Research on Several Typical Distribution Discrete Data Sampling Methods for Evaluating Measurement Uncertainty Using Monte Carlo Method

Yi Yang *, Tao Wang, Yuzhou Chang, Wei Zhang, Rong Zou

Hubei Institute of Measurement and Testing Technology, Wuhan, Hubei, China

* Corresponding Author Email: alex_to@126.com

Abstract. The Monte Carlo Method (MCM) is a method of evaluating measurement uncertainty by using random sampling of probability distributions for distribution propagation. The discrete sampling of the probability density function (PDF) of the input quantity is a key and difficult step in the Monte Carlo method. This article proposes a probability density function (PDF) discrete sampling method based on EXCEL software by analyzing the probability density functions and their corresponding cumulative probability density function relationships of typical distribution types such as normal distribution, rectangular distribution, triangular distribution, trapezoidal distribution, and arcsine distribution, and develops an application program. The experimental verification comparison results show that the standard deviation of the method used in this paper for sampling discrete data is consistent with the theoretical calculation results, and the sampling of discrete data is efficient and accurate.

Keywords: Metrology, Monte Carlo method, measurement uncertainty, discrete data sampling.

1. Introduction

In the evaluation of measurement uncertainty, Monte Carlo method (MCM) is a method that uses random sampling of probability distributions to propagate the distribution and achieve measurement uncertainty evaluation [1]. The commonly used method for evaluating measurement uncertainty is the GUM method described in JJF 1059.1-2012 Evaluation and expression of uncertainty in measurement [2]. The MCM method is a supplementary method for evaluating measurement uncertainty using the GUM method. The MCM method is particularly suitable for situations where the measurement model is significantly nonlinear, the input probability density function (PDF) is significantly asymmetric, and the output PDF deviates significantly from normal or t-distribution. The method of evaluating measurement uncertainty using MCM is to first discretize the probability density function (PDF) of the input variable X, calculate the discrete value of the probability distribution of the output variable through the measurement model, and then directly obtain the best estimate, standard uncertainty, and inclusion interval of the output variable from the discrete distribution value of the output variable. References [3-5] all used the MCM method to evaluate the measurement uncertainty of different measured variables, but the relevant studies did not clearly describe the calculation process of the discrete values of the input variable probability distribution.

2. Typical Distribution Probability Density Function

Probability density distribution function, also known as probability density function or distribution density function, is a mathematical function that describes the probability density distribution of a continuous random variable. It provides a measure of the likelihood of a random variable occurring near a specific value point. The probability density distribution function is an important tool for describing the probability distribution of continuous random variables, providing a detailed characterization of the distribution of random variable values. By using the probability density distribution function, we can calculate the probability of a random variable falling within any interval, and understand the value characteristics and distribution of the random variable. The common types of input probability density distributions include normal distribution, rectangular distribution, triangular distribution, trapezoidal distribution, arcsine distribution, etc.

2.1. Normal Distribution

A normal distribution, also known as a Gaussian distribution, has a probability density function that follows a bell shaped curve and is symmetric about its mean. The normal distribution is completely determined by two parameters, namely the best estimate and the mean standard uncertainty $u(x)$. The best estimate determines the center position of the distribution, while the standard uncertainty $u(x)$ determines the width or dispersion of the distribution. The probability density function of a normal distribution is shown in equation (1), and the shape of the probability distribution is illustrated in Figure 1(a).

$$g_X(\xi)=\frac{1}{\sqrt{2\pi}u(x)}\exp\left[-\frac{(\xi-x)^2}{2u^2(x)}\right] \tag{1}$$

2.2. Rectangular Distribution

The probability of all possible outcomes of a rectangular distribution is equal. There are two main types of uniform distribution: discrete uniform distribution and continuous uniform distribution. All possible outcomes or subintervals of a rectangular distribution have the same probability. Rectangular distribution is a very simple probability distribution that is easy to understand and calculate. For the input quantity X, its lower limit is a, its upper limit is b, and $a<b$. The probability density function of the rectangular distribution is shown in equation (2), and the shape of the probability distribution is illustrated in Figure 1(b).

$$g_X(\xi)=\begin{cases}1/(b-a), a\le\xi\le b\\ 0, \text{else}\end{cases} \tag{2}$$

2.3. Triangular Distribution

Triangular distribution is a continuous probability distribution, and its probability density function (PDF) takes the shape of a triangle. This distribution is commonly used to represent the probability distribution of variables that only know their minimum, maximum, and most likely values (modes). The probability density function of a triangular distribution can be defined based on these parameters and has specific shapes at the minimum, maximum, and mode values. This distribution is commonly used to simulate certain types of random processes, especially in the absence of complete probability distribution information. Assuming that the quantity X is the sum of two independent quantities X_1 and X_2, X_i follows a rectangular distribution with a lower limit of a_i and an upper limit of b_i, and b_1-a_1=b_2-a_2. The probability density function of the triangular distribution is shown in equation (3), and the shape of the probability distribution is illustrated in Figure 1(c).

$$g_X(\xi)=\begin{cases}(\xi-a)/\omega^2, a\le\xi\le x\\ (b-a)/\omega^2, x\le\xi\le b\\ 0, \text{else}\end{cases} \tag{3}$$

$$\text{where, } x=(a+b)/2, \omega=(b-a)/2$$

2.4. Trapezoidal Distribution

A trapezoidal distribution is a specific form of probability distribution, characterized by shape parameters, position parameters, and scale parameters. The probability density is uniform within a certain interval, and gradually decreases to zero outside the interval. The triangular distribution is a continuous probability distribution, and its probability density function (PDF) presents a trapezoidal shape. Assuming that the quantity X is the sum of two independent quantities X_1 and X_2, X_i follows a rectangular distribution with a lower limit of a_i and an upper limit of b_i. The probability density function of the trapezoidal distribution is shown in equation (4), and the shape of the probability distribution is illustrated in Figure 1 (d).

$$g_X(\xi)=\begin{cases}(\xi-x+\lambda_2)/(\lambda_2^2-\lambda_1^2), x-\lambda_2\le\xi\le x-\lambda_1\\ 1/(\lambda_1+\lambda_2), x-\lambda_1<\xi<x+\lambda_1\\ (x+\lambda_2-\xi)/(\lambda_2^2-\lambda_1^2), x+\lambda_1<\xi\le x+\lambda_2\\ 0,\text{else}\end{cases} \tag{4}$$

where, $\lambda_1=|(b_1-a_1)-(b_2-a_2)|/2, \lambda_2=(b-a)/2$, and,
$0\le\lambda_1\le\lambda_2, x=(a+b)/2$

2.5. Arcsine Distribution

Arcsine distribution is a continuous probability distribution, characterized by its probability density function (PDF) being related to the inverse sine function of the sine function. Anyway, the probability density function of string distribution usually presents a symmetrical pattern between 0 and 1, increasing first and then decreasing. Assuming that the quantity X varies sinusoidally with an unknown phase Φ between the lower limit a and the upper limit b, Φ follows a rectangular distribution R (0, 2π). The probability density function of the arcsine distribution is shown in equation (5), and the shape of the probability distribution is illustrated in Figure 1(e).

$$g_X(\xi)=\begin{cases}(1/\pi)\left[(b-\xi)(\xi-a)\right]^{-1/2}, a<\xi<b\\ 0,\text{else}\end{cases} \tag{5}$$

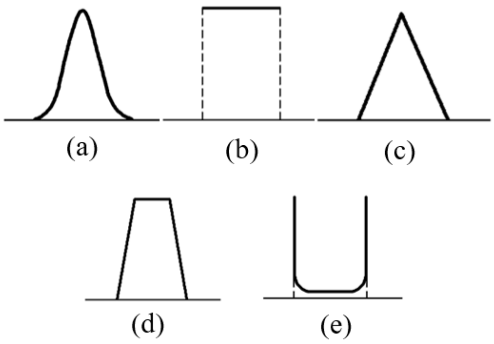


Figure 1. Schematic diagram of several typical probability distribution shapes

3. Implementation of Discrete Data Sampling

3.1. The Method of Sampling Discrete Data

Uses EXCEL software [6] to conduct research on discrete data sampling methods for several typical distribution types, including normal distribution, rectangular distribution, triangular distribution, trapezoidal distribution, and arcsine distribution. The specific method is as follows.

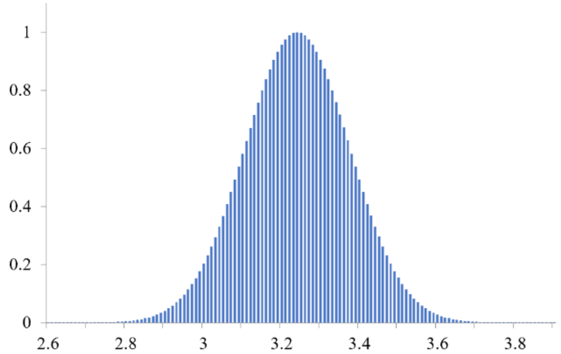
For a normal distribution, the "NORM. INV (probability, mean, standard_dev)" function in EXCEL can be called to implement discrete data sampling. Among them, probability corresponds to the probability of a normal distribution, mean is the arithmetic mean of the distribution, and standard_dev is the standard deviation of the distribution.

For rectangular distributions, the "RANDBETWEN (bottom, top)" function in EXCEL can be called to implement discrete data sampling. Among them, bottom is the smallest integer that the function will return, and top is the largest integer that the function will return.

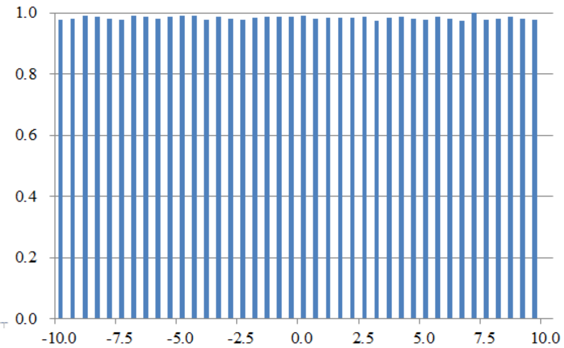
There are no directly callable functions in EXCEL software for distribution types such as triangular distribution, trapezoidal distribution, and arcsine distribution. For a certain distribution type, if the probability density function of a random variable x is known, its cumulative distribution function can be obtained through integration. That is to say, the cumulative distribution function is the integral of the probability density function from negative infinity to the current value point. Similarly, if the cumulative distribution function of the random variable X is known, its probability density function can be obtained by taking the derivative. The probability density function is the derivative of the cumulative distribution function. Based on this, the inverse function of the cumulative distribution function of the random variable in different intervals can be calculated first according to its cumulative distribution function. Then use EXCEL software to generate a random number between (0, 1), which is the probability of the occurrence of the random variable x. Then, based on the size of the random number, the value of the random number can be substituted into the inverse function of the cumulative distribution function of the corresponding interval to calculate the value of the random variable.

3.2. Discrete Data Sampling

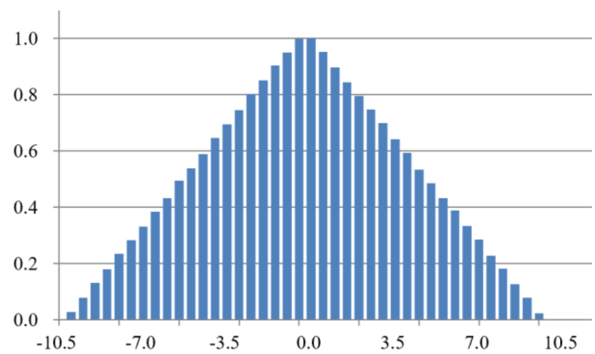
Using the above method, sample discrete data of several typical distribution types such as normal distribution, rectangular distribution, triangular distribution, trapezoidal distribution, and arctangent distribution, with a sample size of $M=10^6$. The probability distribution of obtaining discrete data with different distributions is shown in Figure 2. The horizontal axis in the figure represents the value of the random variable X, and the vertical axis represents the normalized frequency of occurrence of the random variable X. Comparing Figure 2 with Figure 1, it can be seen that the random number distribution obtained from sampling has good consistency with its distribution type.



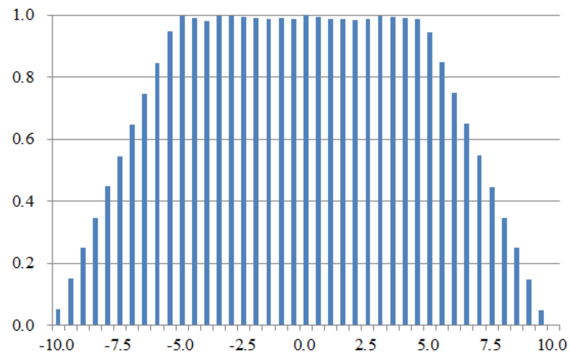
(a) Normal distribution



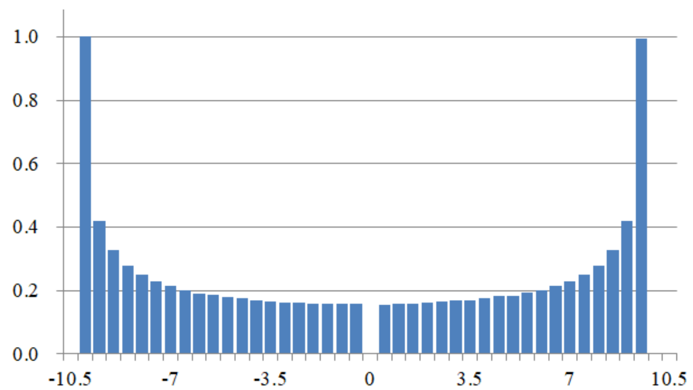
(b) Rectangular distribution



(c) Triangular distribution



(d) Trapezoidal distribution



(e) Arcsine distribution

Figure 2. Normalized frequency plots of several typical probability distributions for discrete data

3.3. Validation of Discrete Data

To further verify the correctness of the method proposed in this paper, a comparison was made between the standard deviation and theoretical standard deviation results of more discrete sampling data, as shown in Table 1. It can be seen that the difference between the two is very small, indicating that the method proposed in this article is correct and reliable.

Table 1. Comparison of discrete data standard deviation and theoretical standard deviation

Distribution type	Calculation formula for theoretical standard deviation	Theoretical standard deviation	Standard deviation of discrete data	Standard deviation of discrete data
Normal distribution	/	0.134	0.133	-0.75%
Rectangular distribution	$a/\sqrt{3}$	5.774	5.772	-0.03%
Triangular distribution	$a/\sqrt{6}$	4.082	4.084	0.05%
Trapezoidal distribution	$a/(\sqrt{6}/\sqrt{1+\beta^2})$	4.564	4.569	0.11%
Arcsine distribution	$a/\sqrt{2}$	7.071	7.070	-0.01%

4. Conclusion

By analyzing the probability density functions and their corresponding cumulative probability density function relationships of typical distribution types such as normal distribution, rectangular distribution, triangular distribution, trapezoidal distribution, and arctangent distribution, a probability density function (PDF) discrete sampling method based on EXCEL software is proposed, and an application program is developed. The experimental verification comparison results show that the standard deviation of the method used in this paper for sampling discrete data is consistent with the theoretical calculation results, and the sampling of discrete data is efficient and accurate. The research content can serve as an efficient and practical tool for evaluating measurement uncertainty using Monte Carlo method.

References

[1] JJF 1059.2-2012 Monte Carlo Method for Evaluation of Measurement Uncertainty [S]. Beijing: China Quality Inspection Press, 2012.

[2] JJF 1059.1-2012 Evaluation and Expression of Uncertainty in Measurement [S]. Beijing: China Quality Inspection Press, 2012.

[3] Liu Yuan-yuan, Yang Jian, Zhao Xi-yong, et al. Comparative Analysis of Uncertainty Measurement Evaluation with GUM and MCM [J]. ACTA METROLOGICA SINICA, 2018, 39(1), 135-139.

[4] Cheng Yin-bao, Chen Xiao-huai, Wang Zhong-yu, et al. Uncertainty Analysis and Evaluation of Form Measurement Task for CMM [J]. ACTA METROLOGICA SINICA, 2020, 41(2), 134-138.

[5] Li Yuan-feng, Meng Ling-chuan, Huang Yao, et al. Research on Uncertainty Evaluation Methods of Straightness of Straight Edge Based on Monte Carlo Method [J]. ACTA METROLOGICA SINICA, 2023, 44(4), 540-548.

[6] Microsoft. Excel help file, 2024.