TECHNICAL REPORT

ISO/TR 13587

First edition
2012-07-15

Three statistical approaches for the assessment and interpretation of measurement uncertainty

Trois approches statistiques pour l'évaluation et l'interprétation de l'incertitude de mesure

Reference number
ISO/TR 13587:2012(E)

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

Contents

Page

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of technical committees is to prepare International Standards. Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

In exceptional circumstances, when a technical committee has collected data of a different kind from that which is normally published as an International Standard ("state of the art", for example), it may decide by a simple majority vote of its participating members to publish a Technical Report. A Technical Report is entirely informative in nature and does not have to be reviewed until the data it provides are considered to be no longer valid or useful.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights.

ISO/TR 13587:2012 was prepared by Technical Committee ISO/TC 69, *Applications of statistical methods*, Subcommittee SC 6, *Measurement methods and results*.

This Technical Report is primarily based on Reference [10].

Introduction

The adoption of ISO/IEC Guide 98-3 (GUM) [1] has led to an increasing recognition of the need to include uncertainty statements in measurement results. Laboratory accreditation based on International Standards like ISO 17025 [2] has accelerated this process. Recognizing that uncertainty statements are required for effective decision-making, metrologists in laboratories of all types, from National Metrology Institutes to commercial calibration laboratories, are exerting considerable effort on the development of appropriate uncertainty evaluations for different types of measurement using methods given in the GUM.

Some of the strengths of the procedures outlined and popularized in the GUM are its standardized approach to uncertainty evaluation, its accommodation of sources of uncertainty that are evaluated either statistically (Type A) or non-statistically (Type B), and its emphasis on reporting all sources of uncertainty considered. The main approach to uncertainty propagation in the GUM, based on linear approximation of the measurement function, is generally simple to carry out and in many practical situations gives results that are similar to those obtained more formally. In short, since its adoption, the GUM has sparked a revolution in uncertainty evaluation.

Of course, there will always be more work needed to improve the evaluation of uncertainty in particular applications and to extend it to cover additional areas. Among such other work, the Joint Committee for Guides in Metrology (JCGM), responsible for the GUM since the year 2000, has completed Supplement 1 to the GUM, namely, "Propagation of distributions using a Monte Carlo method" (referred to as GUMS1) [3]. The JCGM is developing other supplements to the GUM on topics such as modelling and models with any number of output quantities.

Because it should apply to the widest possible set of measurement problems, the definition of measurement uncertainty in ISO/IEC Guide 99:2007 [4] as a "non-negative parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used" cannot reasonably be given at more than a relatively conceptual level. As a result, defining and understanding the appropriate roles of different statistical quantities in uncertainty evaluation, even for relatively well-understood measurement applications, is a topic of particular interest to both statisticians and metrologists.

Earlier investigations have approached these topics from a metrological point of view, some authors focusing on characterizing statistical properties of the procedures given in the GUM. Reference [5] shows that these procedures are not strictly consistent with either a Bayesian or frequentist interpretation. Reference [6] proposes some minor modifications to the GUM procedures that bring the results into closer agreement with a Bayesian interpretation in some situations. Reference [7] discusses the relationship between procedures for uncertainty evaluation proposed in GUMS1 and the results of a Bayesian analysis for a particular class of models. Reference [8] also discusses different possible probabilistic interpretations of coverage intervals and recommends approximating the posterior distributions for this class of Bayesian analyses by probability distributions from the Pearson family of distributions.

Reference [9] compares frequentist ("conventional") and Bayesian approaches to uncertainty evaluation. However, the study is limited to measurement systems for which all sources of uncertainty can be evaluated using Type A methods. In contrast, measurement systems with sources of uncertainty evaluated using both Type A and Type B methods are treated in this Technical Report and are illustrated using several examples, including one of the examples from Annex H of the GUM.

Statisticians have historically placed strong emphasis on using methods for uncertainty evaluation that have probabilistic justification or interpretation. Through their work, often outside metrology, several different approaches for statistical inference relevant to uncertainty evaluation have been developed. This Technical Report presents some of those approaches to uncertainty evaluation from a statistical point of view and relates them to the methods that are currently being used in metrology or are being developed within the metrology community. The particular statistical approaches under which different methods for uncertainty evaluation will be described are the frequentist, Bayesian, and fiducial approaches, which are discussed further after outlining the notational conventions needed to distinguish different types of quantities.

Three statistical approaches for the assessment and interpretation of measurement uncertainty

1 Scope

This Technical Report is concerned with three basic statistical approaches for the evaluation and interpretation of measurement uncertainty: the frequentist approach including bootstrap uncertainty intervals, the Bayesian approach, and fiducial inference. The common feature of these approaches is a clearly delineated probabilistic interpretation or justification for the resulting uncertainty intervals. For each approach, the basic method is described and the fundamental underlying assumptions and the probabilistic interpretation of the resulting uncertainty are discussed. Each of the approaches is illustrated using two examples, including an example from ISO/IEC Guide 98-3 (*Uncertainty of measurement — Part 3: Guide to the expression of uncertainty in measurement (GUM:1995)*). In addition, this document also includes a discussion of the relationship between the methods proposed in the GUM Supplement 1 and these three statistical approaches.

2 Normative references

The following referenced documents are indispensable for the application of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 3534-1:2006, *Statistics — Vocabulary and symbols — Part 1: General statistical terms and terms used in probability*

ISO 3534-2:2006, *Statistics — Vocabulary and symbols — Part 2: Applied statistics*

ISO/IEC Guide 98-3:2008, *Uncertainty of measurement — Part 3: Guide to the expression of uncertainty in measurement (GUM:1995)*

ISO/IEC Guide 98-3:2008/Suppl 1:2008, *Uncertainty of measurement — Part 3: Guide to the expression of uncertainty in measurement (GUM:1995) — Supplement 1: Propagation of distributions using a Monte Carlo method*

3 Terms and definitions

For the purposes of this document, the terms and definitions in ISO 3534-1, ISO 3534-2 and the following apply.

3.1
empirical distribution function
empirical cumulative distribution function
distribution function that assigns probability $1/n$ to each of the n items in a random sample, i.e., the empirical distribution function is a step function defined by

$$F_n(x) = \frac{\left|\{x_i \le x\}\right|}{n},$$

where $\{x_1, \ldots, x_n\}$ is the sample and $|A|$ is the number of elements in the set A.

3.2
Bayesian sensitivity analysis
study of the effect of the choices of prior distributions for the parameters of the statistical model on the posterior distribution of the measurand

3.3
sufficient statistic
function of a random sample $X_1,...,X_n$ from a probability density function with parameter θ for which the conditional distribution of $X_1,...,X_n$ given this function does not depend on θ

NOTE A sufficient statistic contains as much information about θ as $X_1,...,X_n$.

3.4
observation model
mathematical relation between a set of measurements (indications), the measurand, and the associated random measurement errors

3.5
structural equation
statistical model relating the observable random variable to the unknown parameters and an unobservable random variable whose distribution is known and free of unknown parameters

3.6
non-central chi-squared distribution
probability distribution that generalizes the typical (or central) chi-squared distribution

NOTE 1 For k independent, normally distributed random variables X_i with mean μ_i and variance σ_i^2, the random variable $X = \sum_{i=1}^{k} (X_i/\sigma_i)^2$ is non-central chi-squared distributed. The non-central chi-squared distribution has two parameters: k, the degrees of freedom (i.e., the number of X_i), and λ, which is related to the means of the random variables X_i by $\lambda = \sum_{i=1}^{k} (\mu_i/\sigma_i)^2$ and called the non-centrality parameter.

NOTE 2 The corresponding probability density function is expressed as a mixture of central χ^2 probability density functions as given by

$$
\begin{aligned}
g_X(\xi) &= \sum_{i=0}^{\infty} \frac{e^{-\lambda/2}(\lambda/2)^i}{i!} g_{Y_{k+2i}}(\xi) \\
&= \frac{e^{\frac{-(\xi+\lambda)}{2}}}{2^{\frac{k}{2}}} \sum_{i=0}^{\infty} \frac{\xi^{\frac{k}{2}+i-1}\lambda^i}{\Gamma\left(\frac{k}{2}+i\right)2^{2i}\, i!},
\end{aligned}
$$

where Y_q is distributed as chi-squared with q degrees of freedom.

4 Symbols (and abbreviated terms)

In 4.1.1 of the GUM, it is stated that Latin letters are used to represent both physical quantities to be determined by measurement (i.e., measurands in GUM terminology) as well as random variables that may take different observed values of a physical quantity. This use of the same symbols, whose different meanings are only indicated by context, can be difficult to interpret and sometimes leads to unnecessary ambiguities or misunderstandings. To mitigate this potential source of confusion, the more traditional notation often used in the statistical literature is employed in this Technical Report. In this notation, Greek letters are used to represent parameters in a statistical model (e.g., measurands), which can be either random variables or constants depending on the statistical approach being used and nature of the model. Upper-case Latin letters are used to represent random variables that can take different values of an observable quantity (e.g., potential measured values), and lower-case Latin letters to represent specific observed values of a quantity (e.g., specific measured values). Since additional notation may be required to denote other physical, mathematical, or statistical concepts, there will still always be some possibility for ambiguity[1]. In those cases the context clarifies the appropriate interpretation.

5 The problem addressed

5.1 The concern in this Technical Report is with a measurement model in which $\mu_1,\ldots,\mu_p$ are input quantities and θ is the output quantity:

$$\theta = f\left(\mu_1,\ldots,\mu_p\right), \tag{1}$$

where f is known as the measurement function. The function f is specified mathematically or as a calculation procedure. In the GUM (4.1, NOTE 1), the same functional relationship is given as

$$Y = f\left(X_1,\ldots,X_p\right) \tag{2}$$

which cannot be easily distinguished from the measurement function evaluated at the values of the corresponding random variables for each observed input.

Using the procedure recommended in the GUM, the p unknown quantities $\mu_1,\ldots,\mu_p$ are estimated by values $x_1,\ldots,x_p$ obtained from physical measurement or from other sources. Their associated standard uncertainties are also obtained from the relevant data by statistical methods or from probability density functions based on expert knowledge that characterize the variables. The GUM (also see 4.5 in Reference [11]) recommends that the same measurement model that relates the measurand θ to the input quantities $\mu_1,\ldots,\mu_p$ be used to calculate y from $x_1,\ldots,x_p$. Thus, the measured value (or, in statistical nomenclature, the estimate) y of θ is obtained as

$$y = f(x_1,\ldots,x_p), \tag{3}$$

that is, the evaluated Y, $y = f(x_1,\ldots x_p)$, is taken to be the measured value of θ. The estimates y, $x_1,\ldots,x_p$ are realizations of $Y, X_1,\ldots,X_p$, respectively.

5.2 In this Technical Report, three statistical approaches are each used to provide (a) a best estimate y of θ, (b) the associated standard uncertainty $u(y)$, and (c) a confidence interval or coverage interval for θ for a prescribed coverage probability (often taken as 95 %).

5.3 When discussing standard uncertainties, distinction is made between evaluated standard uncertainties associated with estimates of various quantities and their corresponding theoretical values. Accordingly, notation such as σ_μ or σ_X will denote theoretical standard uncertainties and notation such as S_X and s_x will denote an evaluated standard uncertainty before and after being observed, respectively.

---

1) For example, not all quantities represented by Greek letters in a statistical model must be parameters of the model. One common example of this type of quantity is the set of unobservable quantities that represent the random measurement errors found in most statistical models (i.e., the ε_i in the model $Y_i = \mu + \varepsilon_i$).

6 Statistical approaches

6.1 Frequentist approach

6.1.1 The first statistical approach to be considered, in which uncertainty can be evaluated probabilistically, is frequentist. The frequentist approach is sometimes referred to as “classical” or “conventional”. However, due to the nature of uncertainty in metrology, these familiar methods must often be adapted to obtain frequentist uncertainty intervals under realistic conditions.

6.1.2 In the frequentist approach, the input quantities $\mu_1,...,\mu_p$ in the measurement model (1) and the output quantity θ are regarded as unknown constants. Then, data related to each input parameter, μ_i, is obtained and used to estimate the value of θ based on the measurement model or the corresponding statistical models. Finally, confidence intervals for θ, for a specified level of confidence, are obtained using one of several mathematical principles or procedures, for example, least-squares, maximum likelihood, or the bootstrap.

6.1.3 Because θ is treated as a constant, a probabilistic statement associated with a confidence interval for θ is not a direct probability statement about its value. Instead, it is a probability statement about how frequently the procedure used to obtain the uncertainty interval for the measurand would encompass the value of θ with repeated use. “Repeated use” means that the uncertainty evaluation is replicated many times using different data drawn from the same distributions. Traditional frequentist uncertainty intervals provide a probability statement about the long-run properties of the procedure used to construct the interval under the particular set of conditions assumed to apply to the measurement process.

6.1.4 In most practical metrological settings, on the other hand, uncertainty intervals are to account for the uncertainty associated with estimates of quantities obtained using measured values (observed data) and also the uncertainty associated with estimates of quantities based on expert knowledge. To obtain an uncertainty interval analogous to a confidence interval, the quantities that are not based on measured values are treated as random variables with probability distributions for their values while those quantities whose values can be estimated using statistical data are treated as unknown constants.

6.1.5 Traditional frequentist procedures for the construction of confidence intervals are then to be modified to attain the specified confidence level after averaging over the potential values of the quantities assessed using expert judgment [5]. Such modified coverage intervals provide long-run probability statements about the procedure used to obtain the interval given probability distributions for the quantities that have not been measured, just as traditional confidence intervals do when all parameters are treated as constants.

6.1.6 Table 1 summarizes interpretations of the frequentist, Bayesian and fiducial approaches to uncertainty evaluation.

Table 1 — Interpretations of the approaches to uncertainty evaluation

Approach	Characterization of quantities in measurement model $\theta = f\left(\mu_1,\ldots,\mu_p\right)$	Uncertainty interval for output quantity θ	Note
Frequentist	θ and the μ_i all unknown constants	Long-run occurrence frequency that interval contains θ	Classical frequentist approach extended to integrate over uncertainties that are not statistically evaluated
Bayesian	θ and the μ_i are random variables. Their probability distributions represent beliefs about the values of the input and output quantities	Coverage interval containing θ based on a posterior distribution for θ	Possible non-uniqueness of interval due to the choice of priors
Fiducial	μ_i regarded as random variables whose distributions are obtained from assumptions on observed data used to estimate μ_i and expert knowledge about μ_i	Coverage interval containing θ based on a fiducial distribution for θ	Non-uniqueness due to the choice of the structural equation

6.2 Bayesian approach

The second approach is called the Bayesian approach. It is named after the fundamental theorem on which it is based, which was proved by the Reverend Thomas Bayes in the mid-1700s [12]. In this approach, knowledge about the quantities in measurement model (1) in Clause 5 is modelled as a set of random variables that follow a joint probability distribution for $\mu_1,\ldots,\mu_p$ and θ. Bayes' theorem then allows these probability distributions to be updated based on the observed data (also modelled using probability distributions) and the interrelationships of the parameters defined by the function f or equivalent statistical models. Then, a probability distribution is obtained that describes knowledge of θ given the observed data. Uncertainty intervals that contain θ with any specified probability can then be obtained from this distribution. Because knowledge of the parameter values is described by probability distributions, Bayesian methods provide direct probabilistic statements about the value of θ and the other parameters, using a definition of probability as a measure of belief.

6.3 Fiducial approach

6.3.1 The fiducial approach was developed by R.A. Fisher [13] in the 1930s. In this approach, a probability distribution, called the fiducial distribution, for θ conditional on the data is obtained based on the interrelationship of θ and the μ_i described by f and the distributional assumptions about the data used to estimate the μ_i. Once obtained, the fiducial distribution for θ can be used to obtain uncertainty intervals that contain θ with any specified probability.

6.3.2 The argument that justifies the process used to obtain the fiducial distribution is illustrated using a simple example. Suppose the values taken by a quantity Y can be described by the equation $Y = \mu + Z$, where μ is the measurand and Z is a quantity characterized by a standard normal random variable. If y is a realized value of Y corresponding to a realized value z of Z, then $\mu = y - z$. Despite Z not being observable, knowledge of the distribution from which z was generated enables a set of plausible values of μ to be determined. The probability distribution for Z can be used to infer the probability distribution for μ. The process of transferring the relationship $\mu = y - z$ to the relation $\mu = y - Z$ is what constitutes the fiducial argument. The fiducial distribution for μ is the probability distribution for the random variable $y - Z$ with y fixed.

6.4 Discussion

When describing the different methods for uncertainty evaluation under each of these statistical approaches, their fundamental underlying assumptions, incorporation of uncertainties obtained using Type A or Type B evaluation, and the probabilistic interpretation of the resulting uncertainty evaluations will be discussed. A description of how the methods used in the GUM relate to the frequentist, Bayesian, or fiducial results will also be given.

7 Examples

7.1 General

Two examples are given to illustrate the approaches. Example 1 is concerned with a physical quantity that is to be corrected for background interference. Table 2 gives the notation used and Subclauses 7.2 to 7.4 define variants of this evaluation problem. Example 2 is the calibration of the length of a gauge block taken from Annex H.1 of the GUM. Because it is more complicated, it is considered in Clause 11, after the three methods for uncertainty evaluation are discussed and illustrated using Example 1.

In later clauses, the three approaches will be applied to these examples.

NOTE The units of the quantities involved are not given when they are immaterial for the example.

Table 2 — Notation for Example 1

Quantity	Symbol
Physical quantity of interest (the measurand)	θ
Quantity detected by measurement method when measuring background (i.e., expected value of B) (Background interference)	β
Quantity detected by measurement method when measuring the physical quantity of interest (i.e., expected value of Y)	$\gamma = \theta + \beta$
Standard deviation of measurement method when measuring the physical quantity of interest (i.e., standard deviation of Y)	σ_Y
Standard deviation of measurement method when measuring background (i.e., standard deviation of B)	σ_B

7.2 Example 1a

Five measured values, obtained independently, of signal plus background are observed. Each measured value is assumed to be a realization of a random variable, Y, having a Gaussian distribution with mean $\gamma = \theta + \beta$ and standard deviation σ_Y. The measured values, y, of the signal plus background are

$$3{,}738, \quad 3{,}442, \quad 2{,}994, \quad 3{,}637, \quad 3{,}874.$$

This data has a sample mean of $\bar{y} = 3{,}537$ and a sample standard deviation of $s_y = 0{,}342$.

Similarly, five measured values, obtained independently, of the background are obtained. These measured values are assumed to be realizations of a random variable, B, having a Gaussian distribution with mean β and standard deviation σ_B. The observed values, b, of the background are

1,410, 1,085, 1,306, 1,137, 1,200.

Because there are measured values for each quantity that is a source of uncertainty, Example 1a has a straightforward statistical interpretation for each approach.

7.3 Example 1b

Example 1b is identical to Example 1a with the exception that the assessment of the background is based on expert knowledge or past experience, rather than on fresh experimental data. In this case, the background β is believed to follow a uniform (or rectangular) distribution with endpoints 1,126 and 1,329. Because expert judgment is applied, the uncertainty associated with a value of the background will be obtained using a Type B evaluation. Thus, Example 1b can be considered closer than Example 1a to a real measurement situation.

7.4 Example 1c

Example 1c is identical to Example 1b except that the signal θ is closer to the background. The data observed for the signal plus background in this case are

1,340, 1,078, 1,114, 1,256, 1,192.

With the signal just above the background, Example 1c illustrates how physical constraints can be incorporated in the evaluation of uncertainty for each approach.

8 Frequentist approach to uncertainty evaluation

8.1 Basic method

8.1.1 In the frequentist context, parameters are unknown constants. Following the convention to denote random variables by upper case letters and observed values of random variables by lower case letters, a confidence interval can be obtained from a *pivotal quantity* for θ, i.e., a function $W(Y,\theta)$ of the (possibly multivariate) data Y and the parameter θ, whose probability distribution is parameter-free (provided such a distribution can be determined.) Then, a $100(1-\alpha)$ % confidence interval for θ can be determined by calculating lower and upper percentiles ℓ_α and u_α to satisfy $P_\theta(\ell_\alpha \le W(Y,\theta) \le u_\alpha) = 1-\alpha$.

8.1.2 For example, let $Y=(Y_1,...,Y_n)$ be random variables, distributed as $N(\mu,\sigma^2)$, with the further random variable $\bar{Y}=\sum_{i=1}^{n} Y_i / n$. If the parameter of interest is μ, then for known σ, $Z=\dfrac{\bar{Y}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$.

is a pivotal quantity. The frequentist confidence interval for μ is

$$\bar{Y} \pm \frac{\sigma}{\sqrt{n}} z_{\alpha/2}, \tag{4}$$

where z_β is the $100\,\beta$ percentile of the standardized normal distribution.

If σ is not known, it can be estimated by the sample standard deviation

$$S = \sqrt{\frac{\sum_{j=1}^{n}\left(Y_j - \bar{Y}\right)^2}{n-1}}$$

Then, the (exact) pivotal quantity for μ is obtained by replacing σ in interval (4) by S:

$$\frac{\bar{Y}-\mu}{S/\sqrt{n}} \sim t(n-1). \tag{5}$$

Thus, a 100($1-\alpha$) % confidence interval for μ based on the Student's *t*-distribution is

$$\bar{Y} \pm \frac{S}{\sqrt{n}} t_{n-1,1-\alpha/2},$$

where $t_{n-1,\beta}$ is the 100β percentile of the *t*-distribution with $n-1$ degrees of freedom.

8.1.3 Instead of *exact* pivotal quantities, which exist only in simple situations, approximate pivotal quantities are commonly employed in applications. For large samples, the central limit theorem can be invoked to obtain approximate confidence intervals based on the normal distribution.

8.1.4 Further methods of obtaining confidence intervals (inverting a test statistic, pivoting a continuous cumulative distribution function, ordering the discrete sample values according to their probabilities, etc.) are discussed in Reference [14]. Some of them are mentioned in Example 1. A computer-intensive method, called the bootstrap, also can be used to construct a confidence interval for pivotal quantities that have unknown distributions. The bootstrap procedure is discussed in 8.2.

8.1.5 Although not explicitly given a frequentist justification from fundamental scientific considerations, the procedures recommended in the GUM can be used to obtain an approximate confidence interval for the measurand. Such confidence intervals are based on an approximate pivotal quantity with an assumed *t*-distribution obtainable from the measurement model (1). Under this procedure, the unknown quantities $\mu_1,\ldots,\mu_p$ are estimated by values $x_1,\cdots,x_p$ obtained from physical measurement or from other sources. Some of the values x_i might be sample means or other functions of data designed to estimate the quantities $\mu_i, i=1,\ldots,m$. Their associated standard uncertainties $u(x_i)$ are also evaluated from the data by statistical methods, typically using the sample standard deviation or using robust rank-based procedures. Such methods are known as Type A evaluations of uncertainty. The degrees of freedom ν_i associated with $u(x_i)$ is determined from the sample size used to estimate μ_i.

8.1.6 Since physical measurements might not always be possible or feasible for some of the μ_i, estimates x_i of μ_i for some i, say $i=m+1,\ldots,p$, are obtained by subjective (or potentially subjective) evaluations, and used together with x_i, for $i=1,\ldots,p$, obtained from Type A evaluations of uncertainty. Thus, non-statistical types of information are used to estimate $\mu_{m+1},\ldots,\mu_p$ using Type B evaluations of uncertainty, including scientific judgment, manufacturer's specifications, or other indirectly related or incompletely specified information.

NOTE Sometimes uncertainties are obtained by both Type A and Type B evaluations of uncertainty.

8.1.7 The GUM recommends that the same measurement model relating the measurand θ to the input quantities $\mu_1,\ldots,\mu_p$ be used to calculate y from $x_1,\ldots,x_p$. Thus, the measured value (or the estimate) y of θ is obtained as

$$y = f(x_1,\ldots,x_m,x_{m+1},\ldots,x_p),$$

that is, the evaluated Y, $y = f(x_1, \ldots x_p)$, is taken to be the measured value of θ.

8.1.8 In the GUM, the law of propagation of uncertainty is used to evaluate the standard uncertainty, $u(y)$. associated with y. The standard uncertainties $u(x_1), \ldots, u(x_p)$ associated with the values $\mathbf{x} = (x_1, \ldots, x_p)$ are used in the Taylor series expansion of the function $f(x_1, \ldots x_p)$ at $\mu_1, \ldots, \mu_p$, whose terms up to first order are

$$f(x_1, \ldots, x_p) \approx f(\mu_1, \ldots, \mu_p) + \sum_{i=1}^{p} c_i (x_i - \mu_i). \tag{6}$$

Denoting $(\mu_1, \ldots, \mu_p)$ by $\boldsymbol{\mu}$, the partial derivatives

$$c_i = \left. \frac{\partial f}{\partial \mu_i} \right|_{\boldsymbol{\mu} = \mathbf{x}}$$

are called sensitivity coefficients. Applying the law of propagation of uncertainty in the GUM gives the approximate standard uncertainty associated with y:

$$u(y) \approx \sqrt{\sum_{i=1}^{p} c_i^2 u^2(x_i) + 2 \sum_{i<j} c_i c_j u(x_i, x_j)}, \tag{7}$$

where $u(x_i, x_j)$ is the covariance between X_i and X_j.

8.1.9 To evaluate the standard uncertainty $u(y)$, the GUM uses the effective degrees of freedom ν_{eff} computed from the Welch-Satterthwaite formula,

$$\nu_{\text{eff}} = \frac{u^4(y)}{\sum_{i=1}^{p} \frac{c_i^4 u^4(x_i)}{\nu(x_i)}}. \tag{8}$$

NOTE Reference [15] discusses a counter-intuitive property according to which in interlaboratory studies a confidence interval based on the Welch-Satterthwaite approximation may be shorter for a between-laboratory difference than for one of its components.

8.1.10 Finally, in order to construct a confidence interval for θ, the approximate pivotal quantity,

$$W(y, \theta) = \frac{y - \theta}{u(y)} \tag{9}$$

is employed. According to the GUM,

$$W(Y, \theta) \sim t(\nu_{\text{eff}}), \tag{10}$$

that is, $W(Y, \theta)$ is an approximately pivotal quantity having a *t*-distribution with ν_{eff} degrees of freedom. The $100(1-\alpha)$ % confidence interval

$$y \pm u(y) t_{\nu_{\text{eff}}, 1-\alpha/2}, \tag{11}$$

for θ can then be recommended as the 100(1 – α) % uncertainty interval for θ. The half-width $t_{\nu_{\text{eff}}, 1-\alpha/2} u(y)$ of this interval is known as the expanded uncertainty associated with y.