Int. J. Metrol. Qual. Eng. **3**, 71–77 (2012)
© Crown, published by EDP Sciences 2012
DOI: 10.1051/ijmqe/2012017

Approaches to evaluating measurement uncertainty

A.B. Forbes*

National Physical Laboratory, TW11 0LW, Teddington, UK

Received: 5 May 2012 / Accepted: 27 June 2012

Abstract. The *Guide to the expression of measurement uncertainty*, (GUM, JCGM 100) and its *Supplement 1: propagation of distributions by a Monte Carlo method*, (GUMS1, JCGM 101) are two of the most widely used documents concerning measurement uncertainty evaluation in metrology. Both documents describe three phases (a) the construction of a measurement model, (b) the assignment of probability distributions to quantities, and (c) a computational phase that specifies the distribution for the quantity of interest, the measurand. The two approaches described in these two documents agree in the first two phases but employ different computational approaches, with the GUM using linearisations to simplify the calculations. Recent years have seen an increasing interest in using Bayesian approaches to evaluating measurement uncertainty. The Bayesian approach in general differs in the assignment of the probability distributions and its computational phase usually requires Markov chain Monte Carlo (MCMC) approaches. In this paper, we summarise the three approaches to evaluating measurement uncertainty and show how we can regard the GUM and GUMS1 as providing approximate solutions to the Bayesian approach. These approximations can be used to design effective MCMC algorithms.

Keywords: Measurement uncertainty; Monte Carlo; Markov chain Monte Carlo

1 Introduction

The evaluation of measurement uncertainty is an integral part of metrology, no measurement result is complete without an associated statement of uncertainty. The VIM [1] defines the uncertainty of measurement as "a non-negative parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used". This definition affords considerable latitude as to how this non-negative parameter is calculated and how it should be interpreted. Other standards documents have been written in order to specify how measurement uncertainty is to be evaluated, in particular, the GUM and its supplement 1 (GUMS1) [2,3]. These documents adopt a strongly probabilistic/statistical approach. Measurements are usually made in order to make inferences or decisions about the quantities involved. We suppose these quantities (e.g., length of a gauge block, mass of a standard) are represented by parameters $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)^T$. By associating a probability distribution with density $p(\boldsymbol{\alpha})$ with these quantities, it is then possible to evaluate the probability that the quantities lie in a given region A through

$$\Pr(\alpha \in A) = \int_A p(\boldsymbol{\alpha})\, d\boldsymbol{\alpha}.$$

It is usual to interpret this distribution as a state of knowledge or state of belief distribution. The density allows us to evaluate the probability associated with statements involving $\boldsymbol{\alpha}$ on which decision can then be made, e.g., does a part conform to its specification. At a later stage, it may be possible to determine if these decisions were correct or not.

In general, there are two statistical modelling paradigms commonly used in science, the classical (or frequentist) paradigm and the Bayesian paradigm [4]. This paper is mainly about how the GUM and GUMS1 can be used as approximate computational approaches in either the classical or Bayesian paradigm as they apply to two related statistical modelling problems in metrology, measurement system characterisation and measurement uncertainty evaluation.

2 The classical paradigm and measurement system characterisation

The classical paradigm for making inferences about quantities on the basis of measurement is based on estimation. We suppose that an experiment involves a set of quantities represented by parameters $\boldsymbol{\alpha}$, and that a set of measurement indications $\boldsymbol{\zeta}$ are recorded. A mathematical/statistical model of the measurement system is used to define the probability density $p(\boldsymbol{\zeta}|\boldsymbol{\alpha})$ that relates to the probability that a set of parameters $\boldsymbol{\alpha}$ could give rise to the set of indications $\boldsymbol{\zeta}$. Viewed as a function of $\boldsymbol{\alpha}$, this density in known as the likelihood. In the classical

* Correspondence: Alistair.Forbes@npl.co.uk

Article published by EDP Sciences

https://doi.org/10.1051/ijmqe/2012017 Published online by Cambridge University Press

approach, a method of analysing the data is designed to return an estimate $\hat{\boldsymbol{\alpha}} = E(\boldsymbol{\zeta})$. The estimation method is characterised, if possible, in terms of a density $p(\hat{\boldsymbol{\alpha}}|\boldsymbol{\alpha})$ that shows how the estimates $\hat{\boldsymbol{\alpha}}$ are distributed given a hypothesised set of parameter values $\boldsymbol{\alpha}$. A good estimation method always has most of the probability mass of this density as close as possible to $\boldsymbol{\alpha}$.

The density $p(\hat{\boldsymbol{\alpha}}|\boldsymbol{\alpha})$ addresses the following issue: if I use this measurement system to measure my calibrated artefact (with associated calibration values $\boldsymbol{\alpha}$) how far are the measurement results ($\hat{\boldsymbol{\alpha}}$) likely to be from the calibrated values. The density depends on the likelihood $p(\boldsymbol{\zeta}|\boldsymbol{\alpha})$ and the definition of the estimation method $\hat{\boldsymbol{\alpha}} = E(\boldsymbol{\zeta})$. One particularly important estimation method is the maximum likelihood estimate [4] that determines the estimate that maximises $p(\boldsymbol{\zeta}|\boldsymbol{\alpha})$, regarded as a function of $\boldsymbol{\alpha}$. (Note that $p(\boldsymbol{\zeta}|\boldsymbol{\alpha})$ is a density with respect to $\boldsymbol{\zeta}$, not $\boldsymbol{\alpha}$.)

3 The Bayesian paradigm and the evaluation of measurement uncertainty

The classical paradigm starts with hypothesised values for the parameters and then characterises the behaviour of the measurement system and estimation method. The density $p(\hat{\boldsymbol{\alpha}}|\boldsymbol{\alpha})$ can be used to make inferences about the estimate, given hypothesised values $\boldsymbol{\alpha}$ for the parameters. It can say nothing about $\boldsymbol{\alpha}$ since these values are fixed inputs to the evaluation process. The Bayesian paradigm is used to address the inverse problem: what can be said about $\boldsymbol{\alpha}$ if the measuring returns indications $\boldsymbol{\zeta} = \boldsymbol{z}$? The Bayesian paradigm, e.g., [4, 5], answers this question by specifying a posterior distribution in terms a density $p(\boldsymbol{\alpha}|\boldsymbol{\zeta})$ that encodes the information about the parameters on the basis of the indications and any prior information. The posterior distribution is arrived at using Bayes' theorem which says

$$p\,(\boldsymbol{\alpha}|\boldsymbol{\zeta}) \propto p(\boldsymbol{\zeta}|\boldsymbol{\alpha})p(\boldsymbol{\alpha})$$

where $p(\boldsymbol{\zeta}|\boldsymbol{\alpha})$ is the likelihood function and $p\,(\boldsymbol{\alpha})$ is the prior density encoding information about the parameters available prior to or independent from the indications. The posterior distribution can be used directly to make inferences about $\boldsymbol{\alpha}$ since it represents the state of knowledge about $\boldsymbol{\alpha}$, following the measurement.

4 Three computational tools

In this section, we describe three computational tools. The first two, the law of propagation of uncertainty (LPU) and the Monte Carlo method (MCM), apply in the case of a possibly multivariate input-output model of the form $\boldsymbol{Y} = \boldsymbol{f}(\boldsymbol{X})$ which expresses $\boldsymbol{Y}$ as a function of input quantities $\boldsymbol{X}$, and the input quantities are associated with a known probability distribution. In this case the distribution associated with the output $\boldsymbol{Y}$ is defined by the input distribution and the functional relationship. The first two computational approaches show how information about the output distribution associated with $\boldsymbol{Y}$ can be calculated or approximated. These computational approaches can be applied to the density $p(\hat{\boldsymbol{\alpha}}|\boldsymbol{\alpha})$, since the estimation method $\hat{\boldsymbol{\alpha}} = E(\boldsymbol{\zeta})$ defines the estimate as a function of the indications $\boldsymbol{\zeta}$ and the latter are associated with the distribution $p\,(\boldsymbol{\zeta}|\boldsymbol{\alpha})$.

The third is the Markov chain Monte Carlo method (MCMC) which is a general method for sampling from a probability distribution and is suitable for sampling from the posterior distribution $p(\boldsymbol{\alpha}|\boldsymbol{\zeta})$ derived in the Bayesian paradigm.

4.1 The law of propagation of uncertainty (LPU)

Let $\boldsymbol{Y} = \boldsymbol{f}(\boldsymbol{X})$ and assume that $\boldsymbol{X}$ is associated with a distribution with mean $\boldsymbol{x}$ and variance matrix $V_{\boldsymbol{X}}$. Then the mean and variance matrix of the distribution associated with $\boldsymbol{Y}$ is approximated [1] by $\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x})$ and $V_{\boldsymbol{Y}} = CV_{\boldsymbol{X}}C^{\mathrm{T}}$, respectively, where C is the matrix of partial derivatives $\partial f_i/\partial X_j$ evaluated at $\boldsymbol{x}$. If $\boldsymbol{f}$ is a linear function then these relationships are exact, irrespective of the shape of the distribution associated with input $\boldsymbol{X}$. If, in addition, the distribution associated with $\boldsymbol{X}$ is a multivariate Gaussian distribution $N(\boldsymbol{x}, V_X)$, then the distribution associated with $\boldsymbol{Y}$ is the multivariate distribution $N(\boldsymbol{y}, V_Y)$. Often, even for nonlinear models and non-Gaussian input distributions, the distribution associated with the output $\boldsymbol{Y}$ is taken to be $N(\boldsymbol{y}, V_{\boldsymbol{Y}})$, in the absence of any other information.

4.2 The Monte Carlo method (MCM)

The Monte Carlo method is a general tool for propagating distributions [2] through an input-output model $\boldsymbol{Y} = \boldsymbol{f}(\boldsymbol{X})$. If $\{\boldsymbol{x}_q q = 1, M\}$ is a sample from the distribution associated with $\mathbf{X}$, then $\{\boldsymbol{y}_q = \boldsymbol{f}(\boldsymbol{x}_q)\}$ is a sample from the distribution associated with $\mathbf{Y}$. The mean and variance matrix associated with the output distribution can be estimated by the mean and variance of the sample. Thus, the mean of the distribution associated with the kth output quantity is estimated by

$$\bar{y}_k = \frac{1}{M}\sum_{q=1}^{M} y_{k,q},$$

and the corresponding standard deviation is estimated by

$$s_k^2 = \frac{1}{M-1}\sum_{q=1}^{M}(y_{k,q} - \bar{y}_k)^2.$$

4.3 Markov chain Monte Carlo (MCMC)

The MCMC approach [5, 6] is similar to the MCM already described in that the aim is to produce a sample

https://doi.org/10.1051/ijmqe/2012017 Published online by Cambridge University Press

distributed according to the desired distribution. There are a number of MCMC algorithms designed to achieve this. Here, we describe one such algorithm, a Metropolis-Hastings independence chain (MHIC) [6]. The MHIC algorithm applies if we wish to sample according to a distribution $p(\boldsymbol{\alpha})$ (the target distribution) and there is an approximating distribution $p_0(\boldsymbol{\alpha})$ (the proposal distribution) from which it is easy to sample. Suppose that $\{\boldsymbol{a}_{0,q} : q = 1, \ldots, M\}$ is a sample according to the proposal distribution $p_0(\boldsymbol{\alpha})$. The aim of the MHIC scheme is to replicate those samples that correspond to areas where the proposal distribution under-represents the probability mass of the target distribution at the expense of rejecting samples where the proposal over-represents the probability mass of the target distribution. The accept-reject step is as follows, starting with $\boldsymbol{a}_1 = \boldsymbol{a}_{0,1}$, assuming that $p(\boldsymbol{a}_{0,1}) > 0$. If $\boldsymbol{a}_{q-1}$ has been assigned then $\boldsymbol{a}_q$ is set to $\boldsymbol{a}_{0,q}$ with probability $\min\{1, r_q\}$, where

$$r_q = \frac{p(\boldsymbol{a}_{0,q})\, p_0(\boldsymbol{a}_{q-1})}{p(\boldsymbol{a}_{q-1})\, p_0(\boldsymbol{a}_{0,q})}.$$

Thus, if $u_q \in R(0,1)$ is a random draw from a uniform distribution, then $\boldsymbol{a}_q = \boldsymbol{a}_{0,q}$ if $u_q < r_q$, otherwise $\boldsymbol{a}_q = \boldsymbol{a}_{q-1}$. If this process is run a number of times in parallel then the initial samples will be from the proposal distribution and the later samples will be from the target distribution. Usually, multiple chains are used so that some assessment of the convergence can be made [5]. The closer the proposal distribution is to the target distribution the faster the convergence.

5 Example: exponential decay

In this section, we show how these three computational tools can be applied in a simple example. The example concerns the response model, $\eta = \phi(t, \alpha_1, \alpha_2, \lambda) = \alpha_1 \exp\{-\lambda t\} + \alpha_2$. We wish to determine estimates of α_1, α_2 and λ from measurements of the response η at times $t = 0$, $t = t_0$ and $t = 2t_0$ Writing $\alpha_3 = \exp\{-\lambda t_0\}$, the three responses can be written as

$$\eta_1 = \phi_1(\boldsymbol{\alpha}) = \alpha_1 + \alpha_2,$$
$$\eta_2 = \phi_2(\boldsymbol{\alpha}) = \alpha_1\alpha_3 + \alpha_2,$$

and

$$\eta_3 = \phi_3(\boldsymbol{\alpha}) = \alpha_1\alpha_3^2 + \alpha_2$$

or, in vector notation, $\boldsymbol{\eta} = \boldsymbol{\phi}(\boldsymbol{\alpha})$. The form of these equations allows us to solve for $\boldsymbol{\alpha}$ in terms of $\boldsymbol{\eta}$: $\alpha_3 = \frac{\eta_2 - \eta_3}{\eta_1 - \eta_2}$, $\alpha_1 = \frac{\eta_1 - \eta_2}{1 - \alpha_3}$, and $\alpha_2 = \eta_1 - \alpha_1$. In vector terms, $\boldsymbol{\alpha} = \boldsymbol{\psi}(\boldsymbol{\eta})$, with $\boldsymbol{\psi}$ the inverse of $\boldsymbol{\phi}$.

We suppose the responses are measured independently with a measurement system that produces indications characterised by $\zeta_i = \eta_i + \epsilon_i$, where ϵ_i is drawn from the Gaussian distribution $N(0, \sigma^2)$ with zero mean and standard deviation σ. Another way of expressing this measurement model is that $\boldsymbol{\zeta}|\boldsymbol{\alpha} \sim N(\boldsymbol{\phi}(\boldsymbol{\alpha}), \sigma^2 I)$, i.e., given $\boldsymbol{\alpha}$, the indications are expected to be normally distributed about $\boldsymbol{\phi}(\boldsymbol{\alpha})$. The measurement model allows us to define the likelihood

$$p(\boldsymbol{\zeta}|\boldsymbol{\alpha}) \propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{3}\left(\zeta_i - \phi_i(\boldsymbol{\alpha})\right)^2\right\}. \qquad (1)$$

Given indications $\boldsymbol{\zeta}$, an estimate of the quantities $\boldsymbol{\alpha}$ is given by $\hat{\boldsymbol{\alpha}} = \boldsymbol{\psi}(\boldsymbol{\zeta})$. Thus, if we hypothesize a set of values for $\boldsymbol{\alpha}$, $\boldsymbol{\zeta}$ is assigned the distribution $N(\boldsymbol{\phi}(\boldsymbol{\alpha}), \sigma^2 I)$. Since the quantities $\boldsymbol{\zeta}$ have been assigned a distribution, the distribution for the estimate $\hat{\boldsymbol{\alpha}}$, given $\boldsymbol{\alpha}$, is also defined

$$p(\hat{\boldsymbol{\alpha}}|\boldsymbol{\alpha}) = p(\boldsymbol{\phi}(\hat{\boldsymbol{\alpha}})|\boldsymbol{\alpha})\, |J(\hat{\boldsymbol{\alpha}})|.$$

The first term on the right is the likelihood $p(\boldsymbol{\zeta}|\boldsymbol{\alpha})$ evaluated at $\boldsymbol{\zeta} = \boldsymbol{\phi}(\hat{\boldsymbol{\alpha}})$ and $J(\hat{\boldsymbol{\alpha}})$ is the Jacobian matrix of partial derivatives $\partial\phi_i/\partial\alpha_j$ evaluated at $\hat{\boldsymbol{\alpha}}$. This result comes from the general formula for defining a distribution for a reparametrization of a set of quantities [7], Section 4.7.

While we can write down the density $p(\hat{\boldsymbol{\alpha}}|\boldsymbol{\alpha})$ for this model, the presence of the determinant of the Jacobian matrix is a nonlinear function which means that this distribution is not a standard distribution and, as a consequence, we cannot write down its mean nor its standard deviation, for example. However, the relationship $\hat{\boldsymbol{\alpha}} = \boldsymbol{\psi}(\boldsymbol{\zeta})$ expresses $\hat{\boldsymbol{\alpha}}$ as a function of $\boldsymbol{\zeta}$ and in situations such as this, we can use the two computational approaches described above, the law of propagation of uncertainty (LPU), and the Monte Carlo method (MCM).

5.1 LPU approach

The LPU approach is implemented as follows. Given, $\boldsymbol{\alpha}$, evaluate $\boldsymbol{\eta} = \boldsymbol{\phi}(\boldsymbol{\alpha})$. The vector of indications $\boldsymbol{\zeta}$ is associated with the distribution $N(\boldsymbol{\eta}, \sigma^2 I)$ which has mean $\boldsymbol{\eta}$ and variance matrix $\sigma^2 I$. The estimate is given by $\hat{\boldsymbol{\alpha}} = \boldsymbol{\psi}(\boldsymbol{\zeta})$ and the LPU method requires us to evaluate the 3×3 matrix C of partial derivatives of $\boldsymbol{\psi}$ with respect to $\boldsymbol{\zeta}$, evaluated at the mean of the input distribution, i.e., at $\boldsymbol{\zeta} = \boldsymbol{\eta}$. The mean of the distribution for $\hat{\boldsymbol{\alpha}}$ is estimated by

$$\boldsymbol{\psi}(\boldsymbol{\eta}) = \boldsymbol{\psi}(\boldsymbol{\phi}(\boldsymbol{\alpha})) = \boldsymbol{\alpha},$$

and the variance matrix is estimated by $C\sigma^2 I C^{\mathrm{T}} = \sigma^2 C C^T$. In fact, the matrix C is perhaps more easily evaluated as $C = J^{-1}$, where J is the matrix of partial derivatives of $\boldsymbol{\phi}$ with respect to $\boldsymbol{\alpha}$. These calculations show that, to first order, the estimate $\hat{\boldsymbol{\alpha}} = \boldsymbol{\psi}(\boldsymbol{\zeta})$ is unbiased, i.e., the expected value of the estimate of $\boldsymbol{\alpha}$ is $\boldsymbol{\alpha}$.

5.2 MCM approach

The MCM approach is implemented as follows. Given $\boldsymbol{\eta} = \boldsymbol{\phi}(\boldsymbol{\alpha})$, sample $\boldsymbol{z}_q$ from $N(\boldsymbol{\eta}, \sigma^2 I)$ and evaluate $\boldsymbol{a}_q = \boldsymbol{\psi}(\boldsymbol{z}_q)$ Then $\{\boldsymbol{a}_q\}$ is a sample from the distribution $p(\hat{\boldsymbol{\alpha}}|\boldsymbol{\alpha})$. As can be seen, the MCM method is conceptually very simple to implement and it corresponds to simulating a large number of repeated experiments for a fixed set of system parameters $\boldsymbol{\alpha}$.

https://doi.org/10.1051/ijmqe/2012017 Published online by Cambridge University Press

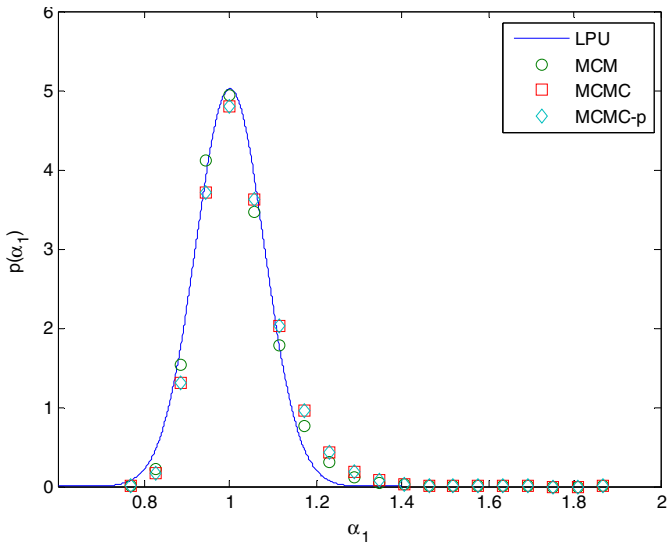


Fig. 1. Four distributions associated with the parameter α_1.

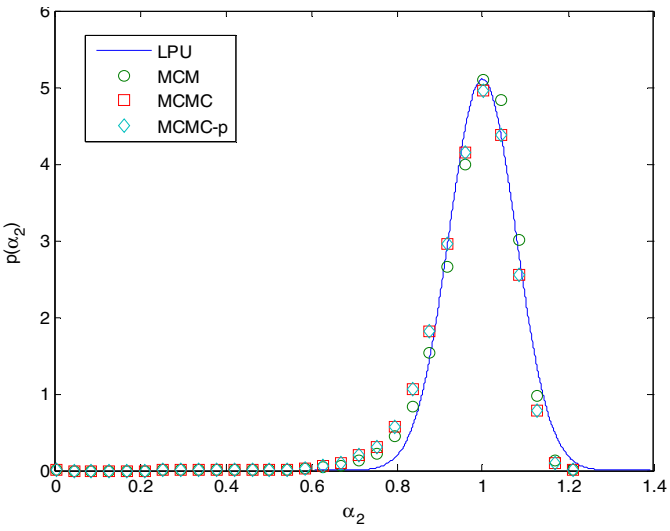


Fig. 2. Four distributions associated with the parameter α_2.

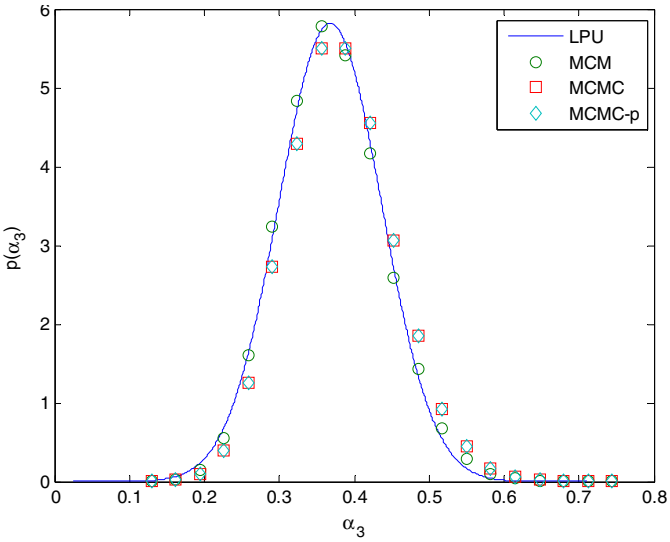


Fig. 3. Four distributions associated with the parameter α_3.

5.3 MCMC approach

The posterior distribution $p\left(\boldsymbol{\alpha}|\boldsymbol{\zeta}\right) \propto p(\boldsymbol{\zeta}|\boldsymbol{\alpha})p(\boldsymbol{\alpha})$ is defined by the likelihood in equation (1) and the prior distribution. In practice, we may know very little about $\boldsymbol{\alpha}$ beforehand except that for physical reasons both $M_1 \geqslant \alpha_1 \geqslant 0$, $M_2 \geqslant \alpha_2 \geqslant 0$ and $1 > \alpha_3 > 0$, where M_1 and M_2 are possibly large upper bounds. Accordingly we define a prior with $p(\boldsymbol{\alpha}) = 1$ if these conditions are satisfied, and equal to zero otherwise.

We can produce a sample from the posterior distribution using the Metropolis-Hastings independence chain algorithm if we can find an approximating distribution that is easy to sample from. Given $\boldsymbol{\zeta}$, we can set $\boldsymbol{\alpha} = \boldsymbol{\psi}\left(\boldsymbol{\zeta}\right)$ and use MCM to the sample from distribution $p\left(\hat{\boldsymbol{\alpha}}|\boldsymbol{\alpha} = \boldsymbol{\psi}\left(\boldsymbol{\zeta}\right)\right)$. With a little algebra and a change of notation, it is straightforward to show that this distribution [8–10] can be written as $p_0\left(\boldsymbol{\alpha}\right) = p\left(\boldsymbol{\zeta}|\boldsymbol{\alpha}\right)\left|J\left(\boldsymbol{\alpha}\right)\right|$, where $\left|J\left(\boldsymbol{\alpha}\right)\right|$ is the determinant of the matrix of partial derivatives of $\boldsymbol{\phi}$ with respect to $\boldsymbol{\alpha}$, as before. Comparing this distribution with posterior distribution we have

$$\frac{p(\boldsymbol{\alpha}|\boldsymbol{\zeta})}{p_0(\boldsymbol{\alpha})} = \frac{p(\boldsymbol{\alpha})}{\left|J\left(\boldsymbol{\alpha}\right)\right|}.$$

So long as this ratio remains relatively constant for the relevant range of $\boldsymbol{\alpha}$, the distribution $p_0(\boldsymbol{\alpha})$ is a close approximation to the posterior distribution. As a consequence, a sample produced by MCM can be used effectively to construct a sample from the posterior distribution using a MHIC algorithm [11, 12].

5.4 Results

Figures 1–3 show the distributions associated with the three parameters α_j, $j = 1, 2, 3$, for data generated for the case $\alpha_1 = \alpha_2 = 1$, $\alpha_3 = \frac{1}{e} \approx 0.37$, and $\sigma = 0.025$. The distributions are calculated in four ways, (1) using LPU to estimate the mean and standard deviation associated with each parameter and then approximating the output distribution with a Gaussian with the same mean and standard deviation, (2) the MCM method, (3) the MCMC approach with a uniform (improper) prior $p(\boldsymbol{\alpha}) \propto 1$ (labelled MCMC), and (4) the MCMC approach with a uniform prior but with bound constraints on each parameter (labelled MCMC-p). The posterior distributions for α_1 and α_2 are distinctly asymmetric. The MCM and MCMC distributions are somewhat different, as to be expected, since the determinant of the Jacobian matrix is not constant over the relevant region, ranging from 0.1 to 0.65. Regarding the MCM approach as sampling from the distribution $p(\hat{\boldsymbol{\alpha}}|\boldsymbol{\alpha})$, the fact that the mean of the distribution associated with the MCM calculations is not the same as the input $\boldsymbol{\alpha}$ in general shows that the estimation method is biased.

https://doi.org/10.1051/ijmqe/2012017 Published online by Cambridge University Press

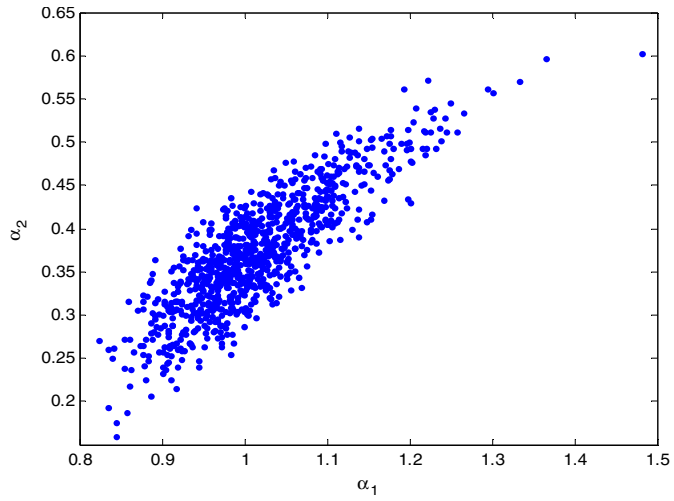


Fig. 4. Representation of the correlation between α_1 and α_2.

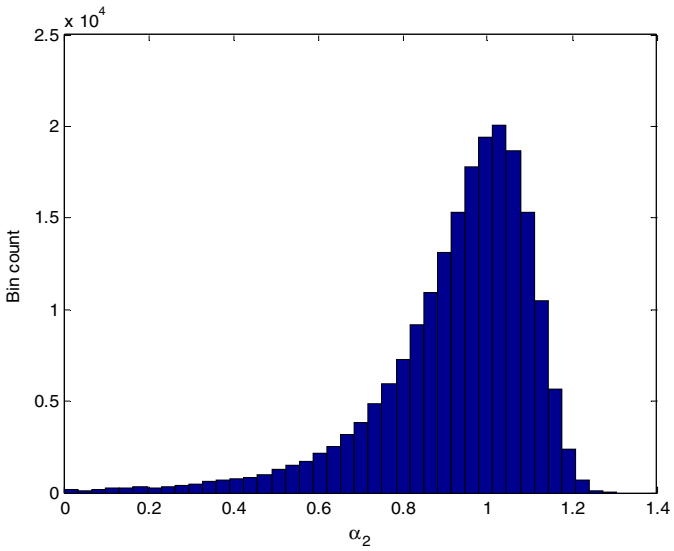


Fig. 5. Histogram of the samples for α_2 for noisier data.

Figure 4 represents the nonlinear correlation between the first two parameters, showing that the posterior distribution is not approximated particularly well by a multivariate Gaussian. For these calculations, there is no difference in the posterior distributions associated with the two choices of priors. The constraints imposed by the prior in calculation (4) do not come into play. For a less accurate measuring system characterised by a larger standard deviation $\sigma = 0.04$, the constraints do come into play. Figure 5 shows the histogram of samples associated with the parameter α_2 and indicates the distribution is slightly truncated at 0.

6 The GUM and GUMS1 approach to uncertainty evaluation

The GUM approach to uncertainty evaluation is based on LPU. Many of the examples are formulated in terms of a measurement response model $\eta = \phi(\alpha, \boldsymbol{\beta})$ which describes the response η to the stimulus α, the measurand, and influence factors $\boldsymbol{\beta}$. It is assumed that there is information about the influence factors $\boldsymbol{\beta}$ in terms of a distribution $p(\boldsymbol{\beta})$ and that information about the measurand α is to be derived from a measurement of the response η.

The response model is a special case of the one considered above where the quantities $\boldsymbol{\alpha}$ and $\boldsymbol{\eta}$ are partitioned according to $\boldsymbol{\alpha} = \begin{bmatrix} \alpha \\ \boldsymbol{\beta} \end{bmatrix}, \boldsymbol{\eta} = \begin{bmatrix} \eta \\ \boldsymbol{\tau} \end{bmatrix}$, with

$$\boldsymbol{\phi}(\boldsymbol{\alpha}) = \begin{bmatrix} \phi(\alpha, \boldsymbol{\beta}) \\ \boldsymbol{\beta} \end{bmatrix},$$

i.e., $\boldsymbol{\tau} = \boldsymbol{\beta}$, and

$$\boldsymbol{\psi}(\boldsymbol{\eta}) = \begin{bmatrix} \psi(\eta, \boldsymbol{\beta}) \\ \boldsymbol{\tau} \end{bmatrix}.$$

If the response η is measured, leading to an indication ζ, according to the measurement model $\zeta|\eta \sim N(\eta, \sigma^2)$, then η is assigned the distribution $N(\zeta, \sigma^2)$. If the mean and variance matrix of the distribution $p(\boldsymbol{\beta})$ is given by $\boldsymbol{b}$ and V_B, respectively then $\boldsymbol{\eta}$ is assigned the mean and variance

$$\begin{bmatrix} \zeta \\ \boldsymbol{b} \end{bmatrix} \text{ and } \begin{bmatrix} \sigma^2 & \mathbf{0} \\ \mathbf{0} & V_B \end{bmatrix}.$$

Respectively, and these are propagated through to estimate the mean and variance of the output $\boldsymbol{\psi}(\boldsymbol{\eta})$.

The special form of this function means that the mean and variance associated with α is estimated by $\hat{\alpha} = \psi(\zeta, \boldsymbol{b})$ and

$$u^2(\hat{\alpha}) = \sigma^2 + \boldsymbol{c}^{\mathrm{T}} V_B \boldsymbol{c},$$

respectively, where $\boldsymbol{c}$ is the vector of partial derivatives of ψ with respect to $\boldsymbol{\beta}$ evaluated at $\eta = \zeta$ and $\boldsymbol{\beta} = \boldsymbol{b}$.

The GUMS1 approach to uncertainty evaluation uses the same formulation of the input-output model as the GUM approach but uses as its computational tool MCM. Given samples $y_q \in N(\zeta, \sigma^2)$ and $\boldsymbol{b}_q$ sampled according to the distribution $p(\boldsymbol{\beta})$, we set $a_q = \psi(y_q, \boldsymbol{b}_q)$. It produces a sample from the posterior distribution

$$p_{\mathrm{MCM}}(\alpha, \boldsymbol{\beta}|\zeta) \propto p(\zeta|\alpha, \boldsymbol{\beta}) \left| \frac{\partial \phi}{\partial \alpha}(\alpha, \boldsymbol{\beta}) \right|.$$

If the term $\left| \frac{\partial \phi}{\partial \alpha}(\alpha, \boldsymbol{\beta}) \right|$ is approximately constant over the range of interest then this distribution approximates well the posterior distribution corresponding to a noninformative $p(\alpha) \propto 1$, often the preferred choice.

7 Example: mass calibration

See, for example [13], Section 9.10 and [14]. The model concerns the calibration of a weight W of mass density ρ_W against a reference weight R of mass density ρ_R. It is assumed that the two masses are nominally the same. The measurement model is of the form

$$\delta m = \frac{m_{W,c}}{1 + (\rho_a - \rho_0)\left(\frac{1}{\rho_W} - \frac{1}{\rho_R}\right)} - m_{R,c},$$

https://doi.org/10.1051/ijmqe/2012017 Published online by Cambridge University Press

where δm is the difference in mass to be measured by a balance operating in air of density ρ_a, $m_{R,c}$ is the conventional mass of reference weight, i.e., the mass ignoring buoyancy effects of "conventional" air whose density is assigned to be $\rho_0 = 1.2$ kg/m^3, $m_{W,c}$ the conventional mass of W, the measurand. The mass difference is measured on a balance giving rise to an indication ζ according to $\zeta|\delta m \sim N(\delta m, \sigma^2)$. The measurement model can be written as a response model $\delta m = \phi(m_{W,c}, \boldsymbol{\beta})$ with $\boldsymbol{\beta} = (m_{R,c}, \rho_a, \rho_W, \rho_R)^T$. Given a prior distribution $p(\boldsymbol{\beta})$ for $\boldsymbol{\beta}$ and a noninformative prior distribution $p(\alpha) \propto 1$ for α, the posterior distribution $p(\alpha, \boldsymbol{\beta}|\zeta)$ is such that

$$p(m_{W,c}, \boldsymbol{\beta}|\zeta) \propto p(\zeta|\delta m)\, p(\boldsymbol{\beta}).$$

In the GUM/GUMS1 approach, we write $m_{W,c} = \psi(\delta m, \boldsymbol{\beta})$, i.e.,

$$m_{W,c} = (m_{R,c} + \delta m)\left(1 + (\rho_a - \rho_0)\left(\frac{1}{\rho_W} - \frac{1}{\rho_R}\right)\right),$$

and assign the distribution $\delta m|\zeta \sim N(\zeta\sigma^2)$ to δm and propagate the distribution

$$p(\delta m, \boldsymbol{\beta}|\zeta) = p(\delta m|\zeta)\, p(\boldsymbol{\beta})$$

through to that for $m_{W,c}$. Applying LPU to implement the propagation, for ρ_a close to ρ_0 and ρ_W close to ρ_R (as is likely to be the case in practice), the sensitivity of $m_{W,c}$ with respect to these four mass densities is also close to zero so that a first order calculation essentially ignores their uncertainty contribution. On the other hand, for this example, the partial derivative $\frac{\partial\phi}{\partial\delta m}$ is approximately constant, and a GUMS1 MCM sample can be used as a very good starting point for the Metropolis-Hastings independence chain algorithm.

The calculations below relate to the calibration of an OIML class F2 weight [14] with a nominal mass of m_0 =100 mg. The masses are associated with Gaussian distributions and the densities with rectangular distributions, the means and standard deviations of which are given in Table 1. The LPU calculation assigns to the measurand, $m_{W,c}$, a Gaussian distribution with mean 100 000.553 mg and standard deviation 0.064 mg. The MCM and MCMC samples were essentially identical for 100 000 samples since $\frac{\partial\phi}{\partial\delta m}$ varied by less than 3 parts in 10^6. The mean and standard deviation of the sample was 100 000.551 mg and 0.073 mg, respectively. Figure 6 shows as histogram associated with the MCM sample for the mass calibration example and a Gaussian with the same mean and standard deviation as the sample. As can be seen, the Gaussian is a very good approximation to the histogram. This is to be expected as the two dominant sources of uncertainty, $m_{R,c}$ and δm, are associated with Gaussian distributions and $m_{W,c} = \psi(\delta m, \boldsymbol{\beta})$ is linear in these parameters. The underestimation of the uncertainty by the LPU method is significant due to the nonlinearity of the model.

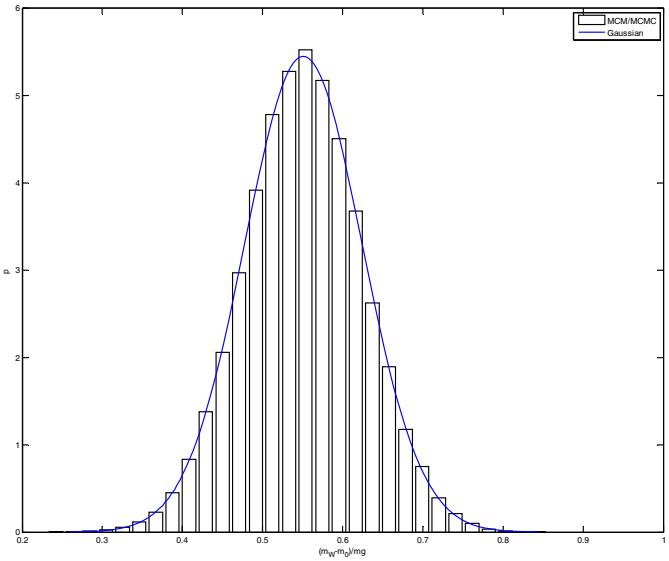


Fig. 6. Histogram associated with the MCM sample for the mass calibration example and a Gaussian with the same mean and standard deviation as the sample.

Table 1. Uncertainty data for mass calibration example.

	Mean	Std	
$m_{R,c}$	99 999.634 mg	0.053 mg	G
δm	0.917 mg	0.029 mg	G
ρ_W	8000 kg/m^3	240 kg/m^3	R
ρ_R	7960 kg/m^3	60 kg/m^3	R
ρ_a	1.170 kg/m^3	0.20 kg/m^3	R

8 Discussion and concluding remarks

We have seen that the GUM and GUMS1 approaches are based on the LPU and MCM computational tools, respectively. In our discussion above we have associated LPU and MCM with the problem of measurement system characterisation, i.e., concerned with the distribution $\boldsymbol{p(\hat{\alpha}|\alpha)}$, not uncertainty evaluation. In fact, we can argue [10, 15, 16] that this is how the GUM and GUMS1 are most easily interpreted, rather than approaches to implementing Bayesian inference. The GUM, especially, was written before Bayesian inference approaches were being considered seriously in metrology and before computational tools such as MCMC were being used for Bayesian inference. Neither the GUM nor GUMS1 make significant or explicit use of prior distributions although any Bayesian approach to evaluating measurement data necessarily has to involve the assignment of a prior distribution.

However, the GUM and GUMS1 approaches have been used extensively and effectively for many years and the reason for their effectiveness relates to the fact that in most situations the distribution $p(\hat{\boldsymbol{\alpha}}|\boldsymbol{\alpha})$ is a good approximation to the desired posterior distribution so that inferences made on this basis will be reliable.

More strongly, assuming an uninformative prior $p(\alpha) \propto 1$, then (a) if the response function $\eta = \phi(\alpha, \boldsymbol{\beta})$ is linear then the LPU approach gives the correct mean and variance of the posterior distribution $p(\alpha|\zeta)$, and (b) if the

https://doi.org/10.1051/ijmqe/2012017 Published online by Cambridge University Press

response function is of the form $\phi(\boldsymbol{\alpha}, \boldsymbol{\beta}) = k_1\alpha + \kappa_2(\boldsymbol{\beta})$, then the MCM approach gives a sample from the posterior distribution $p(\alpha|\zeta)$. If the response function is reasonably constant of the region of interest, then the combined MCM/MCMC approach using the MCM samples as proposals in a Metropolis-Hastings independence chain is proving to be very effective.

Acknowledgements. This work was funded by the UK's National Measurement Office programme for Mathematics and Modelling for Metrology. Colleague Dr. Clare Matthews, NPL, made many helpful suggestions on drafts of this paper.

References

1. JCGM 200, International vocabulary of metrology – basic and general concepts and associated terms (VIM), 2008
2. JCGM 100, Guide to the expression of uncertainty in measurement (GUM), 2008
3. JCGM 101, Guide to the expression of uncertainty in measurement – Supplement 1: Propagation of distribution using the Monte Carlo method (GUMS1), 2008
4. H.S. Migon, D. Gamerman, *Statistical Inference: an Integrated Approach* (Arnold, London, 1999)
5. A. Gelman, J.B. Carlin, H.S. Stern, D.B. Rubin, *Bayesian Data Analysis*, 2nd edn. (Chapman & Hall/CRC, Boca Raton, 2004)
6. D. Gamerman, *Markov Chain Monte Carlo* (Chapman & Hall/CRC, Boca Raton, 1999)
7. G. Grimmett, D. Stirzaker, *Probability and Random Processes*, 3rd edn. (Oxford University Press, Oxford, 2001)
8. C. Elster, W. Wöger, M.G. Cox. Draft GUM Supplement 1 and Bayesian analysis, Metrologia **44**, L31–L32 (2007)
9. C. Elster, B. Toman, Bayesian uncertainty analysis under prior ignorance of the measurand versus analysis using Supplement 1 to the Guide: a comparison. Metrologia **46**, 261–266 (2009)
10. A.B. Forbes, J.A. Sousa, The GUM, Bayesian inference and forward and inverse uncertainty evaluation. Measurement **44**, 1422–1435 (2011)
11. A.B. Forbes, An MCMC algorithm based on GUM Supplement 1 for uncertainty evaluation. Measurement. (in press, DOI: 10.1016/j.measurement.2012.01.018)
12. A.B. Forbes, A two stage MCM/MCMC algorithm for uncertainty evaluation, in *Advanced Mathematical and Computational Tools in Metrology and Testing IX, Göteborg, Sweeden, 2011,* edited by F. Pavese et al. (World Scientific, 2012), pp. 159–170
13. M.G. Cox, P.M. Harris, *Uncertainty Evaluation*, Report No. MS 6 (Software Support for Metrology Best Practice Guide 6) (National Physical Laboratory, Teddington, 2011)
14. J.A. Sousa, A.S. Ribeiro, A.B. Forbes, P.M. Harris, F. Carvalho, L. Bacelar, *The relevance of using a Monte Carlo method to evaluate uncertainty in mass calibration, IMEKO TC3, TC16 and TC22 International Conference Merida, Mexico, 2007*
15. A. Possolo, B. Toman, Assessment of measurement uncertainty via observation equations. Metrologia **44**, 464–475, 2007
16. A.B. Forbes, Nonlinear least squares and Bayesian inference, in *Advanced Mathematical and Computational Tools for Metrology VIII*, edited by F. Pavese et al. (World Scientific, Singapore, 2009), pp. 103–111

https://doi.org/10.1051/ijmqe/2012017 Published online by Cambridge University Press