Journal of Probability
and Statistical Science
20(1), 52-79, August 2022

A new modified Bayesian method for measurement uncertainty analysis and the unification of frequentist and Bayesian inference

Hening Huang

ABSTRACT
This paper proposes a new modification of the traditional Bayesian method for measurement uncertainty analysis. The new modified Bayesian method is derived from the law of aggregation of information (LAI) and the rule of transformation between the frequentist view and Bayesian view. It can also be derived from the original Bayes Theorem in continuous form. We focus on a problem that is often encountered in measurement science: a measurement gives a series of observations. We consider two cases: (1) there is no genuine prior information about the measurand, so the uncertainty evaluation is purely Type A, and (2) prior information is available and is represented by a normal distribution. The traditional Bayesian method (also known as the reformulated Bayes Theorem) fails to provide a valid estimate of standard uncertainty in either case. The new modified Bayesian method provides the same solutions to these two cases as its frequentist counterparts. The differences between the new modified Bayesian method and the traditional Bayesian method are discussed. This paper reveals that the traditional Bayesian method is not a self-consistent operation, so it may lead to incorrect inferences in some cases, such as the two cases considered. In the light of the frequentist-Bayesian transformation rule and the law of aggregation of information (LAI), the frequentist and Bayesian inference are virtually equivalent, so they can be unified, at least in measurement uncertainty analysis. The unification is of considerable interest because it may resolve the long-standing debate between frequentists and Bayesians. The unification may also lead to an indisputable, uniform revision of the GUM (*Evaluation of measurement data - Guide to the expression of uncertainty in measurement* (JCGM 2008)).

Keywords Bayesian method · frequentist method · measurement uncertainty · prior information · probability distribution

1. Introduction

Measurement uncertainty analysis is an important task in many fields of science and engineering. It can be regarded as a statistical inference process that involves three components: (1) information about the measurand, (2) a probability model with one or more unknown parameters, and (3) a statistical method. The information about the measurand may include prior information and the data obtained from current measurement (hence called current information). A probability model describes the variability of the measurand due to measurement errors.

□ Received October 2021, revised December 2021, in final form January 2022.
□ Hening Huang (corresponding author) is affiliated with the Teledyne RD Instruments (retired) 14020 Stowe Drive, Poway, CA 92064, USA, heninghuang1@gmail.com

The probability model commonly used in measurement uncertainty analysis is the normal distribution, which is also known as the law of probability of errors. A statistical method is a tool for inferring the unknown parameters of a probability model based on all available information: prior and current. It should be noted that this three-component inference process takes place for a direct measurement only. An indirect measurement may involve additional components: a measurement model that relates the quantity of interest (the measurand) to the measurable quantities and the law of propagation of uncertainty (LPU). This paper deals with direct measurements only.

In measurement science, an often-encountered problem is to estimate the true value of a physical quantity (the measurand) and the associated uncertainty from a measurement that gives a series of observations. We consider two cases: (1) there is no genuine prior information about the measurand, and (2) prior information is available and is represented by a normal distribution. We assume that the measurand is normally distributed with two unknown parameters: location parameter μ and scale parameter σ. Let $\hat{\mu}$ denote an estimator of μ and $\widehat{\mathrm{SU}}$ denote the standard uncertainty (SU) of $\hat{\mu}$. So, our job is to obtain $\hat{\mu}$ and $\widehat{\mathrm{SU}}$ for Case 1 and Case 2 respectively.

There are two different views and methodologies for the problem considered: frequentist and Bayesian. Frequentists view the unknown parameters μ and σ as fixed constants (denoted by μ_T and σ_T respectively). For Case 1, the classical frequentist point estimation gives the sample mean as $\hat{\mu}$ and the sample standard error with bias correction as the Type A SU (Huang 2014, 2018a). The frequentist Type A SU along with a uncertainty-based measurement quality control criterion (Huang 2015) is recently adopted in ISO:24578:2021(E) (ISO 2021). For Case 2, two frequentist methods: LCD-based (LCD stands for the law of combination of distributions) and least squares, give the same results: $\hat{\mu}$ = inverse-variance weighted-average of the prior mean and the sample mean, and $\widehat{\mathrm{SU}}$ = square root of the combined variance (Huang 2020b) (see table 1).

Table 1. Frequentist and Bayesian (traditional method) solutions to the problem considered

Case	Genuine prior information	Frequentist solution		Bayesian solution (traditional method)	
		$\hat{\mu}$	$\widehat{\mathrm{SU}}$	$\hat{\mu}$	$\widehat{\mathrm{SU}}$
1	None	$\bar{x}_D$	$\dfrac{s_D}{c_4\sqrt{n}}$	$\bar{x}_D$	$\dfrac{\sqrt{n-1}}{\sqrt{n-3}}\dfrac{s_D}{\sqrt{n}}$
2	$N(x_{\text{prior}}, \sigma_{\text{prior}})$	$\dfrac{x_{\text{prior}}\left(\frac{s_D}{c_4}\right)^2 + n\bar{x}\sigma_{\text{prior}}^2}{\left(\frac{s_D}{c_4}\right)^2 + n\sigma_{\text{prior}}^2}$	$\dfrac{\sigma_{\text{prior}}\left(\frac{s_D}{c_4}\right)}{\sqrt{\left(\frac{s_D}{c_4}\right)^2 + n\sigma_{\text{prior}}^2}}$	see section 5	see section 5

Note: $\bar{x}_D$= observed sample mean, s_D = observed sample standard deviation, n = sample size (number of observations), x_{prior} = prior mean, σ_{prior}^2 = prior variance, c_4 = bias correction factor, $c_4 = \sqrt{\frac{2}{n-1}}\Gamma\left(\frac{n}{2}\right)/\Gamma\left(\frac{n-1}{2}\right)$, and $\Gamma(.)$ stands for Gamma function.

In contrast to the frequentist view, Bayesians view the unknown parameters μ and σ as random variables. Also, Bayesians view the measurement as part of an epistemic process that combines prior knowledge (represented by prior distribution) with current knowledge (represented by likelihood function) about the unknown parameters through Bayes Theorem to obtain the updated knowledge (represented by posterior distribution) (Huang 2020a). The mean of the marginal posterior distribution of μ is taken as $\hat{\mu}$ and the standard deviation is the SU of $\hat{\mu}$. For Case 1, a traditional Bayesian solution is obtained by using the Jeffreys prior $1/\sigma$, leading to the scaled and shifted *t*-distribution as the marginal posterior distribution. The Bayesian estimate $\hat{\mu}$ is the same as the frequentist estimate, i.e. the sample mean $\bar{x}_D$; the Bayesian estimate $\widehat{\mathrm{SU}}$, known as the Bayesian Type A SU, is $\frac{\sqrt{n-1}}{\sqrt{n-3}}\frac{s_D}{\sqrt{n}}$ (e.g. Kacker and Jones 2003). Note that when the sample size *n* is less than 4, the Bayesian Type A SU is undefined. The undefined factor $\sqrt{(n-1)/(n-3)}$ may be replaced with an ad hoc factor (Kacker and Jones 2003).

In the author's belief, a valid Type A SU ought to meet the following criteria:

(1) Unbiasedness, also known as conformity. That is, the expectation of Type A SU must be equal to the true SU, the scale parameter $\mathrm{SU} = \sigma_T/\sqrt{n}$ (Huang 2020a).
(2) On the average, Type A SU must meet the -1/2 power law, the physical law for the relationship between the measurement precision and the number of observations (Huang 2018b).
(3) Uncertainty analysis must be consistent with error analysis in the case that the true value about the measurand is known or assumed (Huang 2018b).
(4) Transferability (Wubbeler and Elster 2020). That is, "it should be possible to use directly the uncertainty evaluated for one result as a component in evaluating the uncertainty of another measurement in which the first result is used".

The frequentist Type A SU meets all four criteria, whereas the Bayesian Type A SU does not. Wubbeler and Elster (2020) demonstrated in terms of simple examples that the GUM-S1 type A evaluation of uncertainty (i.e. the Bayesian Type A SU) fails to generally ensure the requirement of transferability. Additionally, the Bayesian Type A SU artificially dilates uncertainty, causing unrealistic estimates of uncertainty when the sample size is small. Moreover, the use of the Bayesian Type A SU in the law of propagation of uncertainty (LPU) does not resolve the well-known Ballico paradox (Huang 2018a). Therefore, the Bayesian Type A SU is essentially invalid and should be abandoned (Huang 2020a).

It is a surprising fact that there is no Bayesian solution to Case 2 reported in the literature, neither analytical nor numerical. This is confirmed by a discussion on ResearchGate in which five statisticians are participated (Huang 2019b). In this study, we provided a numerical solution to Case 2 for several datasets using the traditional Bayesian method. The estimated SUs, which are presented in section 5, are significantly biased with respect to the true SUs, so they are essentially invalid as far as the unbiasedness criterion is concerned.

The application of Bayesian statistics in metrology and measurement science has been studied for more than two decades, but it is still immature. Possolo and Bodnar (2018) stated, "…*Metrologia* alone has published more than 80 articles since 2002 including the word 'Bayes' and its derivatives." Most recently, Wubbeler et al. (2020) proposed a simple method for Bayesian uncertainty evaluation for measurement models that linearly depend on a single input quantity, Demeyer et al. (2021) provided Bayesian uncertainty evaluations for a large class of GUM measurement models, and Stoudt et al. (2021) illustrated the application of empirical Bayes methods to uncertainty evaluations. In particularly, the revision of the GUM adopted the

Bayesian methodology (Bich 2014, Kyriazis 2015, Lira 2016, Bich *et al*. 2016). However, some authors objected to the revision of the GUM based on Bayesian statistics (Willink and White 2011, Attivissimo et al. 2012, Giaquinto et al. 2014, Giaquinto and Fabbiano 2016, White 2016, Willink 2016, Huang 2019c). Willink and White (2011) stated, "It is our view that the GUM should be revised, but not according to the Bayesian philosophy." The first draft of the revised GUM was released in December 2014; it received more than 1000 comments and the feedback was largely negative (Bich *et al*. 2016). According to the news on the website of the working group 1 (WG1) of the Joint Committee for Guides in Metrology (JCGM) dated May 4, 2019, the JCGM-WG1 plans to preserve the existing GUM and to generate several documents including JCGM 108: Bayesian methods, based on the rejected draft of the revised GUM. In the author's opinion, however, Bayesian methods should not be accepted for measurement uncertainty analysis unless they can provide valid solutions to Case 1 and Case 2.

It is important to note that, the traditional (and current) Bayesian Theorem used for measurement uncertainty analysis and other applications is not the original Bayesian Theorem; it is a reformulated Bayesian Theorem, which remarkably deviates from the original Bayesian Theorem. Box and Tiao (1992) provided a derivation of the reformulated Bayesian Theorem from the original Bayesian Theorem. However, as demonstrated later in this paper, his derivation is far from rigorous and is actually faulty. Therefore, the traditional Bayesian method is flawed, which is why it fails to provide a valid estimate of standard uncertainty in Case 1 or Case 2.

The purpose of this study is to develop valid Bayesian solutions to Case 1 and Case 2. The author accepts both the Bayesian view and frequentist view. It has been the author's thought for years that Bayesian and frequentist methods should give the same or approximately the same results in measurement uncertainty analysis. Otherwise, come to think of it, a practitioner would be confused about different results from Bayesian and frequentist methods. The author believes that, at least in measurement uncertainty analysis, Bayesian statistics and frequentist statistics should find some common grounds, and the unification of frequentist and Bayesian inference should be possible.

In the following, section 2 presents a rule of transformation between the frequentist and Bayesian views. Section 3 presents a new modified Bayesian method. Section 4 presents the solution to the problem considered (Case 1 and Case 2) with the new modified Bayesian method. Section 5 compares the new modified Bayesian method with the traditional Bayesian method. Section 6 and 7 present discussion and conclusion respectively.

2. Transformation between the frequentist view and Bayesian view

Suppose we are considering a measurement problem that involves one unknown parameter θ in the probability model (e.g. the normal distribution with unknown mean μ, i.e. $\theta = \mu$). We assume that there exists a true value of θ, denoted by θ_T, which is an unknown constant. It should be mentioned that there are some confusion and misunderstanding about the concept of true values in the literature. Huang (2020a) recently clarifies that the concept of true values is a common ground among the three approaches for computing measurement uncertainties: GUM's confidence interval based, Bayesian, and probability interval based (i.e. the unified theory of measurement errors and uncertainties (Huang 2018a)). That is, both frequentists and Bayesians accept the concept of true values and admit that there exists the true value θ_T (an unknown constant). When discussing a post by Mayo (2012), a guest statistician stated, “Bayesians describe what they know about parameters via probability calculus; this doesn’t rule out the truth being an unknown constant.”

According to frequentist statistics, the sample statistic $\hat{\theta}$ is an estimator of the true value θ_T. The estimator $\hat{\theta}$ is a random variable that gives a fixed value $\hat{\theta}_D$ for a given dataset, i.e. $\hat{\theta}_D$ is a realization of $\hat{\theta}$. For example, in the normal distribution model, μ_T is the true location parameter. The sample mean $\bar{x}$ is an estimator of μ_T, i.e. $\hat{\mu} = \bar{x}$. The observed sample mean $\hat{x}_D$ is a realization of $\hat{\mu} = \bar{x}$ or an estimate of μ_T. Note that throughout this paper, we use the subscript “*T*” to indicate the true value and “*D*” the realization of a random variable for a given dataset, any of which is a constant; the quantity without the subscript “*T*” or “*D*” represents a random variable, unless otherwise stated.

In the frequentist view, a simple (and basic) measurement model is written as

$$\hat{\theta} = \theta_T + \varepsilon \tag{1}$$

where ε is the error of $\hat{\theta}$ relative to the true value θ_T. We assume that the error ε is caused by a random effect (e.g. electronic noise of a measuring instrument) that is zero in the long run. Thus, $\hat{\theta}$ is an unbiased estimator of θ_T

$$\mathrm{E}(\hat{\theta}) = \theta_T + 0 \tag{2}$$

and the variance of $\hat{\theta}$ is the same as the variance of ε

$$\mathrm{Var}\left(\hat{\theta}\right) = 0 + \mathrm{Var}(\varepsilon) \tag{3}$$

The true probability density function (PDF) of $\hat{\theta}$, from the frequentist view, is written as $p[\hat{\theta}|\theta_T, \mathrm{Var}(\varepsilon)]$, where θ_T is the location parameter and $\mathrm{Var}(\varepsilon)$ is the scale parameter. For the normal distribution model, $\hat{\theta} = \hat{\mu} = \bar{x}$, $\theta_T = \mu_T$, and $\mathrm{Var}(\varepsilon) = \sigma_T$. Thus, the true frequentist PDF of $\bar{X}$ is $p[\bar{x}|\mu_T, \sigma_T]$.

Bayesians treat θ as a random variable. In the Bayesian view, the measurement model is written as

$$\theta = \hat{\theta}_D + \varepsilon' \tag{4}$$

where ε' is the error of θ relative to the observed value $\hat{\theta}_D$. Equation (4) can be rewritten as

$$\theta = \hat{\theta}_D + \delta_D + \varepsilon \tag{5}$$

where δ_D is the deviation of the observed value from the true value: $\delta_D = \theta_T - \hat{\theta}_D$.

The expectation of θ is

$$\mathrm{E}(\theta) = \hat{\theta}_D + \delta_D + 0 = \theta_T + 0 \tag{6}$$

and the variance of θ is the same as the variance of ε

$$\mathrm{Var}(\theta) = 0 + \mathrm{Var}(\varepsilon) \tag{7}$$

The true PDF of θ, from the Bayesian view, is written as $p[\theta|\theta_T, \mathrm{Var}(\varepsilon)]$. For the normal distribution model, $\theta = \mu$, $\theta_T = \mu_T$, and $\mathrm{Var}(\varepsilon) = \sigma_T$. Thus, the true Bayesian PDF of μ is $p[\mu|\mu_T, \sigma_T]$.

In Bayesian statistics, the given quantity is just a dataset (i.e. the observed value $\hat{\theta}_D$) only. So Eq. (4) is approximately written as (Huang 2020a)

$$\theta \approx \hat{\theta}_D + \varepsilon \tag{8}$$

Consequently, $\mathrm{E}(\theta) = \theta_T \approx \hat{\theta}_D$. The usual Bayesian PDF of θ is written as $[\theta|\hat{\theta}_D, \mathrm{Var}(\varepsilon)]$, which is an estimate (or approximation) of the true Bayesian PDF: $p[\theta|\theta_T, \mathrm{Var}(\varepsilon)]$. For the normal distribution model, $\theta = \mu$ and $\hat{\theta}_D = \bar{x}_D$. Thus, the usual Bayesian PDF of μ is $p[\mu|\bar{x}_D, \sigma_T]$.

On the other hand, if we simply use $\hat{\theta}_D$ as the estimate of θ_T in Eq. (1), the frequentist view-based measurement model becomes

$$\hat{\theta} \approx \hat{\theta}_D + \varepsilon \tag{9}$$

Consequently, $\mathrm{E}(\hat{\theta}) = \theta_T \approx \hat{\theta}_D$. The true frequentist PDF: $p[\hat{\theta}|\theta_T, \mathrm{Var}(\varepsilon)]$ may be replaced by the estimated frequentist PDF: $p[\hat{\theta}|\hat{\theta}_D, \mathrm{Var}(\varepsilon)]$. For the normal distribution model, $\hat{\theta} = \bar{x}$ and $\hat{\theta}_D = \bar{x}_D$. Thus, the estimated frequentist PDF of $\bar{X}$ is $p[\bar{x}|\bar{x}_D, \sigma_T]$.

It is important to note that either $\hat{\theta}$ or θ stands for a random variable; they can be exchanged with each other, or they can be replaced by a common symbol, such as *Z*. That is, the left side of Eqs. (1), (5), (8), and (9) can be replaced with the common random variable *Z*. Therefore, Eq. (1) and Eq. (5) are mathematically equivalent; Eq. (8) and Eq. (9) are mathematically equivalent. The equivalence is due to the fact that the true value (or its estimate) and the measurement error in these formulas are physical quantities that are independent of the viewpoint (or reference frame). In other words, the true value and measurement error are the common ground for frequentist and Bayesian statistics for measurement uncertainty analysis. Moreover, if θ and $\hat{\theta}$ are replaced with *Z*, the true Bayesian PDF of θ, $p[\theta|\theta_T, \mathrm{Var}(\varepsilon)]$ is identical to the true frequentist PDF of $\hat{\theta}$, $p[\hat{\theta}|\theta_T, \mathrm{Var}(\varepsilon)]$ and the usual Bayesian PDF: $p[\theta|\hat{\theta}_D, \mathrm{Var}(\varepsilon)]$ is identical to the estimated frequentist PDF: $p[\hat{\theta}|\hat{\theta}_D, \mathrm{Var}(\varepsilon)]$. That is, the usual Bayesian PDF of θ, $p[\theta|\hat{\theta}_D, \mathrm{Var}(\varepsilon)]$, is nothing but the estimated frequentist PDF of $\hat{\theta}$, $p[\hat{\theta}|\hat{\theta}_D, \mathrm{Var}(\varepsilon)]$.

The above analysis suggests that the frequentist view can be transformed to the Bayesian view, and vice versa. The rule of the transformation between the two views can be written as

$$\text{frequentist:}\begin{Bmatrix} \hat{\theta} \\ \theta_T \approx \hat{\theta}_D \end{Bmatrix} \leftrightarrow \text{Bayesian:}\begin{Bmatrix} \theta \\ \hat{\theta}_D \end{Bmatrix} \tag{10}$$

This transformation requires that the observed value $\hat{\theta}_D$ is an unbiased estimate of the true value θ_T. Equation (10) is referred to as the frequentist-Bayesian transformation rule hereafter.

It should be pointed out that in the Bayesian view, Eq. (8) is an "epistemic" formulation for the measurement problem considered when no prior knowledge is involved (Huang 2020a). In this situation, θ only represents the current knowledge about the measurand, based on the observed value $\hat{\theta}_D$. However, this knowledge has uncertainty due to the measurement error ε (Huang 2020a). Moreover, $\mathrm{E}(\theta) \approx \hat{\theta}_D$ is the Bayesian estimate of θ_T for a given dataset when no prior information is involved in the analysis. The value of $\mathrm{E}(\theta)$ or $\hat{\theta}_D$ will be different for different datasets. Therefore, the Bayesian estimator of θ_T is $\mathrm{E}(\theta) \approx \hat{\theta}_D$, which is essentially the same as $\hat{\theta}$, the frequentist estimator of θ_T. And the standard uncertainty is $\sqrt{\mathrm{Var}(\varepsilon)}$, regardless of the viewpoint.

In summery, for the simple (and basic) measurement model without involving prior information, the parameter space $\Omega(\theta)$ in the Bayesian view is identical to the sample space $\Omega(\hat{\theta})$ in the frequentist view. That is, the set of possible values of the parameter θ is the same as the set of the possible values of the sample statistic $\hat{\theta}$, which is a specified exhaustive set of possibilities (Loredo 1990). Both the usual Bayesian PDF of θ and the estimated frequentist PDF of $\hat{\theta}$ share the same probability model, e.g. the normal distribution model in which $\theta = \mu$, and $\hat{\theta} = \bar{x}$. The true Bayesian PDF of θ, given θ_T (there may be a special case where $\hat{\theta}_D = \theta_T$) is identical to the true frequentist PDF of $\hat{\theta}$ when θ_T is known; both PDFs come from a common ground: the PDF of measurement error.

However, theoretically, Bayesian methods require the use of prior knowledge (information). This is because Bayesians consider the measurement as part of an epistemic process that combines prior knowledge with current knowledge through Bayes Theorem to obtain the updated knowledge (Huang 2020a).

3. The new modified Bayesian method

According to Shannon's information theory, the information content of an outcome $\{z\}$ that is drawn from the probability distribution $p_j(z)$ is written as (Shannon 1948)

$$I_j(z) = -\log\left[p_j(z)\right] \tag{11}$$

where j is the index of information sources. Information can be viewed as a real, physical quantity (Landauer 1961, Clemen and Winkler 1999, Wile 2012). The information that comes from different (independent) sources is additive to give aggregated information, denoted by $I_{\mathrm{c}}(z)$, as stated by the law of aggregation of information (LAI) (Huang 2020b)

$$I_{\mathrm{c}}(z) = \sum_{j=1}^{N} I_j(z) - I_0 \tag{12}$$

where N is the number of information sources and I_0 is a constant that is due to the scale factor in the law of combination of distribution (LCD). The LCD is written as (Huang 2020b)

$$p_{\mathrm{c}}(z) = \frac{\prod_{j=1}^{N} p_j(z)}{\int \prod_{j=1}^{N} p_j(z)dz} \tag{13}$$

where $\int \prod_{j=1}^{N} p_j(z)dz$ is the scale factor that ensures the integration of $p_{\mathrm{c}}(z)$ is one. So $I_0 = -\log\left[\int \prod_{j=1}^{N} p_j(z)dz\right]$. The LAI and the LCD are virtually equivalent.

It should be pointed out that the essence of the LCD (or LAI) is the way of using probability distributions, regardless of whether a probability distribution is based on the frequentist view of probability or based on the Bayesian view of degree of belief (i.e. subjective probability) (Huang

2020b). In other words, the LAI (or LCD) is applicable to either the information represented by frequency distributions in frequentist statistics or the information represented by state-of-knowledge probability distributions in Bayesian statistics. In the author's opinion, however, it is always preferred that, in measurement uncertainty analysis, a state-of-knowledge probability distribution can be traced back to some kind of relative frequency. In other words, measurement uncertainty analysis should be free of degree of belief or subjective probability whenever possible so that a uniform result can be obtained by different people.

For a problem involving one unknown parameter, we can deal with two random variables from different viewpoints: $\hat{\theta}$ from the frequentist view and θ from the Bayesian view. However, the information contained in the data obtained from current measurement must be the same, regardless of the viewpoint. That is

$$I_{\text{current}}(\hat{\theta}) = I_{\text{current}}(\theta) \tag{14}$$

Equation (14) is valid because we have demonstrated in section 2 that the usual Bayesian PDF of θ, $p[\theta|\hat{\theta}_D, \text{Var}(\varepsilon)]$, is nothing but the estimated frequentist PDF of $\hat{\theta}$, $p[\hat{\theta}|\hat{\theta}_D, \text{Var}(\varepsilon)]$.

Suppose we have the prior information about θ in the Bayesian view, denoted by $I_{\text{prior}}(\theta)$. We assume that the prior information is independent of the current information (measurement). Then, according to the LAI, Eq. (12), the aggregated information (i.e. posterior information), denoted by $I_{\text{post}}(\theta)$, can be written as

$$I_{\text{post}}(\theta) = I_{\text{prior}}(\theta) + I_{\text{current}}(\theta) - I_0 \tag{15}$$

In terms of PDFs

$$p_{\text{post}}(\theta) \propto p_{\text{prior}}(\theta) \cdot p_{\text{current}}(\theta) \tag{16}$$

On the other hand, in the frequentist view, the posterior information can be written as

$$I_{\text{post}}(\hat{\theta}) = I_{\text{prior}}(\hat{\theta}) + I_{\text{current}}(\hat{\theta}) - I_0 \tag{17}$$

or

$$p_{\text{post}}(\hat{\theta}) \propto p_{\text{prior}}(\hat{\theta}) \cdot p_{\text{current}}(\hat{\theta}) \tag{18}$$

which is the same formula proposed by Huang (2020b) that gives the frequentist solution to Case 2 shown in table 1.

Thus, the LAI (or LCD) provides a unified framework for combing prior information with current measurement regardless of the viewpoint.

Equation (16) is the proposed new modified Bayesian method for the one-dimensional problem that only deals with one unknown parameter θ. It can be extended to two or higher dimensions when more than one unknown parameters are involved. For example, if a problem involves two unknown parameters θ_1 and θ_2, the joint posterior PDF of θ_1 and θ_2 can be written as

$$p_{\text{post}}(\theta_1, \theta_2) \propto p_{\text{prior}}(\theta_1, \theta_2) \cdot p_{\text{current}}(\theta_1, \theta_2) \tag{19}$$

The key in using the new modified Bayesian method (MBM) is to obtain the current PDF, such as $p_{\text{current}}(\theta)$ or $p_{\text{current}}(\theta_1, \theta_2)$.

4. Solutions to the problem considered

4.1. The joint current PDFs for the unknown parameters μ and σ

For the problem considered, the probability model is assumed to be normal with two unknown parameters: μ and σ. The elementary statistical analysis of the data obtained from a series of observations gives two sample statistics: sample mean $\bar{x}$ and sample standard deviation s. The (unbiased) measurement model for μ in the frequentist view is written as

$$\bar{x} = \mu_T + \varepsilon \tag{20}$$

where μ_T is the true values of μ and $\bar{x}$ is an unbiased estimator of μ_T. Both $\bar{x}$ and ε (error) are random variables. The true PDF of $\bar{x}$ can be written as

$$p(\bar{x}|\mu_T, \sigma_T) = N\left(\bar{x}|\mu_T, \frac{\sigma_T}{\sqrt{n}}\right) = \frac{\sqrt{n}}{\sqrt{2\pi}\sigma_T} \exp\left(-\frac{n}{2}\frac{(\bar{x}-\mu_T)^2}{\sigma_T^2}\right) \tag{21}$$

where σ_T is the true value of σ; it is a fixed constant.

On the other hand, the (unbiased) measurement model for σ in the frequentist view is written as

$$\frac{s}{c_4} = \sigma_T + \xi \tag{22}$$

where, $\frac{s}{c_4}$ is an unbiased estimator of σ_T. Both $\frac{s}{c_4}$ and ξ (error) are random variables. The true PDF of s/c_4 can be obtained from a slight modification to the distribution of the sample standard deviation given in WolfamMathworld (2020). It is a function of σ_T and n

$$p\left(\frac{s}{c_4}|\sigma_T\right) = \frac{1}{\Gamma(\frac{n-1}{2})} 2^{\frac{3-n}{2}} \sqrt{n}(n-1)^{\frac{n-2}{2}} \frac{(\frac{s}{c_4})^{n-2}}{\sigma_T^{n-1}} \exp\left(-\frac{1}{2}(n-1)\frac{\left(\frac{s}{c_4}\right)^2}{\sigma_T^2}\right) \tag{23}$$

Since $\varepsilon = (\bar{x} - \mu_T)$ and s/c_4 are independent random variables, the true joint PDF of $\bar{x}$ and s/c_4 can be constructed as

$$p\left(\bar{x}, \frac{s}{c_4}|\mu_T, \sigma_T\right) = p(\bar{x}|\mu_T, \sigma_T) \cdot p\left(\frac{s}{c_4}|\sigma_T\right) = N\left(\bar{x}|\mu_T, \frac{\sigma_T}{\sqrt{n}}\right) \frac{1}{\Gamma(\frac{n-1}{2})} 2^{\frac{3-n}{2}} \sqrt{n}(n-1)^{\frac{n-2}{2}} \frac{(\frac{s}{c_4})^{n-2}}{\sigma_T^{n-1}} \exp\left(-\frac{1}{2}(n-1)\frac{\left(\frac{s}{c_4}\right)^2}{\sigma_T^2}\right) \tag{24}$$

According to the frequentist-Bayesian transformation rule, the joint current PDF of μ and σ can be obtained by replacing $\bar{x}$ with μ, $\frac{s}{c_4}$ with σ, μ_T with $\bar{x}_D$, and σ_T with $\frac{s_D}{c_4}$ respectively in Eq. (24)

$$p_{\text{current}}(\mu,\sigma|\text{data}) = N(\mu|\bar{x}_D,\frac{s_D}{c_4\sqrt{n}})\frac{1}{\Gamma\left(\frac{n-1}{2}\right)}2^{\frac{3-n}{2}}\sqrt{n}(n-1)^{\frac{n-2}{2}}\frac{\sigma^{n-2}}{\left(\frac{s_D}{c_4}\right)^{n-1}}\exp\left(-\frac{1}{2}(n-1)\frac{\sigma^2}{\left(\frac{s_D}{c_4}\right)^2}\right) \quad (25)$$

4.2. Solution to Case 1

According to the new modified Bayesian method (MBM), Eq. (19), the joint posterior PDF of μ and σ can be written as

$$p_{\text{post}}(\mu,\sigma|\text{prior},\text{data}) \propto p_{\text{prior}}(\mu,\sigma)\cdot p_{\text{current}}(\mu,\sigma|\text{data}) \quad (26)$$

Since there is no genuine prior information about μ and σ in Case 1, we employ the flat priors over μ and σ. The resulting joint posterior PDF of μ and σ is the same as the joint current PDF $p_{\text{current}}(\mu,\sigma|\text{data})$, Eq. (25). Consequently, the marginal posterior PDF of μ can be written as

$$p_{\text{post}}(\mu|\text{data}) = N(\mu|\bar{x}_D,\frac{s_D}{c_4\sqrt{n}}) = \frac{\sqrt{n}}{\sqrt{2\pi}\frac{s_D}{c_4}}\exp\left(-\frac{n}{2}\frac{(\bar{x}_D-\mu)^2}{\left(\frac{s_D}{c_4}\right)^2}\right) \quad (27)$$

The expectation of $p_{\text{post}}(\mu|\text{data})$ is the Bayesian estimator $\hat{\mu}_{\text{MBM}}$

$$\hat{\mu}_{\text{MBM}} = \text{E}(\mu) = \bar{x}_D \quad (28)$$

and the standard deviation of $p_{\text{post}}(\mu|\text{data})$ is the Bayesian estimator $\widehat{\text{SU}}_{\text{MBM}}$

$$\widehat{\text{SU}}_{\text{MBM}} = \sqrt{\text{Var}(\mu)} = \frac{s_D}{c_4\sqrt{n}} \quad (29)$$

These results are the same as those obtained with the frequentist methods shown in table 1.

4.3. Solution to Case 2

In Case 2, the prior information about μ is represented by a normal PDF: $p_{\text{prior}}(\mu,\sigma) = N(\mu|x_{\text{prior}},\sigma_{\text{prior}})$. According to the new modified Bayesian method (MBM), the joint posterior PDF of μ and σ can be written as

$$p_{\text{post}}(\mu, \sigma|\text{prior}, \text{data}) \propto N(\mu|x_{\text{prior}}, \sigma_{\text{prior}})N(\mu|\bar{x}_D, \frac{s_D}{c_4\sqrt{n}})\sigma^{n-2}\exp\left(-\frac{1}{2}(n-1)\frac{\sigma^2}{\left(\frac{s_D}{c_4}\right)^2}\right) \quad (30)$$

The marginal posterior PDF of μ can be written as

$$p_{\text{post}}(\mu|\text{prior}, \text{data}) \propto N(\mu|x_{\text{prior}}, \sigma_{\text{prior}}) \cdot N(\mu|\bar{x}_D, \frac{s_D}{c_4\sqrt{n}}) \quad (31)$$

which is a normal PDF. The expectation and standard deviation of $p_{\text{post}}(\mu|\text{prior}, \text{data})$ are the Bayesian estimators of the measurand and the SU respectively

$$\hat{\mu}_{\text{MBM}} = \text{E}(\mu) = \frac{x_{\text{prior}}\left(\frac{s_D}{c_4}\right)^2 + n\bar{x}\sigma_{\text{prior}}^2}{\left(\frac{s_D}{c_4}\right)^2 + n\sigma_{\text{prior}}^2} \quad (32)$$

$$\widehat{\text{SU}}_{\text{MBM}} = \sqrt{\text{Var}(\mu)} = \frac{\sigma_{\text{prior}}\left(\frac{s_D}{c_4}\right)}{\sqrt{\left(\frac{s_D}{c_4}\right)^2 + n\sigma_{\text{prior}}^2}} \quad (33)$$

These results are the same as those obtained with the frequentist methods (Huang 2020b) shown in table1.

5. Comparing with the traditional Bayesian method (TBM)

The Bayes Theorem, in its discrete form, relates the conditional probabilities of events *A* and *B*:

$$P(A|B) = \frac{P(A) \cdot P(B|A)}{P(B)} \quad (34)$$

where $P(A|B)$ is the probability of event *A* given event *B*, $P(B|A)$ is the probability of event *B* given event *A*, $P(A)$ is the probability of event *A*, and $P(B)$ is the probability of event *B*. According to Sober (2001), Eq. (34) is true if each quantity (probability) mentioned in it is well defined.

However, the Bayes Theorem in discrete form, Eq. (34), is rarely used in measurement uncertainty analysis. Instead, the Bayes Theorem in continuous form is usually used. For a one-dimensional problem (i.e. one unknown parameter θ), the traditional (and current) Bayes Theorem in continuous form is written as

$$p_{\text{post}}(\theta|\text{prior}, \text{data}) \propto p_{\text{prior}}(\theta) \cdot L(\theta|\text{data}) \quad (35)$$

where $L(\theta|\text{data})$ is the likelihood function of θ given the data. We also refer to Eq. (35) as the traditional Bayesian method (TBM).

The biggest difference between the new modified Bayesian method, Eq. (16), and the traditional Bayesian method, Eq. (35), is that in the former the current measurement (data) is represented by the current PDF, whereas in the latter the current measurement (data) is represented by the likelihood function. In order to gain insight into the difference between the new method and the traditional method, the following subsections present several examples of Case 1 and Case 2, solved with these two methods.

5.1. Examples of Case 1

We first consider a dataset randomly generated from a normal distribution $N(\mu_T = 10, \sigma_T = 0.5)$, which gives four numerical values: 10.00, 9.79, 9.76, and 10.75. The sample mean happens to be the same as the population mean, i.e. $\bar{x}_D = \mu_T = 10$; the sample standard deviation after the bias correction happens to be the same as the population standard deviation, i.e. $\frac{s_D}{c_4} = \sigma_T = 0.5$ (where c_4==0.9213 for n=4). The Type A SU estimated with the frequentist method based on the unbiasedness criterion is the same as the true SU (i.e. Type B SU), i.e. $\frac{s_D}{c_4\sqrt{n}} = \frac{\sigma_T}{\sqrt{n}} = 0.25$.

Since there is no genuine prior information in Case 1, the marginal posterior PDF obtained with the new modified Bayesian method coincides with the current PDF $N\left(\mu \middle| \bar{x}_D, \frac{s_D}{c_4\sqrt{n}}\right) = N(\mu|10,0.25)$ that is the true posterior distribution at n=4. Consequently, $\hat{\mu}_{\text{MBM}}$=10 and $\widehat{\text{SU}}_{\text{MBM}} = 0.25$.

On the other hand, according to the traditional Bayesian method, the joint posterior PDF of μ and σ is written as

$$p_{\text{post}}(\mu, \sigma|\text{prior}, \text{data}) \propto p_{\text{prior}}(\mu, \sigma) \cdot L(\mu, \sigma|\text{data}) \tag{36}$$

where $L(\mu, \sigma|\text{data})$ is the joint likelihood function of μ and σ that can be written as (e.g. Edwards 1992, Murphy 2007)

$$L(\mu, \sigma|\text{data}) \propto \sigma^{-n} \exp\left(-\frac{n}{2}\frac{(\bar{x}_D - \mu)^2}{\sigma^2}\right) \exp\left(-\frac{1}{2}(n-1)\frac{s_D^2}{\sigma^2}\right) \tag{37}$$

The solution of the traditional Bayesian method depends on the prior $p_{\text{prior}}(\mu, \sigma)$. We consider two types of priors: Jeffreys prior and flat prior. For this dataset (n=4), the Jeffreys prior results in a scaled and shifted t-distribution (marginal posterior PDF) with $\hat{\mu}_{\text{TBM}}$=10 and $\widehat{\text{SU}}_{\text{TBM}} = 0.374$. The flat prior results in a marginal posterior PDF of μ with $\hat{\mu}_{\text{TBM}}$=10 and $\widehat{\text{SU}}_{\text{TBM}} = 0.603$.

Figure 1 shows the three marginal posterior distributions of μ for this dataset (n=4). It can be seen from figure 1 that the scaled and shifted t-distribution has a lower peak and fatter tails than its normal distribution counterpart. The marginal posterior PDF given by the traditional Bayesian method (TBM) with the flat prior has a lower peak and fatter tails than both the t-distribution and the normal distribution. It can be seen that the two marginal posterior PDFs given by the traditional Bayesian method (TBM) are distorted with respect to the normal distribution, the true posterior distribution.

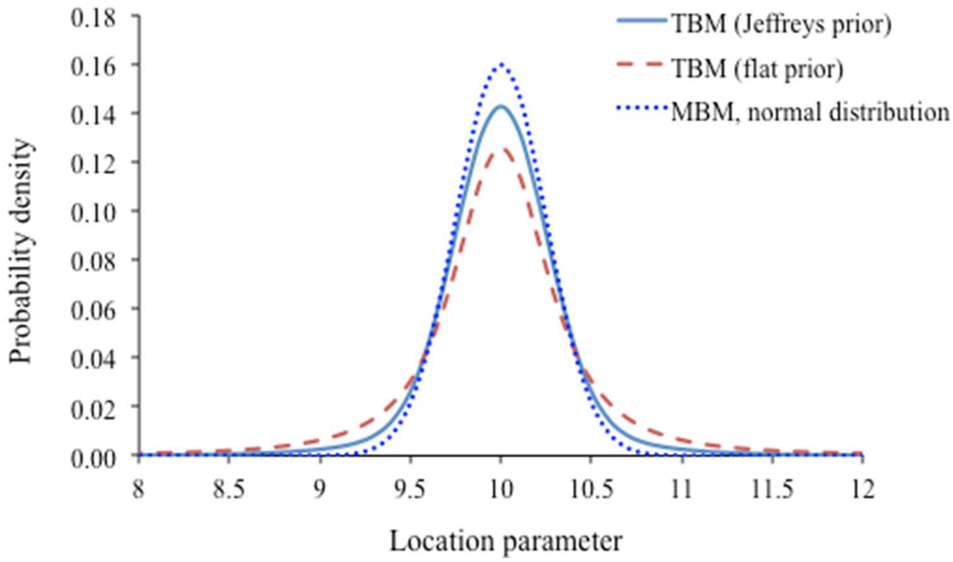

Figure 1. Comparison of the marginal posterior distributions of μ, given by the new modified Bayesian method (MBM) and the traditional Bayesian method (TBM) (with the Jeffreys prior and the flat prior) (Case 1, n=4)

We then consider other two datasets at n=10 and 20, which are generated from the same normal distribution $N(10, 0.5)$. Each of the datasets gives $\bar{x}_D = \mu_T = 10$ and $\frac{s_D}{c_4} = \sigma_T = 0.5$ (c_4=0.9727 at n=10 and c_4=0.9869 at n=20). Table 2 shows the Bayesian Type A SUs estimated with the new modified Bayesian method and the traditional Bayesian method. Additionally, both methods give the same estimate ($\hat{\mu}_{\text{MBM}} = \hat{\mu}_{\text{TBM}}$=10) of the population mean ($\mu_T = 10$), regardless of the sample size.

Table 2. Bayesian Type A SU ($\widehat{\text{SU}}_{\text{MBM}}$ and $\widehat{\text{SU}}_{\text{TBM}}$) for the three datasets considered (Case 1)

Sample size n	New modified Bayesian method	Traditional Bayesian method with Jeffreys prior	Traditional Bayesian method with flat prior
4	0.250	0.374	0.603
10	0.158	0.174	0.188
20	0.112	0.117	0.120

Since these three datasets are drawn from the same normal distribution with $\mu_T = 10$ and $\sigma_T = 0.5$, the true SU is $\sigma_T/\sqrt{n}$= 0.250, 0.158, and 0.112 for n=4, 10, and 20, respectively. Notice from table 2 that the Type A SU estimated with the new modified Bayesian method is the same as the true SU. In contrast, the Type A SU estimated with the traditional Bayesian method, either with Jeffreys prior or flat prior, is positively biased with respect to the true SU. The bias is high when the sample size is small (e.g. n=4); it decreases with increasing the sample size.

5.2. Examples of Case 2

In Case 2, the prior information is represented by a normal PDF: $N(\mu|x_{\text{prior}}, \sigma_{\text{prior}})$. If the current information is represented by the true Bayesian distribution $N(\mu|\mu_T, \sigma_T/\sqrt{n})$, the true posterior distribution is also normal and can be obtained with the LAI (or LCD) or the least-squares method. Accordingly, the true posterior mean is

$$\mu_{\text{true}} = \frac{x_{\text{prior}}\sigma_T^2 + n\mu_T\sigma_{\text{prior}}^2}{\sigma_T^2 + n\sigma_{\text{prior}}^2} \tag{38}$$

and the true posterior SU is

$$\text{SU}_{\text{true}} = \frac{\sigma_{\text{prior}}\sigma_T}{\sqrt{\sigma_T^2 + n\sigma_{\text{prior}}^2}} \tag{39}$$

Again, we consider three datasets at n=4, 10 and 20, which are generated from the same normal distribution $N(\mu_T = 10, \sigma_T = 0.5)$. Each of the datasets gives $\bar{x}_D = \mu_T = 10$ and $\frac{s_D}{c_4} = \sigma_T = 0.5$.

For these three datasets, the posterior mean $\hat{\mu}_{\text{MBM}}$ and posterior SU $\widehat{\text{SU}}_{\text{MBM}}$ estimated with the new modified Bayesian method, Eq. (32) and Eq. (33), coincide with μ_{true} and SU_{true}, Eq. (38) and Eq. (39), respectively.

On the other hand, according to the traditional Bayesian method, the joint posterior PDF of μ and σ is written as

$$p_{\text{post}}(\mu, \sigma|\text{prior}, \text{data}) \propto N(\mu|x_{\text{prior}}, \sigma_{\text{prior}})\sigma^{-n}\exp\left(-\frac{n}{2}\frac{(\bar{x}_D-\mu)^2}{\sigma^2}\right)\exp\left(-\frac{1}{2}(n-1)\frac{s_D^2}{\sigma^2}\right) \tag{40}$$

Equation (40) does not lead to an analytical solution of the marginal posterior PDF of μ. We employed a numerical method, known as probability domain simulation (PDS) developed by Huang and Fergen (1995), to solve Eq. (40).

We consider two sceneries in calculating the marginal posterior PDF of μ using the PDS. In the first scenery, we assume that $x_{\text{prior}} = 10$ (fixed) and σ_{prior} varies from 0.1 to 1.0. The PDS solution gives $\hat{\mu}_{\text{TBM}} = 10$, the same as μ_{true}, regardless of the sample size. The PDS solution gives $\widehat{\text{SU}}_{\text{TBM}}$ that is biased with respect to SU_{true}. The bias depends on the value of σ_{prior}. Figure 2 shows the relative bias of $\widehat{\text{SU}}_{\text{TBM}}$ as a function of σ_{prior}.

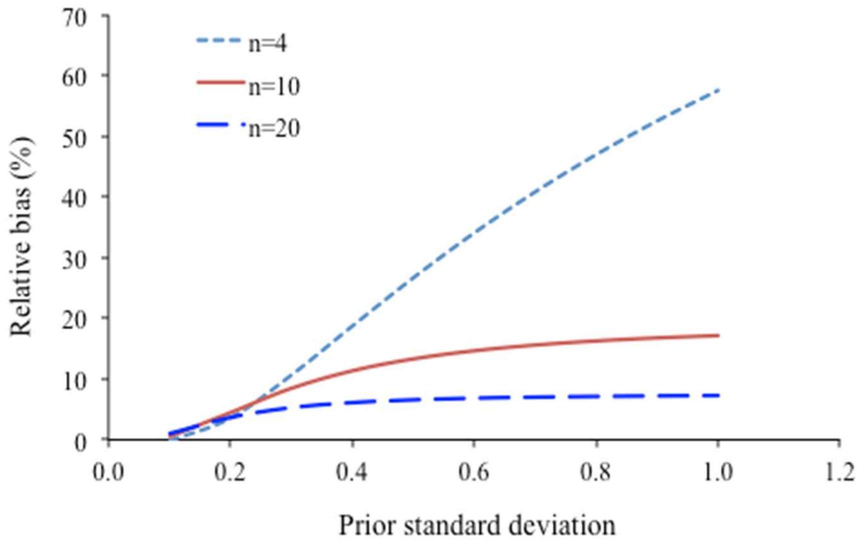

Figure 2. The relative bias of $\widehat{\mathrm{SU}}_{\mathrm{TB}}$ as a function of the prior standard deviation for Case 2 with the datasets at *n*=4, 10, and 20 (the first scenery: $x_{\mathrm{prior}} = 10$)

In the second scenery, we assume that $\sigma_{\mathrm{prior}} = 0.5$ (fixed) and x_{prior}varies from 9 to 11. The PDS solution gives $\hat{\mu}_{\mathrm{TBM}}$ that is biased with respect to μ_{true} and $\widehat{\mathrm{SU}}_{\mathrm{TBM}}$ that is biased with respect to $\mathrm{SU}_{\mathrm{true}}$. Both biases depend on the value of x_{prior}. Figure 3 and figure 4 show the relative bias of $\hat{\mu}_{\mathrm{TBM}}$ and $\widehat{\mathrm{SU}}_{\mathrm{TBM}}$, respectively, as a function of x_{prior}.

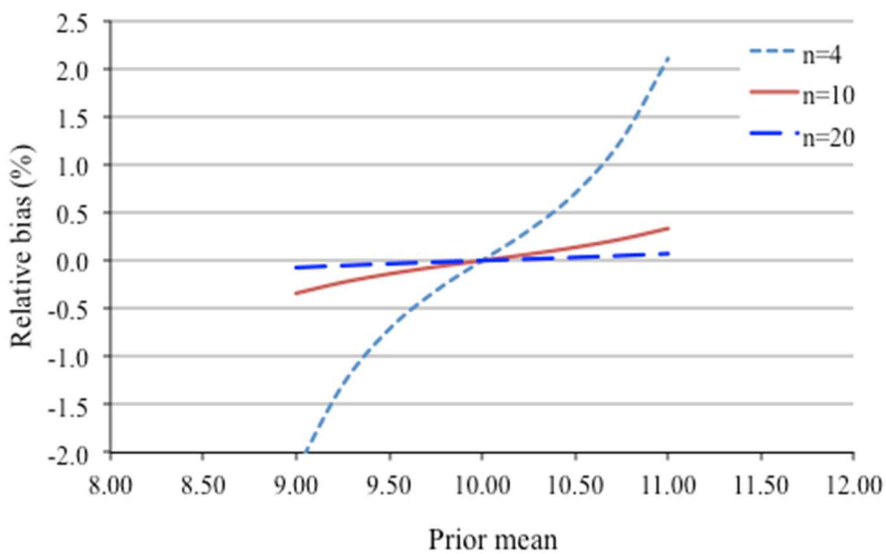

Figure 3. The relative bias of $\hat{\mu}_{\mathrm{TBM}}$ as a function of the prior mean for Case 2 with the datasets at *n*=4, 10, and 20 (the second scenery: $\sigma_{\mathrm{prior}} = 0.5$)

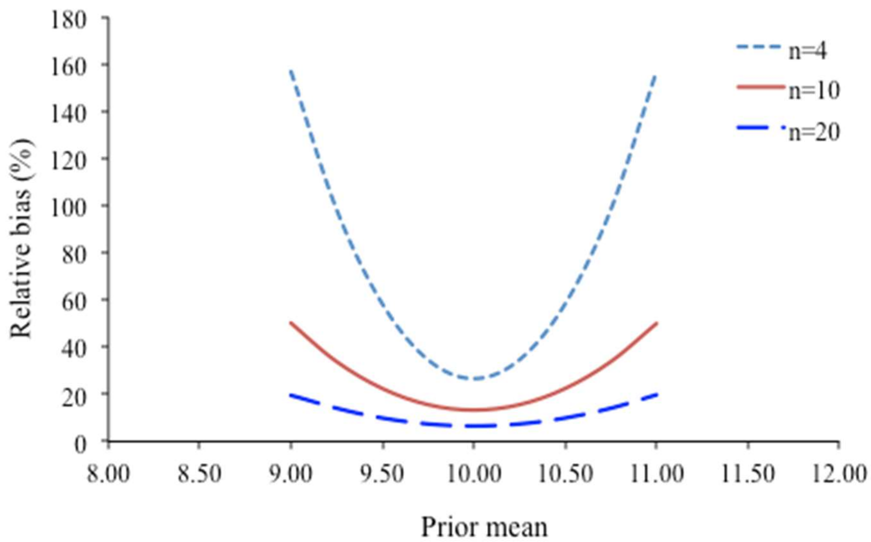

Figure 4. The relative bias of $\widehat{\mathrm{SU}}_{\mathrm{TBM}}$ as a function of the prior mean for Case 2 with the datasets at *n*=4, 10, and 20 (the second scenery: $\sigma_{\mathrm{prior}} = 0.5$)

It is interesting to note from figure 3 that the relative bias of $\hat{\mu}_{\mathrm{TBM}}$ is antisymmetric about $x_{\mathrm{prior}} = 10$. It is also interesting to note from figure 4 that the relative bias of $\widehat{\mathrm{SU}}_{\mathrm{TBM}}$ is symmetric about $x_{\mathrm{prior}} = 10$.

The results of these examples indicate that the traditional Bayesian method produce significant biases with respect to the true values in inferring SU in both Case 1 and Case 2. Thus, the traditional Bayesian method is essentially invalid as far as the unbiasedness criterion is concerned. In contrast, the new modified Bayesian method produces the unbiased estimates of SU in both Case 1 and Case 2, the same as its frequentist counterparts.

6. Discussion

6.1. The new modified Bayesian method is a self-consistent operation

One of the important axioms of addition states, “Only numerical measures of magnitudes of the same nature when expressed in terms of a common unit of measure can be added” (Clark 1902). We designate this statement as “the principle of self-consistent operation”. This principle is self-evidently true because it does not make sense to add the quantities of different nature together.

The new modified Bayesian method is a self-consistent operation because it operates entirely on PDFs and the negative of the log-probability is the Shannon information that has a common unit, e.g. “bit” based on the binary logarithm. Equation (15) can be rewritten as

$$\text{posterior information} = \text{prior information} + \text{current information} \tag{41}$$

We ignore the constant I_0 in Eq. (41) because it does not affect the discussion in this section.

It is interesting to note that Eq. (41) can be regarded as a counterpart of the Method of Support proposed by Edwards (1992, p36)

$$\text{posterior support} = \text{prior support} + \text{experimental support} \tag{42}$$

where "support" stands for "support function", which is definded as "the naural logarithm of the likelihood function" (Edwards 1992, p.32). In symbols, Eq. (42) is written as

$$\ln\left[L_{\text{post}}(\theta|\text{prior}, \text{data})\right] = \ln[L_{\text{prior}}(\theta)] + \ln[L(\theta|\text{data})] \tag{43}$$

Edwards' Method of Support is a self-consistent operation because it operates entirely on likelihoods, i.e. the two quantities of the right side of Eq. (42) or Eq. (43) have the same nature. Edwards' Method of Support reduces to the method of maximum likelihood in the case of no genuine prior information by using a uniform prior support function (Edwards 1992).

It is important to note that the support function is not the Shannon information because likelihood function is not probability distribution. Fisher (1921) stated, "… probability and likelihood are quantities of an entirely different nature." Edwards (1992) stated, "… this [likelihood] function in no sense gives rise to a statistical distribution." In fact, the likelihood function supplies a nature order of preference among the possibilities under consideration (Edwards 1992). Consequently, the mode of a likelihood function corresponds to the most preferred parameter value for the given data. Thus, Edwards' Method of Support or the method of maximum likelihood is a likelihood-based inference procedure that utilizes the mode only for point estimation of unknown parameters; it does not utilize the entire curve of likelihood functions. In contrast, a probability-based inference, either frequentist or Bayesian, usually requires the use of the entire curve of probability distributions for inference.

6.2. The traditional Bayesian method is not a self-consistent operation

Using the log-transformation, the traditional Bayes Theorem, Eq. (35), can be rewritten as

$$\ln\left[p_{\text{post}}(\theta|\text{prior}, \text{data})\right] = \ln[p_{\text{prior}}(\theta)] + \ln[L(\theta|\text{data})] \tag{44}$$

That is,

$$\text{posterior information} = \text{prior information} + \text{experimental support} \tag{45}$$

Recall that log-probability is the Shannon information that can be regarded as a physical quantity, whereas log-likelihood is not the Shannon information (although it might be regarded as an another physical quantity). That is, prior information (log-probability) and experiment support (log-likelihood) are the quantities of an entirely different nature; they should not be added together according to the principle of self-consistent operation. Therefore, the traditional Bayes Theorem violates the self-consistent operation principle. However, a Gaussian distribution model with known variance is an exception because the likelihood function of μ has the same mathematical form as the current PDF of μ (both are Gaussian).

In fact, Eq. (35) is known as the reformulated Bayes Theorem by some authors, e.g. Lira and Kyriazis (1999), because it is not the original Bayes Theorem. The original Bayes Theorem in continuous form, which simply follows the axioms of conditional probability (distribution), is written as, e.g. formula (1.2.4) in Box and Tiao (1992, p10) (in their notations)

$$p(\boldsymbol{\theta}|\mathbf{y}) = c \cdot p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}) \tag{46}$$

where $\mathbf{y}$ is the vector of n observations, $p(\mathbf{y}|\boldsymbol{\theta})$ is the (frequentist) probability distribution of $\mathbf{y}$, $\boldsymbol{\theta}$ is the parameter vector, and c is the normalizing constant to ensure $p(\boldsymbol{\theta}|\mathbf{y})$ integrates to one. The statement of Sober (2001) about the Bayes Theorem in discrete form, Eq. (34), also applies to Eq. (46). That is, Eq. (46) is true if each probability distribution mentioned in it is well defined. Note that the original Bayes Theorem in continuous form, Eq. (46), is a self-consistent operation because it operates entirely on probability distributions. So is the Bayes Theorem in discrete form, Eq. (34), which operates entirely on probabilities.

However, the original Bayes Theorem, Eq. (46), is not operational unless $p(\mathbf{y}|\boldsymbol{\theta})$ can be converted to a function of $\boldsymbol{\theta}$, instead of a function of $\mathbf{y}$. According to Box and Tiao (1992), "Now given the data $\mathbf{y}$, $p(\mathbf{y}|\boldsymbol{\theta})$ in (1.2.4) may be regarded as a function not of $\mathbf{y}$ but of $\boldsymbol{\theta}$. When so regarded, following Fisher (1922), it is called the likelihood function of $\boldsymbol{\theta}$ for given $\mathbf{y}$ and can be written $L(\boldsymbol{\theta}|\mathbf{y})$." Consequently, $p(\mathbf{y}|\boldsymbol{\theta})$is replaced with the likelihood function $L(\boldsymbol{\theta}|\mathbf{y})$, leading to the traditional (or reformulated) Bayes Theorem, e.g. formula (1.2.5) in Box and Tiao (1992, p10)

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto L(\boldsymbol{\theta}|\mathbf{y})p(\boldsymbol{\theta}) \tag{47}$$

To the author's knowledge, the above is a formal derivation of the traditional (i.e. reformulated) Bayes Theorem, Eq. (47), from the original Bayes Theorem, Eq. (46). However, this derivation is far from rigorous because it simply assumes that $p(\mathbf{y}|\boldsymbol{\theta})$ is equivalent to $L(\boldsymbol{\theta}|\mathbf{y})$ without a rigorous justification. In fact, this derivation is faulty because likelihood function is not probability distribution as emphasized by Fisher (1921) and Edwards (1992). A likelihood function is actually a distorted mirror of its probability distribution counterpart. Appendix shows an example of the distortion. Therefore, it is wrong to replace $p(\mathbf{y}|\boldsymbol{\theta})$ with $L(\boldsymbol{\theta}|\mathbf{y})$, and Eq. (47) is methodologically flawed. Due to this flaw, the traditional Bayesian method may lead to incorrect inferences (e.g. significant biases) as shown in section 5.

In summary, the original Bayes Theorem, Eq. (46), is valid, upon which the Bayesian statistics should be built. The reformulated Bayes Theorem, Eq. (47), deviates markedly from the original Bayes Theorem, so it is invalid. It is wrong to replace the probability distribution $p(\mathbf{y}|\boldsymbol{\theta})$ with the likelihood function $L(\boldsymbol{\theta}|\mathbf{y})$. However, as the example in Appendix shows, the discrepancy between the likelihood function $L(\boldsymbol{\theta}|\mathbf{y})$ and its probability distribution counterpart $p(\mathbf{y}|\boldsymbol{\theta})$ is serious only when the sample size is small (say n<5). $L(\boldsymbol{\theta}|\mathbf{y})$ will approach $p(\mathbf{y}|\boldsymbol{\theta})$ when the sample size is large enough (say n>30). Perhaps this is the reason why this flaw has been overlooked for more than 250 years. This flaw is the root cause of the inherent bias of the traditional Bayesian method; it is corrected in the new modified Bayesian method.

6.3. Deriving the new modified Bayesian method from the original Bayes Theorem

The new modified Bayesian method derived in section 3 can also be derived directly from the original Bayes Theorem, Eq. (46). It is important to note that, in $p(\mathbf{y}|\boldsymbol{\theta})$, $\mathbf{y}$ is a random vector and $\boldsymbol{\theta}$ are the values of k parameters θ_1, $\theta_2 \ldots \theta_k$, which are constants. That is, $\mathbf{y} \sim p(\mathbf{y}|\boldsymbol{\theta})$ depends on the values of the parameters $\boldsymbol{\theta}$ (Box and Tiao 1992); it is the sampling distribution of $\mathbf{y}$ (Loredo 1990). On the other hand, in $p(\boldsymbol{\theta}|\mathbf{y})$ (also in $p(\boldsymbol{\theta})$), $\boldsymbol{\theta}$ is a random vector and $\mathbf{y}$ are the observed values (constants). That is, $\boldsymbol{\theta} \sim p(\boldsymbol{\theta}|\mathbf{y})$ depends on the observed values $\mathbf{y}$; it is the distribution of the parameter vector $\boldsymbol{\theta}$. Apparently, the notations used by Box and Tiao (1992) in Eq. (46) and Eq. (47) are confusing because the same symbol $\boldsymbol{\theta}$ (or $\mathbf{y}$) represents the quantities with different natures: random and constant. In contrast, the notations used in this paper are clear about which quantity is a random variable, which is a parameter (constant), and which is the observed value (constant). In our notations, the original Bayes Theorem, Eq. (46), can be rewritten as

$$p_{post}\left(\boldsymbol{\theta}\middle|\text{prior}, \widehat{\boldsymbol{\theta}}_D\right) = c \cdot p_{\text{current}}(\widehat{\boldsymbol{\theta}}|\boldsymbol{\theta}_T) p_{\text{prior}}(\boldsymbol{\theta}) \tag{48}$$

where $\widehat{\boldsymbol{\theta}}_D$ are the observed values (i.e. the realizations of the sample statistics $\widehat{\boldsymbol{\theta}}$) and $\boldsymbol{\theta}_T$ are the true values of the parameters of the probability distributions of $\widehat{\boldsymbol{\theta}}$; both $\widehat{\boldsymbol{\theta}}_D$ and $\boldsymbol{\theta}_T$ are constants. Note that $p_{\text{current}}(\widehat{\boldsymbol{\theta}}|\boldsymbol{\theta}_T)$ is the true sampling distribution when $\boldsymbol{\theta}_T$ are known. In this situation, Eq. (48) will lead to the true posterior distribution as demonstrated in the examples of Case 2 in section 5. Since $\boldsymbol{\theta}_T$ are usually unknown, we replace them with their estimates $\widehat{\boldsymbol{\theta}}_D$ based on the point estimation method. Thus, $p_{\text{current}}(\widehat{\boldsymbol{\theta}}|\boldsymbol{\theta}_T) \approx p_{\text{current}}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}_D)$. Note that $p_{\text{current}}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}_D)$ is the estimated probability distribution of $\widehat{\boldsymbol{\theta}}$. According to the proposed frequentist-Bayesian transformation rule, Eq. (10), we substitute $\widehat{\boldsymbol{\theta}}$ with $\boldsymbol{\theta}$ in $p_{\text{current}}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}_D)$. Thus, Eq. (48) becomes

$$p_{post}\left(\boldsymbol{\theta}\middle|\text{prior}, \widehat{\boldsymbol{\theta}}_D\right) = c \cdot p_{\text{current}}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}_D) p_{\text{prior}}(\boldsymbol{\theta}) \tag{49}$$

Therefore, the new modified Bayes Theorem, Eq. (49), is an approximation of the original Bayes Theorem, Eq. (46).

For the problem considered with two unknown parameters, $\boldsymbol{\theta}=(\mu, \sigma)$ and $\widehat{\boldsymbol{\theta}}_D=(\bar{x}_D, \frac{s_D}{c_4})$, Eq. (49) reduces to

$$p_{\text{post}}\left(\mu, \sigma\middle|\text{prior}, \left(\bar{x}_D, \frac{s_D}{c_4}\right)\right) = c \cdot p_{\text{current}}(\mu, \sigma|\bar{x}_D, \frac{s_D}{c_4}) \cdot p_{\text{prior}}(\mu, \sigma) \tag{50}$$

which is essentially identical to Eq. (26).

6.4. In the absence of genuine prior information

If there is no genuine prior information about the unknown parameter θ, the prior information content of θ is zero, i.e. $I_{\text{prior}}(\theta)=0$. The posterior information of θ, Eq. (15), reduces to

$$I_{\text{post}}(\theta) = I_{\text{current}}(\theta) \tag{51}$$

Accordingly, the posterior PDF of θ is the same as the current PDF of θ. That is, Eq. (16) reduces to

$$p_{\text{post}}(\theta) = p_{\text{current}}(\theta) \tag{52}$$

Equation (52) can also be obtained by using a flat prior in Eq. (16). A flat prior is essentially equivalent to and is known as the locally uniform prior (Box and Tiao 1992). According to Jaynes' maximum entropy principle, maximum entropy is a way to assign a prior probability distribution (Jaynes 1988). It is well known that a uniform distribution is the least informative distribution, for which entropy is maximized. Therefore, a flat prior should be adopted in the absence of genuine prior information.

Moreover, Eq. (51) or (52) is consistent with the common sense that, in the case that there is no genuine prior information, the statistical inference (e.g. measurement uncertainty analysis) should rely on current information (data) itself. In other words, we would like the data to speak for themselves and we should believe in the light of the data. Thus, in philosophy and methodology, the new modified Bayesian method does not require any noninformative priors, and even flat priors are considered redundant. This completely eliminates the debate about how to choose noninformative priors when using the traditional Bayesian method.

In contrast, in the traditional Bayesian method, a prior is required, even if there is no genuine prior information. If a flat prior is used, Eq. (35) becomes:

$$p_{\text{post}}(\theta|\text{prior}, \text{data}) = \text{standardized } L(\theta|\text{data}) \tag{53}$$

However, Eq. (53) is wrong because the likelihood function $L(\theta|\text{data})$ is not a probability distribution as addrssed by Fisher (1921) and Edwards (1992). Thus, not only Eq. (53) conflicts with the common sense, but also it is methodologically flawed. In fact, the choice of a flat prior has long been, and still is, a matter of dispute in Bayesian statistics (Box and Tiao 1992). Bayesians often use improper noninformative priors such as the Jeffreys priors. However, the validity of the Jeffreys priors has been an argument even among Bayesians. D'Agostini (1998), a leading proponent of Bayesian methods in particle physics, argued "…it is rarely the case that in physical situations the status of prior knowledge is equivalent to that expressed by the Jeffreys priors, …". The use of the Jeffreys prior $1/\sigma$ in the traditional Bayesian method results in the scaled and shifted t-distribution that gives the invalid Bayesian Type A SU as discussed in the introduction section. Moreover, Huang (2018c, 2018d) revealed that the t-based inference for measurement uncertainty is invalid because of the "t-transformation distortion"; the t-based interval, whether it is derived from the Bayesian method or from the frequentist method, is actually misused in uncertainty estimation.

6.5. On the probability domain simulation (PDS)

The PDS naturally operates with the discretized formula of the new modified Bayesian method (or the traditional Bayesian method). When both μ and σ are unknown, the discretized formula is written as

$$p_{\text{post}}(\mu_i, \sigma_j) = \frac{p_{\text{prior}}(\mu_i,\sigma_j)\cdot p_{\text{current}}(\mu_i,\sigma_j)}{\sum_{i=1}^{m}\sum_{j=1}^{n} p_{\text{prior}}(\mu_i,\sigma_j)\cdot p_{\text{current}}(\mu_i,\sigma_j)} \tag{54}$$

where m is the number of the intervals for μ and n is the number of the intervals for σ. The size and number of the intervals for a parameter should be selected to cover a large range of the parameter with high precision. For example, if $\bar{x}_D = 10$ and $s_D/\sqrt{n} = 2$, we may use 100 intervals with the size 0.2, leading to the μ range from 0 to 20.

Equation (54) is a two-dimensional PDS; it can be easily implemented with Excel spreadsheets. It outputs the joint posterior PDF of μ and σ in the form of an $m \times n$ matrix, from which the marginal posterior PDF of μ or σ can be readily obtained.

The PDS may be an effective alternative to the Markov Chain Monte Carlo (MCMC) sampling. The MCMC sampling is used with great generality to produce samples from Bayesian posterior distributions (Robert and Casella 2011). However, the MCMC method requires considerable familiarity with specialized tools for statistical computing (Possolo 2015, Possolo and Bodnar 2018). Additionally, the MCMC method in general associates with computational difficulty and lack of transparency. Further study of comparing the PDS with the MCMC sampling is needed to evaluate the potential of the PDS as a general numerical procedure for Bayesian methods.

6.6. The unification of frequentist and Bayesian inference

Huang (2020b) presented a formula for combining prior information with current measurement based on the frequentist sampling theory and the LCD

$$p_{\text{post}}(x) = \frac{p_{\text{prior}}(x)\cdot p_{\text{current}}(x)}{\int p_{\text{prior}}(x)\cdot p_{\text{current}}(x)dx} \tag{55}$$

where x represents a random sample drawn from the sampling distributions: prior, current, and posterior PDF, and $p_{\text{current}}(x)$ is the *estimated* sampling distribution of x, i.e. $p_{\text{current}}(x) = p_{\text{current}}(x|\hat{\theta}_D)$. Note that Eq. (55) is identical to Eq. (18) if we change the symbol x to $\hat{\theta}$.

In fact, Eq. (55) is mathematically equivalent to the new modified Bayesian method, Eq. (16), because x can be replaced with θ. Thus, both the frequentist and Bayesian views and methodologies end up with the same formula for combining prior information with current measurement.

Equation (55) leads to a frequentist solution to Case 2 (Huang 2020b)

$$p_{\text{post}}(x|\text{prior, data}) \propto N(x|x_{\text{prior}}, \sigma_{\text{prior}}) \cdot N(x|\bar{x}_D, \frac{s_D}{c_4\sqrt{n}}) \tag{56}$$

which is identical to the solution of the new modified Bayesian method, Eq. (31), if x is replaced with μ.

Therefore, in the light of the frequentist-Bayesian transformation rule and the LAI (or LCD), the frequentist and Bayesian inference are virtually equivalent so they can be unified, at least in measurement uncertainty analysis. The unification may resolve the long-standing debate between frequentists and Bayesians. The unification may also shed light on the revision of the GUM, which may leads to an indisputable, uniform revision of the GUM.

Perhaps the unification is a reunion because the original Bayes Theorem, Eq. (46), which is merely a statement of conditional probability (Box and Tiao 1992), is based on both the frequentist view and Bayesian view. Refer to Eq. (46), where $\boldsymbol{\theta}$ in the conditional probability distribution $p(\boldsymbol{\theta}|\mathbf{y})$ (also in $p(\boldsymbol{\theta})$) are the unknown parameters that are treated as random variables based on the Bayesian view, whereas $\boldsymbol{\theta}$ in the conditional probability distribution $p(\mathbf{y}|\boldsymbol{\theta})$ are the true values (constants) of the parameters based on the frequentist view. Recall that $p(\mathbf{y}|\boldsymbol{\theta})$ is actually the probability distribution of $\mathbf{y}$ according to the frequentist sampling theory (Loredo 1990). Thus, the original Bayes Theorem in fact tells us that the posterior distribution of unknown parameters $\boldsymbol{\theta}$ is proportional to the product of the (frequentist) probability distribution of $\mathbf{y}$ and the (Bayesian) prior distribution of $\boldsymbol{\theta}$. Therefore, the original Bayes Theorem may have suggested the equivalence of or the transferability between the frequentist view and Bayesian view. Indeed, we have demonstrated in section 2 that, the frequentist view can be transformed to the Bayesian view and vice versa through the proposed frequentist-Bayesian transformation rule. The frequentist-Bayesian transformation rule can make the original Bayes Theorem, Eq. (46), operational, either based on the "pure" Bayesian view, which leads to Eq. (49), or based on the "pure" frequentist view, which leads to (in our notations)

$$p_{\text{post}}(\hat{\boldsymbol{\theta}}|\text{prior}, \hat{\boldsymbol{\theta}}_D) = c \cdot p_{\text{current}}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}_D) p_{\text{prior}}(\hat{\boldsymbol{\theta}}) \tag{57}$$

In the author's opinion, the separation of the Bayesian inference and frequentist inference is due to the error of using the likelihood function $L(\boldsymbol{\theta}|\mathbf{y})$ as a substitute for the conditional probability distribution $p(\mathbf{y}|\boldsymbol{\theta})$. The left side of the reformulated Bayes Theorem, Eq. (47), is a mixture of likelihood and probability. Likelihood is useful in its own way as in Edwards' Method of Support or in the MLE; it should not be mixed with probability. The new modified Bayesian method corrects this error so that the frequentist and Bayesian inference may reunite after being separated for more than 250 years. Further investigation on the unification and its impact is warranted.

7. Conclusion

We have demonstrated that, for the simple (and basic) measurement model considered, the frequentist view can be transformed to the Bayesian view and vice versa through the proposed frequentist-Bayesian transformation rule. This is because the true value and the measurement error are physical quantities that are independent of the viewpoint (or reference frame). In the case that there is no genuine prior information, the usual Bayesian PDF of an unknown parameter with a given dataset is nothing but the estimated probability distribution (i.e. frequentist PDF) in which the unknown parameters are estimated with the point estimation method.

The proposed new modified Bayesian method is a self-consistent operation because it operates entirely on PDFs. As a result, it gives the correct inferences for the problem considered (Case 1 and Case 2): same solutions as its frequentist counterparts. In contrast, the traditional Bayesian method, i.e. the reformulated Bayes Theorem, is not a self-consistent operation because it operates on likelihood function and PDF. This is a flaw of the traditional Bayesian method. As a result, the traditional Bayesian method gives the incorrect inferences: invalid estimates of standard uncertainty (SU) in Case 1 and Case 2. A likelihood function is a distorted mirror of its probability distribution counterpart. The use of likelihood functions in Bayes Theorem is the root cause of the inherent bias of the traditional Bayesian method. However, the original Bayes Theorem, either in continuous or discrete form, is a self-consistent operation because it operates entirely on probability distributions or probabilities.

In the light of the frequentist-Bayesian transformation rule and the LAI (or LCD), the frequentist and Bayesian inference are virtually equivalent so that they can be unified, at least in measurement uncertainty analysis. The unification may resolve the long-standing debate between frequentists and Bayesians. The unification may also shed light on the revision of the GUM, which may lead to an indisputable, uniform revision of the GUM. Further investigation on the unification and its impact is warranted.

The proposed new modified Bayesian method is a special case of the LAI where only two information sources (prior and current) are involved. It is not limited to the measurement science because the LAI is a universal law. Further studies are warranted to extend the new modified Bayesian method to different fields of science.

It is the author's hope that this paper will draw practitioners' and statisticians' attention to the limitation (or flaw) of the traditional Bayesian method. The traditional (or reformulated) Bayes Theorem, Eq. (47), deviates markedly from the original Bayes Theorem, Eq. (46). Bayesian statistics should not be built upon the reformulated Bayes Theorem; it should be built upon the original Bayes Theorem, Eq. (46), or upon the new modified Bayes Theorem, Eq. (48), which is a valid approximation of the original Bayes Theorem.

Appendix: A likelihood function is a distorted mirror of its probability distribution counterpart: an example

Consider a normal distribution for which the mean μ_T is known. The likelihood of σ, given n observations, can be written as (Box and Tiao 1992) (in our notations)

$$L(\sigma|s_D) \propto \sigma^{-n}\exp\left(-\frac{1}{2}(n-1)\frac{s_D^2}{\sigma^2}\right) \tag{56}$$

The standardized likelihood of σ can be written as

$$L'(\sigma|s_D) = \frac{\sigma^{-n}\exp\left(-\frac{1}{2}(n-1)\frac{s_D^2}{\sigma^2}\right)}{\int_0^\infty \sigma^{-n}\exp\left(-\frac{1}{2}(n-1)\frac{s_D^2}{\sigma^2}\right)d\sigma} \tag{57}$$

provided that the integral is finite.

On the other hand, the probability distribution counterpart of $L'(\sigma|\text{data})$ is the sampling distribution of s (WolfamMathworld 2020) that can be rewritten as

$$p(s|\sigma_T) = \frac{s^{n-2}\exp\left(-\frac{1}{2}(n-1)\frac{s^2}{\sigma_T^2}\right)}{\int_0^\infty s^{n-2}\exp\left(-\frac{1}{2}(n-1)\frac{s^2}{\sigma_T^2}\right)ds} \tag{58}$$

Note that $L'(\sigma|s_D) \neq p(s|\sigma_T)$. In order to visualize the difference between $L'(\sigma|s_D)$ and $p(s|\sigma_T)$, we calculated their numerical values by assuming that $s_D/c_4 = \sigma_T = 1$ for n=2, 4, and 10. Figure 5, 6, and 7 show the comparison between $L'(\sigma|s_D)$ and $p(s|\sigma_T)$ at n=2, 4, and 10 respectively.

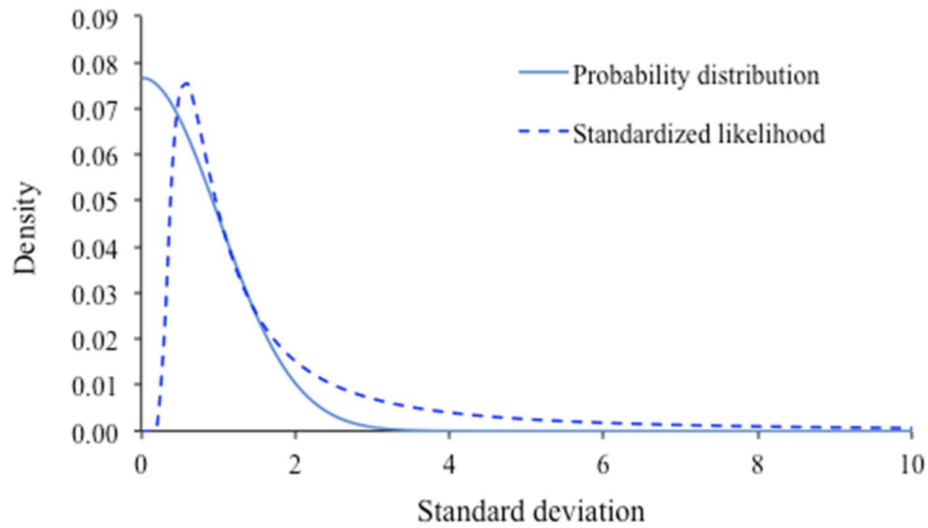

Figure 5. Comparison between the standardized likelihood function of σ and its probability distribution counterpart (n=2)

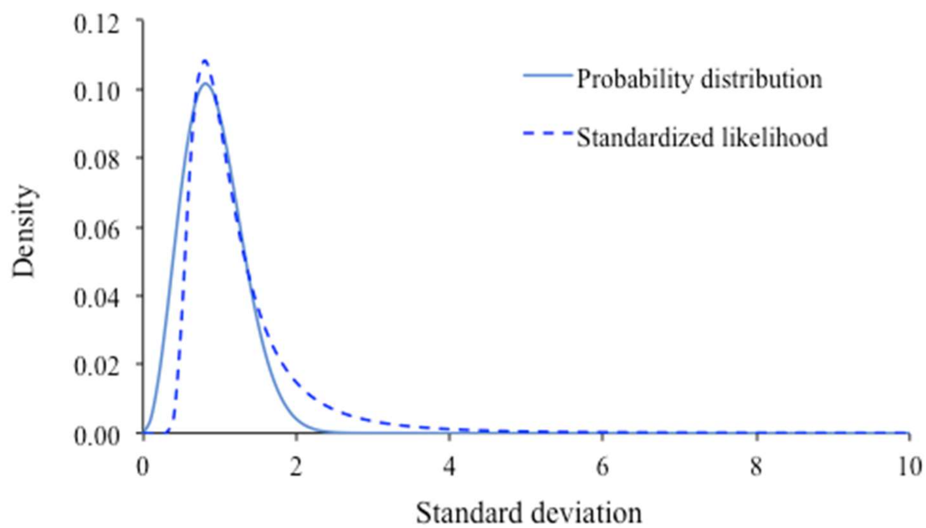

Figure 6. Comparison between the standardized likelihood function of σ and its probability distribution counterpart (n=4)

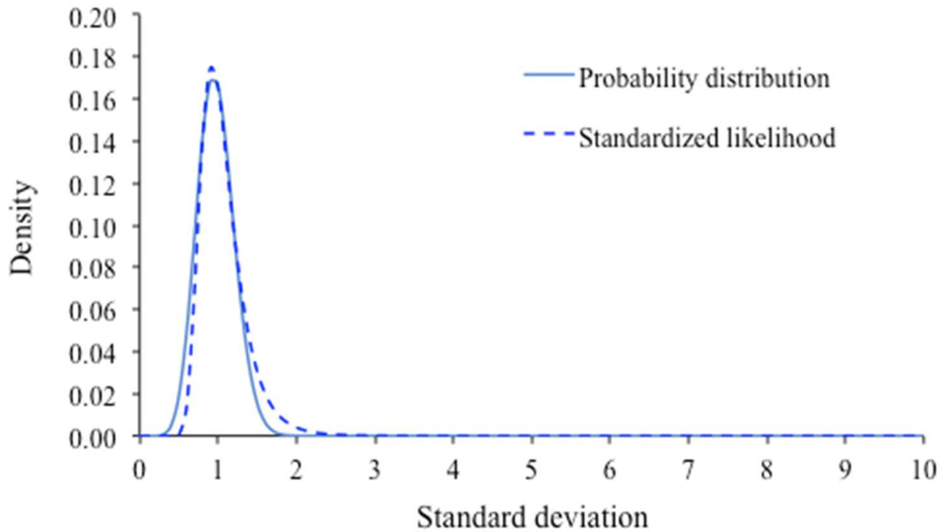

Figure 7. Comparison between the standardized likelihood function of σ and its probability distribution counterpart (n=10)

At n=2, the $L'(\sigma|s_D)$ curve and the $p(s|\sigma_T)$ curve are significantly different. The $L'(\sigma|s_D)$ curve has a fatter and longer tail, which means that there are many occurrences of σ far from the central part of the standardized likelihood function. At n=4, the pattern of the $L'(\sigma|s_D)$ and $p(s|\sigma_T)$ curves are similar, but the $L'(\sigma|s_D)$ curve has a fatter and longer tail. At n=10, the $L'(\sigma|s_D)$ and $p(s|\sigma_T)$ curves are closer.

Therefore, the standardized likelihood function of σ is a distorted mirror of its probability distribution counterpart, the sampling distribution of s. The distortion is severe when the sample size is small; it decreases with increasing sample size. When the sample size is large enough, say n>30, $L'(\sigma|s_D)$ will approach $p(s|\sigma_T)$.

In addition, it should be mentioned that for the example (n=4) of Case 1 in section 5.1, the marginal posterior distribution of μ obtained with the traditional Bayesian method (TBM) with the flat prior is a standardized likelihood function of μ, whose probability distribution counterpart is the normal distribution. Then, it can be seen from figure 1 that this standardized likelihood function is distorted markedly with respect to the normal distribution. The distortion will decrease with increasing sample size, and this standardized likelihood function will approach the normal distribution when the sample size is large enough.

References

[1] Attivissimo F, Giaquinto N, and Savino M (2012) A Bayesian paradox and its impact on the GUM approach to uncertainty *Measurement* **45** 2194–2202 DOI:10.1016/j.measurement.2012.01.022

[2] Bich W (2014) Revision of the 'Guide to the Expression of Uncertainty in Measurement' Why and how *Metrologia* **51** S155

[3] Bich W, Cox M and Michotte C (2016) Towards a new GUM—an update *Metrologia* **53** S149–159

[4] Box G E P and Tiao G C (1992) *Bayesian Inference in Statistical Analysis* Wiley New York
[5] Clark L H 1902 *An Arithmetic for Review Classes in High Schools and Normal Schools* Benj. H. Sanborn & Co. Boston
[6] Clemen R T and Winkler R L (1999) Combining probability distributions from experts in risk analysis *Risk Analysis* **19** 187-203
[7] D'Agostini G (1998) Jeffreys priors versus experienced physicist priors: arguments against objective Bayesian theory *Proceedings of the 6th Valencia International Meeting on Bayesian Statistics* (Alcossebre, Spain, May 30th-June 4th)
[8] Demeyer S, Fischer N, and Elster C (2021) Guidance on Bayesian uncertainty evaluation for a class of GUM measurement models *Metrologia* **58** 014001
[9] Edwards, A W F (1992) *Likelihood* (expanded edition) Johns Hopkins University Press Baltimore
[10] Fisher R A (1921) On the 'Probable Error' of a coefficient of correlation deduced from a small sample *Metron* **I** part 4, 3-32
[11] Giaquinto N and Fabbiano L (2016) Examples of S1 coverage intervals with very good and very bad long-run success rate *Metrologia* **53** S65-S73
[12] Giaquinto N, Fabbiano L, Trotta A and Vacca G (2014) About the frequentist and the Bayesian approach to uncertainty *20th IMEKO TC4 International Symposium and 18th International Workshop on ADC Modelling and Testing Research on Electric and Electronic Measurement for the Economic Upturn* (Benevento, Italy, September 15-17)
[13] Huang H (2014) Uncertainty-based measurement quality control *Accreditation and Quality Assurance* **19** 65-73
[14] Huang H (2015) Statistical quality control of streamflow measurements with moving-boat acoustic Doppler current profilers *Journal of Hydraulic Research* **53** 820-827 https://doi.org/10.1080/00221686.2015.1074947
[15] Huang H (2018a) A unified theory of measurement errors and uncertainties *Measurement Science and Technology* **29** 125003 https://doi.org/10.1088/1361-6501/aae50f
[16] Huang H (2018b) More on the *t*-interval method and mean-unbiased estimator for measurement uncertainty estimation *Cal Lab the International Journal of Metrology* **25** 24-33 available from https://www.researchgate.net/publication/326331247_More_on_the_t-Interval_Method_and_Mean-Unbiased_Estimator_for_Measurement_Uncertainty_Estimation
[17] Huang H (2018c) Uncertainty estimation with a small number of measurements, Part I: new insights on the *t*-interval method and its limitations *Measurement Science and Technology* **29** https://doi.org/10.1088/1361-6501/aa96c7
[18] Huang H (2018d) Uncertainty estimation with a small number of measurements, Part II: a redefinition of uncertainty and an estimator method *Measurement Science and Technology* **29** https://doi.org/10.1088/1361-6501/aa96d8
[19] Huang H (2019a) Signal content index (SCI): a measure of the effectiveness of measurements and an alternative to *p*-value for comparing two means *Measurement Science and Technology* https://doi.org/10.1088/1361-6501/ab46fd
[20] Huang H (2019b) On the Bayesian estimator of normal mean, discussion on ResearchGate https://www.researchgate.net/post/On_the_Bayesian_estimator_of_normal_mean

[21] Huang H (2019c) Why the scaled and shifted *t*-distribution should not be used in the Monte Carlo method for estimating measurement uncertainty? *Measurement* **136** 282-288 https://doi.org/10.1016/j.measurement.2018.12.089

[22] Huang (2020a) Comparison of three approaches for computing measurement uncertainties *Measurement* **163** https://doi.org/10.1016/j.measurement.2020.107923

[23] Huang H (2020b) Two simple and practical methods for combining prior information with current measurement in uncertainty analysis, accepted for publication in *Cal Lab the International Journal of Metrology* **27** 22-32 available from https://www.researchgate.net/publication/344502279_Two_simple_and_practical_methods_for_combining_prior_information_with_current_measurement_in_uncertainty_analysis

[24] Huang H and Fergen R E (1995) Probability-domain simulation - A new probabilistic method for water quality modeling. *WEF Specialty Conference "Toxic Substances in Water Environments: Assessment and Control"* (Cincinnati, Ohio, May 14-17, 1995) available from https://www.researchgate.net/publication/285589247_Probability-domain_simulation_-_A_new_probabilistic_method_for_water_quality_modeling

[25] International Organization of Standards (ISO) (2021) ISO:24578:2021(E), Hydrometry — Acoustic Doppler profiler — Method and application for measurement of flow in open channels from a moving boat, first edition, 2021-3

[26] Jaynes E T (1988) The Relation of Bayesian and Maximum Entropy Methods, in *Maximum-Entropy and Bayesian Methods in Science and Engineering (Vol. 1)*, Kluwer Academic Publishers, p. 25-29

[27] Joint Committee for Guides in Metrology (JCGM) (2008) *Evaluation of measurement data - Guide to the expression of uncertainty in measurement* (GUM 1995 with minor corrections), Sevres, France

[28] 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

[29] Kyriazis G A (2015) Contributions to the revision of the 'Guide to the Expression of Uncertainty in Measurement' *J. Phys. Conf. Ser*. 575 012039

[30] Landauer R (1961) Irreversibility and heat generation in the computing process *IBM Journal of Research and Development* **5** 183-191

[31] Lira I and Kyriazis G (1999) Bayesian inference from measurement information *Metrologia* **36** 163-169 DOI:10.1088/0026-1394/36/3/1

[32] Lira I (2016) The GUM revision: the Bayesian view toward the expression of measurement uncertainty *European Journal of Physics* **37** 1-16

[33] Loredo T (1990) From Laplace to Supernova SN 1987A: Bayesian Inference in Astrophysics *Maximum Entropy and Bayesian Methods* 81-142 Kluwer Academic Press, Dordrecht, The Netherlands

[34] Mayo D (2012) Dennis Lindley's "Philosophy of Statistics" available from https://errorstatistics.com/2012/07/12/dennis-lindleys-philosophy-of-statistics/

[35] Murphy K P (2007) Conjugate Bayesian analysis of the Gaussian distribution, Last updated October 3, 2007, available from https://www.cs.ubc.ca/~murphyk/Papers/bayesGauss.pdf

[36] Possolo A (2015) *Simple Guide for Evaluating and Expressing the Uncertainty of NIST Measurement Results* DOI:10.6028/NIST.TN.1900 NIST Technical Note 1900, National Institute of Standards and Technology

[37] Possolo A and Bodnar O (2018) Approximate Bayesian evaluations of measurement uncertainty *Metrologia* **55** 147–157 https://doi.org/10.1088/1681-7575/aaa5be

[38] Robert C and Casella G (2011) A short history of Markov Chain Monte Carlo: subjective recollections from incomplete data *Stat. Sci.* **26** 102–15
[39] Shannon C E (1948) A Mathematical Theory of Communications *The Bell System Technical Journal* **27** 379-423, 623-656
[40] Sober E (2001) Bayesianism - its scope and limits *Proceedings of the British Academy* **113** DOI: 10.5871/bacad/9780197263419.003.0002
[41] Stoudt S, Pintar A, and Possolo A (2021) Uncertainty Evaluations from Small Datasets accepted by *Metrologia*
[42] White D R (2016) In pursuit of a fit-for-purpose uncertainty guide *Metrologia* **53** S107–24
[43] Wile J L (2012) Information is a real, physical quantity *Proslogion* available from https://blog.drwile.com/information-is-a-real-physical-quantity/
[44] Willink R (2016) What can we learn from the GUM of 1995? *Measurement* **91** 692-698
[45] Willink R and White D R (2011) Disentangling classical and Bayesian approaches to uncertainty analysis, Measurement Standards Laboratory, PO Box 31310, Lower Hutt 5040, New Zealand
[46] WolfamMathworld (2020) Standard deviation distribution https://mathworld.wolfram.com/StandardDeviationDistribution.html
[47] Wubbeler G and Elster C (2020) On the transferability of the GUM-S1 type A uncertainty *Metrologia* **57** DOI: 10.1088/1681-7575/ab50d6
[48] Wubbeler G, Marschall M, and Elster C (2020) A simple method for Bayesian uncertainty evaluation in linear models *Metrologia* **57** DOI: 10.1088/1681-7575/aba3b8