March 16, 2011

Only non-informative Bayesian prior distributions agree with the GUM Type A evaluations of input quantities

Raghu Kacker[1], Ruediger Kessel[1], Kensei Ehara[2], and Klaus-Dieter Sommer[3]
[1]National Institute of Standards and Technology, Gaithersburg, Maryland 20899, USA
[2]National Metrology Institute of Japan, Tsukuba, Ibaraki 305-8563, Japan
[3]Physikalisch-Technische Bundesanstalt, D-38116 Braunschweig, Germany

Contact Email: raghu.kacker@nist.gov

Abstract
The Guide to the Expression of Uncertainty in Measurement (GUM) is self-consistent when Bayesian statistics is used for the Type A evaluations and the standard deviation of posterior distribution is used as the Bayesian Type A standard uncertainty. We present the case that there are limitations on the kind of Bayesian statistics that can be used for the Type A evaluations of input quantities of the measurement function. The GUM recommends that the (central) measured value should be an unbiased estimate of the corresponding (true) quantity value. Also, the GUM uses the expected value of state-of-knowledge probability distributions as the (central) measured value for both the Type A and the Type B evaluations of input quantities. It turns out that the expected value of a Bayesian posterior distribution used as a Type A (central) measured value for an input quantity can be unbiased only when a non-informative prior distribution is used for that input quantity. Metrologically, this means that only the current observations without any additional information should be used to determine a Type A (central) measured value for an input quantity of the measurement function.

1. Introduction

The Guide to the Expression of Uncertainty in Measurement (GUM) uses conventional (sampling theory) statistics for the Type A (statistical) evaluations of input quantities of the measurement function. The estimates from conventional statistics have sampling probability interpretation. The Type B (non-statistical) evaluations have state-of-knowledge probability interpretation. Thus the Type A and the Type B evaluations have different probabilistic interpretations and they cannot be combined logically. So the GUM declares the Type A evaluations of input quantities to be the parameters of state-of-knowledge probability distributions [1, section 4.1.6]. The discussion of section 2 reminds us that this declaration is insufficient to make the GUM self-consistent. It is well known that the GUM is self-consistent when Bayesian statistics is used for the Type A evaluations and the standard deviation of posterior distribution is used as the Bayesian Type A standard uncertainty. We present the case that there are limitations on the kind of Bayesian statistics that can be used for the Type A evaluations of input quantities of the measurement function. Use of Bayesian statistics for the Type A evaluation of an input quantity requires specification of a prior distribution for that input quantity. The prior distribution can be informative or non-informative. For example, the scaled and shifted *t*-distribution discussed in the Supplement 1 to the GUM (GUM-S1) [2, section 6.4] is a Bayesian posterior distribution determined from non-informative prior distributions for the parameters of normally distributed observations. As discussed in section 3, the GUM recommends that the (central) measured value should be an unbiased estimate of the corresponding (true) quantity value. Also the GUM uses the expected value of state-of-knowledge probability distributions as the (central) measured value for both the Type A and the Type B evaluations. As discussed in section 4, the expected value of a Bayesian posterior distribution used as a Type A (central) measured value for an input quantity can be unbiased only when a non-informative prior distribution is used for that quantity. Concluding remarks appear in section 5.

2. Use of Bayesian statistics the Type A evaluations of input quantities makes the GUM self-consistent

Suppose the measurement function is

$$\mathrm{Y} = \mathrm{X}_1 + \mathrm{X}_2 \tag{1}$$

where X_1 is a statistical parameter representing the expected value of a series of observations (subject to random effects) depending on the value Y for a measurand and X_2 is the required correction for all recognized significant systematic effects in the (central) measured value for X_1. We use upright symbols (such as X_1, X_2, and Y) for the (true) values of physical quantities, statistical parameters, and required corrections for systematic effects. As in the GUM, we assume that the measurand is defined with sufficient completeness; thus, for all practical purposes the (true) value Y is unique. The constants X_1, X_2, and Y have unknown values. In metrology, almost no quantity has completely unknown value because some information is often available. Therefore by 'unknown value' we really mean 'insufficiently known value'. Operationally, the measurement function (1) is expressed as

$$Y = X_1 + X_2 \tag{2}$$

where X_1, X_2, and Y represent random variables with state-of-knowledge probability distributions about the unknown values X_1, X_2, and Y respectively. We use italic symbols (such as X_1, X_2, and Y) for variables with state-of-knowledge probability distributions about the quantity values X_1, X_2, and Y, respectively. The GUM uses the same symbols for the quantities and for the variable with state-of-knowledge distributions about the quantities. Since we discuss the sampling theory concept of unbiased estimates, it is useful to use different symbols for quantities and variables.

Suppose X_1 is evaluated by statistical analysis of a series of n independent observations $q_1, q_2, \ldots, q_n$ obtained under the same conditions. Therefore, the corresponding measured value x_1 and the standard uncertainty $u(x_1)$ are Type A evaluations [1, section 4.2]. Suppose X_2 is evaluated by non-statistical methods and the corresponding measured value x_2 and the standard uncertainty $u(x_2)$ are Type B evaluations [1, section 4.3]. A measured value y and the standard uncertainty $u(y)$ for Y are determined from the results (x_1, $u(x_1)$) and (x_2, $u(x_2)$) using the measurement function (2) [1, section 5]. Thus

$$y = x_1 + x_2 \tag{3}$$

and

$$u^2(y) = u^2(x_1) + u^2(x_2) \tag{4}$$

When the probability distributions represented by (x_1, $u(x_1)$) and (x_2, $u(x_2)$) for X_1 and X_2, respectively, are correlated, the correlation coefficient between them must be included in the expression (4) for $u^2(y)$ [1, section 5.2]. In the GUM, a result of measurement such as (x_1, $u(x_1)$) indicates a range of values which could reasonably be attributed to the corresponding input quantity X_1; the range of values is expressed as ($x_1 \pm k \cdot u(x_1)$), where k is a chosen coverage factor [1, section 6]. To distinguish the particular measured value x_1 from the other values in the interval ($x_1 \pm k \cdot u(x_1)$) which could also be attributed to X_1, we refer to it as the *central* measured value.

Suppose the measurement procedure is in a state of statistical control; thus, the observations $q_1, q_2, \ldots, q_n$ may be regarded as independent realizations (random draws) from the same Gaussian (normal) sampling probability distribution $N(X_1, \sigma^2)$ with expected value X_1 and some fixed variance σ^2 (or standard deviation σ). A sampling probability density function (pdf) describes the relative frequencies of occurrence for all realizations possible in contemplated replications of the observation process conditionally on the values of one or more statistical parameters. Sampling pdfs used in metrology have a finite expected value and a finite variance. The expected value X_1 of the sampling pdf of the observations is unknown. Indeed the purpose of measurement is to assign a value to X_1 and hence to Y. The variance σ^2 may be known or unknown. For example when the observations are indications from an established measurement system which has been used for a long time, substantial historical data might be available to know the value of σ^2 for all practical purposes. In that case the known value of σ^2 rather than its statistical estimate discussed below should be used. A Type A statistical estimate of σ^2 determined from a few observations is not very reliable [1, section E.4]

Suppose

$$q_A = \frac{1}{n}\Sigma_i q_i \tag{5}$$

$$s^2 = \frac{1}{n-1}\Sigma_i (q_i - q_A)^2 \tag{6}$$

$$s(q_A) = \frac{s}{\sqrt{n}} \tag{7}$$

are the arithmetic mean, the experimental variance, and the experimental standard deviation of the mean of the n observations. A function of the observations is called a *statistic*. Thus q_A, s^2, and $s(q_A)$ are statistics. The sampling distribution of the observations induces sampling distributions for the statistics. If $q_1, q_2, \ldots, q_n$ have joint normal sampling distribution, $N(q_1, q_2, \ldots, q_n \mid X_1, \sigma^2)$, then the sampling distributions of q_A and s^2 are independent, q_A has normal distribution $N(X_1, \sigma^2/n)$, and s^2 has the multiple $\sigma^2/(n-1)$ of the chi-square (χ^2) distribution with degrees of freedom $(n-1)$ denoted by $[\sigma^2/(n-1)]\cdot\chi^2_{(n-1)}$ [3, section 5.3]. The GUM [1, section 4.2] indentifies the Type A measured value x_1 and the standard uncertainty $u(x_1)$ as $x_1 \equiv q_A$ and $u(x_1) \equiv s(q_A)$, respectively. The statistics q_A and $s^2(q_A)$ are point (single value) estimates of the expected values of the sampling pdfs of X_1 and σ^2/n in conventional statistics. Thus the GUM uses conventional statistics for the Type A evaluations.

The GUM [1, section 4.1.6] declares the Type A evaluations $x_1 \equiv q_A$ and $u(x_1) \equiv s(q_A)$ to be the parameters of a state-of-knowledge probability distribution represented by the result $(q_A, s(q_A))$ for the variable X_1. This declaration imparts a common probabilistic interpretation to the Type A and the Type B evaluations of the input quantities of the measurement function, so they can be combined logically. The GUM [1, section G.3] identifies the state-of-knowledge distribution assigned by the result $(q_A, s(q_A))$ to the variable X_1 as the scaled-and-shifted Student's (central) t-distribution with degrees of freedom $(n-1)$, denoted here by $t_{(n-1)}(q_A, s^2(q_A))$. The expected value and the variance of $t_{(n-1)}(q_A, s^2(q_A))$ are q_A and $[(n-1)/(n-3)]\cdot s^2(q_A)$, respectively [2, section 6.4.9.4]. Thus the expected value $E(X_1)$ is equal to q_A and the standard deviation $S(X_1)$ is equal to $\sqrt{[(n-1)/(n-3)]}\cdot s(q_A)$. We note that the GUM Type A standard uncertainty $u(x_1) \equiv s(q_A)$ is not the standard deviation $S(X_1)$ of the t-distribution $t_{(n-1)}(q_A, s^2(q_A))$ represented by the result $(q_A, s(q_A))$.

Suppose that the Type B evaluations x_2 and $u(x_2)$ are quantified by assigning to the variable X_2, a state-of-knowledge pdf which is a rectangular on the interval $(-a, a)$ for some $a > 0$ with expected value $E(X_2) = 0$ and standard deviation $S(X_2) = a/\sqrt{3}$. The GUM [1, section 4.3.7] indentifies x_2 as $x_2 = E(X_2) = 0$ and $u(x_2)$ as $u(x_2) = S(X_2) = a/\sqrt{3}$. We note that the GUM is not self-consistent: a Type B standard uncertainty $u(x_2)$ is the standard deviation $S(X_2) = a/\sqrt{3}$ of the assigned rectangular distribution but a Type A standard uncertainty $u(x_1) \equiv s(q_A)$ is not the standard deviation $S(X_1) = \sqrt{[(n-1)/(n-3)]}\cdot s(q_A)$ of the assigned t distribution. Thus the GUM declaration [1, section 4.1.6] of the Type A evaluations from conventional statistics as parameters of state-of-knowledge probability distributions is insufficient to make the GUM self-consistent.

The GUM is self-consistent when Bayesian statistics is used for the Type A evaluation of an input quantity, the state of knowledge about that input quantity is expressed as a Bayesian posterior distribution, and the standard deviation of the posterior distribution is used as the Bayesian Type A standard uncertainty [4], [5], [6], and [7]. In the following paragraphs we review the use of Bayesian statistics for the GUM Type A evaluation of X_1 from normally distributed observations $q_1, q_2, \ldots, q_n$. In Bayesian statistics, the state of knowledge about unknown statistical parameters is expressed in terms of prior and posterior pdfs, see for example [8] and [9]. A prior pdf represents *a priori* state of knowledge about a parameter before seeing the realized observations. A posterior pdf represents *a posteriori* state of knowledge about a parameter in view of the realized observations. The sampling pdf of the observations is interpreted as being proportional to something called a *likelihood function* for the unknown parameters conditional on the observations. A likelihood function represents the information about the unknown parameters contained in the realized observations. The prior pdfs and the likelihood function are combined using the Bayes's theorem to determine posterior pdfs for the parameters. The Bayes's theorem states that posterior pdf is proportional to the product of the likelihood function and the prior pdf [8] and [9]. Thus the joint normal sampling pdf $\mathrm{N}(q_1, q_2, \ldots, q_n \mid X_1, \sigma^2)$ of the observations $q_1, q_2, \ldots, q_n$ conditional on the parameters X_1 and σ^2 is interpreted as being proportional to the likelihood function $l(X_1, \sigma^2 \mid q_1, q_2, \ldots, q_n)$ for the parameters X_1 and σ^2 conditional on the realized observations. Suppose the joint prior pdf for the parameters X_1 and σ^2 is $p(X_1, \sigma^2)$. Then according to the Bayes's theorem, the joint posterior pdf $p(X_1, \sigma^2 \mid q_1, q_2, \ldots, q_n)$ for X_1 and σ^2 conditioned on the realized observations is proportional to $l(X_1, \sigma^2 \mid q_1, q_2, \ldots, q_n) \times p(X_1, \sigma^2)$. A (marginal) posterior pdf for X_1 alone is obtained from the joint pdf $p(X_1, \sigma^2 \mid q_1, q_2, \ldots, q_n)$ by integrating out σ^2.

Two board categories of prior distributions are: informative and non-informative. An informative prior distribution represents available knowledge about the value of an unknown statistical parameter as a probability distribution excluding the information supplied by the current observations. A non-informative prior distribution is used when only the information supplied by the current observations and no prior information (whether available or not) is to be used in determining the posterior distribution. A non-informative prior distribution is a function of the unknown parameter which does not integrate to one. Thus a non-informative prior distribution is not a proper pdf. Bayesian statistical evaluations based on suitably chosen non-informative improper prior distributions are identical to or close to the estimates from conventional statistics albeit with state-of-knowledge interpretation [8] and [9]. This is to be expected because the likelihood function is same in conventional statistics and in Bayesian statistics.

The expected value and the standard deviation of a Bayesian posterior pdf are commonly referred to as *posterior expected value* and *posterior standard deviation* respectively. We will discuss two situations: σ^2 is known and σ^2 is unknown. If σ^2 is known and the prior pdf for X_1 is taken as the non-informative function $p(X_1) \propto 1$ then the Bayesian posterior pdf $p(X_1 \mid q_1, q_2, \ldots, q_n)$ of X_1 is the normal distribution $\mathrm{N}(q_\mathrm{A}, \sigma^2/n)$ [9, section 2.9]. Thus the posterior expected value and the posterior standard deviation of X_1 are, respectively, $\mathrm{E}(X_1 \mid q_1, q_2, \ldots, q_n) = q_\mathrm{A}$ and $\mathrm{S}(X_1 \mid q_1, q_2, \ldots, q_n) = \sigma/\sqrt{n}$. If σ^2 is unknown and the joint prior pdf for X_1 and σ^2 is taken as the non-informative function $p(X_1, \sigma^2) \propto 1/\sigma^2$ then the Bayesian posterior pdf $p(X_1 \mid q_1, q_2, \ldots, q_n)$ of X_1 is the scaled and shifted *t*-distribution $t_{(n-1)}(q_\mathrm{A}, s^2(q_\mathrm{A}))$ [9, section 3.2]. Thus the posterior expected value and the posterior standard deviation of X_1 are, respectively, $\mathrm{E}(X_1 \mid q_1, q_2, \ldots, q_n) = q_\mathrm{A}$ and $\mathrm{S}(X_1 \mid q_1, q_2, \ldots, q_n) = \sqrt{[(n-1)/(n-3)]}\cdot s(q_\mathrm{A})$. Note that in both cases the posterior expected value is $\mathrm{E}(X_1 \mid q_1, q_2, \ldots, q_n) = q_\mathrm{A}$.

The most basic expression of uncertainty in the GUM is the standard uncertainty. When Bayesian statistics is used for the Type A evaluation of an input quantity, the posterior distribution represents the state of knowledge about the input quantity and the posterior standard deviation is used as the standard uncertainty. A Type B evaluation of standard uncertainty is the standard deviation of a state-of-knowledge pdf assigned to an input quantity. Thus when Bayesian statistics is used for the Type A evaluations, both the Type A and the Type B evaluations of standard uncertainty are the standard deviations of state-of-knowledge distributions. The use of standard deviation as the basic expression of uncertainty corresponds to the use of the expected value as the central measured value. In addition we note that the GUM recommends propagating the uncertainties through a linear Taylor series approximation of the measurement function about the expected values of state-of-knowledge pdfs for the input quantities [1, section E.3.1]. Thus in the GUM a central measured value for an input quantity is always the expected value of a state-of-knowledge pdf for that quantity.

Note 2.1: Use of Bayesian statistics for the Type A evaluations not only makes the GUM self-consistent but also simplifies the expression of uncertainty in measurement. A Type A standard uncertainty $s(q_A)$ determined from conventional statistics is an incomplete expression without an accompanying statement of its associated degrees of freedom $n - 1$ which indicates the statistical uncertainty in $s^2(q_A)$ from limited number n of observations. The Bayesian standard uncertainty $u(x_1) = \sqrt{[(n-1)/(n-3)]}\cdot s(q_A)$ is a complete expression of uncertainty because it has no statistical uncertainty. The built in factor $\sqrt{[(n-1)/(n-3)]}$ accounts for the limited number n of observations. Thus the use of Bayesian statistics greatly simplifies the expression of uncertainty in measurement by eliminating altogether the need of calculating degrees of freedom [6].

Note 2.2: The sampling distributions of the mean q_A and the experimental variance s^2 are statistically independent. This independence and that q_A and s^2 are sufficient statistics [8, section 2.9] makes possible the Bayesian statistics results discussed in this paper. It turns out that only for normally distributed observations $q_1, q_2, \ldots, q_n$, the mean q_A and the experimental variance s^2 are independently distributed [10, chapter 4]. So the sampling distribution of the observations must be normal for the Bayesian statistics results discussed in this paper to apply. This means that the measurement procedure must be in a state of statistical control.

3. The GUM recommends that a Type A central measured value for an input quantity should be unbiased

In conventional statistics a desirable property of a point estimate is that it is unbiased. *Unbiased estimate*: A statistic x is said to be an unbiased estimate of a parameter of true value X if the expected value of the sampling distribution of x is X, i.e. $\mathrm{E}(x) = \mathrm{X}$ (equivalently, the expected value of x with respect to the joint sampling distribution of the observations is X). The mean q_A and the experimental variance of the mean $s^2(q_A)$ are, respectively, unbiased estimates of X_1 and σ^2/n [3, chapter 7].

Whether σ^2 is known or unknown, the Bayesian posterior expected value of X_1 determined from suitably chosen non-informative improper prior distributions is $E(X_1 \mid q_1, q_2, \ldots, q_n) = q_A$. The expected value of the posterior expected value $E(X_1 \mid q_1, q_2, \ldots, q_n)$ with respect to the joint normal sampling distribution of the observations $q_1, q_2, \ldots, q_n$ is

$$\mathrm{E}(\mathrm{E}(X_1 \mid q_1, q_2, ..., q_n)) = \mathrm{E}(q_\mathrm{A}) = \mathrm{X}_1 \tag{8}$$

Thus the posterior expected value $E(X_1 \mid q_1, q_2, \ldots, q_n) = q_A$ determined from suitably chosen non-informative prior distributions (for X_1 and σ^2) is an unbiased estimate of X_1.

The GUM [1, section 3.2.4] states the following:

> It is assumed that the result of a measurement has been corrected for all recognized significant systematic effects and that every effort has been made to identify such effects.

We believe this statement implies that the GUM subsumes and recommends that a Type A central measured value (referred to as the result of a measurement in the GUM) should be an unbiased estimate of the corresponding (true) quantity value.

The GUM [1, section 3.2.3] states the following:

> It is assumed that, after correction, the expectation or expected value of the error arising from a systematic effect is zero.

We think this statement means that the measured value y determined from a linear expression (such as (3)) of one Type A evaluated measured value (such as x_1) and one or more corrections for systematic effects is regarded as an unbiased evaluation of the corresponding measurand Y.

Note 3.1: Consider the measurement function $\mathrm{Y} = \mathrm{W}_1/\mathrm{W}_2$ where the measured values w_1 and w_2 for the input quantities W_1 and W_2 are Type A evaluations. The values w_1 and w_2 should be unbiased. However the measured value $y = w_1/w_2$ for Y will not be unbiased. In general, a measured value y determined from a non-linear measurement function of one or more input quantities evaluated by statistical methods may not be unbiased. This is a limitation of the GUM.

Note 3.2: The concept of unbiasedness is meaningful only for the Type A evaluations. The Type B evaluations are determined using non-statistical methods and statistical criteria such as unbiasedness are not relevant.

4. A Bayesian posterior expected value based on a proper prior pdf for an input quantity cannot be unbiased

We will discuss posterior expected value of X_1 determined from independent normally distributed observations $q_1, q_2, \ldots, q_n$ and a proper prior pdf for X_1. First we consider the case of known σ^2 and then we consider the case of unknown σ^2 with a proper prior pdf. Suppose σ^2 is known and the prior pdf for X_1 is taken as normal N(μ_0, τ_0^2), where μ_0 and τ_0^2 are specified (known) constants, then the posterior state-of-knowledge pdf $p(X_1 \mid q_1, q_2, \ldots, q_n)$ of X_1 is normal N(μ_1, σ_1^2), where

$$\mu_1 = \left(\frac{1}{\tau_0^{\ 2}} + \frac{n}{\sigma^2} \right)^{-1} \left(\frac{1}{\tau_0^{\ 2}} \mu_0 + \frac{n}{\sigma^2} q_{\mathrm{A}} \right) \tag{9}$$

and

$$\sigma_1^{\ 2} = \left(\frac{1}{\tau_0^{\ 2}} + \frac{n}{\sigma^2} \right)^{-1} \tag{10}$$

[9, section 2.6]. The posterior expected value μ_1 of X_1 based on the prior pdf N(μ_0, τ_0^2) is a weighted mean of μ_0 and q_{A}. The expected value of μ_1 with respect to the sampling distribution of q_{A} is

$$\mathrm{E}(\mu_1) = \left(\frac{1}{\tau_0^{\ 2}} + \frac{n}{\sigma^2} \right)^{-1} \left(\frac{1}{\tau_0^{\ 2}} \mu_0 + \frac{n}{\sigma^2} \mathrm{X}_1 \right) \tag{11}$$

which differs from X_1. Thus the posterior expected value μ_1 is not an unbiased estimate of X_1. We note that as τ_0^2 tends to ∞ (in which case the prior pdf N(μ_0, τ_0^2) of X_1 becomes less and less informative), μ_1 tends to q_{A} the estimate discussed in section 2 based on a non-informative improper prior distribution for X_1. A result similar to (10) is shown in [11, section 3].

Now suppose that σ^2 is unknown. A simple closed form of the posterior pdf for X_1 is obtained when the prior pdfs for X_1 and σ^2 are dependent, conditionally on σ^2 the distribution of X_1 is normal N(μ_0, σ^2/κ_0) and σ^2 has the multiple σ_0^2 of inverse chi-square distribution with ν_0 degrees of freedom. This means that $\sigma_0^2/\ \sigma^2$ has chi-square distribution and that the prior distribution for X_1 is worth κ_0 independent normal observations of variance σ^2. Then the posterior state-of-knowledge pdf $p(X_1 \mid q_1, q_2, \ldots, q_n)$ of X_1 is the scaled and shifted t-distribution

$$t_{\nu_1}(\mu_1, \frac{\sigma_1^{\ 2}}{\kappa_1}) \tag{12}$$

where

$$\nu_1 = \nu_0 + n \tag{13}$$

$$\kappa_1 = \kappa_0 + n \tag{14}$$

$$\mu_1 = \frac{\kappa_0}{\kappa_0 + n}\mu_0 + \frac{n}{\kappa_0 + n} q_{\mathrm{A}} \tag{15}$$

$$\nu_1 \sigma_1^{\ 2} = \nu_0 \sigma_0^{\ 2} + (n-1)s^2 + \frac{\kappa_0 n}{\kappa_0 + n}(q_{\mathrm{A}} - \mu_0)^2 \tag{16}$$

[9, section 3.3]. The posterior expected value μ_1 of X_1 is a weighted mean of μ_0 and q_{A}, where q_{A} is an unbiased estimate of X_1. Thus μ_1 is not an unbiased estimate of X_1. We note that as κ_0 tends to 0 (in which case the prior pdf $\mathrm{N}(\mu_0, \sigma^2/\kappa_0)$ of X_1 becomes less and less informative), μ_1 tends to q_{A} the estimate discussed in section 2 based on non-informative improper prior distributions for X_1 and σ^2.

A general result about Bayesian posterior expected values is as follows: *A Bayesian posterior expected value determined using a proper prior pdf cannot be an unbiased estimate of the corresponding parameter* [3, section 7.5.2]. In Bayesian statistics, unbiasedness of Bayesian point estimates is not important. However in metrology, a central measure value identified as the posterior expected value is recommended to be unbiased [1, section 3.2.4].

Note 4.1: In the first example discussed in this section the prior and the posterior pdfs for X_1 had the same form namely normal distributions. In the second example the joint prior and posterior distributions for X_1 and σ^2 had the same form [9, section 3.3]. Such prior distributions are referred to as conjugate prior distributions [9, section 2.6]. When conjugate prior pdfs are used, the posterior expected value often turns out to be a weighted average of the mean of observations (such as q_{A}) and the expected value of the prior pdf (such as μ_0) [3, section 7.2.3]. Such weighted mean cannot be an unbiased estimate of the expected value of the mean of observations with respect to the joint sampling distribution of the observations.

5. Concluding remarks

It is widely recognized that the use of Bayesian statistics for the Type A evaluations makes the GUM self-consistent and greatly simplifies the expression of uncertainty in measurement by eliminating altogether the need of calculating degrees of freedom. We note that there are limitations on the kind of Bayesian statistics that can be used for the Type A evaluations of input quantities. A central measured value for an input quantity in the GUM is always the expected value of a state-of-knowledge probability distribution for that quantity. Further the GUM subsumes and recommends that a Type A central measured value should be an unbiased estimate of the corresponding (true) quantity value. A Bayesian posterior expected value determined using a proper prior distribution for a quantity cannot be an unbiased estimate. Therefore the GUM recommendation that a Type A measured value for an input quantity should be unbiased is satisfied only when a non-informative improper prior distribution is used for that input quantity. Metrologically, this means that only the current observations without any additional information should be used to determine a Type A central measured value for an input quantity of the measurement function.

Note 5.1: In metrology, a prior pdf for the value of a physical quantity is often based on an alternate measurement (or a previous measurement) determined using a different method perhaps by another metrologist. In such cases, the object of current measurement may be to determine an independent measured value for that quantity; therefore, prior pdf cannot be used. The current measured value may be compared with an alternate measured value. When the uncertainties are reasonable and the two measured values are significantly different, one would suspect that the value of the measurand may have changed. In that case the prior distribution and the current observations may correspond to different measurands.

Note 5.2: Non-informative improper prior distributions are needed only for the Type A evaluations. A Type B evaluation is frequently determined from a state-of-knowledge pdf assigned to the input quantity. A Type B state-of-knowledge pdf may be interpreted as a Bayesian prior distribution. Such prior distributions are proper pdfs.

Note 5.3: The principles underlying non-informative prior distributions are discussed in text books such as the following: [9, section 2.9] and [12, section 5.6.2]. Useful non-informative improper prior distributions for many commonly occurring situations in metrology have been determined by analytical methods. The computational software for numerical Bayesian analysis is not very useful for finding suitable non-informative improper prior distributions.

Acknowledgment
- Charles Hagwood, Division reader

References

[1] BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, OIML 1995 *Guide to the Expression of Uncertainty in Measurement* 2nd ed (Geneva, Switzerland: International Organization for Standardization) ISBN 92-67-10188-9 (2008 version available at http://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf)
[2] BIPM/JCGM 2008 *Evaluation of measurement data – Supplement 1 to the "Guide to the expression of uncertainty in measurement" – Propagation of distributions using a Monte Carlo method* (Sèvres: Bureau International des Poids et Mesures, Joint Committee for Guides in Metrology) (available at http://www.bipm.org/utils/common/documents/jcgm/JCGM_101_2008_E.pdf)
[3] Casella G and Berger R L 2002 *Statistical Inference* 2nd ed (Pacific Grove, CA, USA: Duxbury/Thomson Learning)
[4] Weise K and Wöger W 1992 A Bayesian theory of measurement uncertainty *Meas. Sci. Technol.* **3** pp 1-11
[5] Lira I and Wöger W 2001 Bayesian evaluation of the standard uncertainty and coverage probability in a simple measurement model *Meas. Sci. Technol.* **12** pp 1172-1179

[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** pp 235-248
[7] Kacker R N 2006 Bayesian alternative to the ISO-GUM's use of the Welch-Satterthwaite formula *Metrologia* **43** pp 1-11
[8] Lee P M 1997 *Bayesian Statistics: An Introduction* 2nd ed (New York, USA: Oxford University Press)
[9] Gelman A Carlin J B Stern H S and Rubin, D B 2004 *Bayesian Data Analysis* 2nd ed (Boca Raton, FL, USA: Chapman & Hall/CRC)
[10] Mathai A M and Pederzoli G 1977 *Characterizations of the normal probability law* (New York, USA: Wiley)
[11] Elster C 2007 Calculation of uncertainty in the presence of prior knowledge *Metrologia* **44** pp 111-116
[12] Bernardo J M and Smith A F M 1994 *Bayesian Theory* (New York, USA: Wiley)