XVII IMEKO World Congress
Metrology in the 3rd Millennium
June 22–27, 2003, Dubrovnik, Croatia

DETERMINATION OF UNCERTAINTY ASSOCIATED WITH QUANTIZATION ERRORS USING THE BAYESIAN APPROACH

Hiroshi Sato and Kensei Ehara

National Metrology Institute of Japan (NMIJ) / AIST, Tsukuba, Japan

Abstract – In practice, quantization of a measured quantity often significantly influences observation values. A typical example is found in measurements using digital instruments. In some cases, due to the quantization, no dispersion is observed among repeated measurements. The type A evaluation then gives zero standard uncertainty. In such a case, the most common practice is to assume, as an a priori distribution in type B evaluation, a uniform distribution, the width of which is given by the quantization interval, and take the width divided by square root of 12 as the standard uncertainty.

This practice, however, is justified only when the population standard deviation is exactly zero. But generally this condition does not hold true even if the sample standard deviation appears to be zero. In the present study, we use the Bayesian approach to evaluate the uncertainty of a measurement based on quantized data with due consideration to the difference between the standard deviation of the apparent sample and the population standard deviation.

We assume that the quantity before quantization obeys a normal distribution having the average μ and standard deviation σ. A measurement data corresponds to a value of the quantity after quantization. Based on a specific combination of *n* repeated measurements, we can construct the probability density $p(\mu, \sigma)$ using the Bayesian method. The standard deviation of the function, $\tilde{p}(\mu) = \int p(\mu,\sigma)d\sigma$, in terms of μ gives the uncertainty of the measurement result. We have shown that when all of the measurement data take the same value, the conventional type B evaluation described the above results in an underestimate of the uncertainty, if the number of data is less than five. Analysis is also conducted in cases in which not all of the data take the same value.

Keywords: GUM, quantization uncertainty, Bayesian approach

1. INTRODUCTION

Since issued, the Guide to the Expression of Uncertainty in Measurement (GUM) has been a useful common scale used in a number of fields to indicate the reliability of measurement [1]. However, several common problems still occur in these fields. One of these problems is the determination of uncertainty associated with quantization.

Concerning the uncertainty associated with quantization, the GUM describes an evaluation based on the resolution of the indicating device (scale interval Δ) (GUM F.2.2.1). According to this description, the type B evaluation is often conducted using $\Delta/\sqrt{12}$ as uncertainty for the scale interval Δ. In most cases, this evaluation is formulary. The uncertainty associated with quantization influences observation values when the data is uniform rather than when the data is dispersive. If all of the measurement data is identical, for example, if, due to the quantization, no dispersion is observed among repeated measurements. The type A evaluation then gives zero standard uncertainty. In such a case, the most common practice is to conduct the type B evaluation. This evaluation, however, may not always give appropriate uncertainty because the population standard deviation may not be zero even if the sample standard deviation appears to be zero. Under these circumstances, we need to review the selection of indicating device. If the necessary accuracy is well satisfied or an actual benefit cannot be expected from a new indicating device, the best estimation of uncertainty under the current conditions is preferable.

In the present study, we take the Bayesian approach to evaluate the uncertainty of measurement data associated with quantization. By using the Bayesian theorem, we can estimate the useful uncertainty from empirical information [2]. In the population of measurement data before quantization, the probability distribution having a parameter (μ, σ) is assumed as a probability variable. By using the Bayesian approach with the parameter (μ, σ) as the variable {μ, σ}, the average dispersion is estimated from the characteristics of prior probability on the variable plane and posterior probability estimated by n-times repetitive measurement.

2. QUANTIZATION

2.1. Resolution of indication device

Measurement data can be obtained from the indication device of some instruments. Most instruments almost have a digital indication device. Generally, the quantization error is not a significant problem. This means that the influence due to the resolution of the indication is negligible compared to the combined standard uncertainty. In contrast, when the resolution is not sufficient, errors due to quantization occur.

Consider $s(q_k)$, the experimental standard deviation of repeated observations q_k of a normally distributed random variable having parameter (μ, σ). When the data is quantized.

$$u^2(\delta) = s^2(q_k) - \sigma^2 \quad (1)$$

$u(\delta)$ is the uncertainty of quantization error

$s(q_k)$ is the experimental standard deviation

If we obtain enough data for the estimation of the parameter, the mean including error is as shown in Figure 1, and the standard uncertainty associated with quantization error was estimated like as Figure 2.

2.2. Problems

These figures show results for ideal occasions. In fact, due to limited sampling, the estimation value of uncertainty can be larger at the small observation number or small experimental standard deviation.

The data were sampled from a distribution having parameter (μ, σ) that we cannot actually know. Especially, in the case of the experimental standard deviation under Δ, some significant influence on the estimation occurs since that the data does not dispersion due to quantization.

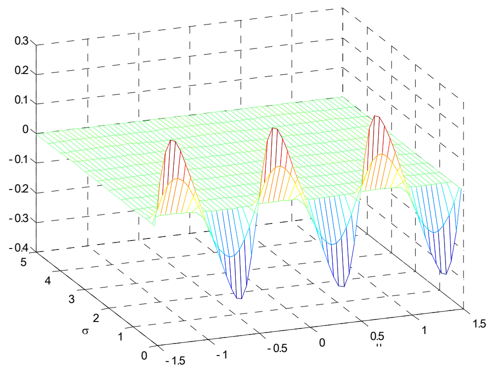

Fig.1 Error of best estimation of μ

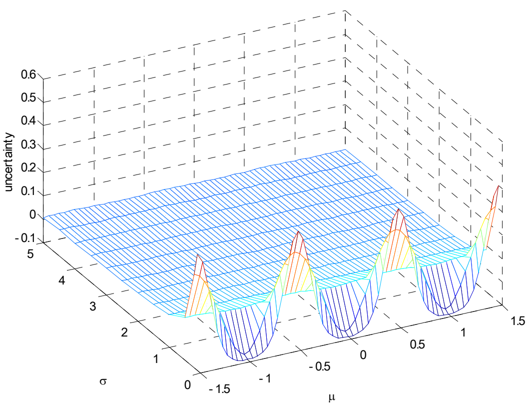

Fig.2 Error of best estimation of σ

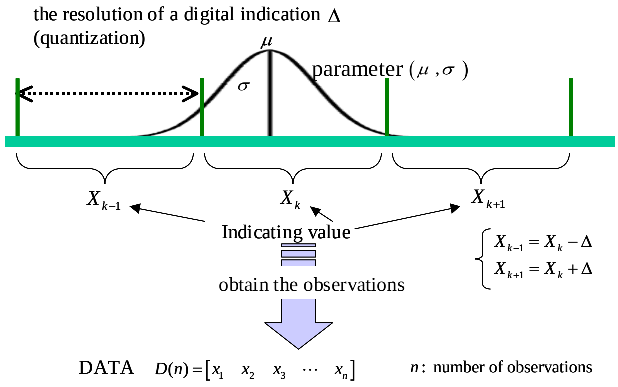

Fig.3 Quantization of observations

3. THE BAYESIAN APPROACH

3.1. Assumptions

Measurement data can be considered as a sample from a distribution having parameter (μ, σ) that we cannot know. An acquired quantity value is handled as measurement data after quantization. Since the sample data is from an unknown probability scatter, the estimation is simplified on the basis of the following assumptions:

Assumptions

[A1] observations are samples from a normal distribution

[A2] This event has indicating values for not more than three ranged values

[A3] prior probabilities of variable; $p\{\mu, \sigma\}$ are identical

3.2. Event

The mean value from observed values is taken as a measured value. In the present study, from the number of observations and the number of indicated values, we obtain a mean value and its uncertainty using the Bayesian Approach (referred to hereafter as B.A.).

From Assumption [A2], an event can be described as follows:

$$E = [a \quad b \quad c] \quad (2)$$

a is the number of the indicating X_{k-1} observations

b is the number of the indicating X_k observations

c is the number of the indicating X_{k+1} observations

Observed values obtained thus far are digitized output (Fig. 3). We pursue the probability of event 'E' if we assume data to be acquired from a normal distribution.

Assuming a probability density function of normal distribution $N(\mu, \sigma)$ shows the probability mass of the quantized value that can be obtained as an observed value is pursued as follows:

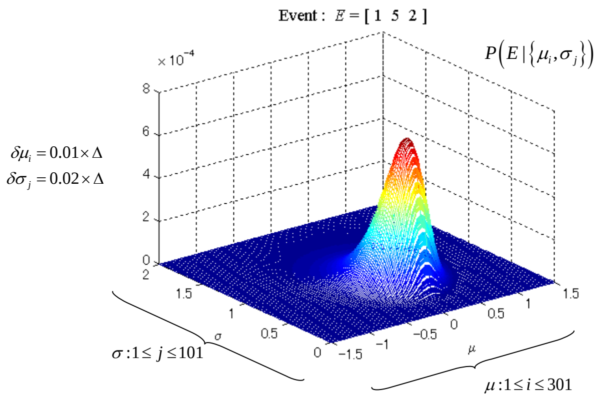

Fig.4 Probability mass of the calculation-space

$$P(X_k \mid \{\mu,\sigma\}) = \int_{X_k-\Delta/2}^{X_k+\Delta/2} N(\mu,\sigma)dx = \frac{1}{\sigma\sqrt{2\pi}}\int_{X_k-\Delta/2}^{X_k+\Delta/2} e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx \quad (3)$$

An event '*E*' indicates that the types of data value are three continuous values. And a, b, and c are the number of the continuous values.

The probability of appearing is given as the next equation by the number of each indicated value.

$$P(E \mid \{\mu,\sigma\}) = \left[\int_{X_{k-1}-\Delta/2}^{X_{k-1}+\Delta/2} N(\mu,\sigma)dx\right]^a \cdot \left[\int_{X_k-\Delta/2}^{X_k+\Delta/2} N(\mu,\sigma)dx\right]^b \cdot \left[\int_{X_{k+1}-\Delta/2}^{X_{k+1}+\Delta/2} N(\mu,\sigma)dx\right]^c \quad (4)$$

Now, we consider that the parameter { μ,σ } is a random variable in a set-space. The space setting affects accuracy of the final estimations and should be meshed to calculate the posterior probability. These meshed intervals also affect the accuracy of the final estimations. Figure 4 shows an example of calculation of the event $E=[1\quad 5\quad 2]$. In the space, the posterior probabilities for each argument in the setting and meshed space are estimated based on Bayes' theorem. Expectation of mean and dispersion of mean when event *E* occurred can be decided based on the characterized probability density function estimated from this space.

We assume that all the prior probability is equal, so that the expectation of mean is not a complicated equation. We meshed the space by δμ and δσ.

The probability density of the posterior probability $p(\{\mu,\sigma\} \mid E)$ from which one population is selected is shown as follows:

$$p(\{\mu_i,\sigma_j\} \mid E) = \frac{P(E \mid \{\mu_i,\sigma_j\})\cdot p(\{\mu_i,\sigma_j\})}{\sum_i\sum_j P(E \mid \{\mu_i,\sigma_j\})\cdot p(\{\mu_i,\sigma_j\})\cdot\delta\mu_i\cdot\delta\sigma_j} \quad (5)$$

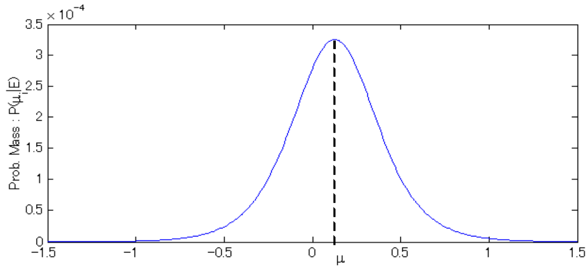

Fig.5 Probability mass of the μ axis

The probability function $p(\{\mu,\sigma\})$ is obtained from the prior probability density function from which the 2-variable { μ,σ } is selected.

3.3. Expectation of Mean and Variance

We calculated the expectation of mean from the event of the observed data. Using the probability density function of the posterior probability $p(\mu_i \mid E)$ integral for the σ axis, we estimate the expectation. Figure 5 shows the probability mass of the μ axis of the example.

$$p(\mu_i \mid E) = \frac{\sum_j P(E \mid \{\mu_i,\sigma_j\})\cdot p(\{\mu_i,\sigma_j\})\cdot\delta\sigma_j}{\sum_i\sum_j P(E \mid \{\mu_i,\sigma_j\})\cdot p(\{\mu_i,\sigma_j\})\cdot\delta\mu_i\cdot\delta\sigma_j} \quad (6)$$

$$\mathrm{E}[\mu \mid E] = \sum_i \mu_i\cdot p(\mu_i \mid E)\cdot\delta\mu_i = \frac{\sum_i \mu_i\cdot P(E,\mu_i)\cdot\delta\mu_i}{\sum_i P(E,\mu_i)\cdot\delta\mu_i} \quad (7)$$

where,

$$P(E,\mu_i) = \sum_j P(E \mid \{\mu_i,\sigma_j\})\cdot p(\{\mu_i,\sigma_j\})\cdot\delta\sigma_j \quad (8)$$

if we set the constant $\delta\mu_i$, the equation is as follows.

$$\mathrm{E}[\mu \mid E] = \frac{\sum_i \mu_i\cdot P(E,\mu_i)}{\sum_i P(E,\mu_i)} \quad (9)$$

Using the mean expectation estimated from (9), we can describe the variance of mean as follows:

$$\mathrm{V}[\mu \mid E] = \sum_i(\mu_i - \mathrm{E}[\mu \mid E])^2\cdot p(\mu_i \mid E)\cdot\delta\mu_i = \frac{\sum_i(\mu_i - \mathrm{E}[\mu \mid E])^2\cdot P(E,\mu_i)\cdot\delta\mu_i}{\sum_i P(E,\mu_i)\cdot\delta\mu_i} \quad (10)$$

From (10), a standard uncertainty of mean expectation $\mathrm{E}[\mu \mid E]=\hat{\mu}$ follows as

$$u(\hat{\mu}) = \sqrt{\frac{\sum_i(\mu_i - \mathrm{E}[\mu \mid E])^2\cdot P(E,\mu_i)}{\sum_i P(E,\mu_i)}} \quad (11)$$

4. ALGORITHM

4.1. Categorization of events

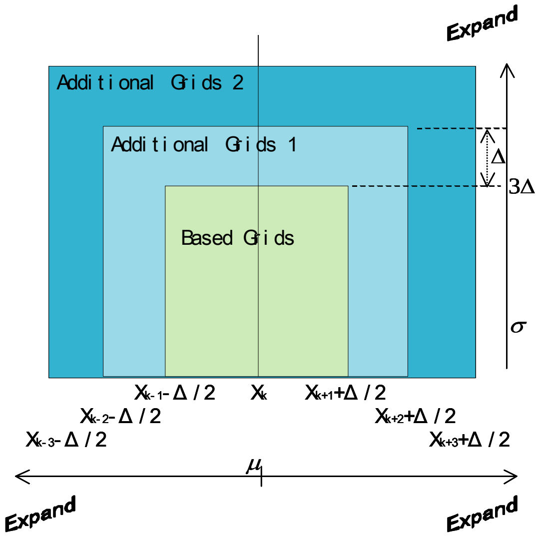

Fig.6 Calculation-space

For a space having the variable {μ、σ}, we calculate the posterior probability density changing continuously in the space and estimate the uncertainty from the calculated value. Figure 6 shows an outline of the algorithm.

$$\mu \text{ axis} \begin{cases} \delta\mu_i = 0.01\Delta \\ \mu_{\max} = 1.5\Delta \end{cases}$$

$$\sigma \text{ axis} \begin{cases} \delta\sigma_j = 0.01\Delta \\ \sigma_{\min} = 0.0001\Delta \\ \sigma_{\max} = 3\Delta \end{cases}$$

The above initial value is expanded by Δ in the ∞ direction from the σ axis and by Δ in the ∞ and −∞ directions from the μ axis. The space is expanded until the following criterion of convergence is satisfied.

4.2. Criterion of convergence

If the number of expansions is h, the space is expanded until the estimated mean uncertainty satisfies the following criterion:

$$u(\hat{\mu})_{h+1} - u(\hat{\mu})_h \geq 5.0\times10^{-5} \qquad (12)$$

Based on the mean uncertainty estimated under the criterion, an accuracy up to the third decimal place can be expected.

If the number of measurements n is small, however, a convergence failure is anticipated. In such a case, the maximum range of expansion is $\mu_{\max} < 100\Delta$.

5. ESTIMATIONS

5.1. Categorization of events

Based on these assumptions, the event E is classified into the following three cases:

[Case 1] All of the indicated values are identical.

[Case 2] Two types of adjacent indicated values are identical.

[Case 3] Three types of continuous indicated values are identical.

The uncertainty associated with quantization is discussed for each of the above cases.

For numeric operations, we used general-purpose calculation software, MATLAB Ver. 6.1 (Cybernet Systems Co., Ltd.).

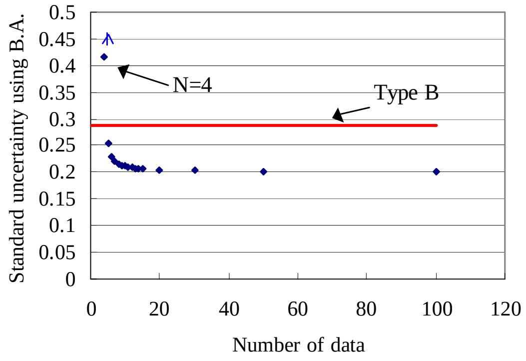

Fig.7 Standard uncertainty using B.A.; Case 1

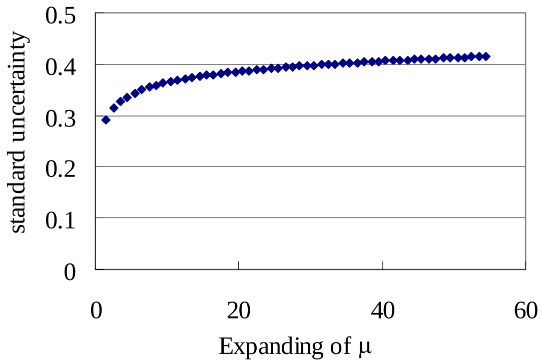

Fig. 8 Standard deviation of mean using B.A. at *N*=4, Case 1 =[0 4 0]

5.2. Case 1

In some cases, no apparent dispersion is observed among repeated measurements on account of the quantization. The type A evaluation then gives zero standard uncertainty. In such a case, the most common practice is to assume, as an a priori distribution in type B evaluation, a uniform distribution, the width of which is given by the quantization interval, and take the width divided by square root of 12 as the standard uncertainty in [1].

Therefore, we estimated the standard uncertainty of this case using the B.A. The results are shown in Fig.7. Depending on measurement data number *n*, an uncertainty of mean shows the estimation becoming small. When the measurement data number is less than five according to the estimation, the uncertainty estimation may be underestimated when using the conventional type B estimation.

When *N*=4, convergence criteria; equation. (11), was not attained, and the value given in Fig. 7 shows just the value at the maximum calculation-space. Figure 8 shows the increase of the estimation value depending on the algorithm at *N*=4.

5.3. Case 2

Similar to Case 1, the difference from these estimations of mean is gradually extending. These results are shown in Fig. 9. By increasing *N*, the difference of mean is getting large gradually.

Figure 10 shows the experimental standard deviation of mean with respect to the standard uncertainty using B.A. for Case 2, from *N*=4 to 15, 20, 30, 50 and 100. The standard uncertainty is estimated to be larger than the experimental STD of mean, as in the case for the estimation of mean. Particularly, at *N*=4, the estimation was very large. When *N*=4, the convergence criteria was not attained, and the value given in Fig. 10 shows just the value at the maximum calculation-space.

5.4. Case 3

It was not able to find the difference of the mean values between these estimation methods, given in Fig. 11.

Figure 12 shows the experimental standard deviation of the mean with respect to the standard uncertainty using B.A. for Case 3, from *N*=4 to 15, 20, 30 and 50. The standard uncertainty of *N*=4 using B.A. was estimated for very large uncertainty. As in the case for *N*=4 in Case 1 and Case 2, the estimation of uncertainty was estimated to be very large. When *N*=4, the convergence criteria was not attained, and the value given in Fig. 12 shows just the value at the maximum calculation-space.

5.5. Discussion

We assumed the prior probabilities of variable; $p\{\mu, \sigma\}$ are identical. The convergent criteria was not attained at *N*=4 in any cases. It was due to the calculation algorithm did not use the goodness of Bayesian approach positively. We consider that the algorithm should include some information of the prior probabilities as a variable. After this consideration, it is possible that the uncertainty is estimated more exactly.

Though numerous problems remain unresolved at *N*<5, it is considered that the approach is a new method to estimate of quantization uncertainty.

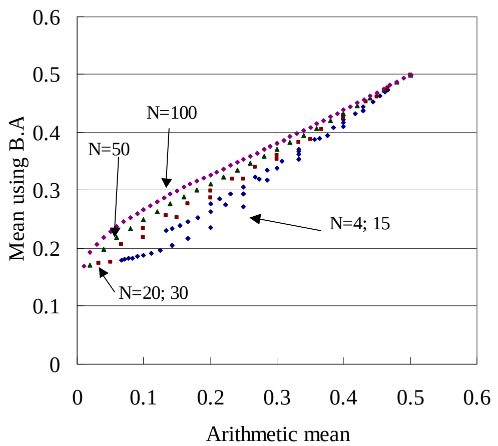

Fig. 9 Arithmetic mean vs. mean using B.A.; Case 2

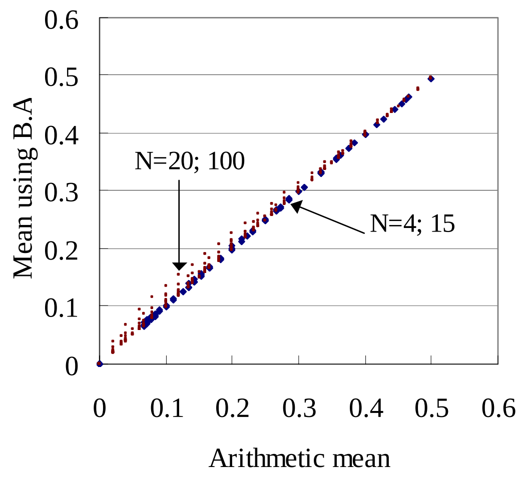

Fig. 11 Arithmetic mean vs. mean using B.A.; Case 3

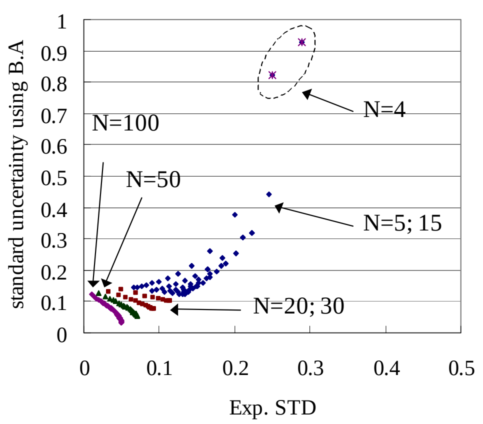

Fig. 10 Experimental standard deviation of mean vs. standard uncertainty using B.A.; Case 2

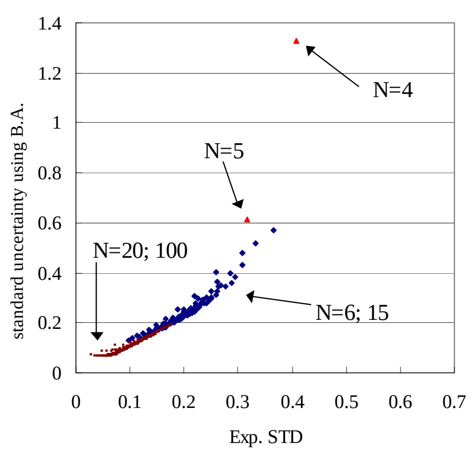

Fig. 12 Experimental standard deviation of mean vs. standard uncertainty using B.A.; Case 3

6. CONCLUSIONS

We calculated the estimation using the Bayesian approach from *N*=4 to 15, 20, 30, 50. In all cases the estimation of standard uncertainty as *N*=4 was larger than the experimental standard deviation of mean or the conventional type B evaluation.

We have shown that when the values of all of the measurement data are identical, the conventional type B evaluation described above results in underestimation of the uncertainty, if the number of data is less than five.

In estimation of the uncertainty associated with quantization, *N*=4 showed a long-tail posterior probability density function, where the mean value is not stable. Consequently, the mean uncertainty cannot be obtained in a space in which a solution of convergence is set. Therefore, we propose to obtain at least five data sets, not only when evaluating the uncertainty associated with quantization, but for all cases in which experimental data showing little dispersion is anticipated.

REFERENCES

[1] Guide to the Expression of Uncertainty in Measurement 2nd edition; BIPM, IEC, IFCC, ISO, IUPAC, IUPAC, IUPAP, OIML, International Organization for Standardization, 1995

[2] Calculation of Measurement Uncertainty Using Prior Information, S. D. Phillips, Journal of Research of the NIST Vol. 103-6 625-632, 1998

Authors:
Dr. Hiroshi Sato, Metrological Statistics and Particle Measurement Section, NMIJ / AIST
Address: 1-1-1-c3, Umezono Tsukuba, Ibaraki, 305 8563, Japan
Phone: +81 29 861 4051, Fax: +81 29 861 4070
e-mail: sato-hiroshi@aist.go.jp
Dr. Kensei Ehara, Metrological Statistics and Particle Measurement Section, NMIJ / AIST
Address: 1-1-1-C3, Umezono Tsukuba, Ibaraki, 305 8563, Japan
Phone: +81 29 861 4051, Fax: +81 29 861 4070
e-mail: ehara.kensei@aist.go.jp