Measurement 88 (2016) 194–201

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

Removing divergence of JCGM documents from the GUM (1993) and repairing other defects

Raghu N. Kacker [a,*], Rüdiger Kessel [b], James F. Lawrence [a,c]

[a] National Institute of Standards and Technology, Gaithersburg, MD 20899, USA
[b] Physikalisch-Technische Bundesanstalt, Braunschweig D-38116, Germany
[c] George Mason University, Fairfax, VA 22030, USA

ARTICLE INFO

Article history:
Received 11 December 2015
Received in revised form 21 March 2016
Accepted 25 March 2016
Available online 29 March 2016

Keywords:
Bayesian statistics
Coverage interval
GUM-S1
Uncertainty in measurement
VIM3

ABSTRACT

The mission of the Joint Committee for Guides in Metrology (JCGM) is to maintain and promote the use of the Guide to the Expression of Uncertainty in Measurement (GUM) and the International Vocabulary of Metrology (VIM, second edition). The JCGM has produced the third edition of the VIM (referred to as VIM3) and a number of documents; some of which are referred to as supplements to the GUM. We are concerned with the Supplement 1 (GUM-S1) and the document JCGM 104. The signal contribution of the GUM is its operational view of the uncertainty in measurement (as a parameter that characterizes the dispersion of the values that could be attributed to an unknown quantity). The operational view promulgated by the GUM had disconnected the uncertainty in measurement from the unknowable quantities true value and error. The GUM-S1 has diverged from the operational view of the uncertainty in measurement. Either the disparities should be removed or the GUM-S1 should not be referred to as a supplement to the GUM. Also, the GUM-S1 has misinterpreted the Bayesian concept of a statistical parameter and the VIM3 definitions of coverage interval and coverage probability are mathematically defective. We offer practical suggestions for revising the GUM-S1 and the VIM3 to remove their divergence from the GUM and to repair their defects.

Published by Elsevier Ltd.

1. Introduction

The Joint Committee for Guides in Metrology (JCGM) was formed in 1997 [1] to maintain and promote the use of the Guide to the Expression of Uncertainty in Measurement (GUM, 1993) [2] and the International Vocabulary of Metrology (VIM, 1993, second edition, now referred to as VIM2) [3]. The JCGM has two working groups. The JCGM working group 1 (JCGM WG1) has produced a number of documents; some of which are referred to as supplements to the GUM. We are concerned with the supplement 1, the GUM-S1 published in 2008 and the document JCGM 104 published in 2009. The GUM-S1 is entitled Supplement 1 to the GUM – Propagation of distributions using a Monte Carlo (MC) method [4] and the JCGM 104 is entitled Introduction to the GUM and related documents [5]. The concept and definitions in the JCGM 104 apply to the GUM-S1. The JCGM working group 2 (JCGM WG2) published in 2008 the third edition of the VIM, identified as VIM3 or JCGM 200 [6].

The GUM is not completely consistent with either conventional or Bayesian statistical concepts [7]. However, the GUM can be made fully consistent with Bayesian concepts by using for the Type A (statistical) evaluations Bayesian statistics (with non-informative prior distributions) [8]. Then the GUM concept (from conventional statistics) of

* Corresponding author.
E-mail addresses: raghu.kacker@nist.gov (R.N. Kacker), ruediger.kessel@ptb.de (R. Kessel), lawrence@gmu.edu (J.F. Lawrence).

http://dx.doi.org/10.1016/j.measurement.2016.03.057
0263-2241/Published by Elsevier Ltd.

quantifying uncertainty of the uncertainty in measurement by 'degrees of freedom' would vanish. An introductory reference for conventional statistical inference is [9]. Bayesian inference based on the use of Bayes' rule (theorem) is described in many textbooks and articles such as the following [10–21]. In this paper, phrases displayed in the *italic* font are direct quotes from a cited reference; sometimes additional words are inserted in parentheses to clarify the intended meaning.

To a statistician, the term Bayesian implies the use of Bayes' rule. However, metrologists use the term Bayesian in a broad sense meaning only that the state of knowledge about an unknown quantity is expressed by a subjective (personal degree of belief) probability distribution. The GUM-S1 does not use the Bayes' rule, yet it considers the input and output probability distributions of its Monte Carlo method to be Bayesian. For example, the GUM-S1 makes the following statements: *(i) A coverage interval is sometimes known as a credible interval or Bayesian interval, (ii) the use of probability distributions in Type B evaluation is a feature of Bayesian inference, (iii) in the Bayesian context of this Supplement, concepts such as the reliability, or the uncertainty, of an uncertainty are not necessary* [4, Sections 3.12, 5.1.2, 6.4.9.4 Note 2]. We use the term Bayesian in the broad sense of metrologists, unless stated otherwise.

The GUM-S1 is aligned with Bayesian concepts. The GUM-S1 identifies the true value of an unknown quantity (measurand) as a statistical parameter and regards that parameter as a random variable [4, Section 5.1.1, c], [5, Section 3.17]. A probability distribution for this random variable describes the probabilities of the unknown true value lying in different intervals [5, Section 3.17]. The GUM-S1 introduced an expression of uncertainty in measurement called a coverage interval. A coverage interval is defined to be an *interval containing the value of a quantity with a stated probability, based on the information available* [4, Section 3.12], and its associated coverage probability is the *probability that the value of a quantity is contained within a specified coverage interval.* In these GUM-S1 definitions, 'the value of a quantity' refers to 'the true value' even though the adjective 'true' is suppressed to create an appearance of fealty to the GUM [4, Section 3.12 Note 4]. In the GUM-S1, the true value of a quantity is assumed to be essentially unique [4, Section 1], [6, Section 2.11 Note 3], [22, Section 2.5]. A real physical quantity involved in measurement has a set of multiple true values rather than a single true value, because of the inherently incomplete amount of detail to which the quantity can be specified [2, Annex D including Figures D.1 and D.2], [6, Section 2.11 Note 1], [22, Section 2.4]. Fundamental constants of nature are exceptions. So the VIM3 gives the following general definitions: a coverage interval is an *interval containing the set (range) of true quantity values of a measurand with a stated probability, based on the information available* [6, Section 2.36], and the coverage probability is the *probability that the set (range) of true quantity values of a measurand is contained within a specified coverage interval* [6, Section 2.37]. Uncertainty arising from the finite amount of detail in the definition of a quantity is called definitional uncertainty [6, Section 2.27]. A quantity is considered to have an essentially unique true value when the definitional uncertainty is believed to be negligible [6, Section 2.11 Note 3].

The GUM-S1 has misinterpreted the Bayesian concept of a statistical parameter and the VIM3 definitions of coverage interval and coverage probability are mathematically defective. These defects should be repaired. The GUM-S1 concepts of coverage interval and coverage probability do not agree with the operational view of the uncertainty in measurement promulgated by the GUM. Since the GUM-S1 is a supplement to the GUM it should be revised to agree with the GUM.

In Section 2, we describe the Bayesian concepts of a statistical parameter, a probability distribution (expressing the state of knowledge about the value of that parameter), and an interval estimate for that value. In Section 3, we review various concepts and terms introduced by the GUM and identify the operational view of the uncertainty in measurement as the signal contribution of the GUM. In Section 4, we show that the GUM-S1 misinterprets the Bayesian concept of a statistical parameter, and in Section 5, we show that the VIM3 definitions of coverage interval and coverage probability are mathematically defective. In Section 6, we demonstrate that the GUM-S1 concepts of coverage interval and coverage probability do not agree with the GUM. In Section 7 we offer suggestions for revising the GUM-S1 to remove its divergence from the GUM. Summary and concluding remarks appear in Section 8.

2. Bayesian concepts of parameter, probability distribution, and interval estimate

In this section we are concerned with statistical inference for the unknown value of a statistical parameter based on the Bayes' rule. It is often said that in conventional statistical inference a statistical parameter has a fixed value but in Bayesian statistical inference the statistical parameter is treated as a random variable with a probability distribution which describes the possible variation of that parameter [9, Section 7.2.3]. This is a widespread misinterpretation of the Bayesian concept of a statistical parameter. In Bayesian inference also, the value of a parameter is fixed. That fixed value is the target of statistical inference. What changes is a probability distribution (over the possible values for that parameter) expressing the state of knowledge about that fixed value.

Statistician Dennis Lindley was a leading expert and advocate of Bayesian inference. We quote Lindley from the Ref. [15, p. 301]: "*The parameter is also uncertain. Indeed, it is that uncertainty that is the statistician's main concern. The recipe says that it also should be described by a probability .... In so doing we depart from the conventional attitude. It is often said that the parameters are assumed to be random quantities. This is not so. It is the axioms that are assumed, from which the randomness property is deduced.*"

Physicist E.T. Jaynes was also a leading expert and advocate of Bayesian inference. We quote Jaynes from the Ref. [12, p. 11]: "*For decades Bayesians have been accused of "supposing that an unknown parameter is a random variable"; and we have denied hundreds of times,* *with increasing vehemence, that we are making any such assumption. We have been unable to comprehend why our denials have no effect, and that charge continues to be made. Sometimes, in our perplexity, it has seemed to us that there are two basically different kinds of mentality in statistics; those who see the point of Bayesian inference at once, and need no explanation; and those who never see it, however much explanation is given. But a Seminar talk by Professor George Barnard, given in Cambridge in February 1984, provided a clue to what has been causing this Tower of Babel situation. Instead of merely repeating the old accusation (that we could only deny still another time), he expressed the orthodox puzzlement over Bayesian methods in a different way, more clearly and specifically than we had ever heard it put before. Barnard complained that Bayesian methods of parameter estimation, which present our conclusions in the form of a posterior distribution, are illogical; for "How could the distribution of a parameter possibly become known from data which were taken with only one value of the parameter actually present?" This extremely revealing comment finally gave some insight into what has been causing our communication problems. Bayesians have always known that orthodox terminology is not well adapted to expressing Bayesian ideas; but at least this writer had not realized how bad the situation was. Orthodoxians trying to understand Bayesian methods have been caught in a semantic trap by their habitual use of the phrase "distribution of the parameter" when one should have said "distribution of the probability". Bayesians had supposed this to be merely a figure of speech; i.e. that those who used it did so only out of force of habit, and really knew better. But now it seems that our critics have been taking that phraseology quite literally all the time. Therefore, let us belabor still another time what we had previously thought too obvious to mention. In Bayesian parameter estimation, both the prior and posterior distributions represent not any measurable property of the parameter, but only our own state of knowledge about it. The width of the distribution is not intended to indicate the range of variability of the true values of the parameter, as Barnard's terminology led him to suppose. It indicates the range of values that are consistent with our prior information and data, and which honesty therefore compels us to admit as possible values. What is "distributed" is not the parameter, but the probability."*

The set of all possible (permissible) values for a statistical parameter is called parameter space. In Bayesian statistical inference, a 'new' random variable with a probability distribution over the parameter space is introduced. Its probability distribution expresses the state of knowledge (uncertainty) about the fixed value of the parameter. Bayesian statisticians refer to this new random variable as non-observable to distinguish it from random observations (with assumed sampling probability distributions) which are observable. A non-observable random variable has no counterpart in conventional statistical inference; therefore, it is a new random variable.

Bayesian authors tend to use the same lowercase Greek letters such as θ, η, μ for three different things (i) the fixed value of a parameter which is unknown and the target of statistical inference, (ii) possible values for the parameter which form the parameter space, and (iii) a random variable with a probability distribution over the parameter space [10–21]. Even though the probability distribution of the random variable is defined over the parameter space, it is not the parameter. The parameter has a fixed value. To help clarify the concepts, in this paper we will use an uppercase Greek letter Θ for a (non-observable) random variable with a probability distribution expressing the state of knowledge about the fixed value of parameter, lowercase letter θ for its possible values in the parameter space, and the symbol $\tau[\Theta]$ for the fixed value of parameter [22, Section 3]. The fixed value $\tau[\Theta]$ is one of the possible values θ (a particular one) of the random variable Θ.

Bayesian statistical inference starts with a prior probability distribution $\pi(\Theta)$ for Θ defined over the parameter space which represents the state of knowledge (uncertainty) about $\tau[\Theta]$ before the current information [21,22]. The prior distribution is updated using the Bayes' rule in view of all available information $\boldsymbol{I}$ and all assumptions $\boldsymbol{A}$ made to obtain a posterior probability distribution $\pi(\Theta|\boldsymbol{I},\boldsymbol{A})$ for Θ defined over the parameter space which represents the state of knowledge about $\tau[\Theta]$ after the current information. The information $\boldsymbol{I}$ includes current observations and all available information (data) about the other parameters. The assumptions $\boldsymbol{A}$ include the mathematical or computational model of measurement, likelihood function (assumed sampling distributions for the observed data), and probability distributions for the unknown parameters. From the posterior distribution one can obtain an interval estimate for $\tau[\Theta]$ of desired probability. For a specified interval (θ_l,θ_h) of the parameter space, the fraction of posterior probability density function which corresponds to that interval is called the posterior probability $\Pr\{\theta_l < \Theta < \theta_h|\boldsymbol{I},\boldsymbol{A}\}$ of the interval (θ_l,θ_h). A Bayesian interval estimate (θ_l,θ_h) for the fixed value $\tau[\Theta]$ is the highest probability density (HPD) interval under the posterior distribution $\pi(\Theta|\boldsymbol{I},\boldsymbol{A})$ having a specified probability. The posterior probability $\Pr\{\theta_l < \Theta < \theta_h|\boldsymbol{I},\boldsymbol{A}\}$ is the probability under $\pi(\Theta|\boldsymbol{I},\boldsymbol{A})$ of the values in the interval (θ_l,θ_h) regarded as a subset of the parameter space [14, Section 5.1.5], [21,22]. A Bayesian interval estimate (θ_l,θ_h) is often referred to as a credible interval, especially to distinguish it from the corresponding conventional (frequentist) confidence interval. We will describe two practical interpretations of the Bayesian posterior distribution $\pi(\Theta|\boldsymbol{I},\boldsymbol{A})$ and the probability $\Pr\{\theta_l < \Theta < \theta_h|\boldsymbol{I},\boldsymbol{A}\}$ of an interval (θ_l,θ_h).

A theoretical interpretation of a Bayesian posterior probability distribution $\pi(\Theta|\boldsymbol{I},\boldsymbol{A})$ for Θ is that it describes the probability of each possible value θ of Θ being the fixed but unknown value $\tau[\Theta]$ (evading the issue of zero probability for a single value) [13, Section 1.2]. This is the logical basis for the interpretation of the posterior probability $\Pr\{\theta_l < \Theta < \theta_h|\boldsymbol{I},\boldsymbol{A}\}$ as the conditional probability based on the information $\boldsymbol{I}$ and the assumptions $\boldsymbol{A}$ that 'the fixed value $\tau[\Theta]$ lies within the fixed interval (θ_l,θ_h)' [16,21].

An operational interpretation of a Bayesian posterior probability distribution $\pi(\Theta|\boldsymbol{I},\boldsymbol{A})$ for Θ is that it describes for each possible value θ of Θ the probability with which the value θ could be attributed (assigned) to the unknown value $\tau[\Theta]$ (again evading the issue of zero probability for a single value). According to the operational interpretation, the posterior probability $\Pr\{\theta_l < \Theta < \theta_h|\boldsymbol{I},\boldsymbol{A}\}$ is the conditional probability based on the information $\boldsymbol{I}$ and the assumptions $\boldsymbol{A}$ of 'the values in the fixed interval (θ_l, θ_h) that could be attributed to the fixed value $\tau[\Theta]$' [22, Note 3.1].

Whether or not a Bayesian interval estimate (θ_l, θ_h) captures the fixed but unknown value $\tau[\Theta]$ depends on the qualities of the underlying information $\boldsymbol{I}$ and the assumptions $\boldsymbol{A}$. The information $\boldsymbol{I}$ and the assumptions $\boldsymbol{A}$ on which a posterior distribution $\pi(\Theta\,|\,\boldsymbol{I},\boldsymbol{A})$ is conditioned are subject to deficiencies in the model of measurement, the assumed likelihood function, and the assumed probability distributions for the unknown quantities [22, Notes 3.3, 3.4]. Uncertainty arising from possible deficiencies in the underlying information $\boldsymbol{I}$ and the assumptions $\boldsymbol{A}$ is not included in a Bayesian posterior probability distribution or in an interval estimate obtained from that distribution. When the authenticity of the information $\boldsymbol{I}$ or the validity of the assumptions $\boldsymbol{A}$ is questionable, an interval estimate (θ_l, θ_h) may not include $\tau[\Theta]$. When the true value $\tau[\Theta]$ is unknowable, the truth of the claim that an interval estimate (θ_l, θ_h) includes $\tau[\Theta]$ cannot be known. The operational interpretation of a Bayesian probability distribution is mute about the success of an interval estimate (θ_l, θ_h) in capturing the unknown value $\tau[\Theta]$.

Note 2.1: A key feature of Bayesian inference is that it requires a prior probability distribution for the fixed but unknown value of a parameter. The prior distribution is updated using the current information to obtain a posterior distribution. High echelon measurement laboratories receive periodically the same physical artifacts for re-calibration. Thus they have historical records of the results of measurement for the same artifacts. But metrologists do not use the previous results of measurement as prior information to obtain current results of measurement: "the purpose of current measurement is to make sure that the value of the artifact did not change; therefore we do not use historical results as prior information". In metrology, how beneficial is a prior distribution that is not based on factual information (data)? Also, how can a metrologist be sure that the measurand did not change during the time period between obtaining the data on which a Bayesian prior probability distribution might be based and obtaining the current measurements (excluding non-informative improper prior distributions)?

3. Operational view of uncertainty in measurement is signal contribution of GUM

The most visible concepts and terms introduced by the GUM are as follows: (i) Type A and Type B evaluations of uncertainty, (ii) measurement equation, (iii) linear propagation of uncertainties, (iv) and standard measurement uncertainty. Type A and Type B are new labels for an old classification of methods for evaluating uncertainty. This classification came into being because certain components of uncertainty could not be evaluated by statistical methods so technical judgment was needed. Type A (statistical methods) and Type B (other methods) are labels that were used in the recommendations of the CIPM (International Committee for Weights and Measures) on which the GUM is based [2, Section 0.7]. However, it is noteworthy that the GUM bestows on a Type B evaluation the same respect that is traditionally given to a Type A evaluation. Indeed the GUM states the following: *One may therefore conclude that Type A (statistical) evaluations of standard uncertainty are not necessarily more reliable than Type B (other) evaluations, and that in many practical measurement situations where the number of observations is limited, the components obtained from Type B evaluations may be better known than the components obtained from Type A evaluations* [2, Section E.4.3]. A Type B evaluation is obtained from an assigned probability distribution expressing the state of knowledge based on the pool of available information (data) and scientific judgement. A Type B probability distribution is intended to be minimally subjective and maximally based on data [2, Section 4.3]. The GUM declares a Type A evaluation (based on formulas from conventional statistics) as a parameter of a probability distribution expressing the state of knowledge [2, Section 4.1.6]. Therefore a probability distribution represented by a Type A evaluation, and a Type B probability distribution have the same interpretation: both express the state of knowledge. The declaration of a Type A evaluation as expressing the state of knowledge is required in the GUM to legitimatize combining Type A and Type B evaluations. This declaration of the GUM can be made authentic by using Bayesian statistics for the Type A evaluations, and interpreting a Type B distribution as a Bayesian probability distribution [8].

The concept of a measurement equation (referred to as a mathematical model of the measurement in the GUM and a measurement function in the VIM3) is a new contribution of the GUM. A measurement equation $Y = f(X_1, \ldots, X_N)$ describes a method for determining a measured value and its associated uncertainty for the measurand Y from the measured values and uncertainties for various input quantities $X_1, \ldots, X_N$ [23]. There is a one-to-one correspondence between the sources of uncertainty considered and the input quantities of the measurement equation. A measurement equation is based on technical knowledge of the measurement procedure and practical experience. An input quantity may have its own measurement equation.

The GUM-S1 gives an impression that linear propagation of uncertainties (LPU) method for combining components of uncertainty is the main contribution of the GUM. This is a narrow view of the GUM. The LPU method, also called as the root-sum-of-squares (RSS) method or the delta method, was widely used before the GUM was published. In fact this method was introduced by Gauss for independent components in the year 1823 [21, Section 5.5]. The GUM formally recognized a previously used method for combining standard deviations.

The GUM promulgated an operational perspective of the uncertainty in measurement [2, Section E.5.1]. According to the GUM (i) a measured value (referred to as a result of measurement in the GUM) is a value attributed to an unknown quantity [2, Section B.2.11], and (ii) measurement uncertainty is a parameter (associated with a measured value) that characterizes the dispersion of the values that could reasonably be attributed to an unknown quantity, conditional on the presently available knowledge [2, Section B.2.18]. Uncertainty in measurement expressed as a standard deviation is called standard measurement uncertainty [2, Section 2.3.1]. It follows that measurement is the process of assigning (attributing) values to an unknown quantity.

The GUM does not explicitly define the coverage probability of a result of measurement expressed as an interval. However, it is clear from various citations in the GUM that the coverage probability of an interval is the probability of the values in that interval which could be attributed to the unknown quantity [2, Sections 0.4, 2.3.5 Note 1, 3.3.7, 6.2.2]. The GUM does not explicitly state its interpretation of a probability distribution represented by a measured value and its associated uncertainty. In view of the definitions of a measured value and its associated uncertainty in measurement (and the GUM interpretation of coverage probability), a probability distribution in the GUM can have no meaning other than that it describes the probabilities of the values that could be attributed to an unknown quantity. The GUM interpretation of a probability distribution corresponds to the operational interpretation of a Bayesian probability distribution (discussed in Section 2); therefore, we will refer to it as an operational interpretation.

The operational view of the uncertainty in measurement is the signal contribution of the GUM because it overturned the pre-GUM paradigm of metrology which was concerned with true value and error to the modern concept of evaluated uncertainty from the recognized sources that are judged to be significant [23]. Uncertainty in measurement is an evaluated quantity; it does not include uncertainty from unrecognized sources and components of uncertainty that are judged to be negligible (this judgment can sometimes be wrong). The operational view, promulgated by the GUM, disconnected the uncertainty in measurement from the unknowable quantities true value and error [1, Section E.5.1].

Note 3.1: In metrology, a true quantity value is unknowable. The link of a result of measurement (a measured value with its associated uncertainty) with the unknown target of measurement is assured by the calibration of the measurement system using artifacts of assigned reference values. The reference values serve as surrogates for the unknowable true values of artifacts. High accuracy measurement techniques used to assign reference values are assessed by inter-comparison.

4. GUM-S1 misinterprets Bayesian concept of statistical parameter

The JCGM 104 states the following. *The true values of the input quantities $X_1, \ldots, X_N$ are unknown. In the approach advocated (in the GUM and the GUM-S1) $X_1, \ldots, X_N$ are characterized by (univariate) probability distributions and treated mathematically as random variables. These distributions describe the respective probabilities of their true values lying in different intervals, and are assigned based on available knowledge concerning $X_1, \ldots, X_N$* [5, Section 3.17]. If the symbols $X_1, \ldots, X_N$ refer to the input quantities then the JCGM 104 declares that an input quantity is treated (mathematically) as a random variable. If the symbols $X_1, \ldots, X_N$ refer to the true values of the input quantities then the JCGM 104 declares that a true quantity value is treated as a random variable. Thus the JCGM 104 treats either a 'quantity' or a 'true quantity value' as a random variable. In either case a probability distribution for that random variable describes the probabilities of different intervals containing the true value. This is the GUM-S1 definition of a probability distribution. It corresponds to the theoretical interpretation of a Bayesian probability distribution (discussed in Section 2).

The GUM-S1, recommends a Monte Carlo Method (MCM) to determine a numerical probability density function (PDF) for the output random variable Y corresponding to a mathematically specified joint probability distribution for the input random variables $X_1, \ldots, X_N$ of the measurement equation $Y = f(X_1, \ldots, X_N)$ [4, Section 7]. According to the GUM-S1, the numerical PDF for Y determined by MCM is used to obtain *(1) the expectation of Y, taken as an estimate y of the quantity, (2) the standard deviation of Y, taken as the standard uncertainty u(y) associated with y, and (3) a coverage interval containing Y with a specified probability (the coverage probability)* [4, Section 5.1.1, c]. In the parts (1) and (2) of this statement, Y is a random variable with a PDF having an expected value and a standard deviation. In the part (3), Y is the true value of the measurand (in view of the GUM-S1 definitions of coverage interval and coverage probability). Therefore the GUM-S1 uses the symbol Y for the true value of the measurand and treats it as a random variable with a PDF (expressing the probabilities of different intervals containing the true value). The GUM-S1 is aligned with Bayesian concepts [4, Sections 3.12, 5.1.2, 6.4.9.4 Note 2]. From that viewpoint, the GUM-S1 identifies the true value of a quantity as a statistical parameter and treats that parameter as a random variable. As discussed in Section 2, this is a misinterpretation of the Bayesian concept of a statistical parameter: the value of a parameter is always fixed. What changes is a probability distribution which expresses the state of knowledge about that fixed value.

The GUM uses an uppercase Roman letter such as X_i or Y for two different things: (i) a physical quantity (specified by its description) and (ii) the corresponding random variable (with a probability distribution describing the probabilities of the values that could be attributed to that quantity) [2, Section 4.1.1 Note 1]. This use of one symbol for these two different things does not cause communication difficulties in the GUM. The GUM-S1 uses an uppercase Roman letter such as X_i or Y for three different things: (i) a quantity, (ii) the true value of that quantity, and (iii) the corresponding random variable expressing the state of knowledge about that quantity. Using one symbol for 'a quantity' as well as 'the true value' of that quantity confuses two different concepts [6, Sections 1.1, 1.19, 2.11].

The misinterpretation of the Bayesian concept of a statistical parameter in the GUM-S1 can be repaired by introducing a new symbol for the true value of a quantity and making it clear that the true value is fixed, rather than a random variable [22, Note 4.2].

5. VIM3 definitions of coverage interval and coverage probability are mathematically defective

The VIM3 definitions of a coverage interval and its associated coverage probability are stated in Section 1. Suppose a physical quantity X has an interval (range) of true values $\tau[X]$. Suppose the interval $\tau[X]$ is believed to be so small that the state of knowledge about $\tau[X]$ may be adequately represented by a measured value (with evaluated uncertainty) or by a univariate probability distribution attributed to the random variable X. Suppose (x_l, x_h) is a result of measurement for the quantity X expressed as an interval. A single point on the number line has zero width; an interval can have any positive width, large or small. Now imagine that the interval (x_l, x_h) is so tiny that its width is even less than that of the interval $\tau[X]$. Then the interval of true values $\tau[X]$ cannot be contained within the interval (x_l, x_h). Consequently, the coverage probability for that tiny interval would be zero regardless of the probability distribution used to represent the quantity X. Thus the VIM3 definitions of a coverage interval and coverage probability require that intervals of width less than the range of true values $\tau[X]$ have probability zero.

A number of probability distributions to express the state of knowledge about a quantity are described in the GUM-S1 [4, Section 6.4]. These distributions (which include normal and rectangular distributions) have continuous probability densities. The coverage probability of an interval is the fraction of the probability density function (PDF) corresponding to that interval. Every interval in the support of a continuous probability distribution has a positive probability no matter how tiny that interval may be. In particular, every probability distribution described in the GUM-S1 would assign a positive (non-zero) probability to a tiny interval whose width is less than that of the interval $\tau[X]$. Thus none of the probability distributions described in the GUM-S1 satisfies the requirement of the VIM3 definitions of coverage interval and coverage probability that intervals of width less than the range of true values $\tau[X]$ have probability zero. Indeed no probability distribution for the variable X can be specified which satisfies this requirement of the VIM3 definitions. Thus the VIM3 definitions of coverage interval and coverage probability break down for intervals of width smaller than the range of true values. Therefore these definitions are mathematically defective.

The VIM3 concept of a coverage interval is well-defined only for those quantities which have single (unique) true values. Therefore the VIM3 definitions of coverage interval and coverage probability can be repaired by replacing the phrase 'the set of true quantity values' with the phrase 'the single true quantity value'.

6. GUM-S1 concepts of coverage interval and coverage probability do not agree with GUM

Suppose a quantity X has as an essentially unique true value $\tau[X]$ and (x_l, x_h) is a result of measurement for X expressed as an interval. The GUM-S1 refers to the interval (x_l, x_h) as a coverage interval and looks at it in terms of whether it contains or it does not contain the unknown true value $\tau[X]$. A claim that $\tau[X]$ is contained within the interval (x_l, x_h) is either true or false. The true value $\tau[X]$ is unknowable even in principle [6, Section 2.11]. Therefore the truth of the claim that $\tau[X]$ is contained within (x_l, x_h) can never be known. The GUM-S1 coverage probability of an interval (x_l, x_h) is a degree of belief assigned to the truth of the claim that the fixed true value $\tau[X]$ is contained within the fixed interval (x_l, x_h). The first edition of the International Vocabulary of Metrology (dated 1984, now referred to as the VIM1) defined the uncertainty in measurement as *an estimate characterizing the range of values within which the true value of a measurand lies* [24]. The GUM-S1 concept of a coverage interval (with its associated coverage probability) is essentially a pre-GUM concept of uncertainty from the VIM1. Thus the GUM-S1 has restored an essentially pre-GUM concept of uncertainty in measurement which was concerned with capturing (covering) the true value of the measurand in a computed interval.

The concept of a coverage interval introduced by the GUM-S1 does not exist in the GUM or the VIM2. The GUM defined the uncertainty in measurement as a parameter that characterizes the dispersion of the values that could reasonably be attributed to the measurand. The uncertainty in measurement does not refer to the true value of the measurand. In the GUM, the coverage probability of an interval (x_l, x_h) is the probability of the values in that interval which could be attributed to the measurand X. The GUM view of the coverage probability of an interval (x_l, x_h) is silent about the relationship between that interval and the true value $\tau[X]$ of the measurand X. The operational view of the uncertainty in measurement and the concept of coverage probability promulgated by the GUM had disconnected the uncertainty in measurement from the true value of the measurand. Thus the GUM-S1 concepts of coverage interval and coverage probability do not agree with the GUM.

7. Suggestions for revising GUM-S1 to remove its divergence from GUM

The divergence of the GUM-S1 from the GUM concept of uncertainty in measurement arises primarily from adopting the interpretation of a probability distribution that it describes the probabilities of the unknown true value lying in different intervals (see, Sections 1 and 4). The divergence of the GUM-S1 from the GUM can be removed by adopting the operational interpretation of a probability distribution that it describes the probabilities of the values that could be attributed to an unknown quantity (see, Sections 3 and 6).

The adoption of operational interpretation of a probability distribution will render the GUM-S1 concept of a coverage interval superfluous. We suggest that the GUM-S1 concept of a coverage interval should be replaced with an interval result; that is, a result of measurement expressed as an interval. An interval result represents a range of values that could reasonably be attributed to a quantity conditional on the presently available knowledge. The coverage probability of an interval result is the probability of the values in that interval that could be attributed to the unknown quantity. An interval result and its associated coverage probability are well-defined whether the unknown quantity has a unique true value, an essentially unique true value, or a set of multiple true values with a non-negligible definitional uncertainty [6, Section 2.11 Note 3]. The term interval result is analogous to the statistical term interval estimate. A confidence interval and a Bayesian credible interval are examples of interval estimate.

The adoption of operational view of a probability distribution and the replacement of the concept of a coverage interval with the concept of an interval result should remove the divergence of the GUM-S1 from the GUM while maintaining its alignment with Bayesian concepts. Enactment of this proposal will not affect procedures for determining results of measurement, magnitudes of evaluated uncertainties, and the assigned coverage probabilities.

8. Summary and concluding remarks

The GUM-S1 has misinterpreted the Bayesian concept of a statistical parameter by identifying the true value of the measurand as a statistical parameter and treating that parameter as a random variable. The value of a parameter is always fixed. What changes is a probability distribution which expresses the state of knowledge about that fixed value. The VIM3 definitions of a coverage interval and its coverage probability are mathematically defective because they break down for a result (of measurement) expressed as an interval of width smaller than the range of true values. We have offered simple suggestions to repair these two defects. The JCGM 104 is not a helpful introduction to the operational view of the uncertainty in measurement established by the GUM [22]; therefore, this document should be withdrawn.

The GUM-S1 introduced the idea of a coverage interval as the dominant expression of uncertainty in measurement. A coverage interval is defined as an interval containing the true value of the measurand with a stated probability (called coverage probability). The GUM-S1 idea of a coverage interval is essentially a pre-GUM concept of uncertainty which was concerned with covering the true value of the measurand by a computed interval. The GUM promulgated an operational view of the uncertainty in measurement as a parameter that characterizes the dispersion of the values that could be attributed to the measurand. In the GUM, the coverage probability of a result expressed as an interval is the probability of the values in that interval which could be attributed to the measurand. The operational perspective of the GUM had disconnected the uncertainty in measurement from the true value of the measurand. Thus the GUM-S1 concept of coverage interval does not agree with the GUM. Either the divergence of the GUM-S1 concept of a coverage interval from the GUM should be removed or the GUM-S1 should not be referred to as a supplement to the GUM. We have offered practical suggestions for revising the GUM-S1 to remove its divergence from the GUM while maintaining its alignment with Bayesian concepts.

Disclaimer

This article reflects only the views of the authors on the topics discussed, and does not necessarily reflect the official positions that their employers (NIST and PTB) may have about those topics or about the GUM and the JCGM documents.

Acknowledgments

The authors are grateful to the referees and the editor for their comments and suggestions which have greatly improved the draft that was submitted. The following also provided very useful comments on earlier drafts: Richard Allen, Michael Mascagni, Tony Kearsley, Marcel Reginatto, Tim Burns, and Ron Boisvert.

References

[1] JCGM. <http://www.bipm.org/en/committees/jc/jcgm/; http://www.iso.org/sites/JCGM/JCGM-introduction.htm>.
[2] Guide to the Expression of Uncertainty in Measurement (GUM) (BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, and OIML); 1993. ISBN 92-67-10188-9. The JCGM has made the GUM available in electronic form at <http://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf>.
[3] International vocabulary of basic and general terms in metrology, 2nd ed., (VIM2) (BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, and OIML); 1993. ISBN 92-67-01075-1.
[4] Supplement 1 to the GUM – propagation of distributions using Monte Carlo method (GUM-S1); 2008. Electronic version JCGM 101 from the JCGM is available at <http://www.bipm.org/utils/common/documents/jcgm/JCGM_101_2008_E.pdf>.
[5] An introduction to the "Guide to the expression of uncertainty in measurement" and related documents; 2009. Electronic version JCGM 104 from the JCGM is available at <http://www.bipm.org/utils/common/documents/jcgm/JCGM_104_2009_E.pdf>.
[6] International vocabulary of metrology – basic and general concepts and associated terms, third edition (VIM3); 2008. Electronic version JCGM 200 from the JCGM is available at <http://www.bipm.org/utils/common/documents/jcgm/JCGM_200_2012.pdf>.
[7] L.J. Gleser, Assessing uncertainty in measurement, Stat Sci 13 (1998) 277–290.
[8] R.N. Kacker, A. Jones, On use of Bayesian statistics to make the guide to the expression of uncertainty in measurement consistent, Metrologia 40 (2003) 235–248.
[9] G. Casella, R.L. Berger, Statistical inference, 2nd ed., Duxbury, Pacific Grove (CA), 2002.
[10] D.V. Lindley, Introduction to probability and Statistics from a Bayesian viewpoint, (1965a Part 1: probability), (1965b Part 2: inference), Cambridge University Press, UK, 1965.
[11] G.E.P. Box, G.C. Tiao, Bayesian inference in statistical analysis, Addison-Wesley, Reading (MA), 1973.
[12] E.T. Jaynes, Bayesian methods: general background, in: J.H. Justice (Ed.), Maximum entropy and Bayesian methods in applied statistics, Cambridge University Press, UK, 1985, pp. 1–25.
[13] J.O. Berger, Statistical decision theory and Bayesian analysis, 2nd ed., Springer, New York, 1985.
[14] J.M. Bernardo, A.F. Smith, Bayesian theory, Wiley, Chichester (England), 1994.
[15] D.V. Lindley, The philosophy of statistics, Statistician 49 (2000) 293–337.
[16] Bernardo JM. In: Viertl R, editor. Bayesian statistics. A chapter in Encyclopedia of Life Support Systems (EOLSS) Probability and Statistics. Oxford (UK): UNESCO; 2003. An updated and abridged pdf file of this chapter is available at <http://www.uv.es/bernardo/BayesStat.pdf>.
[17] S.J. Press, Subjective and objective Bayesian statistics, 2nd ed., Wiley, Hoboken (NJ), 2003.
[18] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Rubin, Bayesian data analysis, 2nd ed., Chapman & Hall/CRC, Boca Raton (FL), 2003.
[19] E.T. Jaynes, Probability theory: the logic of science, Cambridge University Press, UK, 2003.
[20] N.D. Singpurwalla, Reliability and risk: a Bayesian perspective, Wiley, New York, 2006.
[21] A. Possolo, B. Toman, Tutorial for Metrologists on the probabilistic and statistical apparatus underlying the GUM and related documents, NIST Statistical Engineering Division, Gaithersburg (MD, USA), 2011.

20899. Electronic version from NIST is available at <http://www.nist.gov/itl/sed/gsg/upload/TutorialMetrologists2011Nov22.pdf>.
[22] R.N. Kacker, Probability distributions and coverage probability in GUM, JCGM documents, and statistical inference, Measurement 65 (2015) 61–70.
[23] R.N. Kacker, K.D. Sommer, R. Kessel, Evolution of modern approaches to express uncertainty in measurement, Metrologia 44 (2007) 513–529.
[24] International vocabulary of basic and general terms in metrology, 1st ed. (VIM1) (BIPM, IEC, ISO, OIML); 1984.