Institute of Physics Publishing Metrologia
Metrologia 43 (2006) S167–S177 doi:10.1088/0026-1394/43/4/S02

Comparison of ISO-GUM, draft GUM Supplement 1 and Bayesian statistics using simple linear calibration

Raghu Kacker[1], **Blaza Toman**[1] **and Ding Huang**[2]

[1] National Institute of Standards and Technology, Gaithersburg, MD 20899, USA
[2] Navy Metrology Engineering and Calibration Program, Patuxent River, MD 20670, USA

E-mail: raghu.kacker@nist.gov

Received 24 October 2005
Published 4 August 2006
Online at stacks.iop.org/Met/43/S167

Abstract
We compare three approaches for quantifying uncertainty through a measurement equation: the International Organization for Standardization (ISO) *Guide to the Expression of Uncertainty in Measurement* (GUM), draft GUM Supplement 1 and Bayesian statistics. For illustration, we use a measurement equation for simple linear calibration that includes both Type A and Type B input variables. We consider three scenarios: (i) the measurement equation is linear with one Type B input variable having a normal distribution, (ii) the measurement equation is non-linear with two Type B input variables each having a normal distribution and (iii) the measurement equation is non-linear with two Type B input variables each having a rectangular distribution. We consider both small and large uncertainties for the Type B input variables. We use each of the three approaches to quantify the uncertainty in measurement for each of the three scenarios. Then we discuss the merits and limitations of each approach.

1. Introduction

The *Guide to the Expression of Uncertainty in Measurement* [1][3] has become a *de facto* international standard for evaluating and expressing uncertainty in measurement. The *Guide* is published by the International Organization for Standardization (ISO) in the names of seven international organizations and it is commonly identified as the ISO-GUM. The ISO-GUM propagates the estimates and their associated standard uncertainties for various input quantities through a linear approximation of the measurement equation to determine an estimate and its associated standard uncertainty for the value of the measurand. A working group of the Joint Committee for Guides in Metrology (JCGM) of the Bureau International des Poids et Mesures (BIPM) has developed a draft Supplement 1 to the ISO-GUM (draft GUMS1) [2]. The draft GUMS1 propagates probability distributions assigned to various input quantities through a numerical simulation of the measurement equation to determine a probability distribution for the value of the measurand. The draft GUMS1 is currently in circulation for review and comments from the member organizations of the JCGM and national measurement institutes (NMIs). Another approach is Bayesian statistics [3], which regards the quantities involved in a measurement process, including the measurand, as parameters having state-of-knowledge probability distributions. Bayesian statistics is an approach to update, on the basis of current measurement data, a prior probability distribution (representing the state of knowledge before measurement) about a statistical parameter to obtain a posterior probability distribution (representing the state of knowledge after measurement).

In section 2, we describe a simple linear calibration model using the terminology of the ISO-GUM. In sections 3, 4 and 5, we use the three approaches, the ISO-GUM, the draft GUMS1 and Bayesian statistics, to quantify the uncertainty

[3] The *Guide* is published in the names of seven international scientific organizations: International Bureau of Weights and Measures (BIPM), International Electro-technical Commission (IEC), International Federation of Clinical Chemistry (IFCC), International Organization for Standardization (ISO), International Union of Pure and Applied Chemistry (IUPAC), International Union of Pure and Applied Physics (IUPAP), and International Organization of Legal Metrology (OIML).

associated with the estimate of a measurand determined from the simple linear calibration model of section 2. In section 6, we discuss the results from the three approaches and in section 7 we discuss the merits and limitations of each approach. Concluding remarks appear in section 8.

2. Simple linear calibration model

Suppose the object of measurement is to determine, on the basis of the n measurements $x_1, \ldots, x_n$, an estimate for the value Y of a measurand and its associated uncertainty. Suppose the measurements $x_1, \ldots, x_n$ are regarded as independent realizations from a sampling distribution that is approximately normal $N(X, \sigma^2)$ with expected value X and variance σ^2, both unknown quantities. There is no additional *a priori* knowledge about the values of X and σ^2. An estimate for X and its associated standard uncertainty determined from statistical analysis of $x_1, \ldots, x_n$ are Type A evaluations. A probability distribution for X determined from the statistical analysis of $x_1, \ldots, x_n$ is a Type A state-of-knowledge distribution. (The concepts of Type A and Type B input variables and probability distributions are discussed in section 3.) To determine an estimate for Y, some assumption about the relationship between X and Y is required. Such assumption is generally based on scientific judgment. Suppose X and Y are related by the following simple linear equation

$$X = B_0 + B_1 Y. \tag{1}$$

We will suppose that the states of knowledge about the input quantities, the intercept B_0 and the slope B_1, may be expressed as Type B probability distributions based on scientific judgment and other information. The probability distributions for X, B_0 and B_1 are assumed to be mutually independent. A simple linear calibration model determined from the equation (1) is

$$Y = \frac{X - B_0}{B_1} = \left(\frac{-B_0}{B_1}\right) + \left(\frac{1}{B_1}\right) X, \tag{2}$$

where Y, X, B_0 and B_1 are regarded as variables with state-of-knowledge probability distributions. We will use the calibration model (2) as the measurement equation for the measurand Y. We will not consider any additional inputs for evaluating Y. Note that the variables $(1/B_1)$ and $(-B_0/B_1)$ in the measurement equation (2) are, respectively, multiplicative and additive corrections applied to X to determine Y.

As in the ISO-GUM, we use the symbols x and $u(x)$ for the expected value $E(X)$ and standard deviation $S(X)$ of a state-of-knowledge probability distribution for X. Likewise, we use the symbols b_0 and $u(b_0)$ for the expected value $E(B_0)$ and standard deviation $S(B_0)$ of a state-of-knowledge distribution for B_0. We use the symbols b_1 and $u(b_1)$ for the expected value $E(B_1)$ and standard deviation $S(B_1)$ of a state-of-knowledge distribution for B_1.

We will consider the following three scenarios.

Scenario 1. The slope B_1 is known to be one. That is, the measurement equation (2) is simply $Y = X - B_0$. The state-of-knowledge distribution for B_0 is normal $N(0, u^2(b_0))$.

Table 1. The data sets for scenario 1 (normal distribution for B_0).

Data sets	x	s	b_0	$u(b_0)$
1	100.521	1.502	0	0.250
2	100.521	1.502	0	4.000

Scenario 2. The state-of-knowledge distribution for B_0 is $N(0, u^2(b_0))$ and the state-of-knowledge distribution for B_1 is $N(1, u^2(b_1))$.

Scenario 3. The state-of-knowledge distribution for B_0 is $R(-u(b_0)\sqrt{3}, u(b_0)\sqrt{3})$, a rectangular (uniform) distribution on the interval $(-u(b_0)\sqrt{3}, u(b_0)\sqrt{3})$ with expected value zero and variance $u^2(b_0)$. The state-of-knowledge distribution for B_1 is $R(1 - u(b_1)\sqrt{3}, 1 + u(b_1)\sqrt{3})$ with expected value one and variance $u^2(b_1)$.

Our reasons for considering these three scenarios are as follows. The measurement equation for Y is linear in scenario 1 and non-linear for B_1 in scenarios 2 and 3. Thus we investigate both linear and non-linear measurement equations. The expected values and standard deviations for B_0 and B_1 are identical in scenarios 2 and 3; the only difference is the form of probability distributions. The differences between the results for scenarios 2 and 3 would indicate the sensitivity of the results to the form of probability distributions. We will consider two different standard deviations, one small and one large, for the probability distributions for B_0 and B_1. The differences between the results for small and large standard deviations would indicate the sensitivity of the results to the size of standard deviations for the input variables.

We will use the following data for the Type A evaluations of x and $u(x)$: the number of measurements is $n = 5$ and the measurements are 102.221 90, 99.294 46, 101.596 21, 100.811 06 and 98.679 92. (We regard these measurements as experimental measurements. However, they were numerically generated from a normal distribution.) The arithmetic mean $x = \sum_i x_i/n$, the estimated standard deviation $s = \sqrt{[\sum_i (x_i - x)^2/(n-1)]}$ and the estimated standard deviation of the mean $s(x) = s/\sqrt{n}$ of these measurements are $x = 100.521$, $s = 1.5023$ and $s(x) = 0.6718$, respectively.

The expected value for B_0 is $b_0 = 0$. We will consider the following two values for the standard deviation for B_0: $u(b_0) = 0.25$ and $u(b_0) = 4.00$. The expected value for B_1 is $b_1 = 1$. We will consider the following two values for the standard deviation for B_1: $u(b_1) = 0.05$ and $u(b_1) = 0.20$. Thus in scenario 1, for the two parameters X and B_0, we have the data sets given in table 1, and in scenarios 2 and 3, for the three parameters X, B_0 and B_1, we have the data sets given in table 2.

Comment 1. In many practical applications the relationship between X and Y is either $X - Y = B_0$ or $X/Y = B_1$. In the former case B_0 is additive bias (systematic error) in the measurements $x_1, \ldots, x_n$ and $B_1 \equiv 1$. In the latter case B_1 is fractional (multiplicative) bias in the measurements $x_1, \ldots, x_n$ and $B_0 \equiv 0$.

3. Review of the ISO-GUM

The ISO-GUM is based on the concept of a measurement equation

$$Y = f(Q_1, \ldots, Q_N) \tag{3}$$

Table 2. The data sets for scenario 2 (normal distributions for B_0 and B_1) and scenario 3 (rectangular distributions for B_0 and B_1).

Data sets	x	s	b_0	$u(b_0)$	b_1	$u(b_1)$
1	100.521	1.502	0	0.250	1	0.050
2	100.521	1.502	0	0.250	1	0.200
3	100.521	1.502	0	4.000	1	0.050
4	100.521	1.502	0	4.000	1	0.200

that mathematically represents the process (ingredients and recipe) for determining the estimate y and its associated standard uncertainty $u(y)$ from the estimates and their associated standard uncertainties for various input quantities $Q_1, \ldots, Q_N$. In the ISO-GUM the same symbols are used for the input and output quantities as well as for the corresponding input and output variables having state-of-knowledge probability distributions concerning the input and output quantities. In the measurement equation (3) the inputs $Q_1, \ldots, Q_N$ and the output Y are regarded as variables with state-of-knowledge probability distributions.

The uncertainties evaluated by statistical methods are referred to as Type A and the uncertainties evaluated by other means are referred to as Type B [1, sections 2.3.2 and 2.3.3]. The terms Type A and Type B apply to the methods of evaluation. The expected values, standard deviations, correlation coefficients and probability distributions of the input variables may be evaluated by statistical methods or determined by other means. The Type A and Type B classification naturally applies to the methods of evaluating the expected values, standard deviations, correlation coefficients and probability distributions of the input variables. We may refer to an input variable as Type A or Type B depending on whether its probability distribution or the parameters of its probability distribution are determined by statistical methods or by other means. We may also refer to the input quantities estimated by statistical analyses of the current measurements as Type A, and the input quantities estimated by other means as Type B. The Type A and Type B classification of the input variables and the input quantities is useful in our discussion of the draft GUMS1 and Bayesian statistics in subsequent sections.

A common Type A evaluation of an input quantity Q_i is the arithmetic mean of a series of m_i measurements that may reasonably be regarded as independent realizations from the same sampling distribution which is assumed to be normal with expected value Q_i and some unknown standard deviation σ_i. Suppose the arithmetic mean, the experimental (estimated) standard deviation and the estimated standard deviation of the mean of the m_i measurements are q_i, s_i and $s(q_i) = s_i/\sqrt{m_i}$, respectively. Then q_i is a sampling theory estimate of its expected value Q_i and $s(q_i)$ is the estimated standard deviation of the mean q_i. The ISO-GUM recommends the sampling theory estimate $s(q_i)$ as the standard uncertainty $u(q_i)$ associated with q_i [1, section 4.2], i.e. $u(q_i) = s(q_i)$. However, the ISO-GUM regards q_i and $u(q_i)$ as the expected value and approximate standard deviation of a state-of-knowledge probability distribution for the input variable Q_i, i.e. $E(Q_i) = q_i$ and $S(Q_i) \approx u(q_i) = s(q_i)$ [1, section 4.1.6]. Depending on the number m_i of independent measurements, the expression $u(q_i) = s(q_i)$ is uncertain because of the statistical reason of limited sampling [1, section E.4.3]. The uncertainty concerning $u(q_i)$ arising from a limited number of measurements is quantified by degrees of freedom. The degrees of freedom associated with $u(q_i)$ are $\nu_i = m_i - 1$.

A Type B evaluation of an input quantity Q_j is commonly obtained by assigning a state-of-knowledge probability distribution to the variable Q_j [1, section 4.3]. Then the estimate q_j is the expected value $E(Q_j)$ and the standard uncertainty $u(q_j)$ is the standard deviation $S(Q_j)$ of the assigned distribution. For example, if a rectangular distribution on the interval $(-a, a)$ is assigned to Q_j, then $E(Q_j) = q_j = 0$ and $S(Q_j) = u(q_j) = a/\sqrt{3}$.

The estimate y is determined by substituting the estimates $q_1, \ldots, q_N$ for the input quantities in the measurement equation $Y = f(Q_1, \ldots, Q_N)$. Thus,

$$y = f(q_1, \ldots, q_N). \tag{4}$$

The standard uncertainties $u(q_1), \ldots, u(q_N)$ associated with the input estimates $q_1, \ldots, q_N$ are components of uncertainty in determining y. The measurement equation (3) is approximated about y by a linear Taylor series as

$$Y \approx Y_{\text{linear}} = y + \sum_i c_i(Q_i - q_i), \tag{5}$$

where $c_1, \ldots, c_N$ are partial derivatives of Y with respect to $Q_1, \ldots, Q_N$ evaluated at $q_1, \ldots, q_N$, respectively. The partial derivatives $c_1, \ldots, c_N$ are referred to as sensitivity coefficients. If q_i and $u(q_i)$ were the expected value and standard deviation of a state-of-knowledge distribution for Q_i, then the variance of Y_{linear} would give the following expression for propagating the uncertainties associated with the input values

$$u^2(y) = \sum_i c_i^2 u^2(q_i) + 2 \sum_{(i<j)} c_i c_j u(q_i) u(q_j) r(q_i, q_j), \tag{6}$$

where $r(q_i, q_j)$ is the correlation coefficient between Q_i and Q_j for $i, j = 1, \ldots, N$ and $i \neq j$.

The correlation coefficients are Type A or Type B depending on whether they are determined by statistical analyses or by other means. If q_i and $u(q_i)$ were the expected value and standard deviation of Q_i, for $i = 1, \ldots, N$, then the estimate y and the standard uncertainty $u(y)$ would be the expected value and the standard deviation of Y_{linear}. The ISO-GUM regards y and $u(y)$ as approximate expected value and standard deviation of a state-of-knowledge probability distribution for Y.

When it is necessary to express the uncertainty as an interval, multiply the standard uncertainty $u(y)$ by a coverage factor k to obtain the expanded uncertainty $U = ku(y)$ and the uncertainty interval $[y \pm U] \equiv [y \pm ku(y)]$. The coverage probability of an uncertainty interval $[y \pm ku(y)]$ is the fraction of a state-of-knowledge distribution for Y that is encompassed by this interval. The ISO-GUM [1, section 6.2.2] is very clear that the interval $[y \pm ku(y)]$ is not to be interpreted as a confidence interval of sampling theory. To the extent that a state-of-knowledge probability distribution for Y represented by y and $u(y)$ is incompletely determined the coverage probability of $[y \pm ku(y)]$ cannot be stated [1, section 2.3].

The ISO-GUM does not specify a value for the coverage factor k. However, a commonly used value of k is two [4, 5].

The ISO-GUM [1, section 7.2.3 and note of section 4.3.4] stresses that when the uncertainty is expressed as an interval $[y \pm U]$, the coverage factor k should be explicitly stated in order that the standard uncertainty $u(y)$ may be recovered from the expanded uncertainty $U = ku(y)$.

The ISO-GUM [1, section 8, step 7] suggests selecting—when possible—the coverage factor k_p on the basis of the coverage probability p required of the interval $[y \pm k_p u(y)]$. The caveat 'when possible' is very important because a coverage factor k_p that yields a specified coverage probability p can be determined in very special conditions only. The ISO-GUM [1, section 6.3.2] is very clear that the estimate y and standard uncertainty $u(y)$ are themselves insufficient for determining the coverage factor k_p for a required coverage probability p. Guidance on specifying the coverage factor k_p based on a t-distribution with effective degrees of freedom determined from the Welch–Satterthwaite (W–S) formula are given in Annex G of the ISO-GUM. In metrology, the desired coverage probability p is generally set as 95% [4, 5].

3.1. The Type A evaluations determined from Bayesian statistics

The ISO-GUM's definition of the Type A evaluations [1, section 2.3.2] does not specify the statistical methodology that may be used. Sampling theory (frequentist statistics) and Bayesian statistics are two common statistical methodologies. The use of Bayesian statistics fully agrees with the definition of the Type A evaluations. Although the illustrations in the ISO-GUM use sampling theory, nowhere does the ISO-GUM state that only sampling theory should be used for the Type A evaluations. Thus the ISO-GUM does not preclude use of Bayesian statistics for the Type A evaluations.

The ISO-GUM is consistent only when the Type A evaluations are interpreted as parameters of state-of-knowledge probability distributions [6]. Then the Type A and the Type B evaluations have common probabilistic interpretation and they may be combined through a measurement equation. The authors of the ISO-GUM made it consistent by declaring [1, section 4.1.6] that the sampling theory estimates q_i and $u(q_i) = s(q_i)$ be interpreted as the parameters (expected value and approximate standard deviation) of a state-of-knowledge probability distribution for the input variable Q_i. Bayesian estimates are parameters of state-of-knowledge probability distributions. Therefore, when Bayesian statistics is used for the Type A evaluations there is no need of the ISO-GUM's superficial declaration.

The Bayesian standard uncertainty associated with the arithmetic mean q_i of m_i independent normally distributed measurements based on a well-known set of independent non-informative prior distributions for Q_i and σ_i is $u_{\text{Bayes}}(q_i) = \sqrt{(m_i - 1)/(m_i - 3)} \times s(q_i) = \sqrt{(m_i - 1)/(m_i - 3)} \times u(q_i)$ [3, 6, 7]. The arithmetic mean q_i and uncertainty $u_{\text{Bayes}}(q_i)$ are the expected value $E(Q_i)$ and standard deviation $S(Q_i)$ of a Bayesian posterior state-of-knowledge distribution for Q_i, which is a scaled and shifted t-distribution with degrees of freedom $\nu_i = m_i - 1$. When Bayesian statistics is used, q_i and $u_{\text{Bayes}}(q_i)$ are the exact expected value and standard deviation of a state-of-knowledge distribution for Q_i. Consequently, y and $u(y)$ are the expected value and standard deviation of Y_{linear} defined by (5). Thus the expression (6) for propagating the uncertainties—which is fundamental to the ISO-GUM—is better justified with Bayesian statistics than with sampling theory for the Type A evaluations.

The sampling theory estimate $u(q_i) = s(q_i)$ may be regarded as an approximation to the Bayesian estimate $u_{\text{Bayes}}(q_i)$ based on non-informative prior distributions for Q_i and σ_i [6, 7]. The approximation $u(q_i) = s(q_i)$ is poor when m_i is small but improves as m_i increases. Thus sampling theory estimates may still be used provided they are regarded as approximations to the Bayesian estimates.

As indicated in the ISO-GUM [1, section E.4.3], the sampling theory estimate $u(q_i) = s(q_i)$ is uncertain when the number m_i of measurements is limited. The uncertainty in $u(q_i) = s(q_i)$ arising from the limited number m_i of measurements, may be large for practical values of m_i. Therefore, $u(q_i) = s(q_i)$ is an incomplete expression of the uncertainty associated with q_i without an accompanying statement of its degrees of freedom. The degrees of freedom represent the uncertainty in $u(q_i) = s(q_i)$ arising from the limited number m_i of measurements. Similarly, $u(y)$ defined by (6) is an incomplete expression of the uncertainty associated with y without an accompanying statement of its effective degrees of freedom. Unlike a standard uncertainty determined from sampling theory, a Bayesian standard uncertainty has no uncertainty arising from a limited number of measurements. Thus the uncertainty $u_{\text{Bayes}}(y)$ obtained from (6) by using Bayesian statistics for the Type A evaluations is a complete expression of the uncertainty associated with y. In particular, a Bayesian standard uncertainty does not carry degrees of freedom.

Annex G of the ISO-GUM describes an approach (applicable in some special cases) for specifying a coverage factor k_p that yields an uncertainty interval $[y \pm k_p u(y)]$ with an approximate coverage probability p. The approach is based on calculating effective degrees of freedom from the Welch–Satterthwaite formula. An alternative to the ISO-GUM's approach based on Bayesian statistics is proposed in [7]. The use of Bayesian statistics greatly simplifies the expression of uncertainty by eliminating altogether the need for calculating the effective degrees of freedom from the W–S formula.

The object of this paper is to compare the ISO-GUM, the draft GUMS1, and Bayesian statistics using, for illustration, the measurement equation for simple linear calibration. The draft GUMS1 propagates probability distributions and Bayesian statistics updates prior probability distributions using current measurements. The standard uncertainties determined from the draft GUMS1 and Bayesian statistics are exact standard deviations of completely specified probability distributions. In particular, there is no uncertainty arising from the limited number of measurements in the standard uncertainties determined from the draft GUMS1 and Bayesian statistics. A standard uncertainty determined from the ISO-GUM where sampling theory is used for the Type A evaluations is an approximate expression that is incomplete without an accompanying statement of its effective degrees of freedom. Thus the standard uncertainties determined from the ISO-GUM using sampling theory for the Type A evaluations are qualitatively different entities from the standard uncertainties determined from the draft GUMS1 and Bayesian statistics. Thus it is difficult to compare the ISO-GUM using sampling theory for the Type A evaluations on the one hand and the draft GUMS1 and Bayesian statistics on the other hand. To make the standard uncertainties determined from the ISO-GUM comparable to the standard uncertainties determined from the draft GUMS1 and Bayesian statistics, in this paper we use Bayesian statistics for the Type A evaluations.

3.2. Application of the ISO-GUM to simple linear calibration

A Type A evaluation of the standard uncertainty associated with the mean $x = 100.521$ based on sampling theory is $s(x) = 0.672$ with four degrees of freedom [1, section 4.2]. The corresponding Bayesian standard uncertainty based on exactly the same information is determined as follows. Since there is no *a priori* information about X and σ^2, we use independent non-informative prior distributions for X and σ^2. It is convenient to use the following improper non-informative prior distributions: the distribution for X is proportional to one and the distribution for σ^2 is proportional to $1/\sigma^2$. These improper distributions yield an expression for Bayesian standard uncertainty that is very similar to the sampling theory estimate $s(x)$. The measurements $x_1, \ldots, x_n$ are assumed to be independent with a normal $N(X, \sigma^2)$ sampling distribution. This provides a likelihood function for X and σ^2 given the measurements $x_1, \ldots, x_n$. It can then be shown using the Bayes's theorem [3] that the Bayesian posterior probability distribution for $(X - x)/s(x)$, where $s(x) = s/\sqrt{n}$, is the Student's t-distribution with degrees of freedom $(n - 1)$. Here x and $s(x)$ are regarded as known quantities. It follows that the distribution for X is a scaled-and-shifted t-distribution with degrees of freedom $n - 1$ that has been scaled by $s(x)$ and shifted by x. The expected value and standard deviation of a t-distribution with degrees of freedom $(n - 1)$ are, respectively, zero and $[(n - 1)/(n - 3)]^{1/2}$ [8]. So the expected value and the standard deviation of X are $E(X) = x$ and $S(X) = [(n - 1)/(n - 3)]^{1/2} \times s(x)$, respectively. The uncertainty $u(x) = S(X) = [(n - 1)/(n - 3)]^{1/2} \times s(x)$ is a Bayesian standard uncertainty associated with x. A Bayesian state-of-knowledge probability distribution for X is a scaled-and-shifted t-distribution with expected value x and standard deviation $u(x)$.

For the data in tables 1 and 2, $x = 100.521$, $s = 1.502$, $s(x) = 0.672$, $[(n-1)/(n-3)]^{1/2} = 1.414$ and $u(x) = 0.950$. The sampling theory estimate $s(x) = 0.672$ may be regarded as an approximation with four degrees of freedom for the Bayesian uncertainty $u(x) = 0.950$. As discussed in the ISO-GUM [1, section E.4.3], since $s(x)$ has only four degrees of freedom it is not very reliable. The Bayesian uncertainty $u(x) = 0.950$ has no uncertainty arising from limited number of measurements. Vaguely, we may think of the multiplicative factor $[(n - 1)/(n - 3)]^{1/2} = 1.414$ built in the Bayesian uncertainty $u(x) = 0.950$ as accounting for the fact that $s(x) = 0.672$ is based on four degrees of freedom. In this paper we use $u(x) = 0.950$ as the standard uncertainty associated with the mean $x = 100.521$.

An estimate of Y based on the measurement equation (2) is

$$y = \frac{x - b_0}{b_1} = \left(\frac{-b_0}{b_1}\right) + \left(\frac{1}{b_1}\right) x. \tag{7}$$

Since $b_0 = 0$ and $b_1 = 1$, the estimate y for Y under all three scenarios is $y = (x - b_0)/b_1 = x = 100.521$.

As discussed in the appendix, the relative standard uncertainty $u_r(y) = u(y)/|y|$ associated with y defined by equation (7) is

$$u_r^2(y) = u_r^2(x - b_0) + u_r^2(b_1), \tag{8}$$

where we use the symbol $u(x - b_0)$ for the standard uncertainty $S(X - B_0) = [u^2(x) + u^2(b_0)]^{1/2}$ and the symbol $u_r(x - b_0)$ for the relative standard uncertainty $S(X - B_0)/|E(X - B_0)| = u(x - b_0)/|x - b_0|$. Thus the standard uncertainties under the three scenarios are as follows.

Scenario 1. By substituting $b_1 = 1$ and $u(b_1) = 0$ in equation (8), we get the following expression for the standard uncertainty $u(y)$

$$u(y) = y\sqrt{u_r^2(x - b_0)} = (x - b_0)\sqrt{u^2(x - b_0)}/(x - b_0)$$
$$= \sqrt{u^2(x) + u^2(b_0)}. \tag{9}$$

Scenario 2. The relative standard uncertainty $u_r(y)$ is given by equation (8). Thus the standard uncertainty $u(y)$ is given by

$$u(y) = y\sqrt{u_r^2(x - b_0) + u_r^2(b_1)}$$
$$= y\sqrt{(u^2(x) + u^2(b_0))/(x - b_0)^2 + u^2(b_1)/b_1^2}. \tag{10}$$

Scenario 3. The expected values b_0 and b_1 and the standard deviations $u(b_0)$ and $u(b_1)$ are identical in scenarios 2 and 3. Therefore, the standard uncertainty $u(y)$ for scenario 3 is also given by equation (10).

The estimate y, standard uncertainty $u(y)$ and the limits of the 2-standard uncertainty interval $[y \pm 2u(y)]$ for scenarios 1, 2 and 3 are displayed in tables 3, 4 and 5, respectively. The quantities displayed in table 3 are determined from equation (9) and the quantities displayed in tables 4 and 5 are determined from equation (10).

The last column in tables 3, 4 and 5 displays the coverage probabilities p associated with the uncertainty intervals $[y \pm 2u(y)]$ based on numerical simulation of the probability distributions for Y discussed in section 4. In table 3 for scenario 1 (normal distribution for B_0), the coverage probabilities are about 95%. In table 4 for scenario 2 (normal distributions for B_0 and B_1), when the standard deviation $u(b_1)$ is small $[u(b_1) = 0.05]$ the coverage probabilities are about 95% and when $u(b_1)$ is large $[u(b_1) = 0.20]$ the coverage probabilities are about 92%. In table 5 for scenario 3 (rectangular distributions for B_0 and B_1), the coverage probabilities exceed 95% when the standard deviation $u(b_1)$ is small $[u(b_1) = 0.05]$, but the coverage probabilities drop in the range of 91% to 92% when $u(b_1)$ is large $[u(b_1) = 0.20]$.

Table 3. The estimate y, standard uncertainty $u(y)$ and the limits of the uncertainty interval $[y \pm 2u(y)]$ for scenario 1 (normal distribution for B_0) based on the ISO-GUM.

	$u(x)$	$u(b_0)$	y	$u(y)$	$y - 2u(y)$	$y + 2u(y)$	$100 \times p$
1	0.950	0.250	100.521	0.982	98.556	102.486	95.3
2	0.950	4.000	100.521	4.111	92.298	108.743	95.4

Table 4. The estimate y, standard uncertainty $u(y)$ and the limits of the uncertainty interval $[y \pm 2u(y)]$ for scenario 2 (normal distributions for B_0 and B_1) based on the ISO-GUM.

	$u(x)$	$u(b_0)$	$u(b_1)$	y	$u(y)$	$y - 2u(y)$	$y + 2u(y)$	$100 \times p$
1	0.950	0.250	0.050	100.521	5.121	90.278	110.763	95.3
2	0.950	0.250	0.200	100.521	20.128	60.264	140.777	92.3
3	0.950	4.000	0.050	100.521	6.493	87.534	113.507	95.3
4	0.950	4.000	0.200	100.521	20.520	59.480	141.561	92.4

Table 5. The estimate y, standard uncertainty $u(y)$ and the limits of the uncertainty interval $[y \pm 2u(y)]$ for scenario 3 (rectangular distributions for B_0 and B_1) based on the ISO-GUM.

	$u(x)$	$u(b_0)$	$u(b_1)$	y	$u(y)$	$y - 2u(y)$	$y + 2u(y)$	$100 \times p$
1	0.950	0.250	0.050	100.521	5.121	90.278	110.763	99.3
2	0.950	0.250	0.200	100.521	20.128	60.264	140.777	91.3
3	0.950	4.000	0.050	100.521	6.493	87.534	113.507	96.5
4	0.950	4.000	0.200	100.521	20.520	59.480	141.561	91.8

4. The draft GUM Supplement 1

The basic algorithm of the draft GUMS1 [2] to evaluate the estimate y and its associated uncertainty is as follows.

(1) Define the measurement equation $Y = f(Q_1, \ldots, Q_N)$ as in the ISO-GUM. The draft GUMS1 addresses two situations: (i) the state-of-knowledge probability distributions for the input variables $Q_1, \ldots, Q_N$ are all mutually independent and (ii) the joint state-of-knowledge probability distribution for $Q_1, \ldots, Q_N$ is a multivariate normal distribution. In case (i), the joint pdf for $Q_1, \ldots, Q_N$ is the product of the individual univariate pdfs for $Q_1, \ldots, Q_N$ and the data may be numerically simulated independently for each input variable. In case (ii), the data are numerically simulated from the joint multivariate normal distribution.

(2) Define a joint probability distribution for $Q_1, \ldots, Q_N$. When $Q_1, \ldots, Q_N$ are independently distributed, assign a state-of-knowledge probability distribution to each of the input variables $Q_1, \ldots, Q_N$ then define the joint pdf as a product of the independent pdfs.

The draft GUMS1 [2, clause 4, table 1] lists the following six probability distributions for the input variables useful for some common circumstances:

(i) Normal distribution with mean x and variance $u(x)$.
(ii) Exponential distribution with expected value x.
(iii) Scaled-and-shifted t-distribution with degrees of freedom $n - 1$ that has been scaled by $s(x)$ and shifted by x.
(iv) Multivariate normal distribution with expected value $\boldsymbol{x}$ and variance-covariance matrix $\boldsymbol{V}(\boldsymbol{x})$.
(v) Rectangular distribution with end points a_- and a_+.
(vi) Scaled-and-shifted arcsine distribution with end points a_- and a_+.

(3) Generate M simulated samples $(q_1^{(r)}, \ldots, q_N^{(r)})$, for $r = 1, \ldots, M$, from the joint probability distribution for $Q_1, \ldots, Q_N$. The draft GUMS1 recommends $M = 10^6$.

(4) Calculate the M simulated value $y^{(r)} = f(q_1^{(r)}, \ldots, q_N^{(r)})$ for Y.

(5) Calculate the estimate y and standard uncertainty $u(y)$. Following the ISO-GUM, y is defined as the arithmetic mean and $u(y)$ as the standard deviation of the M simulated values $y^{(1)}, \ldots, y^{(M)}$ for Y. In addition to y, one may calculate other measures of centrality such as the median, y_{median}, and the mode, y_{mode} (when the distribution has a single mode).

(6) Calculate an uncertainty interval $[y_{\text{low}}, y_{\text{high}}]$ for Y by determining the limits y_{low} and y_{high} such that the interval $[y_{\text{low}}, y_{\text{high}}]$ encompasses the desired fraction p of the distribution for Y. Generally the desired p is 95%.

The draft GUMS1 refers to the uncertainty interval $[y_{\text{low}}, y_{\text{high}}]$ as coverage interval. When the measurement equation is non-linear, a simulated distribution for Y defined by $y^{(1)}, \ldots, y^{(M)}$ is asymmetric about y; therefore, the interval $[y_{\text{low}}, y_{\text{high}}]$ is also asymmetric about y. The draft GUMS1 discusses two forms of the interval $[y_{\text{low}}, y_{\text{high}}]$. The first form is the interval $[y_{(0.025)}, y_{(0.975)}]$, where $y_{(0.025)}$ and $y_{(0.975)}$ are the 0.025th quantile (2.5% percentile) and 0.975th quantile (97.5% percentile) of the simulated distribution for Y. The interval $[y_{(0.025)}, y_{(0.975)}]$ excludes equal probability 0.025 (2.5%) on each side and it is therefore a probabilistically symmetric interval. The second form is the shortest width interval $[y_{\text{low}}, y_{\text{high}}]$ having the coverage probability $p = 95\%$. The draft GUMS1 seems to prefer the shortest width interval $[y_{\text{low}}, y_{\text{high}}]$. However, in this paper we will discuss probabilistically symmetric intervals $[y_{(0.025)}, y_{(0.975)}]$ which require fewer calculations.

Comment 2. Unlike the ISO-GUM, the draft GUMS1 does not distinguish between the Type A and Type B state-of-knowledge probability distributions for the input variables. We believe this distinction is important because some probability distributions are determined from statistical analyses of the current measurements (Type A) and some by other means (Type B). We may identify the scaled-and-shifted t-distribution (third in the list of the six distributions of the draft GUMS1 [2, clause 4, table 1]) as a Type A Bayesian posterior distribution based on non-informative prior distributions and normally distributed measurements as indicated in the section 3.2 of this paper and [3, 6, 7]. The other five distributions are maximum entropy state-of-knowledge distributions, which may be identified as Type B distributions if their parameters are specified by non-statistical methods.

Comment 3. The BIPM/JCGM presents the draft GUMS1 as a generalization of the ISO-GUM. The ISO-GUM applies for all probability distributions for the input variables that have the same expected values, standard deviations and correlation coefficients. The GUMS1 applies to specific probability distributions for the input variables. Thus the GUMS1 is not a generalization of the ISO-GUM.

Comment 4. In the draft GUMS1, if a multivariate distribution other than multivariate normal is to be used, then means for numerical random sampling from it would need to be provided.

4.1. Application of the draft GUMS1 to simple linear calibration

Following the draft GUMS1 [2, clause 4, table 1], the input variable X is assigned a scaled-and-shifted t-distribution with degrees of freedom $n - 1$ that has been scaled by $s(x)$ and shifted by x. This assignment of scaled-and-shifted t-distribution to X is in agreement with section 3.2, where we saw that this distribution is a Type A Bayesian posterior distribution for X based on non-informative prior distributions for X and σ^2 and normally distributed measurements $x_1, \ldots, x_n$. The input variables B_0 and B_1 are assigned Type B normal and rectangular distributions. The parameters of the normal and rectangular distributions B_0 and B_1 are set to agree with those specified in section 2 for the three scenarios.

We generated one million (10^6) simulated random samples from the joint probability distribution for X, B_0 and B_1 for each of the data sets in tables 1 and 2 using the software S-PLUS [9]. We then calculated the corresponding values for Y using the measurement equation (2). Tables 6, 7 and 8 display the estimate y, standard uncertainty $u(y)$ and the limits of the uncertainty interval $[y_{(0.025)}, y_{(0.975)}]$ based on the draft GUMS1 for the data sets in tables 1 and 2.

The results displayed in tables 6, 7 and 8 are inexact because of the randomness inherent in numerical simulation. However for comparison with the results from the other two approaches, we will regard them as exact. The coverage probabilities associated the uncertainty intervals $[y_{(0.025)}, y_{(0.975)}]$ displayed in tables 6, 7 and 8 are 95% by definition.

In tables 7 and 8, the mean y is larger than the median y_{median}. This indicates that the distribution for Y is asymmetric with respect to y with a longer tail on the right side. Consequently, the uncertainty intervals $[y_{(0.025)}, y_{(0.975)}]$ displayed in tables 7 and 8 are not symmetric about y.

Table 6. The estimate y, standard uncertainty $u(y)$ and the limits of the probabilistically symmetric uncertainty interval $[y_{(0.025)}, y_{(0.975)}]$ for scenario 1 (normal distribution for B_0) based on the draft GUMS1.

	$u(x)$	$u(b_0)$	y	y_{median}	$u(y)$	$y_{(0.025)}$	$y_{(0.975)}$
1	0.950	0.250	100.522	100.522	0.984	98.602	102.441
2	0.950	4.000	100.529	100.524	4.116	92.479	108.604

5. Bayesian statistics

In Bayesian statistics [3] the quantities involved in a measurement process are regarded as statistical parameters. In particular, the value Q of the measurand is regarded as a statistical parameter along with the other statistical parameters $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_m)$. Let $\boldsymbol{\theta} = (Q, \boldsymbol{\tau})$ be a vector valued statistical parameter. The state of knowledge about $\boldsymbol{\theta}$ before current measurement data are available is represented by a pdf $p(\boldsymbol{\theta})$ referred to as a prior distribution. Lack of prior knowledge is represented by using non-informative prior distributions. It is sometimes convenient to use non-informative prior distributions that are not proper probability distributions. The link between the statistical parameter $\boldsymbol{\theta}$ and the current measurement data is represented by a likelihood function $l(\boldsymbol{\theta}|\mathrm{data})$ conditional on the data. The likelihood function $l(\boldsymbol{\theta}|\mathrm{data})$ is determined from the sampling probability distribution, $g(\mathrm{data}|\boldsymbol{\theta})$, for the measurement data given $\boldsymbol{\theta}$. The state of knowledge about $\boldsymbol{\theta}$ after measurement data are available is represented by a pdf $p(\boldsymbol{\theta}|\mathrm{data})$ referred to as a posterior distribution. The posterior distribution is obtained using the Bayes' theorem which states that the posterior pdf is proportional to the product of the prior pdf and the likelihood function. In symbols $p(\boldsymbol{\theta}|\mathrm{data}) \propto l(\boldsymbol{\theta}|\mathrm{data})\, p(\boldsymbol{\theta})$. Substituting $\boldsymbol{\theta} = (Q, \boldsymbol{\tau})$, we have

$$p(Q, \boldsymbol{\tau}|\mathrm{data}) \propto l(Q, \boldsymbol{\tau}|\mathrm{data})\, p(Q, \boldsymbol{\tau}). \tag{11}$$

Suppose the prior distributions for Q and $\boldsymbol{\tau}$ are independent, i.e. $p(Q, \boldsymbol{\tau}) = p(Q)p(\boldsymbol{\tau})$, where $p(Q)$ and $p(\boldsymbol{\tau})$ are prior distributions for Q and $\boldsymbol{\tau}$ before measurements are available. Then the posterior distribution $p(Q|\mathrm{data})$ for the value of the measurand after measurements is obtained by integrating out $\boldsymbol{\tau}$. Thus

$$p(Q|\mathrm{data}) = \int p(Q, \boldsymbol{\tau}|\mathrm{data})\mathrm{d}\boldsymbol{\tau}$$
$$\propto \int l(Q, \boldsymbol{\tau}|\mathrm{data})\, p(Q)\, p(\boldsymbol{\tau})\mathrm{d}\boldsymbol{\tau}. \tag{12}$$

The right side of (12) is normalized so it integrates to one, making $p(Q|\mathrm{data})$ a pdf. Often, the posterior distribution (12) is determined by numerical analysis such as a Markov Chain Monte Carlo using the software BUGS [10].

The posterior distribution $p(Q|\mathrm{data})$ represents all that is known about the value Q of the measurand based on the prior distributions $p(Q)$ and $p(\boldsymbol{\tau})$, and the likelihood function $l(Q, \boldsymbol{\tau}|\mathrm{data})$. In accordance with the ISO-GUM, the expected value and the standard deviation of the posterior distribution $p(Q|\mathrm{data})$ are the estimate q and its associated standard uncertainty $u(q)$ for Q. The interval $[q_{(0.025)}, q_{(0.975)}]$, where $q_{(0.025)}$ and $q_{(0.975)}$ are the 2.5% and 97.5% percentiles of the posterior distribution $p(Q|\mathrm{data})$, may be taken as a

Table 7. The estimate y, standard uncertainty $u(y)$ and the limits of the probabilistically symmetric uncertainty interval $[y_{(0.025)}, y_{(0.975)}]$ for scenario 2 (normal distributions for B_0 and B_1) based on the draft GUMS1.

	$u(x)$	$u(b_0)$	$u(b_1)$	y	y_{median}	$u(y)$	$y_{(0.025)}$	$y_{(0.975)}$
1	0.950	0.250	0.050	100.774	100.523	5.169	91.351	111.628
2	0.950	0.250	0.200	105.142	100.516	25.097	72.216	165.284
3	0.950	4.000	0.050	100.781	100.522	6.546	88.675	114.331
4	0.950	4.000	0.200	105.149	100.527	25.443	71.453	166.079

Table 8. The estimate y, standard uncertainty $u(y)$ and the limits of the probabilistically symmetric uncertainty interval $[y_{(0.025)}, y_{(0.975)}]$ for scenario 3 (rectangular distributions for B_0 and B_1) based on the draft GUMS1.

	$u(x)$	$u(b_0)$	$u(b_1)$	y	y_{median}	$u(y)$	$y_{(0.025)}$	$y_{(0.975)}$
1	0.950	0.250	0.050	100.784	100.531	5.152	92.592	109.813
2	0.950	0.250	0.200	104.874	100.536	22.101	75.602	149.867
3	0.950	4.000	0.050	100.786	100.541	6.528	89.022	113.732
4	0.950	4.000	0.200	104.876	100.535	22.512	73.897	151.509

probabilistically symmetric uncertainty interval with coverage probability 95%. In Bayesian numerical analyses, the median, q_{median}, of the posterior distribution is also usually calculated. The expected value q and the median q_{median} are identical when the pdf $p(Q|\text{data})$ is symmetric. When the pdf $p(Q|\text{data})$ is asymmetric, one may prefer uncertainty intervals that are not probabilistically symmetric.

5.1. Application of Bayesian statistics to simple linear calibration

There are two approaches for doing a Bayesian analysis for simple linear calibration.

(1) Bayesian analysis treating all unknown quantities as statistical parameters.

(2) Bayesian analysis applied to the Type A input quantities only.

We will illustrate both approaches.

5.1.1. Bayesian analysis treating all unknown quantities as statistical parameters. The measurements $x_1, \ldots, x_n$ are mutually independent and have the same normal sampling distribution with expected value X and variance σ^2. The expected value X is related to the value Y of the measurand by the relationship $X = B_0 + B_1 Y$. This approach regards B_0, B_1, σ^2 and Y as statistical parameters and X as a function of B_0, B_1 and Y. The likelihood function $l(B_0, B_1, \sigma^2, Y, |x_1, \ldots, x_n)$ of B_0, B_1, σ^2 and Y given the measurements $x_1, \ldots, x_n$ is taken as the joint sampling pdf of $x_1, \ldots, x_n$, where each x_i has the normal $N(B_0+B_1 Y, \sigma^2)$ distribution. We treat the Type B state-of-knowledge distributions for B_0 and B_1 specified in section 2 as the prior probability distributions $p(B_0)$ and $p(B_1)$. The posterior pdf $p(Y|x_1, \ldots, x_n)$ for Y based on equation (12) is

$$p(Y|x_1,\ldots,x_n) = \left[\iiint l(B_0, B_1, \sigma^2, Y|x_1,\ldots,x_n) \times p(B_0)p(B_1)p(\sigma^2)p(Y)\,\mathrm{d}B_0\,\mathrm{d}B_1\,\mathrm{d}\sigma^2\right] \times\left[\iiiint l(B_0, B_1, \sigma^2, Y|x_1,\ldots,x_n)p(B_0)p(B_1) \times p(\sigma^2)p(Y)\,\mathrm{d}B_0\,\mathrm{d}B_1\,\mathrm{d}\sigma^2\,\mathrm{d}Y\right]^{-1}, \quad (13)$$

Table 9. The estimate y, median y_{median}, standard uncertainty $u(y)$ and the limits of the probabilistically symmetric uncertainty interval $[y_{(0.025)}, y_{(0.975)}]$ for scenario 1 (normal distribution for B_0) based on Bayesian analysis treating all unknown quantities as statistical parameters.

	$u(x)$	$u(b_0)$	y	y_{median}	$u(y)$	$y_{(0.025)}$	$y_{(0.975)}$
1	0.950	0.250	100.511	100.510	0.974	98.588	102.409
2	0.950	4.000	100.343	100.332	4.172	92.172	108.544

where $p(\sigma^2)$ and $p(Y)$ are prior distributions for σ^2 and Y, respectively. The denominator in (13) is a normalizing constant that makes $p(Y|x_1, \ldots, x_n)$ a Bayesian posterior pdf. The posterior distribution (13) was numerically evaluated using the software BUGS [10]. This software requires that all prior probability distributions should be proper probability distributions. So the prior distribution for Y was set as normal with expected value 100 and variance 10 000; the large variance makes this distribution practically non-informative. It is more convenient to regard $\gamma = 1/\sigma^2$ as a statistical parameter. The prior distribution for γ was set as a gamma distribution [8] with both scale and shape parameters as 0.0001; the expected value of this distribution is one and variance 10 000, which makes it practically non-informative.

The expected value y, median y_{median}, standard deviation $u(y)$ and the percentiles $y_{(0.025)}$ and $y_{(0.975)}$ of the posterior pdf $p(Y|x_1, \ldots, x_n)$ for the three scenarios and data of tables 1 and 2 are given in tables 9, 10 and 11. These results are based on 100 000 numerical iterations using BUGS [10].

We will regard the results displayed in tables 9, 10 and 11 as exact even though they are determined by numerical simulation using BUGS [10]. The coverage probabilities associated with the uncertainty intervals $[y_{(0.025)}, y_{(0.975)}]$ are by definition 95%. In tables 10 and 11, the mean y is larger than the median y_{median}; therefore, the posterior distribution for Y is asymmetric with respect to y with a longer tail on the right side. The uncertainty intervals $[y_{(0.025)}, y_{(0.975)}]$ displayed in tables 10 and 11 are not symmetric about y.

5.1.2. Bayesian analysis applied to the Type A input quantities only. In the measurement equation (2), there is only one Type A input quantity and it is the expected value X of the measurement data. As noted in section 3.2,

Table 10. The estimate y, median y_{median}, standard uncertainty $u(y)$ and the limits of the probabilistically symmetric uncertainty interval $[y_{(0.025)}, y_{(0.975)}]$ for scenario 2 (normal distributions for B_0 and B_1) based on Bayesian analysis treating all unknown quantities as statistical parameters.

	$u(x)$	$u(b_0)$	$u(b_1)$	y	y_{median}	$u(y)$	$y_{(0.025)}$	$y_{(0.975)}$
1	0.950	0.250	0.050	100.760	100.525	5.150	91.319	111.559
2	0.950	0.250	0.200	101.396	98.338	20.099	70.963	152.971
3	0.950	4.000	0.050	100.366	100.126	6.490	88.238	113.687
4	0.950	4.000	0.200	101.176	98.070	20.404	70.255	153.282

Table 11. The estimate y, median y_{median}, standard uncertainty $u(y)$ and the limits of the probabilistically symmetric uncertainty interval $[y_{(0.025)}, y_{(0.975)}]$ for scenario 3 (rectangular distributions for B_0 and B_1) based on Bayesian analysis treating all unknown quantities as statistical parameters.

	$u(x)$	$u(b_0)$	$u(b_1)$	y	y_{median}	$u(y)$	$y_{(0.025)}$	$y_{(0.975)}$
1	0.950	0.250	0.050	100.813	100.594	5.048	92.766	109.655
2	0.950	0.250	0.200	104.012	99.481	21.547	75.838	152.970
3	0.950	4.000	0.050	100.233	99.975	6.416	88.747	148.811
4	0.950	4.000	0.200	104.108	99.724	21.827	73.881	149.759

a Bayesian posterior distribution for X is a scaled-and-shifted t-distribution with degrees of freedom $n-1$ that has been scaled by $s(x)$ and shifted by x. The Type B probability distributions for B_0 and B_1 are as specified in section 2 under the three scenarios. Thus a pdf for Y may be determined by propagating the probability distributions for X, B_0 and B_1 through the measurement equation (2) by numerical simulation as suggested in the draft GUMS1. The pdf for Y so determined can reasonably be interpreted as a state-of-knowledge pdf $p(Y|x_1, \ldots, x_n)$. The probability distributions for X, B_0 and B_1 are identical to those assigned in section 4.1 on application of the draft GUMS1 to simple linear calibration. Therefore, the results of Bayesian analysis applied to the Type A input quantity X in simple linear calibration are identical to those presented in tables 6, 7 and 8 for the draft GUMS1.

Comment 5. The measurement equation may have several Type A input variables or be a system of equations. The 'Bayesian analysis applied to the Type A input quantities only' illustrated in section 5.1.2 is applicable to all measurement equations. When Bayesian statistics is used for the Type A input quantities only, the ISO-GUM may be regarded as an extension of Bayesian statistics to incorporate non-statistical (Type B) evaluations.

6. Discussion of the results

Tables 3, 6 and 9 display the results from the three approaches for scenario 1 where the measurement equation is linear. The results in these tables are identical subject to the vicissitudes of numerical simulation. This affirms that the ISO-GUM yields the correct expected value and standard deviation for the value of the measurand when the measurement equation is linear.

Tables 4 and 5 display the results based on the ISO-GUM for scenarios 2 and 3, where the measurement equation is non-linear and the expected values and standard deviations for X, B_0 and B_1 are identical except for the forms of their distributions. The results in these tables are identical. This indicates that the results based on the ISO-GUM apply for all probability distributions for the input variables that have the same expected values, standard deviations and correlation coefficients. In this sense the results based on the ISO-GUM are *robust* with respect to the forms of probability distributions for the input variables. The coverage probabilities of the uncertainty intervals $[y \pm ku(y)]$ for a fixed coverage factor k depend on the form of probability distributions for Y as indicated in the last columns of tables 4 and 5.

The results in table 7 are exact for scenario 2 (normal distributions for B_0 and B_1), and table 4 displays the corresponding approximate results based on linear approximation of the measurement equation used in the ISO-GUM. The results in rows one and three of these tables are for the case where the uncertainty $u(b_1)$ for the non-linear input variable B_1 is small $[u(b_1) = 0.05]$. We note that the corresponding results in rows one and three of tables 4 and 7 are similar. The results in rows two and four are for the case where the uncertainty $u(b_1)$ for the non-linear input variable B_1 is large $[u(b_1) = 0.20]$. We note that the corresponding results in rows two and four of tables 4 and 7 are different. A similar pattern is observed in comparing exact and approximate results in tables 8 and 5 for scenario 3 (rectangular distributions for B_0 and B_1). This indicates that the linear approximation of the measurement equation used in the ISO-GUM may be adequate when the non-linear input variables have small uncertainties; but, the results from the ISO-GUM are poor approximations when the non-linear input variables have large uncertainties.

The uncertainty intervals $[y_{(0.025)}, y_{(0.975)}]$ in table 8 for rectangular distribution are narrower than the corresponding intervals in table 7 for normal distribution. Similarly, the uncertainty intervals $[y_{(0.025)}, y_{(0.975)}]$ in table 11 are narrower than the intervals in table 10. This is because rectangular distribution assigns all of its probability to a finite central region.

The standard uncertainties $u(y)$ in tables 7 and 8 are correct values for scenarios 2 and 3 subject to the vicissitudes of numerical simulation. The corresponding values in tables 4 and 5 determined from the ISO-GUM are underestimates. The underestimation is small when $u(b_1)$ is small $[u(b_1) = 0.05]$; however, when $u(b_1)$ is large $[u(b_1) = 0.20]$, the underestimation is large.

On comparing tables 10 and 7, we note that the standard uncertainties $u(y)$ determined from the Bayesian analysis (based on the likelihood function (13)) of section 5.1.1 are smaller. The differences are large when $u(b_1)$ is large [$u(b_1) = 0.20$]. Likewise, the standard uncertainties in table 11 are smaller than in table 8. The results from the two Bayesian approaches differ because they are based on different assumptions. In the analysis of section 5.1.1, the prior distribution for X depends on the prior distributions for B_0, B_1 and Y through the relationship $X = B_0 + B_1Y$. In the analysis of section 5.1.2, the prior distributions for X, B_0 and B_1 are mutually independent (which is what we had assumed in section 2) and the prior distribution for X is non-informative.

7. Merits and limitations of the three approaches

In this section we discuss the merits and limitations of each of the three approaches.

7.1. Merits of the ISO-GUM

The primary expression of uncertainty in the ISO-GUM is standard uncertainty. A standard uncertainty is both internally consistent and transferable [1, section 0.4]. In this sense, standard uncertainty is a fundamental expression of uncertainty in measurement.

When the measurement equation is linear, the estimate y and standard uncertainty $u(y)$ determined from the ISO-GUM are correct values for all state-of-knowledge probability distributions for the input variables $Q_1, \ldots, Q_N$ that have the specified expected values, q_i, standard deviations, $u(q_i)$, and correlation coefficients, $r(q_i, q_j)$. In this sense, y and $u(y)$ are the robust estimate and standard uncertainty for Y.

The estimate y and standard uncertainty $u(y)$ determined from the ISO-GUM may be reasonable when all non-linear input variables have small uncertainties.

The ISO-GUM requires simple calculations that are familiar to most metrologists.

7.2. Limitations of the ISO-GUM

When the measurement equation is non-linear and one or more of the input variables have large uncertainties, the standard uncertainty $u(y)$ determined from a linear approximation of the measurement equation based on the ISO-GUM may be a poor approximation for the standard deviation $S(Y)$ for Y.

An uncertainty interval is a secondary expression of uncertainty in the ISO-GUM. It is determined from the standard uncertainty after the latter has been evaluated [1, section 8, step 7]. Since the ISO-GUM propagates the estimates and standard uncertainties rather than probability distributions for the input variables, it does not yield an uncertainty interval with a specific coverage probability[4].

[4] In some cases when the assumptions that underlie the central limit theorem are satisfied, an approximate coverage probability may be determined from normal distribution.

7.3. Merits of the draft GUMS1

The primary expression of uncertainty in the draft GUMS1 is an uncertainty interval with a stated coverage probability. The draft GUMS1 propagates probability distributions assigned to the input variables. Therefore it yields uncertainty intervals for any desired coverage probability.

The draft GUMS1 may be used (when applicable) to determine uncertainty intervals with desired coverage probabilities for various choices of the input probability distributions. The draft GUMS1 may also be used to assess the coverage probabilities of the uncertainty intervals based on the ISO-GUM for various input probability distributions.

7.4. Limitations of the draft GUMS1

When the current measurement data and available information about the measurement process are limited, it may not be possible to reliably specify the probability distribution for each input variable. Some input variables may also be known to be correlated and it may be possible to specify their correlation coefficients but it may not be possible to specify their joint probability distribution. Thus, requirements for using the draft GUMS1 may not always be met.

The probability distribution and uncertainty interval determined from the draft GUMS1 are likely to be interpreted as exact and reliable, even though the joint probability distribution for the input variables is generally an approximation.

The draft GUMS1 applies when the input variables are independent or their joint distribution is multivariate normal. Sometimes not all input variables are independently distributed and there is no basis to claim that their joint distribution is multivariate normal.

The draft GUMS1 does not mention that different approaches are needed to specify the Type A and the Type B input probability distributions. The Type A distributions are determined from statistical analyses of the current data and the Type B distributions are determined by other means. Therefore, different approaches are required for specifying the two types of distributions.

The draft GUMS1 needs to be expanded to include wider ranges of univariate and multivariate distributions for both the Type A and the Type B input variables.

7.5. Merits of Bayesian statistics

The merits of Bayesian statistics in the context of the ISO-GUM are discussed in section 3.1.

7.6. Limitations of Bayesian statistics

Bayesian analysis often requires numerical simulation because closed form solutions are often not available.

8. Concluding remarks

1. The ISO-GUM's declaration of the Type A evaluations (determined from sampling theory) as expressions of the

state-of-knowledge is not needed when Bayesian statistics is used for the Type A evaluations. The sampling theory estimates may be regarded as approximations to the corresponding Bayesian estimates.

2. The ISO-GUM yields a robust estimate and standard uncertainty for the value of the measurand when the measurement equation is linear.
3. The draft GUM Supplement 1 may be used to assess the uncertainty intervals based on the ISO-GUM and to determine uncertainty intervals with required coverage probabilities for various choices of the input probability distributions.
4. We hope that a revised draft of the GUM Supplement 1 would classify the input probability distributions into Type A and Type B because different approaches are required for specifying the two types of input distributions.
5. The use of Bayesian statistics for the Type A evaluations greatly simplifies the expression of uncertainty by eliminating the need for counting degrees of freedom.
6. For the Type A standard uncertainty associated with the arithmetic mean x of m independent normally distributed measurements $x_1, \ldots, x_m$, the ISO-GUM recommends the expression $u(x)$ where $u(x) = s/\sqrt{m}$, and s is the experimental standard deviation $s = \sqrt{[\sum_i (x_i - x)^2/(m-1)]}$. The expression $u(x)$ is incomplete without stating its degrees of freedom $m-1$, which represents the uncertainty in $u(x)$ corresponding to the number m of measurements. For the same situation we recommend the Bayesian uncertainty $u_{\text{Bayes}}(x)$, where $u_{\text{Bayes}}(x) = \sqrt{(m-1)/(m-3)} \times s/\sqrt{m}$. The Bayesian standard uncertainty is a complete expression. A Bayesian uncertainty is never uncertain. In particular, $u_{\text{Bayes}}(x)$ carries no degrees of freedom. We note that $u(x) = 1/\sqrt{m-1} \times \sqrt{\sum_i (x_i - x)^2/m}$ and $u_{\text{Bayes}}(x) = 1/\sqrt{m-3} \times \sqrt{\sum_i (x_i - x)^2/m}$. Thus the only difference is the divisor $\sqrt{m-1}$ used in the ISO-GUM and the divisor $\sqrt{m-3}$ used in the Bayesian uncertainty. The Bayesian uncertainty requires at least four independent measurements. As noted in the ISO-GUM [1, table E.1], the standard uncertainty $u(x) = s/\sqrt{m}$ is unreliable when $m = 2$ or 3. A judiciously determined Type B standard uncertainty may be more reliable than the Type A uncertainty $u(x) = s/\sqrt{m}$ based on only two or three measurements [1, section 4.3.2].
7. When Bayesian statistics is used for the Type A input variables, the ISO-GUM may be regarded as an extension of Bayesian statistics to incorporate non-statistical (Type B) evaluations.

Acknowledgments

This paper was developed after a panel discussion organized by Ding Huang at the 2004 meeting of the National Conference of Standards Laboratories International (www.ncsli.org) held in the Salt Lake City, Utah, USA. Albert Jones, Will Guthrie and Fern Hunt provided useful comments on an earlier draft of this paper. We thank the referee whose comments have improved the paper.

Disclaimer

Certain software are identified in this paper. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that these software are necessarily the best available for the purpose.

Appendix

A linear approximation of the measurement equation (2) about y is

$$Y \approx Y_{\text{linear}} = y + c_x(X - x) + c_{b0}(B_0 - b_0) + c_{b1}(B_1 - b_1), \quad (14)$$

where $c_x = 1/b_1$, $c_{b0} = (-1/b_1)$ and $c_{b1} = (-1/b_1)y$ are the first order partial derivatives of Y with respect to X, B_0 and B_1, respectively, evaluated at $X = x$, $B_0 = b_0$ and $B_1 = b_1$. The expected value and variance of Y_{linear} are

$$E(Y_{\text{linear}}) = y = (x - b_0)/b_1, \quad (15)$$

and

$$V(Y_{\text{linear}}) = u^2(y) = c_x^2 u^2(x) + c_{b0}^2 u^2(b_0) + c_{b1}^2 u^2(b_1). \quad (16)$$

Equation (16) simplifies to $u^2(y) = (1/b_1)^2[u^2(x) + u^2(b_0)] + (1/b_1)^2 y^2 u^2(b_1)$. By dividing both sides by y^2, we get $[u(y)/y]^2 = (1/b_1 y)^2[u^2(x) + u^2(b_0)] + [u(b_1)/b_1]^2 = [u^2(x) + u^2(b_0)]/(x - b_0)^2 + [u(b_1)/b_1]^2$. The ratio $u_r(y) = u(y)/|y|$ is the relative standard uncertainty associated with y and the ratio $u(b_1)/|b_1|$ is the relative standard uncertainty associated with b_1. Using the symbol $u(x - b_0)$ for the standard uncertainty $S(X - B_0) = [u^2(x) + u^2(b_0)]^{1/2}$ and the symbol $u_r(x - b_0)$ for the relative standard uncertainty $S(X - B_0)/|E(X - B_0)| = u(x - b_0)/|x - b_0|$, equation (16) simplifies to the propagation of relative standard uncertainties formula

$$u_r^2(y) = u_r^2(x - b_0) + u_r^2(b_1). \quad (17)$$

References

[1] 1995 *Guide to the Expression of Uncertainty in Measurement* 2nd edn (Geneva: International Organization for Standardization) ISBN 92-67-10188-9
[2] 2004 Draft GUM Supplement 1: numerical methods for the propagation of distributions, BIPM Joint Committee on Guides in Metrology
[3] Lee P M 1997 *Bayesian Statistics* 2nd edn (New York: Oxford University Press)
[4] 1999 *Expression of the Uncertainty of Measurement in Calibration* Publication Reference EA-4/02 European Co-operation for Accreditation
[5] 1995 *Quantifying Uncertainty in Analytical Measurement* EURACHEM/CITAC Guide 2nd edn
[6] Kacker R N and Jones A T 2003 On use of Bayesian statistics to make the Guide to the Expression of Uncertainty in Measurement consistent *Metrologia* **40** 235–48
[7] Kacker R N 2006 Bayesian alternative to the ISO-GUM's use of the Welch–Satterthwaite formula *Metrologia* **43** 1–11
[8] Evans M, Hastings N and Peacock B 2000 *Statistical Distributions* 3rd edn (New York: Wiley)
[9] S-PLUS 2000 Data Analysis Products Division Mathsoft Inc, Seattle WA, USA http://www.insightful.com/
[10] BUGS 2004 MRC Biostatistics Unit, Cambridge, UK http://www.mrc-bsu.cam.ac.uk/bugs/