DETERMINATION OF TEMPERATURE PROBABILITY DISTRIBUTION IN MEASUREMENT UNCERTAINTY

C. Y. Fang¹, C. K. Sung¹, C. Y. Hung², and K. W. Lui²

¹Department of Power Mechanical Engineering, National Tsing Hua University
Hsinchu, Taiwan 300, R.O.C.
²Center for Measurement Standards/ITRI, Hsinchu, Taiwan 300, R.O.C.

ABSTRACT

Temperature is one of the uncertainty sources in precision dimensional measurement. It is better to be evaluated by a Type B evaluation instead of a Type A evaluation, since the latter takes times and costs. A probability density function of temperature change is derived from mathematical model. Simulated data and statistical graphic analysis are used to verify the temperature probability distribution by goodness-fitting test. Results show that it has high confidence in the evaluation of the temperature uncertainty to be represented by a U-shaped distribution.

INTRODUCTION

Precision dimensional measurement is highly sensitive to the environmental temperature. A temperature of 20 °C is recommended to be the standard temperature in order to obtain the best measurement results [1], but temperature control has its limitations. The grade E of class AA laboratory proposed by JIMS 5 is delimited at ±0.5 °C [2]. Thus, measurement uncertainty due to temperature should be estimated. According to ISO GUM [3], the temperature uncertainty is usually estimated by Type B evaluation, but not by Type A evaluation due to cost consideration.

One of the main procedures for Type B evaluation is to hypothesize the probability distribution expressed by temperature effects and then deduce the Type B uncertainty from this probability distribution. If the hypothesis of the probability distribution is not appropriate, this will cause inappropriate uncertainty estimation. Tarapčik made an attempt at finding a probability distribution [4] by numerical simulation, which is mostly applied in normal and rectangular distributions, but there is no further description on the probability distribution of temperature changes. For the measurement uncertainty evaluation, the commonly used probability distributions are normal distribution, U-shaped distribution, rectangular distribution, and triangular distribution. In this paper, the probability density function of laboratory environmental temperature changes is derived from mathematical model. Simulated data and statistical analysis are used to test and verify its probability distribution by goodness of fit test. Proof of the probability distribution generated by temperature changes can be provided for users having high confidence data on Type B uncertainty evaluation and boost the quality of uncertainty evaluation. The size of laboratory and air conditioning performance may be also the influential factors to the uncertainty evaluation, but these factors will not be discussed in this paper.

PROBABILITY DENSITY FUNCTION OF THE TEMPERATURE DISTRIBUTION

The environmental temperature of a dimensional measurement laboratory is recommended to keep at (20 ± A) °C. For temperature correction in precision measurements, it is usually to observe the temperature at a specific time period (0 ≤ t ≤ b) which is a rectangular distribution. The room temperature variation of a temperature-stabilized laboratory against time will perform as a sine function. The temperature at time t can be represented as Y(t)=Asinωt, where A is the half bound of the temperature variation range and ω is the sampling rate (ω=2π/b). Y(t) can be rewritten to give the equation of t, which is then used to derive the probability density function (pdf). We get

\[ t = \frac{\sin^{-1}(\frac{Y(t)}{A})}{\omega} \]

and

\[ \frac{dt}{dY} = \frac{1}{\omega \sqrt{A^2 - Y^2}} \]

After a variable transform, we obtain the probability density function, f_Y(y), as follow.

\[ f_Y(y) = \frac{1}{\pi \sqrt{A^2 - Y^2}} \]  \hspace{1cm} (1)

Equation (1) performs a U-shaped distribution as shown in Fig. 1 and it is subjected to mathematical integration to obtain the probability sum, which is

\[ \int f_Y(y) dy = \int \frac{1}{\pi \sqrt{A^2 - Y^2}} dy = 1 \]

It meets the basic assumption of probability
density functions. Therefore, it is known that its probability density function is expressed in a U-shaped distribution when the laboratory temperature is in a sine distribution.

![Figure 1. U-shaped probability density function.](image1)

**VISUAL OBSERVATION OF TEMPERATURE PROBABILITY DISTRIBUTION**

The environmental temperature of a dimensional laboratory is usually controlled at (20 ± 1) °C. The laboratory temperature changes in a sine function with time can be expressed as \( T(t) = 20 \pm \sin \omega t \). We collect one measured temperature data every minute for 8 hours, which is the usual business hour for most laboratories. We can use statistical charts to help observing the distribution of the collected data. We calculate the frequency of data points by EXCEL Pivot Table Reports and then draw a bar chart as shown in Fig. 3 by using statistical table tools. Figure 3 shows data tending to be U-shaped, which meets the derived results in last section.

![Figure 2. Bar chart of temperature probability distribution.](image2)

However, the charts can only facilitate the determination of data distribution patterns since the probability distribution in Fig. 2 may also tends to be a rectangular distribution. Therefore, statistical methods should be used to confirm its probability distribution, so as to minimize the probability misestimate and increase the reliability of uncertainty estimation.

**STATISTICAL TESTING OF TEMPERATURE PROBABILITY DISTRIBUTION**

The most common and powerful statistical methods for goodness of data fitting are Chi-square test, Kolmogorov-Smirnov test (K-S test), and A-D test. The K-S test is the most typical method among these three statistical testing models. As our collected temperature data are continuous, probability histograms are used to determine the probability distribution. Prior to draw a chart, it is necessary to determine the number of groups and its range. We can group the collected data into \( k \) groups by Stauge grouping formula, \( k = 1 + 3.32 \cdot \log N \), where \( N \) is number of collected data. Since there are 480 sets of collected data, a group number \((k)\) of 10 is obtained. The range \((R)\) of the entire group of data is 2.0 °C. We get \( R/k \) to be 0.2 °C.

Prior to the test, we assume that the data values may appear as Beta, normal or rectangular probability distribution. We substitute the collected data into the software for testing. Table 1 summarized the results obtained from these three statistical tests at a statistical significance level of 5%. We can firstly determine the data probability distribution by observing the mean, variance, the skewness and kurtosis of the probability distribution. Table 1 shows that the data values are close to Beta distribution.

![Table 1](image3)

According to Table 1, the three distributions tested by chi-square test are all rejected. This implies that the testing results of the three above-mentioned distributions are not retained. However, according to the results of K-S test and the A-D test, there is no sufficient evidence to reject the temperature data not to be expressed in a Beta distribution, i.e. U-shaped distribution. Such findings conform the previous hypothesis of this study. This means that its probability density function appear to be a Beta distribution when laboratory’s temperature changes in a sine function with time. Figures 3 to 5 are comparisons of data histogram with their probability density function. The histograms in the charts are the distribution of collected data, whereas the curves show the probability density function fitted by the distributions. Figure 3 fits the above-mentioned hypotheses the best among these charts, whereas the probability density functions of Fig. 4 and Fig. 5 clearly do not meet these data histograms. Therefore, it can be assured that this data histogram is expressed in a Beta distribution.
Comparison of Input Distribution and Beta(0.50, 0.50) \* 2.00 + 19.0

Values in 10^1

Input
Beta

**FIGURE 3. Beta distribution test.**

Comparison of Input Distribution and Uniform(19.00, 21.00)

Values in 10^1

Input
Uniform

**FIGURE 4. Uniform distribution test.**

Comparison of Input Distribution and Normal(20.00, 0.71)

Values in 10^1

Input
Normal

**FIGURE 5. Normal distribution test.**

GROUP DATA DIFFERENCES

Probability distribution of data is not determined only based on charts, but also the objective data and scientific methods. Otherwise, it may cause misestimate. If data are equally grouped into 50 sets by one’s intuition instead of statistical grouping methods, the histograms of Beta distribution test and Uniform distribution test will be respectively drawn as shown in Figures 6 and 7. Data values should be expressed in a U-shaped distribution according to the original hypothesis, but they tend to a rectangular distribution in Fig. 7. If users are not observant enough, they will misuse the Type B uncertainty estimation for probability distribution and thereby underestimate the uncertainty. Although the Type B uncertainty estimation is obtained by non-statistical methods, it is practically better to use statistical methods so as to avoid underestimation or misestimate.

**FIGURE 6. Beta distribution test (k = 50).**

**FIGURE 7. Uniform distribution test (k = 50).**

TIME INTERVAL OF DATA COLLECTION

If the data acquisition interval is now 0.5 minute, 960 sets of data will be collected. The histogram and the probability density function are shown as in Fig. 8. It is found that Fig. 8 and Fig. 3 are not exactly the same, but both give a U-shaped distribution curve. Data can also be drawn in P-P chart as shown in Figure 9. In Fig. 9, the horizontal axis is the probabilities calculated from the collected data and the vertical axis is the probability values of the Beta probability distribution. When the curve closes to diagonal, the original hypothesis (i.e. the data values approach to a Beta probability distribution) becomes more accurate. This implies that the collected data really tend to a Beta distribution (or a U-shaped distribution).

**FIGURE 8. Beta distribution test (N = 960).**
When the temperature probability distribution is tested by histograms, the test results appear to be unrelated to the data-sampling interval. In other words, it is not necessary to make too much consideration on the sampling interval. However, if Fig. 8 is compared with Fig. 3, Fig. 3 appears to be smoother and easily visually observed to be a Beta distribution (U-shaped distribution). Thus, The more the data sampling, the easier is the visual observation of data.

Therefore, too much consideration in sampling size is not necessary. However, when more data are collected, it is easier to determine its probability distribution and even visual observation can achieve the determination of probability distribution. While evaluating the uncertainty for precision dimensional measurements, Type B evaluation is usually used for temperature, but its probability distribution can only be used as a hypothesis. After applying distribution tests and verification, it is relieved to estimate the temperature effects to be an U-shaped distribution in the future. Moreover, if environmental monitoring can be computerized, the environmental temperature and relative humidity information of the laboratory can be easily accessed online in a real-time basis. By the high efficient and speedy computer computation capability, instantaneous temperature probability distribution can be immediately computed to provide the users with more convenient and accurate data to boost measurement quality.

CONCLUSIONS
This paper first proves that temperature stably changes in a sine function with time through the derivation of mathematical formulae, and its probability density function is a U-shaped distribution. Then, we use numerical simulation with actual data, histograms, and statistical tests to test and verify that the probability density function to be truly a U-shaped distribution. While actually executing environmental monitoring, another factor affecting the test results is the sampling size. This study shows that sampling size does not make a great impact on the results of probability distribution derivative.

REFERENCES
1. ISA RP52.1, Recommended Environments for Standards Laboratories, 1975.

TABLE 1. Statistical test of Beta, normal, and uniform distributions

<table>
<thead>
<tr>
<th>Statistics/Distribution Source</th>
<th>Beta(0.50,0.50) * 2.00 + 19.00</th>
<th>Normal(20.00,0.71)</th>
<th>Uniform(19.00,21.00)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>19</td>
<td>18.9999949</td>
<td>18.999949</td>
</tr>
<tr>
<td>Maximum</td>
<td>21</td>
<td>21.000051</td>
<td>21.000051</td>
</tr>
<tr>
<td>Mean</td>
<td>20.002079</td>
<td>20.002079</td>
<td>20</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.708434</td>
<td>0.708434</td>
<td>0.708434</td>
</tr>
<tr>
<td>Variance</td>
<td>0.501879</td>
<td>0.501879</td>
<td>0.501879</td>
</tr>
<tr>
<td>Skewness</td>
<td>-4.16E-03</td>
<td>-3.92E-03</td>
<td>0</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>1.490279</td>
<td>1.489988</td>
<td>1.8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Test</th>
<th>Beta</th>
<th>Normal</th>
<th>Uniform</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chi-Square</td>
<td>55.026011</td>
<td>583.08534</td>
<td>180.651536</td>
</tr>
<tr>
<td>Kolmogorov-Smirnov Confidence</td>
<td>Rejected</td>
<td>Rejected</td>
<td>Rejected</td>
</tr>
<tr>
<td>Anderson-Darling Confidence</td>
<td>&gt;0.15 *</td>
<td>Rejected</td>
<td>Rejected *</td>
</tr>
<tr>
<td>Anderson-Darling Confidence</td>
<td>0.635193</td>
<td>15.04502</td>
<td>33.235429</td>
</tr>
</tbody>
</table>