains its lower and upper bounds, which can correspond to the ends of the interval [*c*, *d*] [32]. The bounds of the function *g*(*x*) can be determined based on the values of the function *g*(*x*) at the ends of the interval [*a*, *b*] and on the extremes of the function *g*(*x*). The extremes of the function *g*(*x*) are determined based on the roots of the equation $g'(x)=0$, where $g'(x)=0$ is a derivative of the function *g*(*x*). The extremes of the functions $g_1(x)$, $g_2(x)$ and $g_3(x)$ can be determined based on the formula:

$$g_1'(x) = \frac{\mathrm{d}}{\mathrm{d}x} g_1(x), \quad g_2'(x) = \frac{\mathrm{d}}{\mathrm{d}x} g_2(x), \quad g_3'(x) = \frac{\mathrm{d}}{\mathrm{d}x} g_3(x), \tag{13}$$

and the roots of the equations: $g'_1(x)=0$, $g'_2(x)=0$, $g'_3(x)=0$.

According to (11), the errors of the estimation of parameters (1)-(3) have the form:

$$\delta_{\tilde{\bar{X}}} = \frac{|\tilde{\bar{X}} - \bar{X}|}{|\bar{X}|} 100\%, \tag{14}$$

$$\delta_{\overline{X^2}} = \frac{|\overline{X^2} - \overline{X^2}|}{\overline{X^2}} 100\%, \tag{15}$$

$$\delta_{\tilde{V}ar[X]} = \frac{|\tilde{V}ar[X] - Var[X]|}{Var[X]} 100\%. \tag{16}$$

From (11) it follows that the accuracy of the estimation of parameters (1)-(3) will increase along with the number *N* of points p_i.

Metrol. Meas. Syst., Vol. XX (2013), No. 2, pp. 249–262.

3.1. Estimation of uniform distribution parameters

Let X be a random variable with uniform distribution in the interval $[-A_u+A_0, A_u+A_0]$, $A_u \in \mathbf{R}_+$, $A_0 \in \mathbf{R}$. The variable X has the density [27]:

$$f_X(x) = \begin{cases} \dfrac{1}{2A_u}, & |x - A_0| \le A_u, \\ 0, & |x - A_0| > A_u. \end{cases} \tag{17}$$

Based on (1)-(3), and (17), we obtain:

$$\bar{X} = A_0, \quad \overline{X^2} = \frac{A_u^2}{3} + A_0^2, \quad Var[X] = \frac{A_u^2}{3}. \tag{18}$$

Parameters (1)-(3) are estimated based on (9). The ends of the interval $[a, b]$ are selected based on the formula:

$$a = -A_u + A_0, \quad b = A_u + A_0. \tag{19}$$

Making use of (13) and (17), we determine the roots of the equations $g_1'(x)=0$, $g_2'(x)=0$ and $g_3'(x)=0$, as well as the values of the functions $g_1(x)$, $g_2(x)$ and $g_3(x)$ at the ends of the interval $[a, b]$. Such quantities are used to determine the ends of the interval $[c, d]$. Estimating (1), we assume:

$$c = \begin{cases} a f_X(a), & a < 0, \\ 0, & a \ge 0, \end{cases} \qquad d = \begin{cases} b f_X(b), & b \ge 0, \\ 0, & b < 0, \end{cases} \tag{20}$$

while (2):

$$c = 0, \quad d = \begin{cases} a^2 f_X(a), & |a| \ge |b|, \\ b^2 f_X(b), & |a| < |b|, \end{cases} \tag{21}$$

and (3):

$$c = 0, \quad d = \frac{1}{4}(a-b)^2 f_X(a) \text{ or } d = \frac{1}{4}(a-b)^2 f_X(b). \tag{22}$$

In the case of uniform distribution, the values of the functions $g_1(x)$ and $g_2(x)$ can only be nonnegative or nonpositive. If there exists $c>0$ or $d<0$, then parameters (1) and (2) can be estimated based on (10). Then estimating (1), we assume:

$$c = a f_X(a), \quad d = b f_X(b), \tag{23}$$

while (2):

$$c = \begin{cases} a^2 f_X(a), & a > 0, \\ b^2 f_X(b), & b < 0, \\ 0, & a \le 0 \wedge b \ge 0, \end{cases} \qquad d = \begin{cases} a^2 f_X(a), & |a| \ge |b|, \\ b^2 f_X(b), & |a| < |b|. \end{cases} \tag{24}$$

In the case of the function $g_3(x)$ the coefficients $c=0$ and $d \ge 0$. This means that the parameter (3) can be estimated based on (9) or (10) and both formulas give the same results.

Example values of 100 averaged results of errors (14)-(16) determined based on (9) and (18), for $A_u=1$, $A_0=10$, $N=10^6$, are: $\bar{\delta}_{\tilde{\bar{X}}}=0.028\%$, $\bar{\delta}_{\overline{X^2}}=0.041\%$, $\bar{\delta}_{\widetilde{Var}[X]}=0.11\%$. In the considered example, $c>0$ exists and for mean $c=9$ and mean square $c=81$. Then estimating errors from (10) and (18), we obtain $\bar{\delta}_{\tilde{\bar{X}}}=0.0087\%$, $\bar{\delta}_{\overline{X^2}}=0.016\%$. The obtained results show that (10) makes it possible to increase the accuracy of the estimation of parameters (1) and (2).

3.2. Estimation of Gaussian distribution parameters

Let X be a continuous Gaussian random variable with parameters $\sigma_X \in \mathbf{R}_+$ and $\mu_X \in \mathbf{R}$. The variable X has the density [27]:

$$f_X(x)=\frac{1}{\sigma_X\sqrt{2\pi}}\mathrm{e}^{-\frac{(x-\mu_X)^2}{2\sigma_X^2}}. \tag{25}$$

Based on (1)-(3), and (25), we obtain:

$$\bar{X}=\mu_X,\quad \overline{X^2}=\sigma_X^2+\mu_X^2,\quad Var[X]=\sigma_X^2. \tag{26}$$

Parameters (1)-(3) are estimated based on (9). The ends of the interval $[a, b]$ assume the values:

$$a=-\infty,\quad b=+\infty. \tag{27}$$

In practice, a and b have finite values. They can be selected either arbitrarily or from the formula:

$$a=-6\sigma_X+\mu_X,\quad b=6\sigma_X+\mu_X. \tag{28}$$

The shape of formula (28) follows from the "6-sigma" rule [33]. Using (4)-(6), it can be verified that for $0.01\le\sigma_X\le100$ and $0<\mu_X\le100$, quantities (1)-(3) are computed in the interval $[a, b]$ with errors not greater than $1.0\cdot10^{-5}\%$.

Making use of (12), (13) and (25), we determine the ends of the interval $[c, d]$. Estimating (1), we assume:

$$\begin{aligned} c&=\frac{\mu_X-\sqrt{4\sigma_X^2+\mu_X^2}}{2}f_X\left(\frac{\mu_X-\sqrt{4\sigma_X^2+\mu_X^2}}{2}\right),\\ d&=\frac{\mu_X+\sqrt{4\sigma_X^2+\mu_X^2}}{2}f_X\left(\frac{\mu_X+\sqrt{4\sigma_X^2+\mu_X^2}}{2}\right), \end{aligned} \tag{29}$$

while (2):

$$c=0,\quad d=\begin{cases} x_{\max}^2 f_X(x_{\max}), & x_{\max}^2 f_X(x_{\max})\ge x_{\min}^2 f_X(x_{\min}),\\ x_{\min}^2 f_X(x_{\min}), & x_{\max}^2 f_X(x_{\max})< x_{\min}^2 f_X(x_{\min}), \end{cases} \tag{30}$$

and (3):

$$c=0,\quad d=2\sigma_X^2 f_X\left(\mu_X\pm\sqrt{2}\sigma_X\right), \tag{31}$$

where:

$$x_{\max}=\frac{\mu_X+\sqrt{8\sigma_X^2+\mu_X^2}}{2},\quad x_{\min}=\frac{\mu_X-\sqrt{8\sigma_X^2+\mu_X^2}}{2}. \tag{32}$$

Metrol. Meas. Syst., Vol. XX (2013), No. 2, pp. 249–262.

Example values of 100 averaged results of errors (14)-(16) determined based on (9) and (26), for σ_X=1, μ_X=2 and N=10^6 are: $\overline{\delta}_{\tilde{\bar{X}}}$=0.17%, $\overline{\delta}_{\overline{X^2}}$=0.17%, $\overline{\delta}_{\tilde{V}ar[X]}$=0.13%.

3.3. Estimation of Student distribution parameters

Let a random variable X have Student distribution with parameter $n\in\mathbf{N}$. The variable X has the density [27]:

$$f_X(x)=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n}\,\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \tag{33}$$

where n is the number of degrees of freedom, $\Gamma(\cdot)$ is a gamma function [32].

Based on (1)-(3), and (33), we obtain that for n>2:

$$\overline{X}=0,\quad \overline{X^2}=\frac{n}{n-2},\quad Var[X]=\frac{n}{n-2}. \tag{34}$$

Parameters (1)-(3) are estimated based on (9). The ends of the interval [a, b] assume the form:

$$a=-\infty,\quad b=+\infty. \tag{35}$$

In practice, a and b have finite values. They can be selected either arbitrarily or from the formula:

$$a=-6\sqrt{\frac{n}{n-2}},\quad b=6\sqrt{\frac{n}{n-2}}. \tag{36}$$

The shape of formula (36) has been established based on the "6-sigma" rule, approximating Student distribution by Gaussian distribution with parameters $\sigma_Y=\sqrt{n/(n-2)}$ and μ_Y=0 [33]. Making use of (4)-(6), it can be verified that for 3≤n≤100, quantity (1) is, in the interval [a, b], computed with the error $\Delta_{\overline{X}_{[a,b]}}=\left|\overline{X}_{[a,b]}-\overline{X}\right|=0$, whereas quantities (2) and (3) are computed with errors from the interval [9.4·10^{-5}%, 21%]. The large spread of errors is due to inaccurate approximation of the Student distribution by Gaussian distribution. For example, if n=3 then quantities (2) and (3) are computed with an error of 21%. If n=5, 10, 100 then the error is equal respectively to 3.6%, 0.25% and 9.4·10^{-5}%.

Making use of (12), (13), and (33), we determine the ends of the interval [c, d]. Estimating (1), we assume:

$$c=-f_X(\pm 1),\quad d=-c, \tag{37}$$

while (2) and (3):

$$c=0,\quad d=\frac{2n}{n-1}f_X\left(\pm\sqrt{\frac{2n}{n-1}}\right). \tag{38}$$

Example values of 100 averaged results of the error $\Delta_{\tilde{\bar{X}}}=\left|\tilde{\bar{X}}-\overline{X}\right|$ and errors (15) and (16) determined based on (9) and (34), for n=10, N=10^6, are: $\overline{\Delta}_{\tilde{\bar{X}}}$=0.0017, $\overline{\delta}_{\overline{X^2}}$=0.27%, $\overline{\delta}_{\tilde{V}ar[X]}$=0.26%.

3.4. Estimation of Simpson distribution parameters

Let a random variable X have Simpson (triangular) distribution in the interval $[-A_t+A_0,\ A_t+A_0]$, $A_t\in\mathbf{R}_+$, $A_0\in\mathbf{R}$. The variable X has the density [27]:

$$f_X(x)=\begin{cases}\dfrac{1}{A_t}-\dfrac{1}{A_t^2}\left|x-A_0\right|, & \left|x-A_0\right|\le A_t,\\ 0, & \left|x-A_0\right|>A_t.\end{cases}\tag{39}$$

Based on (1)-(3) and (39), we obtain:

$$\bar{X}=A_0,\quad \overline{X^2}=\frac{A_t^2}{6}+A_0^2,\quad Var[X]=\frac{A_t^2}{6}.\tag{40}$$

Parameters (1)-(3) are estimated based on (9). The ends of the interval $[a, b]$ are selected based on the formula:

$$a=-A_t+A_0,\quad b=A_t+A_0.\tag{41}$$

Making use of (12), (13), (39), and (41), we determine the ends of the interval $[c, d]$. Estimating (1), we assume:

$$c=\begin{cases}\dfrac{a+b}{2}f_X\left(\dfrac{a+b}{2}\right), & a<0\wedge b<0,\\ \dfrac{a}{2}f_X\left(\dfrac{a}{2}\right), & a<0\wedge b\ge 0,\\ 0, & a\ge 0\wedge b\ge 0,\end{cases}\qquad d=\begin{cases}\dfrac{a+b}{2}f_X\left(\dfrac{a+b}{2}\right), & a\ge 0\wedge b\ge 0,\\ \dfrac{b}{2}f_X\left(\dfrac{b}{2}\right), & a<0\wedge b\ge 0,\\ 0, & a<0\wedge b<0,\end{cases}\tag{42}$$

while (2):

$$c=0,\quad d=\begin{cases}\left(\dfrac{a+b}{2}\right)^2 f_X\left(\dfrac{a+b}{2}\right), & |a|\ge 0\wedge|b|\ge 0\wedge\left(a\ge\dfrac{b}{3}\vee a\ge 3b\right),\\ \dfrac{4a^2}{9}f_X\left(\dfrac{2a}{3}\right), & |a|\ge|b|\wedge a<3b,\\ \dfrac{4b^2}{9}f_X\left(\dfrac{2b}{3}\right), & |a|<|b|\wedge a<\dfrac{b}{3},\end{cases}\tag{43}$$

and (3):

$$c=0,\quad d=\frac{1}{9}(a-b)^2 f_X\left(\frac{a+5b}{6}\right)\ \text{or}\ d=\frac{1}{9}(a-b)^2 f_X\left(\frac{5a+b}{6}\right).\tag{44}$$

Example values of 100 averaged results of errors (14)-(16) determined based on (9) and (40), for $A_t=1$, $A_0=2$, $N=10^6$, are: $\bar{\delta}_{\tilde{\bar{X}}}=0.077\%$, $\bar{\delta}_{\overline{X^2}}=0.084\%$, $\bar{\delta}_{\tilde{V}ar[X]}=0.076\%$.

3.5. Estimation of trapezoidal distribution parameters

Let a random variable X have trapezoidal distribution in the interval $[-A_{tz}+A_0,\ A_{tz}+A_0]$, $A_{tz}\in\mathbf{R}_+$, $A_0\in\mathbf{R}$. The random variable X has the density [2]:

Metrol. Meas. Syst., Vol. XX (2013), No. 2, pp. 249–262.

$$f_X(x) = \begin{cases} u\,\dfrac{x - A_0 + A_{tz}}{v - A_0 + A_{tz}}, & A_0 - A_{tz} \le x < v, \\ u, & v \le x \le w, \\ u\,\dfrac{A_0 + A_{tz} - x}{A_0 + A_{tz} - w}, & w < x \le A_0 + A_{tz}, \\ 0, & |x - A_0| > A_{tz}, \end{cases} \tag{45}$$

where:

$$u = \frac{2}{2A_{tz} + w - v}, \tag{46}$$

with $-A_{tz}+A_0 \le v \le w \le A_{tz}+A_0$, $v, w \in \mathbf{R}$.

In the case when $-A_{tz}+A_0=v$ and $A_{tz}+A_0=w$, trapezoidal distribution transforms to uniform distribution in the interval $[-A_{tz}+A_0, A_{tz}+A_0]$. If $v=w=A_0$, then trapezoidal distribution assumes the form of Simpson distribution in the interval $[-A_{tz}+A_0, A_{tz}+A_0]$.

Based on (1)-(3) and (45), we obtain [34]:

$$\begin{aligned} \bar{X} &= \frac{u}{6}\left(t^2 - s^2 + 3w^2 - 3v^2 + 3sv + 3tw\right), \\ \overline{X^2} &= \frac{u}{12}\left(s^3 + t^3 - 4v^3 + 4w^3 - 4s^2v + 4t^2w + 6sv^2 + 6tw^2\right), \\ Var[X] &= \frac{1}{48}(s + 2w - 2v + t)^2 + \frac{1}{24}(s^2 + t^2) - \frac{1}{144}\frac{(s^2 - t^2)^2}{(s + 2w - 2v + t)^2}, \end{aligned} \tag{47}$$

where $s=v+A_{tz}-A_0$, $t=A_{tz}+A_0-w$.

Parameters (1)-(3) are estimated based on (9). The ends of the interval $[a, b]$ are selected based on the formula:

$$a = -A_{tz} + A_0, \quad b = A_{tz} + A_0. \tag{48}$$

Making use of (12), (13), (45), and (48), we determine the ends of the interval $[c, d]$. Estimating (1), we assume:

$$\begin{aligned} c &= \min\left\{0,\ v f_X(v),\ w f_X(w),\ \frac{a}{2} f_X\left(\frac{a}{2}\right),\ \frac{b}{2} f_X\left(\frac{b}{2}\right)\right\}, \\ d &= \max\left\{0,\ v f_X(v),\ w f_X(w),\ \frac{a}{2} f_X\left(\frac{a}{2}\right),\ \frac{b}{2} f_X\left(\frac{b}{2}\right)\right\}, \end{aligned} \tag{49}$$

while (2):

$$c = 0, \quad d = \max\left\{0,\ v^2 f_X(v),\ w^2 f_X(w),\ \frac{4a^2}{9} f_X\left(\frac{2a}{3}\right),\ \frac{4b^2}{9} f_X\left(\frac{2b}{3}\right)\right\}, \tag{50}$$

and (3):

$$\begin{aligned} c = 0,\ d = \max\Bigg\{&0,\ (v - E[X])^2 f_X(v),\ (w - E[X])^2 f_X(w), \\ &\frac{4}{9}(a - E[X])^2 f_X\left(\frac{2a + E[X]}{3}\right),\ \frac{4}{9}(b - E[X])^2 f_X\left(\frac{2b + E[X]}{3}\right)\Bigg\}. \end{aligned} \tag{51}$$

Example values of 100 averaged results of the errors (14)-(16) determined based on (9) and (47), for A_{tz}=1, A_0=2, v=1.5, w=2, N=10^6, are: $\overline{\delta}_{\tilde{\bar{X}}}$=0.069%, $\overline{\delta}_{\overline{X^2}}$=0.067%, $\overline{\delta}_{\tilde{Var}[X]}$=0.077%.

3.6. Estimation of exponential distribution parameters

Let a random variable X have exponential distribution with parameters $\lambda \in \mathbf{R}_+$ and $\Delta \in \mathbf{R}$. The random variable X has the density [27]:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda(x-\Delta)}, & x \geq \Delta, \\ 0, & x < \Delta. \end{cases} \tag{52}$$

Based on (1)-(3), and (52), we obtain:

$$\bar{X} = \frac{1}{\lambda} + \Delta, \quad \overline{X^2} = \Delta^2 + \frac{2\Delta}{\lambda} + \frac{2}{\lambda^2}, \quad Var[X] = \frac{1}{\lambda^2}. \tag{53}$$

Parameters (1)-(3) are estimated based on (9). The ends of the interval [a, b] assume the values:

$$a = \Delta, \quad b = +\infty. \tag{54}$$

In practice, b has a finite value. The value of b can be selected either arbitrarily or from the formula:

$$b = -\frac{6\ln(\alpha)}{\lambda} + \Delta, \tag{55}$$

where $\alpha \in \mathbf{R}_+$ is the significance level [1].

The shape of formula (55) follows from a mathematical relation describing exponential distribution quantiles [27]. Making use of (4)-(6), it can be verified that for Δ=0, $0.01 \leq \alpha \leq 0.1$, $0.01 \leq \lambda \leq 100$, quantities (1)-(3) are computed in the interval [a, b] with errors from the interval [$2.8 \cdot 10^{-8}$%, 0.020%].

Making use of (12), (13), and (52), we determine the ends of the interval [c, d]. Estimating (1), we assume:

$$c = \begin{cases} \Delta f_X(\Delta), & \Delta < 0, \\ 0, & \Delta \geq 0, \end{cases} \quad d = \begin{cases} \frac{1}{\lambda} f_X\left(\frac{1}{\lambda}\right), & \Delta < \frac{1}{\lambda}, \\ \Delta f_X(\Delta), & \Delta \geq \frac{1}{\lambda}, \end{cases} \tag{56}$$

while (2):

$$c = 0, \quad d = \begin{cases} \frac{4}{\lambda^2} f_X\left(\frac{2}{\lambda}\right), & \Delta^2 f_X(\Delta) < \frac{4}{\lambda^2} f_X\left(\frac{2}{\lambda}\right), \\ \Delta^2 f_X(\Delta), & \Delta^2 f_X(\Delta) \geq \frac{4}{\lambda^2} f_X\left(\frac{2}{\lambda}\right), \end{cases} \tag{57}$$

and (3):

$$c = 0, \quad d = \frac{1}{\lambda}. \tag{58}$$

Metrol. Meas. Syst., Vol. XX (2013), No. 2, pp. 249–262.

Example values of 100 averaged results of the errors (14)-(16) determined based on (9) and (53), for λ=1, Δ=0, α=0.0027, N=10^6, are: $\bar{\delta}_{\tilde{\bar{X}}}$=0.27%, $\bar{\delta}_{\overline{X^2}}$=0.25%, $\bar{\delta}_{\tilde{V}ar[X]}$=0.43%.

3.7. Estimation of gamma distribution parameters

Let a random variable X have gamma distribution with parameters r, $k\in\mathbf{R}_+$. The random variable X has the density [27]:

$$f_X(x)=\begin{cases}\dfrac{1}{r^k\Gamma(k)}x^{k-1}e^{-\frac{x}{r}}, & x\geq 0,\\ 0, & x<0.\end{cases} \tag{59}$$

If k=1 and r=1/λ, then gamma distribution turns into exponential distribution.
Based on (1)-(3), and (59), we obtain:

$$\bar{X}=kr,\quad \overline{X^2}=k(k+1)r^2,\quad Var[X]=kr^2. \tag{60}$$

Parameters (1)-(3) are estimated based on (9). The ends of the interval [a, b] assume the values:

$$a=0,\quad b=+\infty. \tag{61}$$

In practice, b has a finite value. The values of a and b can be selected either arbitrarily or from the formula:

$$a=\begin{cases}-6\sqrt{k}r+kr, & -6\sqrt{k}r+kr\geq 0,\\ 0, & -6\sqrt{k}r+kr<0,\end{cases}\qquad b=\begin{cases}6\sqrt{k}r+kr, & k\geq 1,\\ 6\sqrt{\dfrac{1}{k}}r+\dfrac{1}{k}r, & k<1.\end{cases} \tag{62}$$

The shape of formula (62) is established based on the "6-sigma" rule, approximating gamma distribution by Gaussian distribution with parameters $\sigma_Y=\sqrt{k}\,r$ and $\mu_Y=kr$ [33]. Making use of (4)-(6), it can be verified that for 0.01≤k≤50 and 0.01≤r≤50, the quantities (1)-(3) are computed in the interval [a, b] with errors from the interval [3.4·10^{-9}%, 4.6%].
Using (12), (13), (59), and (62), we determine the ends of the interval [c, d]. Estimating (1), we assume:

$$c=0,\quad d=kr\,f_X(kr), \tag{63}$$

while (2):

$$c=0,\quad d=(k+1)^2r^2 f_X((k+1)r), \tag{64}$$

and (3), for a sufficiently small ε>0:

$$\begin{aligned}c=0,\ \ d=\max\Big\{&0,\ f_X(x)(x-kr)^2\Big|_{\substack{x=a,\ k\geq 1\\ x=\varepsilon,\ k<1}},\ f_X(b)(b-kr)^2,\\ &\frac{r^2}{4}\left(1-\sqrt{1+8k}\right)^2 f_X\left(\frac{r}{2}\left(1-\sqrt{1+8k}\right)+kr\right),\ \frac{r^2}{4}\left(1+\sqrt{1+8k}\right)^2 f_X\left(\frac{r}{2}\left(1+\sqrt{1+8k}\right)+kr\right)\Big\}.\end{aligned} \tag{65}$$

Example values of 100 averaged results of errors (14)-(16) determined based on (9) and (60), for k=0.5, r=10, N=10^6, ε=10^{-3}, are: $\bar{\delta}_{\tilde{\bar{X}}}$=0.16%, $\bar{\delta}_{\overline{X^2}}$=0.15%, $\bar{\delta}_{\tilde{V}ar[X]}$=1.5%.

3.8. Estimation of arcsine distribution parameters

Let a random variable X have arcsine distribution in the interval $(-A_{\sin}+A_0,\ A_{\sin}+A_0)$, $A_{\sin}\in\mathbf{R}_{+}$, $A_0\in\mathbf{R}$. The random variable X has the density [27]:

$$f_X(x)=\begin{cases}\dfrac{1}{\pi\sqrt{A_{\sin}^2-(x-A_0)^2}}, & |x-A_0|<A_{\sin},\\ 0, & |x-A_0|\ge A_{\sin}.\end{cases} \tag{66}$$

Based on (1)-(3), and (66), we obtain:

$$\overline{X}=A_0,\quad \overline{X^2}=\frac{A_{\sin}^2}{2}+A_0{}^2,\quad Var[X]=\frac{A_{\sin}^2}{2}. \tag{67}$$

Parameters (1)-(3) are estimated based on (9). Due to the fact that the function $f_X(x)$ is not defined at the points $(-A_{\sin}+A_0, 0)$ and $(A_{\sin}+A_0, 0)$, we assume that for a sufficiently small $\varepsilon>0$, the ends of the interval $[a, b]$ can be selected based on the formula:

$$a=-A_{\sin}+A_0+\varepsilon,\quad b=A_{\sin}+A_0-\varepsilon. \tag{68}$$

Making use of (12), (13), (66), and (68), we determine the ends of the interval $[c, d]$. Estimating (1), we assume:

$$c=\begin{cases}af_X(a), & af_X(a)<0,\\ 0, & af_X(a)\ge 0,\end{cases}\qquad d=\begin{cases}bf_X(b), & bf_X(b)\ge 0,\\ 0, & bf_X(b)<0,\end{cases} \tag{69}$$

while (2):

$$c=0,\quad d=\begin{cases}b^2f_X(b), & b^2f_X(b)\ge a^2f_X(a),\\ a^2f_X(a), & b^2f_X(b)<a^2f_X(a),\end{cases} \tag{70}$$

and (3):

$$c=0,\quad d=\frac{1}{4}(a-b)^2f_X(a)\ \text{or}\ d=\frac{1}{4}(a-b)^2f_X(b). \tag{71}$$

In the case of arcsine distribution, values of the functions $g_1(x)$ and $g_2(x)$ can only be nonnegative or nonpositive. If there exists $c>0$ or $d<0$, then parameters (1) and (2) can be estimated based on (10). Then estimating (1), we assume:

$$\begin{gathered}c=\begin{cases}\dfrac{2(a-\varepsilon)(b+\varepsilon)}{a+b}f\left(\dfrac{2(a-\varepsilon)(b+\varepsilon)}{a+b}\right), & af_X(a)\ge 0,\\ af_X(a), & af_X(a)<0,\end{cases}\\ d=\begin{cases}\dfrac{2(a-\varepsilon)(b+\varepsilon)}{a+b}f\left(\dfrac{2(a-\varepsilon)(b+\varepsilon)}{a+b}\right), & bf_X(b)<0,\\ bf_X(b), & bf_X(b)\ge 0,\end{cases}\end{gathered} \tag{72}$$

while (2):

$$c=\begin{cases}x_{\min}^2f_X(x_{\min}), & b^2f_X(b)\ge a^2f_X(a),\\ x_{\max}^2f_X(x_{\max}), & b^2f_X(b)<a^2f_X(a),\end{cases}\qquad d=\begin{cases}b^2f_X(b), & b^2f_X(b)\ge a^2f_X(a),\\ a^2f_X(a), & b^2f_X(b)<a^2f_X(a),\end{cases} \tag{73}$$

Metrol. Meas. Syst., Vol. XX (2013), No. 2, pp. 249–262.

where:

$$x_{\max}=\frac{3}{4}(a+b)+\sqrt{\frac{(a+b)^2}{16}+2\left(\frac{b-a}{2}+\varepsilon\right)^2},\ x_{\min}=\frac{3}{4}(a+b)-\sqrt{\frac{(a+b)^2}{16}+2\left(\frac{b-a}{2}+\varepsilon\right)^2}. \quad (74)$$

In the case of the function $g_3(x)$ the coefficients c=0 and $d\geq0$. This means that the parameter (3) can be estimated based on (9) or (10) and both formulas give the same results.

Example values of 100 averaged results of the errors (14)-(16) determined based on (9) and (67), for $A_{\sin}$=1, A_0=10, $\varepsilon=10^{-5}$, $N=10^5$, are: $\bar{\delta}_{\tilde{\bar{X}}}$=1.9%, $\bar{\delta}_{\overline{X^2}}$=2.1%, $\bar{\delta}_{\tilde{V}ar[X]}$=2.7%. In the considered example, c>0 exists and for mean c=6.34 and mean square c=62.4. Then determining errors from (10) and (67), we obtain: $\bar{\delta}_{\tilde{\bar{X}}}$=1.2%, $\bar{\delta}_{\overline{X^2}}$=1.5%. The obtained results show that (10) makes it possible to increase the accuracy of the estimation of parameters (1) and (2).

4. Conclusion

In the paper, properties of the random variable distribution have been examined. An approach consisting in the use of the Monte Carlo method to estimate distribution parameters has been proposed. An important advantage of the proposed approach is that regardless of the type of distribution, calculations are performed on the basis of data drawn from uniform distribution. In the research, distributions commonly used in measurement uncertainty analysis have been applied. Mathematical formulae facilitating practical application of the presented method have been derived. A simple modification of the method making it possible to increase measurement accuracy has been put forward. The obtained results have shown that, although the Monte Carlo method does not produce high accuracy results, it makes it possible to obtain reliable evaluations of distribution parameters.

References

[1] Evaluation of measurement data. (2008). *Guide to the expression of uncertainty in measurement*, JCGM 100, GUM 1995.

[2] Evaluation of measurement data. (2008). *Supplement 1 to the „Guide to the expression in measurement" - Propagation of distributions using a Monte Carlo method*, JCGM 101.

[3] Evaluation of measurement data. (2011). *Supplement 2 to the „Guide to the expression of uncertainty in measurement" - Extension to any number of output quantities*, JCGM 102.

[4] Krajewski, M. (2012). Construction of uncertainty budget using the Monte Carlo method for analysis of digital signal algorithm properties. *Measurement Automation and Monitoring*, 9, 770-773.

[5] Couto, P.R.G, Damasceno, J.C., et al. (2009). Comparative analysis of the measurement uncertainty of the deformation coefficient of a pressure balance using the GUM approach and Monte Carlo simulation methods. *XIX IMEKO World Congress: Fundamental and Applied Metrology*, 2051-2054.

[6] Fraga, I. C. S., Couto, P. R. G., et al. (2008). Uncertainty budget for primary electrolytic conductivity measurement comparing different methods. *2nd IMEKO TC19: Conference on environmental measurements*, *1st IMEKO TC23: Conference on food and nutritional measurements*.

[7] Kovacević, A., Kovacević, D., Osmokrović, P. (2012). Uncertainty evaluation of conducted emission measurement by the Monte Carlo method and the modified least-squares method. *Progress in electromagnetics research symposium proceedings*, 1173-1179.

[8] Cox, M. G., Siebert, B. R. L. (2006). The use of a Monte Carlo method for evaluating uncertainty and expanded uncertainty. *Metrologia*, 43, 178-188.

[9] Wübbeler, G., Krystek, M., Elster, C. (2008). Evaluation of measurement uncertainty and its numerical calculation by a Monte Carlo method. *Measurement science and technology*, 19(8), 084009.

[10] Kreutzer, Ph., Dorozhovets, N., et al. (2009). Monte Carlo simulation to determine the measurement uncertainty of a metrological scanning probe microscope measurement. *SPIE International Society for Optical Engineering*, 7378, 737816-737816-12.

[11] Lal-Jadziak, J., Sienkowski, S. (2008). Models of bias of mean square value digital estimator for selected deterministic and random signals, *Metrology and Measurement Systems*, 15(1), 55-69.

[12] Lal-Jadziak, J., Sienkowski, S. (2009). Variance of random signal mean square value digital estimator, *Metrology and Measurement Systems*, 16(2), 267-278.

[13] Hammersley, J.M., Handscomb, D.C. (1964). *Monte Carlo methods*. London: Methuen & Co.

[14] Buslenko, N.P., Golenko, D.I., et al. (1967). *The Monte Carlo method*. International series of monographs in pure and applied mathematics, 87, Pergamon Press, Oxford.

[15] Zieliński, R. (1970). *Monte Carlo methods*. Polish Scientific and Technical Publishers.

[16] Jermakow, S.M. (1976). *Monte Carlo method and related problems*. Polish Scientific Publishers.

[17] Jain, S. (1992). *Monte Carlo simulations of disordered system*. World Scientific Publishing Company.

[18] Fishman, G.S. (2003). *Monte Carlo concepts, algorithms, and applications*. Springer-Verlag, New York.

[19] Gentle, J.E. (2005). *Random number generation and Monte Carlo methods*. Springer Science+Business Media.

[20] Dagpunar, J. S. (2007). *Simulation and Monte Carlo with applications in finance and MCMC*. John Wiley & Sons.

[21] Kalos, M.H., Whitlock, P.A. (2008). *Monte Carlo Methods*. John Wiley & Sons.

[22] Rubinstein, R.Y., Kroese, D.P. (2008). *Simulation and the Monte Carlo method*. John Wiley & Sons.

[23] Shlomo, M., Shaul, M. (2009). *Applications of Monte Carlo method in science and engineering*. Cambridge University Press.

[24] Landau, D. P., Binder, K. (2009). *A guide to Monte Carlo simulations in statistical physics*. Cambridge University Press.

[25] Dunn, W.L., Shultis, J.K. (2012). *Exploring Monte Carlo methods*. Elsevier B.V.

[26] Sienkowski, S. (2012). Estimation of random variable distribution parameters by Monte Carlo method. Computer program: http://www.ime.uz.zgora.pl/ssienkowski/apps/soft/imemc.rar. Web page: http://www.ime.uz.zgora.pl/ssienkowski/imemc.html.

[27] Papoulis, A., Pillai, S. U. (2002). *Probability*, *random variables*, *and stochastic processes*. New York: McGraw-Hill.

[28] Wikramaratna, R. (2000). Pseudo-random number generation for parallel Monte Carlo - a splitting approach. *Society for Industrial and Applied Mathematics News*, 33(9).

[29] Mohanty, S., Mohanty, A. K., Carminati, F. (2012). Efficient pseudo-random number generation for Monte-Carlo simulations using graphic processors. *Journal of Physics: Conference Series*, conference 1, 368(012024).

[30] Coddinton, P. D. (1994). Analysis of random number generators using Monte Carlo simulation. *International Journal of Modern Physics C*, 5(3), 547-560.

[31] National Instruments Corporation. (2012). Web page: http://zone.ni.com/reference/en-XX/help/371361J-01/lvanls/gaussian_white_noise.

[32] Korn, G. T. (2000). *Mathematical handbook for scientists and engineers*. New York: McGraw-Hill.

[33] Liptak, B. G. (2005). *Instrument engineer's handbook: process control and optimization Vol.IV*. New York: Taylor and Francis Group.

[34] Kacker, R.N., Lawrence, J.F. (2007), Trapezoidal and triangular distributions for type B evaluation of standard uncertainty, *Metrologia*, 44(2), 117-127.