Informative Bayesian type A evaluation of standard uncertainty

Carlo Carobbi
Department of Information Engineering, University of Florence, Firenze, ITALY

Abstract

The non-informative (improper priors) Bayesian type A evaluation of standard uncertainty consistent with the Supplement 1 and Supplement 2 of the Guide to the expression of Uncertainty in Measurement (GUM) leads to a result that cannot be applied in case of small sample size *n*, namely when $n = 2,3$. An ad hoc and non-rigorous solution to this issue, based on the use of the quantiles of the Student's *t* probability PDF, is provided in the IEC/TR 61000-1-6 (Guide to the assessment of measurement uncertainty), a basic electromagnetic compatibility (EMC) publication. A novel informative Bayesian type A evaluation of standard uncertainty is proposed here that can be easily implemented through a spreadsheet and it is applicable also in the case of small sample size. The aim and the approach are like those in a recent paper published in Metrologia by Cox and Shirono. The apparent limitation that the variance has an upper bound, as assumed by Cox and Shirono, is here removed at the expense of introducing a best guess and a value of the variance that is unlikely to (but it can) be exceeded. In addition to a step-by-step derivation numerical examples are offered to support the viability the proposed type A evaluation. The imminent start of the maintenance of the IEC/TR 61000-1-6 stimulates a critical review and update of the type A evaluation of standard uncertainty in EMC.

1 Introduction

To make the GUM [1] consistent with its Supplement 1 [2] and Supplement 2 [3], only state of knowledge PDFs should be assigned to the input quantities to the measurement model function [4]. This approach is implemented in the GUM type B evaluation of standard uncertainty but not in the type A one. In the GUM theoretical framework, the mean $\bar{y}$ and the standard deviation s of a sample of observation are random variables. In the GUM Supplement 1 and GUM Supplement 2, the same $\bar{y}$ and *s* are known parameters of the state of knowledge PDF of the unknown true value μ of an input quantity Y . In the case of normal probability model for that quantity a Student's *t* PDF is assigned to the true value, shifted by $\bar{y}$ and scaled by s^2/n and having $n-1$ degrees of freedom (*n* is the sample size). By assigning this state of knowledge PDF we have a remarkable inconvenient: the state of knowledge type A evaluation of standard uncertainty is given by

$$(\sigma_\mu)^* = \sqrt{\frac{n-1}{n-3}}\frac{s}{\sqrt{n}} \quad (1)$$

which does not exist when $n = 2,3$ and it is larger than the consolidated GUM type A evaluation given by $s/\sqrt{n}$. This approach was implemented in a first Committee Draft of the revised GUM that was circulated among national metrology institutes and member organizations in December 2014. The reactions to this document were largely negative [5] and (1) never became a standard. A solution to the issue of non-existence of the state of knowledge type A evaluation of standard uncertainty when the sample size is small ($n = 2,3$) was proposed in [6]. The solution is based on recognizing that, independently on the coverage probability *p*, the quantiles of the Student's *t* PDF have the following property for $n \geq 4$ [1, clause G.3.4]

$$\frac{t_p(n-1)}{t_p(\infty)} \approx \sqrt{\frac{n-1}{n-3}} . \quad (2)$$

Since the quantiles $t_p(n-1)$ exist for any *n*, then the state of knowledge type A evaluation in the case of small sample size can be obtained by multiplying $s/\sqrt{n}$ by $\eta(n) = t_p(n-1)/t_p(\infty) > 1$. It is assumed that $p = 0.95$ in order to make the state of knowledge and the GUM evaluations of expanded uncertainty equal when the type A contribution is dominant (the same coverage intervals are obtained). We have therefore $\eta(2) = 6.48$ and $\eta(3) = 2.19$. This is the approach implemented in two (EMC) standards (first in [7], then in [8], both published in 2012).

The use of the coefficient $\eta(n)$ to make wider the uncertain GUM evaluation of standard uncertainty, thus obtaining the "safe" state of knowledge evaluation of standard uncertainty, has merits. It is mathematically simple and its interpretation is immediate: $\eta(n)$ is the price to pay when few measurements are made, indeed the smaller is *n* the larger is $\eta(n)$. However it is also an ad hoc way out, relying on a non-rigorous reverse inductive reasoning (what is valid for $n \geq 4$ is unjustifiably applied also for $n = 2,3$) and on a conventional coverage probability $p = 0.95$ (the standard uncertainty depends on the coverage probability).

The state of knowledge Student's *t* PDF with $n-1$ degrees of freedom, shifted by $\bar{y}$ and scaled by s^2/n results from

Bayesian inference assuming Jeffrey's prior, i.e. no prior information. Any experienced experimenter has however some prior knowledge about the measured quantity and the measurement error (see [9] for a founded criticism of objective Bayesian inference). The consequence of neglecting this prior knowledge is a coverage interval which is much wider, and therefore not acceptable, than the one the experimenter would expect based on experience. If, for example, a deviation of 0.5 dB is found between the results of two repeated EMC conducted emission tests on the same device under test, then the corresponding 95 % coverage probability interval based on the objective and non-informative Student's *t* PDF would be ±6.35 dB, which is evidently an overestimate. It is therefore sensible that an amount of prior information should be accounted for when deriving the state of knowledge PDF through Bayesian inference. In [10] an original solution for the state of knowledge PDF is derived assuming a $1/\sigma^2$ prior PDF for σ^2 truncated at σ_M, the maximum value that σ can take. A closed-form expression is derived for the state of knowledge standard uncertainty which is valid for $n \geq 2$ and that can be numerically evaluated by using a spreadsheet.

A novel informative Bayesian type A evaluation of standard uncertainty is here derived along the same line of reasoning as in [10]. The scope, similarly to [10], is to obtain a result which is applicable in the case of normal probability model, small sample size and that can be easily implemented through a spreadsheet. Bayesian analysis is used to derive the searched result. Prior information is expressed in terms of a best estimate σ_0^2 and an *unlikely exceeded* value σ_M^2 of the variance σ^2 of the normal model. Such information is encoded through a conjugate prior distribution for σ^2, namely a scaled inverse chi-squared PDF. No prior information is instead attached to the mean μ and this is represented, in the Bayesian formulation, by a uniform improper prior PDF. Differently from [10] values of σ^2 larger than the maximum are permissible, i.e. σ_M^2 is not a truncation value for the prior of σ^2 but a value that can be exceeded although with low (but nonzero) probability.

In the next section 2 the result is derived. In section 3 some numerical examples are provided. This work is in the same vein of others of the same author and targeting measurement uncertainty in EMC, see [12-15].

2 Derivation

We set

$$\mu \mid \sigma^2 \sim \text{const.} \tag{3}$$

$$\sigma^2 \sim \text{Inv}-\chi^2\left(\nu_0, \sigma_0^2\right), \tag{4}$$

where ν_0 represents the number of degrees of freedom. The joint prior PDF of the parameters of the normal probability model is therefore

$$p\left(\mu, \sigma^2\right) \propto \left(\sigma^2\right)^{-(\nu_0/2+1)} \exp\left(-\frac{\nu_0 \sigma_0^2}{2\sigma^2}\right). \tag{5}$$

Multiplying (5) by the normal likelihood (observations are y_i, $i = 1, 2, \cdots, n$) yields the posterior PDF

$$p\left(\mu, \sigma^2 \mid y\right) \propto \left(\sigma^2\right)^{-(\nu_0/2+1)} \exp\left(-\frac{\nu_0 \sigma_0^2}{2\sigma^2}\right) \times \\ \times \left(\sigma^2\right)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\left[\left(n-1\right)s^2 + n\left(\bar{y}-\mu\right)^2\right]\right). \tag{6}$$

where $\bar{y}$ and s^2 represent the sample mean and the sample variance, respectively. Integration of (6) with respect to σ^2 gives the marginal posterior of μ

$$p\left(\mu \mid y\right) \propto \left(1 + \frac{n\left(\bar{y}-\mu\right)^2}{\left(n-1\right)s^2 + \nu_0 \sigma_0^2}\right)^{-\frac{n+\nu_0}{2}}. \tag{7}$$

Then from (7)

$$p\left(\mu \mid y\right) = t_{\nu_n}\left(\mu \mid \mu_n, \sigma_n^2/n\right), \tag{8}$$

where

$$\begin{aligned} \nu_n &= (n-1) + \nu_0 \\ \mu_n &= \bar{y} \\ \sigma_n^2 &= \frac{\left(n-1\right)s^2 + \nu_0 \sigma_0^2}{(n-1) + \nu_0} \end{aligned} \tag{9}$$

and $t_{\nu_n}\left(\mu \mid \mu_n, \sigma_n^2/n\right)$ is a Student's *t* PDF, shifted by μ_n and scaled by σ_n^2/n. A similar result was derived in [11]. The main difference here with respect to the derivation in [11] is that in [11] a normal prior PDF is assigned to μ with expected value μ_0 and variance σ^2/κ_0, while here we assume a constant improper prior for μ, i.e. the same assumption as in [10].

The variance of μ is therefore given by

$$\sigma_\mu^2 = \frac{\nu_n}{\nu_n - 2} \frac{\sigma_n^2}{n} \tag{10}$$

which represents here the searched type A evaluation of variance. Note that the well-known case of improper prior for σ^2 (or Jeffrey's prior) is simply obtained from (8), (9) and (10) assuming $\nu_0 = 0$. Further, σ_μ^2 exists provided that $\nu_n > 2$ or from the first of (9) $\nu_0 > 3 - n$. Since in a type A evaluation n is an integer and $n \geq 2$ then, for existence of σ_μ^2 for any value of n it should be $\nu_0 > 1$.

It is worth mentioning that the parameters σ_0^2 and ν_0 of the prior of σ^2, see (4), could be directly available as information related to measurement method validation, measurement method verification or from assurance of the validity of measurement results (e.g. through control charts). These are common forms of prior knowledge in those testing and metrological areas "supported by evidence", namely those areas operating under a quality control regime, such as in the case of EMC testing and calibration. When such information, which essentially relies on a sample of observations, is missing then it can be replaced by a different type of information, which is derived from documental evidence or the experience of the experimenter. In this case the parameter ν_0 can be replaced by a more convenient one, which is not directly linked to a sample size, as follows. Let us assume that we have reason to believe that σ^2 has a low probability α of being larger than σ_M^2. Then, by evaluating the cumulative distribution function of the scaled inverse chi-squared PDF at σ_M^2 we have

$$\frac{\Gamma\left(\frac{\nu_0}{2},\frac{\nu_0\sigma_0^2}{2\sigma_M^2}\right)}{\Gamma\left(\frac{\nu_0}{2}\right)}=1-\alpha, \qquad (11)$$

where $\Gamma(a,b)$ is the incomplete gamma function and $\Gamma(a)$ is the gamma function. For a given ratio σ_M/σ_0 (11) is verified by a corresponding value of ν_0 and vice versa. In other words, (11) defines an implicit relation between σ_M/σ_0 and ν_0. Note that ν_0 is not restricted to be an integer, it is a real positive number. This similarly applies to ν_n, which is real and larger than $n-1$.

3 Numerical application

Let us assume that the experimenter choses $\alpha = 0.05$ in (11). The values of ν_0 corresponding to given values of σ_M/σ_0 can be calculated solving the implicit equation (11). The result is in Table 1 for some values of σ_M/σ_0.

Table 1: Values of σ_M/σ_0 and corresponding values of ν_0 derived through (11).

σ_M/σ_0	ν_0	σ_M/σ_0	ν_0
1.5	12.4924	3.5	2.4600
2	5.4604	4	2.1724
2.5	3.6914	4.5	1.9700
3	2.9049	5	1.8191

Note that the larger is σ_M/σ_0 the wider is the prior (4), the poorer is the available prior information on σ^2 and the smaller is ν_0. Further note that also in the (very unlikely) case that the experimenter takes $\sigma_M = 5\sigma_0$ we have $\nu_0 = 1.8191 > 1$ and σ_μ^2 exists, also in the worst case $n = 2$. More directly: even in the case of almost total ignorance on σ^2 we have a decent number of degrees of freedom ν_0 that assures $\nu_n > 2$ and existence of σ_μ^2.
A least-mean-square fit sixth-order polynomial can be derived (see Figure 1) for a wider range of σ_M/σ_0 than that in Table 1 as

$$\frac{1}{\nu_0} = p_0 + p_1\left(\frac{\sigma_M}{\sigma_0}\right)^1 + p_2\left(\frac{\sigma_M}{\sigma_0}\right)^2 + p_3\left(\frac{\sigma_M}{\sigma_0}\right)^3 + p_4\left(\frac{\sigma_M}{\sigma_0}\right)^4 + p_5\left(\frac{\sigma_M}{\sigma_0}\right)^5 + p_6\left(\frac{\sigma_M}{\sigma_0}\right)^6 \qquad (12)$$

(12) provides an approximation of ν_0 as the solution of the implicit equation (11) over the range spanned by σ_M/σ_0 in Fig. 1 (blue curve, corresponding to $\alpha = 0.05$). The maximum relative deviation between ν_0, as obtained from (11), and its approximation as obtained from (12), is less than 2 %. The values of the coefficients $p_0, p_1, \cdots, p_6$ can be calculated but they are not reported here, for sake of brevity.

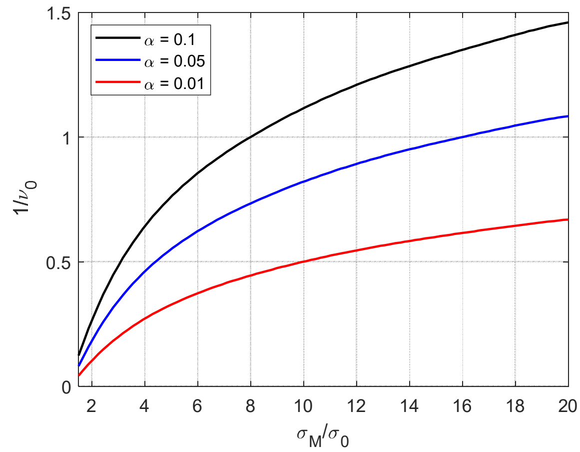

Figure 1: $1/\nu_0$ as a function of σ_M/σ_0 as obtained from the lest-mean-square fit sixth-order polynomial (12) and corresponding to three values of the probability α (see the legend).

In Table 2 several numerical examples are listed to appreciate the features of the proposed solution. Inputs are the values for n, s, σ_0 and σ_M, outputs are ν_0, ν_n, σ_n, σ_μ and $s/\sqrt{n}$. Calculation of the outputs from the inputs was carried out by using a spreadsheet.
Cases 1-4 correspond to increasing values of s assuming $n = 2$, $\sigma_0 = 1$ and $\sigma_M = 3$. Hence, by using (12), $\nu_0 = 2.92$ and $\nu_n = 3.92$ as obtained from the first of (9). σ_n^2 results from a weighted average between s^2 (weighted by $n-1$) and σ_0^2 (weighted by ν_0), see the last of (9). Since $\nu_0 > n-1$ then σ_n is closer to σ_0 than to s. In case 4, s is even larger than σ_M however the weighted average σ_n is again closer to σ_0. In case 5, $n = 10$ so the weight of s^2 is larger than the weight of σ_0^2 and the weighted average σ_n gets closer to s than to σ_0 and larger than σ_M. In cases 6 and 7 we have s larger than σ_M and it is shown, similarly to cases 4 and 5 but for a different ratio σ_M/σ_0, that the weighted average σ_n can get a value larger than σ_M provided that a sufficient large number of

measurements n is available compared with the number of degrees of freedom ν_0. The proposed type A evaluation of standard uncertainty, σ_μ, and the GUM Type A evaluation of standard uncertainty, $s/\sqrt{n}$, are compared in the last two columns of Table 3. If one again considers cases 1 to 4 and 6 notes that a variation of a factor of 8 on s corresponds to a variation of the same factor in $s/\sqrt{n}$ (obvious) but of a factor of only 2.4 in σ_μ. This is because σ_μ results from a weighted average where the weight of s is relatively small. This is an attractive feature of σ_μ, which is relatively insensitive to large variations of s when n is small compared with ν_0. When n is relatively large, as in cases 5 and 7, the two type A evaluations are essentially the same and since $s > \sigma_M$ then σ_μ is larger than σ_M/n, which is not the case in [10], since there σ_M is an upper truncation limit for the prior PDF of σ^2. Finally note that σ_μ exists in all cases.

Table 2: Numerical examples.

Case #	*n*	*s*	σ_0	σ_M	ν_0	ν_n	σ_n	σ_μ	$s/n^{0.5}$
1	2	0.5	1	3	2.92	3.92	0.90	0.91	0.35
2	2	1.5	1	3	2.92	3.92	1.15	1.16	1.06
3	2	2.5	1	3	2.92	3.92	1.53	1.55	1.77
4	2	4	1	3	2.92	3.92	2.20	2.22	2.83
5	10	4	1	3	2.92	11.92	3.51	1.22	1.26
6	2	3	1	2	5.47	6.47	1.50	1.27	2.12
7	10	3	1	2	5.47	14.47	2.44	0.83	0.95

We now consider the following numerical example: $\sigma_0 = 2$, $\nu_0 = 4$, $s = 1$ and $n = 4$. By using these figures the proposed type A evaluation of standard uncertainty is $\sigma_\mu = 0.97$, see (10). In the limiting case where no prior information is available ($\nu_0 = 0$) we have $(\sigma_\mu)^* = \sqrt{\frac{n-1}{n-3}}\frac{s}{\sqrt{n}} = 0.87$. Since $\sigma_\mu > (\sigma_\mu)^*$ it might be argued that while some information is provided a larger uncertainty is obtained, which may appear illogical. Prior information however suggests that uncertainty should be larger than that observed ($\sigma_0 > s$) and prior information is more credible than observation ($\nu_0 > n-1$). Since the proposed type A evaluation is obtained from a weighted average between the variance associated with the prior information and the observation, the weights being the corresponding degrees of freedom, then the result is fully logical and meaningful.

4 References

1. Guide to the Expression of Uncertainty in Measurement, JCGM 100:2008, GUM 1995 with minor corrections, 2008.

2. Supplement 1 to the 'Guide to the Expression of Uncertainty in Measurement' – Propagation of Distributions Using a Monte Carlo Method, JCGM 101:2008, 2008.

3. Supplement 2 to the 'Guide to the Expression of Uncertainty in Measurement – Propagation of Distributions Using a Monte Carlo Method, JCGM 102:2011, 2011.

4. Walter Bich et alii, "Revision of the 'Guide to the Expression of Uncertainty in Measurement'," Metrologia 49 (2012) 702–705.

5. Walter Bich, Maurice Cox and Carine Michotte, "Towards a new GUM—an update," Metrologia 53 (2016) S149–S159.

6. R Kacker and A Jones, "On use of Bayesian statistics to make the Guide to the Expression of Uncertainty in Measurement consistent," Metrologia 40 (2003) 235–248.

7. "Electromagnetic compatibility (EMC) – Part 1-6: General guide to the assessment of measurement uncertainty," Int. Electrotech. Commission, Geneva, Switzerland, Tech. Rep. IEC/TR 61000-1-6, 2012.

8. American National Standard Guide for Electromagnetic Compatibility - Computations and Treatment of Measurement Uncertainty, Amer. Nat. Standard Inst., Washington, DC, USA, ANSI C63.23-2012.

9. Giulio D'Agostini, "Jeffreys Priors versus Experienced Physicist Priors – Arguments against Objective Bayesian Theory," arXiv:physics/9811045 [physics.data-an], Submitted on 23 Nov 1998.

10. M Cox and K Shirono, "Informative Bayesian Type A uncertainty evaluation, especially applicable to a small number of observations," Metrologia 54 (2017), 642–652.

11. A Gelman, J B Carlin, H S Stern, D B Dunson, A Vehtari, D B Rubin, Bayesian Data Analysis, 3rd ed. Boca Raton, FL, USA: CRC Press, 2013, ISBN 9781439840955.

12. Carlo Carobbi and Francesca Pennecchi, "Bayesian conformity assessment in presence of systematic measurement errors," Metrologia 53 (2016) S74–S80.

13. Carlo F. M. Carobbi, "Bayesian Inference in Action in EMC – Fundamentals and Applications", IEEE Transactions on Electromagnetic Compatibility, vol. 59, no. 4, pp. 1114 – 1124, Aug. 2017.

14. Carlo F. M. Carobbi, Sébastien Lalléchère, and Luk R. Arnaut, "Review of Uncertainty Quantification of Measurement and Computational Modeling in EMC Part I: Measurement Uncertainty," IEEE Transactions on Electromagnetic Compatibility, vol. 61, no. 6, pp. 1698 – 1705, Dec. 2019, DOI 10.1109/TEMC.2019.2904973.

15. Sébastien Lalléchère, Carlo F. M. Carobbi, and Luk R. Arnaut, "Review of Uncertainty Quantification of Measurement and Computational Modeling in EMC Part II: Computational Uncertainty," IEEE Transactions on Electromagnetic Compatibility, vol. 61, no. 6, pp. 1699 – 1706, Dec. 2019, DOI 10.1109/TEMC.2019.2904999.