Numerical aspects in the evaluation of measurement uncertainty

Maurice Cox, Alistair Forbes, Peter Harris, and Clare Matthews

National Physical Laboratory
Teddington, United Kingdom
`maurice.cox@npl.co.uk`

Abstract. Numerical quantification of the results from a measurement uncertainty computation is considered in terms of the inputs to that computation. The primary output is often an approximation to the PDF (probability density function) for the univariate or multivariate measurand (the quantity intended to be measured). All results of interest can be derived from this PDF. We consider uncertainty elicitation, propagation of distributions through a computational model, Bayes' rule and its implementation and other numerical considerations, representation of the PDF for the measurand, and sensitivities of the numerical results with respect to the inputs to the computation. Speculations are made regarding future requirements in the area and relationships to problems in uncertainty quantification for scientific computing.

Keywords: Measurement uncertainty, uncertainty quantification, probability density function, Monte Carlo method, sensitivity measure

1 Introduction

Metrologists at National Metrology Institutes (NMIs) and industrial laboratories routinely propagate uncertainties related to input quantities through mathematical models of measurement to provide uncertainties related to output quantities. A traditional approach uses model linearization and normality assumptions. Relevant guidance is available and supporting software exists.

The Joint Committee for Guides in Metrology (JCGM) is responsible for the GUM, the Guide to the expression of uncertainty in measurement [3], and for supporting documents, e.g., references [2,6]. The JCGM is aware of limitations of the traditional approach. To overcome such limitations, current JCGM activity [4,5] characterizes input quantities by probability density functions (PDFs), which are propagated through the model to obtain a joint PDF for the output quantities. Best estimates, covariance matrices and coverage intervals and coverage regions for the output quantities (or measurand), all used by metrologists, can then be obtained.

Models with PDFs for the input quantities can be seen in a broader setting, where possibly many inputs do not simply relate to measurement. The main consideration here, though, remains measurement uncertainty (MU) quantification,

although many of the principles apply more widely. We observe that the result of a computation represents the effect of uncertainty from all sources considered. Numerical methods of solution, especially Monte Carlo (MC) methods, are used.

We regard a numerical calculation as modeled by $\boldsymbol{Y} = \boldsymbol{f}(\boldsymbol{X})$, sometimes known as a *measurement equation* (ME) [22], where $\boldsymbol{X}$ denotes N input quantities, $\boldsymbol{Y}$ the output quantities, and $\boldsymbol{f}$ is a given function, specified by a computational model, that transforms $\boldsymbol{X}$ to $\boldsymbol{Y}$. When $\boldsymbol{X}$ is uncertain, $\boldsymbol{Y}$ is uncertain. Given knowledge about $\boldsymbol{X}$, knowledge is required about $\boldsymbol{Y}$. Prior knowledge of $\boldsymbol{Y}$ may be available. The components of $\boldsymbol{X}$ are characterized by random variables, and we encode available knowledge about $\boldsymbol{X}$ as a PDF.

$p(\boldsymbol{Z})$ denotes the PDF for a quantity $\boldsymbol{Z}$. $\boldsymbol{z}$ denotes an estimate of $\boldsymbol{Z}$ and $\boldsymbol{U_z}$ the associated covariance matrix (sometimes called uncertainty matrix [4]), taken respectively as the expectation $\boldsymbol{E}(\boldsymbol{Z})$ and covariance $\boldsymbol{V}(\boldsymbol{Z})$. $u(z_i)$ denotes the so-called standard uncertainty associated with the ith component of $\boldsymbol{z}$. $U(z_i)$ denotes an expanded uncertainty corresponding to a stipulated coverage probability p.

Consider so-called *aleatory uncertainties* (due to random effects) and *epistemic uncertainties* (due to other effects). Some authors treat aleatory uncertainties as random variables having PDFs, and epistemic uncertainties as intervals with no assumed PDFs. In metrology we encode knowledge of all quantities by PDFs, a view fully consistent with the GUM [3]. The rules of probability calculus can then be employed. In contrast, in some references, e.g., [24], the two types of uncertainty are propagated separately and the results combined.

Numerical analysis has a long history in uncertainty quantification (UQ) when computing in finite-precision arithmetic. Two principal techniques for error propagation are interval analysis [20] and floating-point (FP) error analysis [30]. We will not consider these techniques here, although we recognize the value of FP error analysis, especially in making statements about the numerical stability of algorithms used within the computational model. We strongly distinguish between errors and uncertainties: an error is the difference between the value of a quantity and the true value of that quantity. An uncertainty is a measure of dispersion (such as the standard deviation determined from the PDF) for that quantity.

This paper reviews numerical tools currently used in MU evaluation. Section 2 makes remarks on the process of assembling and using information about the input quantities, that is, uncertainty elicitation. Section 3 reviews the propagation of distributions through a computational model. Section 4 considers the use of Bayes' rule and Monte Carlo Markov chain (MCMC) methods. Section 5 treats numerical considerations in generating Monte Carlo results and Section 6 considers the representation of such results. Section 7 considers sensitivity issues. Section 8 gives concluding remarks and speculations on tools needed for MU in the future in treating more complicated computational models in metrology.

2   Elicitation

Elicitation is the process of obtaining knowledge of an input quantity and transforming that knowledge to a PDF for that quantity. In metrology if the only knowledge about a component X_i of $\boldsymbol{X}$ is the endpoints of an interval for X_i, we use MaxEnt, the maximum entropy principle, to assign a uniform PDF to X_i [4]. If the only knowledge about $\boldsymbol{X}$ is its expectation $\boldsymbol{E}(\boldsymbol{X})$ and covariance $\boldsymbol{V}(\boldsymbol{X})$, we again use MaxEnt to characterize $\boldsymbol{X}$ by a multinormal distribution. Sometimes Bayes' rule is used to assign a PDF, such as when repeated observations of a quantity are available. Reference [4, subclause 6.4] gives PDF assignments for common circumstances in metrology.

Hibbert et al. [12] applied MaxEnt and Bayesian model selection to a fascinating class of decision-making problems in suspected cases of horse-doping. From a large mass of historical data they constructed PDFs for total carbon dioxide concentration in pre-race samples of plasma. Separate PDFs were obtained for 'clean' horses and horses that were subsequently tested positive. Using q leading moments of the data, MaxEnt delivers a PDF based on a set of Lagrangian parameters [19]. Bayesian model selection was used to obtain that value of q that maximized the Bayesian model probability p, thus avoiding model over-fitting. For clean horses, values of $-\log_{10}(p)$ for $q = 2, \ldots, 7$ were approximately $11, 13, 12, 14, 0, 2$, respectively, leading to the conclusion that Bayesian model selection has strongly settled for a moderately complex model of the form $\exp(a_1 X + \cdots + a_6 X^6)$. This model was chosen in preference to a simple model such as $\exp(a_1 X + a_2 X^2)$, which, for $a_2 < 0$, is Gaussian. Measured data for a further horse can be compared with these PDFs and a decision made on whether the horse has been subjected to doping.

3   Propagation of distributions

Obtaining the joint PDF $p(\boldsymbol{Y})$ for $\boldsymbol{Y}$ given the joint PDF $p(\boldsymbol{X})$ for $\boldsymbol{X}$ is known as 'propagation of distributions' [4]. Formally, it constitutes an application of Markov's theorem [9],

$$p(\boldsymbol{Y}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(\boldsymbol{X}) \delta(\boldsymbol{Y} - \boldsymbol{f}(\boldsymbol{X})) \, \mathrm{d}\boldsymbol{X}, \tag{1}$$

where $\delta(\cdot)$ denotes the Dirac delta function. Figure 1 illustrates the principle for a univariate model with $N = 3$ input quantities. $p(Y)$ is indicated as being asymmetric, as generally arises for non-linear models or asymmetric $p(X_i)$.

A quadrature rule can be used, albeit inefficiently, to evaluate the integral (1), so as to provide an approximation $\widehat{p}(\boldsymbol{Y})$ to $p(\boldsymbol{Y})$. $\widehat{p}(\boldsymbol{Y})$ is often obtained in metrology using an MC method [4]. Random draws are made from $p(\boldsymbol{X})$, $\boldsymbol{f}$ evaluated in each case, and the resulting set of values used to form $\widehat{p}(\boldsymbol{Y})$.

See GUM Supplement 1 [4] for details of an MC method that implements the propagation of distributions. These considerations apply when $\boldsymbol{X}$ does not depend on the measurand.

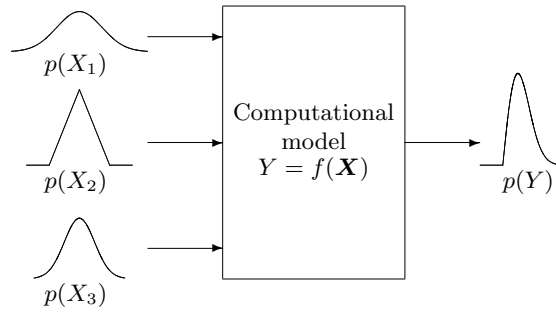**Fig. 1.** Propagation of distributions for $N = 3$ independent input quantities

4   Bayes and MCMC

When observations of a component X_i of $\boldsymbol{X}$ that depend on the value of the measurand are available, an *observation equation* (OE) approach is appropriate. Then Bayes' rule can be used to determine $p(\boldsymbol{Y})$. Let W, one of the X_i, be directly observable. Let $\boldsymbol{X}$ now denote the original $\boldsymbol{X}$ *less* W. Re-express the ME as $\boldsymbol{Y} = \boldsymbol{f}(W, \boldsymbol{X})$ and consider the corresponding OE $W = \phi(\boldsymbol{Y}, \boldsymbol{X})$ [11,22] and observations $W_i \in \mathrm{N}(W, \sigma^2)$. Bayes' rule can be used to update prior knowledge of $\boldsymbol{Y}$, $\boldsymbol{X}$ and σ^2 (regarded as random variables) with observations W_i to give a posterior distribution for these variable, with $p(\boldsymbol{Y})$ obtained by marginalization.

    An MCMC algorithm can be used to obtain $p(\boldsymbol{Y})$. MCMC generates a sequence $\{\boldsymbol{y}_k\}$, in which $\boldsymbol{y}_k$ is obtained from $\boldsymbol{y}_{k-1}$ under an iterative operation. It asymptotically generates draws from $p(\boldsymbol{Y})$. A Metropolis-Hastings algorithm [23, Chapter 6] is an MCMC algorithm that allows $p(\boldsymbol{Y})$ to be specified straightforwardly. One variant, an 'independence chain', can be applied when we have an approximating PDF $\widetilde{p}(\boldsymbol{Y})$ that is easily sampled. Given $\boldsymbol{y}_{k-1}$, $\boldsymbol{y}^*$ is drawn at random from $\widetilde{p}(\boldsymbol{Y})$. Then $\boldsymbol{y}_k$ is set to $\boldsymbol{y}^*$ with acceptance probability

$$\min\{1, r_k\}, \qquad r_k = \frac{p(\boldsymbol{y}^*)\widetilde{p}(\boldsymbol{y}_{k-1})}{p(\boldsymbol{y}_{k-1})\widetilde{p}(\boldsymbol{y}^*)}. \tag{2}$$

To implement the acceptance step, draw v_k from the uniform distribution $\mathrm{R}(0, 1)$ and set $\boldsymbol{y}_k = \boldsymbol{y}^*$ if $v_k < r_k$; otherwise set $\boldsymbol{y}_k = \boldsymbol{y}_{k-1}$. This scheme replicates samples when $\widetilde{p}(\boldsymbol{Y})$ under-represents $p(\boldsymbol{Y})$ at the expense of rejecting samples when $\widetilde{p}(\boldsymbol{Y})$ over-represents $p(\boldsymbol{Y})$. On convergence of the Markov chain, the $\{\boldsymbol{y}_k\}$ are draws from $p(\boldsymbol{Y})$. Implementations typically involve repeat runs (chains) with different initial samples to gauge convergence.

    The ME approach (propagation of distributions with W assigned a PDF based on the sampling distribution) can be analyzed in terms of a Bayesian approach [10]. Let ζ denote an observation. If the OE approach is implemented with prior $p_{\mathrm{OE}}(\boldsymbol{Y}, \boldsymbol{X}) \propto p(\boldsymbol{X})$ to produce posterior distribution $p_{\mathrm{OE}}(\boldsymbol{Y}, \boldsymbol{X}|\zeta)$, then, with $\partial\phi(\boldsymbol{Y}, \boldsymbol{X})/\partial\boldsymbol{Y}$ denoted by $\dot{\phi}(\boldsymbol{Y}, \boldsymbol{X})$, the posterior distribution for

the ME approach is [15]

$$p_{\mathrm{ME}}(\boldsymbol{Y}, \boldsymbol{X}|\zeta) \propto |\dot{\phi}(\boldsymbol{Y}, \boldsymbol{X})| p_{\mathrm{OE}}(\boldsymbol{Y}, \boldsymbol{X}|\zeta),$$

based on a prior distribution

$$p_{\mathrm{ME}}(\boldsymbol{Y}, \boldsymbol{X}) \propto |\dot{\phi}(\boldsymbol{Y}, \boldsymbol{X})| p(\boldsymbol{X}).$$

There will be little difference between the ME and OE approaches if $|\dot{\phi}(\boldsymbol{Y}, \boldsymbol{X})|$ is approximately constant throughout the region of interest.

Consider a Metropolis-Hastings independence chain with $p_{\mathrm{OE}}(\boldsymbol{Y}, \boldsymbol{X}|\zeta)$ playing the role of $p(\boldsymbol{Y})$ and $p_{\mathrm{ME}}(\boldsymbol{Y}, \boldsymbol{X}|\zeta)$ that of $\widetilde{p}(\boldsymbol{Y})$. Expressions (2) then become

$$\min\{1, r_k\}, \qquad r_k = \left| \frac{\dot{\phi}(\boldsymbol{y}_{k-1}, \boldsymbol{x}_{k-1})}{\dot{\phi}(\boldsymbol{y}^*, \boldsymbol{x}^*)} \right|.$$

For models in which $\dot{\phi}(\boldsymbol{Y}, \boldsymbol{X}) \approx$ constant over the region of interest, $r_k \approx 1$ ensuring a high acceptance rate, reflecting the fact that little adjustment has to be made to the sample generated by the ME approach. The chain $\{\boldsymbol{y}_k\}$ will then have low autocorrelation so that sample statistics such as means and standard deviations will converge to their distribution counterparts at rates similar to that for the MC method of GUM Supplement 1.

5   Numerical considerations in generating Monte Carlo results

5.1   Random number generation

In metrology the PDF $p(\boldsymbol{X})$ can often be decomposed into univariate PDFs or joint PDFs involving a smaller number of variables. Procedures for drawing randomly from a variety of PDFs commonly occurring in metrology such as normal, multinormal, t and arcsine are summarized in GUM Supplement 1 [4]. These procedures depend on the quality of an underpinning uniform random number generator (RNG).

An extensive test of the statistical properties of uniform RNGs is carried out by TestU01 [13], a suite containing many individual tests including the so-called Big Crush. Several RNGs passing the Big Crush test are listed by Wichmann and Hill [29], who also considered RNGs on distributed computing systems.

5.2   Practicalities

Rather than carrying out all M Monte Carlo trials and then processing them to obtain a histogram, the following procedure (where an MC trial constitutes making a draw from the input joint PDF and providing the corresponding model value) has benefits in terms of time and memory:

1. Perform a modest number of MC trials, $M_0 = 10^4$, say;

2. Establish a set of bins based on these M trials;
3. Carry out a further $M - M_0$ trials, allocating values to bins or, if a value lies outside the set of bins, storing it *individually.*

The bins, bin frequencies and additional values are used to obtain the required results. This approach is advantageous when computing coverage intervals and regions, which typically involve values in the tails of $p(\boldsymbol{Y})$. Further details are available [8, Annex E].

5.3 Monte Carlo convergence

Suppose M random draws are made from $p(\boldsymbol{X})$ and the corresponding output quantity values are calculated. For any j, the average of these values for the jth component Y_j is a realization of a random variable with expectation $E(Y_j)$ and variance $V(Y_j)/M$. The closeness of agreement between this average and $E(Y_j)$ can be expected to be proportional to $M^{-1/2}$. This 'convergence rate' can be improved using schemes such as Latin Hypercube sampling (LHS) [18] for certain classes of problem.

5.4 Adaptive schemes

The approach in Section 5.2 necessitates specifying M in advance. Thus, the numerical accuracy of the results obtained is unknown *a priori.* The aim of an adaptive scheme is to provide (a) an estimate $\boldsymbol{y}$ of $\boldsymbol{Y}$, (b) an associated covariance matrix $\boldsymbol{V_y}$, and (c) a coverage region for $\boldsymbol{Y}$ for a stipulated coverage probability (or their univariate counterparts for an uncertainty evaluation problem with a single measurand), so as to meet a specified numerical tolerance.

An approach, involving carrying out a sequence of applications of an MC method, is detailed in references [4,5]. It operates in terms of a specified numerical tolerance δ used to assess the 'degree of approximation' required in the elements of the covariance matrix $\boldsymbol{V_y}$. The approach, which utilizes a sequence of batches of, say, $M_0 = 10^4$ MC trials, consists of the following steps:

1. Carry out a batch of MC trials and use the model values to calculate batch results (averages, standard deviations, etc.);
2. Use updating techniques to calculate results for all batches;
3. Regard the computation as having stabilized when the standard deviations of the averages of the batch results is no greater than δ.

The test in step 3 regards the averages as realizations of variables distributed as Student's t, and corresponds to a coverage probability of 95 % [32]. This test is superior to that in reference [4], which is based on regarding the averages as realizations of Gaussian variables.

Adaptive schemes such as that above can be tailored to other sampling procedures such as LHS.

6   Representation of MC results

6.1   General

Suitable representations of MC results are required for (a) visualization pur-
poses, and (b) subsequent use. In terms of (b) the output of one MU evaluation
should be transferable, that is, usable as an input to another evaluation [3]. In
particular it is not always convenient to retain the $M = 10^6$, say, (vector) values
produced by MC and use them subsequently. But, balanced with this, the MC
output is ideal in that it automatically conveys covariance information. Methods
such as kernel density approximation (KDA) can be used for (a) and (b).

6.2   Density approximation

Kernel density estimation (or better approximation) can be used to approximate
a PDF from sampled values from that PDF, but it is not greatly used as yet in
metrological applications. A KDA to a univariate PDF p at the point Y is

$$\widehat{p}_h(Y) = \frac{1}{Mh} \sum_{r=1}^{M} K\left(\frac{Y - y_r}{h}\right).$$

(3)

In expression (3), $y_1, \ldots, y_M$ are sampled values with underlying density p,
and K is a kernel function with unit area. Common kernel functions are Gaussian
and B-spline [27,28]. B-splines have appreciable speed advantages when sam-
pling from $\widehat{p}_h(Y)$ because of their compact support property [7].

In expression (3) h is a smoothing parameter known as the bandwidth and
plays a similar role to that of bin width in a histogram. Too small a value
of h results in spurious oscillatory behavior. Too large a value results in over-
smoothing, losing local detail. A number of methods are used to determine h
given the y_r [27,28]. Figure 2 illustrates such a representation. The bandwidth
in the kernel density approximation and the bin width of the histogram are
identical there.

A (conventional) KDA (3) has the same information content as the data it
represents. With M often of $O(10^6)$, it is possible to produce KDAs with many
fewer terms. It is also possible to describe the sampled values by some parametric
form with adjustable parameters.

If a complete PDF is not required Willink [31] suggests summarizing the
PDF in terms of a model for its quantile function (the inverse of the distribution
function). Willink proposes an asymmetric form of the 'lambda distribution' for
this purpose. The quantile function for this distribution has four free parame-
ters which may be related by non-linear transformations either to the first four
moments of the PDF obtained by Monte Carlo analysis, or, if desired its quan-
tile information. As the extended lambda distribution is defined with respect
to its quantile function, drawing random samples from the resulting model ap-
proximation is straightforward. Should this approximation be inadequate in any
particular case, the more complete KDA can be used.
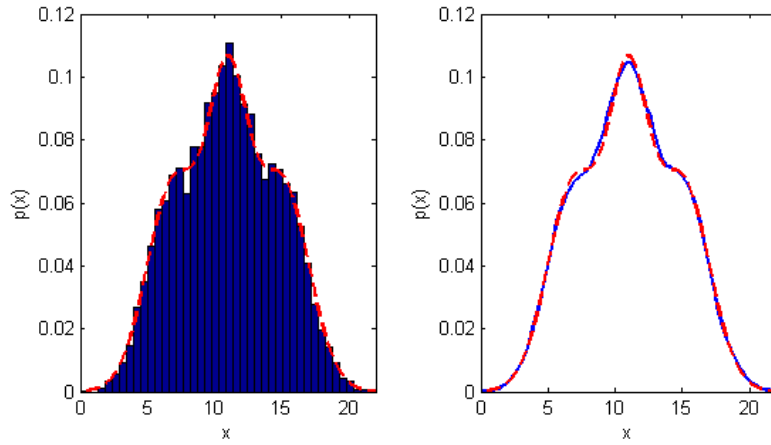
Fig. 2. Representation of a set of observations by a histogram and (right) a KDA, the broken red line showing true density

A bivariate PDF is sometimes represented by a set of contour lines. The contour lines should be faithfully reproduced: as $M \to \infty$, they should converge to the contours of the corresponding PDF. Doing so requires appropriate smoothing [25,28]. Some contour diagrams can be computed directly from a KDA to the corresponding PDF. For others an appropriate smoothing algorithm can be applied to the MC results and the resulting smoothed contours drawn. Figure 3 illustrates the effect of a simple smoothing algorithm.
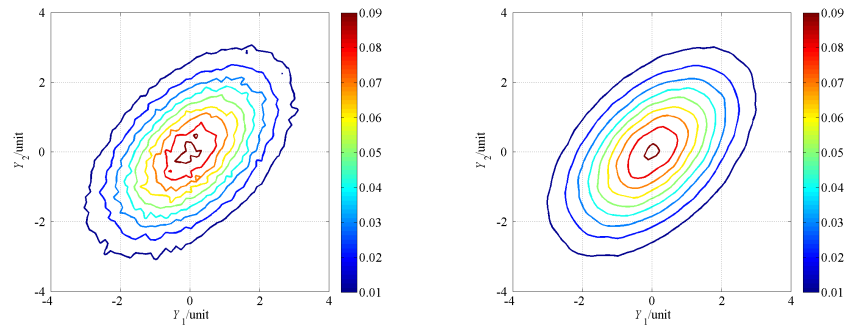
Fig. 3. Contours of the PDF, based on MC results, for a bivariate output quantity and (right) with contour smoothing
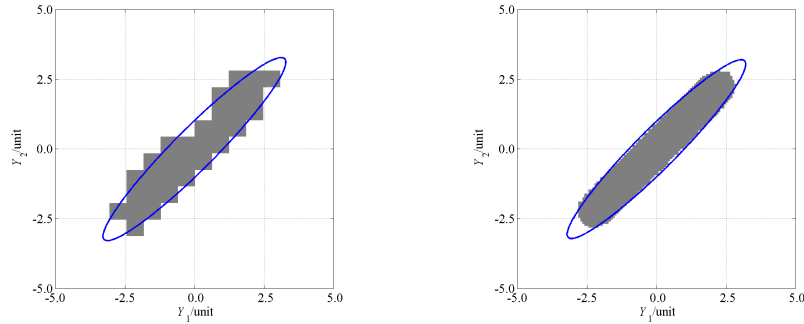
6.3   Coverage regions

In metrology, coverage intervals and regions are often required to accompany measurement results. A procedure included in reference [5], based on work of Possolo [21], for an approximation to the smallest $100p\%$ coverage region is as follows:

1. Construct a (hyper-)rectangular region in the space of the output quantities;
2. Subdivide this rectangular region into a mesh of small rectangles;
3. Assign each output quantity value to the small rectangle containing it;
4. Use the fraction of output quantity values assigned to each rectangle as the approximate probability that Y lies in that rectangle;
5. List the rectangles in terms of decreasing probability;
6. Form the cumulative sum of probabilities for these listed rectangles, stopping when this sum is no smaller than p;
7. Take the selected rectangles as defining the smallest coverage region.

A coverage region that is less disjointed and having a smoother boundary would be expected to be obtained if step 4 were replaced by the use of a more sophisticated approximation to the probability density [26]. In the bivariate case, steps 1 to 4 are also carried out in the initial stages of a typical contouring algorithm used in visualizing an approximation to the PDF for Y.

As a simple example consider the linear bivariate model $Y_1 = X_1 + X_3$, $Y_2 = X_2 + X_3$, with mutually independent $X_1 \sim \mathrm{N}(0, 0.1)$, $X_2 \sim \mathrm{N}(0, 0.1)$ and $X_3 \sim \mathrm{R}(-(5.7)^{1/2}, (5.7)^{1/2})$. Figure 4 shows approximations to the smallest 95% coverage region, obtained using the above procedure, based on a set of rectangles forming a 10×10 and a 100×100 mesh. For comparison the 95% elliptical coverage region for Y based on Gaussian parameters estimated from the model values is shown by a solid line. The assumption of normality yields a conservative coverage region in this example.

Fig. 4. Coverage regions based on a 10×10 mesh and 1 000 points drawn from the PDF for Y, and (right) for a 100×100 mesh and 1 000 000 points

7   Sensitivity

7.1   First-order sensitivity coefficients

Consider a univariate computational model $Y = f(\boldsymbol{X})$. A first-order sensitivity coefficient is $c_i = \partial f / \partial X_i$ evaluated at $\boldsymbol{x}$. Then $u_i(y) = |c_i| u(x_i)$ is the corresponding first-order contribution to $u(y)$. It can readily be obtained using the *complex-step method* [17], which in our opinion deserves greater recognition, particularly when f is complicated. This method is applicable when the real types in the software that implements the model can be replaced by complex types.

The complex-step method is similar to the use of finite differences but complex arithmetic is used to obtain first derivatives. It uses the Taylor expansion of a function f of a complex variable:

$$f(z + w) = \sum_{r=0}^{\infty} \frac{w^r}{r!} f^{(r)}(z),$$

where z and w are complex. Setting $z = x$ and $w = \mathrm{i}h$ where x is real and h is real and small, and taking real and imaginary parts, we have

$$\Re f(x+\mathrm{i}h) = f(x) - \frac{h^2}{2} f''(x) + O(h^4), \qquad \Im f(x+\mathrm{i}h) = hf'(x) - \frac{h^3}{6} f'''(x) + O(h^5),$$

from which

$$f(x) = \Re f(x + \mathrm{i}h), \qquad f'(x) = \frac{1}{h} \Im f(x + \mathrm{i}h),$$

with truncation errors of order h^2. Unlike the use of a finite-difference formula for $f'(x)$, h is chosen to be *very* small with little concern about the loss of significant digits through subtraction cancellation. NPL routinely applies the method with $h = 10^{-100}$ [1], suitable for all but pathologically-scaled problems.

7.2   The use of Monte Carlo

MC can be adapted to provide a non-linear counterpart of a sensitivity coefficient. Hold all input quantities but one, say X_k, at their estimates. Make random draws from $p(X_k)$ and determine an approximation to the PDF for Y that depends only on X_k. The according standard deviation $\widetilde{u}_k(y)$ is an approximation to the component of the standard uncertainty corresponding to X_k.

The use of 'non-linear' sensitivity coefficients in place of 'linear' sensitivity coefficients permits individual non-linear effects to be taken into account. A 'non-linear' sensitivity coefficient $\widetilde{c}_k$ is defined by $\widetilde{c}_k = \widetilde{u}_k(y)/u(x_k)$. It will be equal to the magnitude of the 'linear' sensitivity coefficient c_k for f linear in X_k. The deviation of $\widetilde{u}_k(y)$ from $u_k(y) = |c_k| u(x_k)$ is a measure of the influence of the non-linearity of f with regards to X_k alone. This measure does not account for interaction (i.e., non-additive) terms in f.

8   Concluding remarks and forward look

8.1   Intensive computation beyond GUM

There is in place an infrastructure to deal with MU when expressed as a standard uncertainty or an expanded uncertainty for some coverage probability. The GUM goals of having a universal, internally consistent, and transferable framework are then largely achieved. However, when uncertainty is expressed using PDFs, MU evaluation then relies on (possibly intensive) numerical computation, and generates data that need to be represented in some way. Furthermore, measurement models are becoming more complicated and can only be treated numerically. An example is a radiation-transport calculation, which is itself an MC calculation.

8.2   Need for efficiency

With MU evaluation and uncertainty quantification for scientific computing (UQ) in general there is a need for more efficient techniques for propagating PDFs through computational models. MC is naturally highly parallelizable; NPL uses a grid of PCs to treat the more complex computational models. As all such techniques are based on MC, it is a matter of tuning those techniques appropriately. For some problems the basic technique can hardly be bettered. Approaches such as LHS can give appreciable gains for certain classes of problem.

8.3   Embracing uncertainty quantification

In the future, MU evaluation will embrace more closely concepts used in UQ. Model uncertainty is recognized, being termed *definitional uncertainty* [6]. Elicitation uncertainty is already treated in a small way [3,4]. Numerical uncertainty is considered when computational models constitute FE solvers, for example [16], or in using adaptive schemes [4]. Within general UQ a main aim is to provide probabilistic statements about quantities of interest to inform decision makers. A politician, lawyer or manufacturing production manager needs to consider available evidence and decide a course of action to achieve some goal. Decisions might relate to whether a process remains under statistical control or an athlete is regarded as using a banned substance. Such aspects imply that there are (further) software tools requiring development for MU evaluation.

8.4   Use of MaxEnt

Other metrological applications could benefit from the approach to the horse-doping problem in Section 2. Numerical difficulties, though, can give rise to ill-determined PDFs or prevent a PDF from being obtained at all when MaxEnt is applied to moments as in Section 2 [14]. These difficulties arise when a metrological problem is unreasonable in that it relates to an inadequate measurement of a quantity. It would be useful to have in place a characterization of such problems.

8.5 Efficiency of forms for kernel density approximations

A compact form for an approximation $\widehat{p}(\boldsymbol{Y})$ to the PDF $p(\boldsymbol{Y})$ is desirable when $\widehat{p}(\boldsymbol{Y})$ is to be used as input to a subsequent MU evaluation. The use of kernel density approximations (KDAs) based on B-splines have advantages, delivering computationally efficient representations because of the many properties of B-splines [7]. There is a choice here: either (a) assemble MC results as a histogram and use a KDA, or (b) represent the *ordered* MC results by a suitable approximating cumulative density function (CDF), incorporating monotonicity conditions because of the non-decreasing property of CDFs. For (a), evaluation of the inverse CDF (quantile function), when generating draws from the distribution using a B-spline basis, is very efficient because of the compact support of the B-splines. For (b), the CDF can be differentiated to form the corresponding PDF.

In metrology, the question is raised, and is yet to be addressed, of how calibration certificates of the future can usefully convey results from an MC calculation, particularly in the presence of an asymmetric PDF.

References

1. Al-Mohy, A.H., Higham, N.J.: The complex step approximation to the Fréchet derivative of a matrix function. Numer. Algor. 53, 133–148 (2010)
2. BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, OIML: Evaluation of measurement data — An introduction to the "Guide to the expression of uncertainty in measurement" and related documents. Joint Committee for Guides in Metrology, JCGM 104:2009, `http://www.bipm.org/utils/common/documents/jcgm/JCGM_104_2009_E.pdf`
3. BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, OIML: Evaluation of measurement data — Guide to the expression of uncertainty in measurement. Joint Committee for Guides in Metrology, JCGM 100:2008, `http://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf`
4. BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, OIML: Evaluation of measurement data — Supplement 1 to the "Guide to the expression of uncertainty in measurement" — Propagation of distributions using a Monte Carlo method. Joint Committee for Guides in Metrology, JCGM 101:2008, `http://www.bipm.org/utils/common/documents/jcgm/JCGM_101_2008_E.pdf`
5. BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, OIML: Evaluation of measurement data — Supplement 2 to the "Guide to the expression of uncertainty in measurement" — Models with any number of output quantities. Joint Committee for Guides in Metrology, JCGM 102:2011
6. BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, OIML: International vocabulary of metrology — basic and general concepts and associated terms,. Joint Committee for Guides in Metrology, JCGM 200:2008, `http://www.bipm.org/utils/common/documents/jcgm/JCGM_200_2008.pdf`

7. Cox, M.G.: The numerical evaluation of B-splines. J. Inst. Math. Appl. 10, 134–149 (1972)
8. Cox, M.G., Harris, P.M., Smith, I.M.: Software specifications for uncertainty evaluation. Tech. Rep. MS 7, National Physical Laboratory, Teddington, UK (2010)
9. Cox, M.G., Siebert, B.R.L.: The use of a Monte Carlo method for evaluating uncertainty and expanded uncertainty. Metrologia 43, S178–S188 (2006)
10. Elster, C., Toman, B.: Bayesian uncertainty analysis under prior ignorance of the measurand versus analysis using the Supplement 1 to the Guide: a comparison. Metrologia 46, 261–266 (2009), `http://stacks.iop.org/0026-1394/46/i=3/a=013`
11. Forbes, A.B., Sousa, J.A.: The GUM, Bayesian inference and the observation and measurement equations. Measurement 44, 1422–1435 (2011)
12. Hibbert, D.B., Armstrong, N., Vine, J.H.: Total CO_2 measurements in horses: where to draw the line. Accred. Qual. Assur. 16, 339–345 (2011)
13. L'Ecuyer, P., Simard, R.: TestU01: A software library in ANSI C for empirical testing of random number generators, `http://www.iro.umontreal.ca/~simardr/testu01/tu01.html`
14. Lira, I.: Evaluating the Uncertainty of Measurement. Fundamentals and Practical Guidance. Institute of Physics, Bristol, UK (2002)
15. Lira, I., Grientschnig, D.: Bayesian assessment of uncertainty in metrology: a tutorial. Metrologia 47, R1–R14 (2010), `http://stacks.iop.org/0026-1394/47/i=3/a=R01`
16. Lord, G., Wright, L.: Uncertainty evaluation in continuous modelling. Tech. Rep. CMSC 31/03, National Physical Laboratory, Teddington, UK (2003)
17. Lyness, J.N., Moler, C.B.: Numerical differentiation of analytic functions. SIAM J. Numer. Anal. 4, 202–210 (1967)
18. McKay, M., Conover, W.J., Beckman, R.J.: A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics 42, 55–61 (2000)
19. Mead, L.R., Papanicolaou, N.: Maximum entropy in the problem of moments. J. Math. Phys. pp. 2404–2417 (1984)
20. Moore, R.E.: Interval Analysis. Prentice-Hall, New Jersey (1966)
21. Possolo, A.: Copulas for uncertainty analysis. Metrologia 47, 262–271 (2010)
22. Possolo, A., Toman, B.: Assessment of measurement uncertainty via observation equations. Metrologia 44, 464–475 (2007)
23. Robert, C.P., Casella, G.: Monte Carlo Statistical Methods. Springer-Verlag, New York (1999)
24. Roy, C.J., Oberkampf, W.L.: A comprehensive framework for verification, validation, and uncertainty quantification in scientific computing. Comput. Methods Appl. Mech. Engrg. 200, 2131–2144 (2011)
25. Scott, D.W.: Multivariate Density Estimation: Theory, Practice, and Visualization. John Wiley & Sons, New York (1999)
26. Scott, D.W., Sain, S.R.: Multi-dimensional density estimation. In: Rao, C., Wegman, E. (eds.) Handbook of Statistics, Volume 23, Data Mining and Computational Statistics. pp. 229–261. Elsevier, Amsterdam (2004)
27. Sheather, S.J.: Density estimation. Statist. Sci. 19, 588–597 (2004)
28. Silverman, B.W.: Density Estimation. Chapman and Hall, London (1986)
29. Wichmann, B.A., Hill, I.D.: Generating good pseudo-random numbers. Computational Statistics and Data Analysis 51, 1614–1622 (2006)
30. Wilkinson, J.H.: Notes in Applied Science No. 32. Rounding Errors in Algebraic Processes. Her Majesty's Stationery Office, London (1963)

31. Willink, R.: Representing Monte Carlo output distributions for transferability in uncertainty analysis: modelling with quantile functions. Metrologia 46, 154–166 (2009), `http://stacks.iop.org/0026-1394/46/i=3/a=002`
32. Wübbeler, G., Harris, P.M., Cox, M.G., Elster, C.: A two-stage procedure for determining the number of trials in the application of a Monte Carlo method for uncertainty evaluation. Metrologia 47, 317–324 (2010), `http://stacks.iop.org/0026-1394/47/i=3/a=023`

DISCUSSION

Speaker: Maurice Cox

Mark Campanelli : Has there been any discussion on a "bias discovery" methodology for metrology?

Maurice Cox : Bias is an important concept in measurement. In the *Guide to the expression of uncertainty in measurement* (GUM) [3], concentration is given to *systematic error*, and its characterization by a probability density function (PDF). This distribution is obtained by accounting for available knowledge of the relevant quantity. Bias is regarded as an estimate of a systematic error. In an interlaboratory comparison, nominally the same measurand is measured by participating laboratories and a consensus value determined. The deviations of the measured values provided by the laboratories from the consensus value can be regarded as laboratory biases.

Mark Campanelli : How can the reliability of the Monte Carlo method be checked for computing 99 % coverage intervals for high-dimensional problems?

Maurice Cox : For multivariate problems an adaptive method is given in Supplement 2 to the GUM [6], as outlined in our paper. It constitutes a simple extension of the adaptive scheme in Supplement 1 to the GUM [4]. Further experience needs to be gained before its performance can be judged in obtaining 99 % coverage intervals and 99 % coverage regions. At the moment these guidance documents apply to simple Monte Carlo sampling; different considerations might apply to Latin Hypercube sampling or importance sampling.

Richard Hanson : The use of pure complex "divided differences" requires an algorithm for complex continuation of the function. Is this the case for your application?

Maurice Cox : In metrology, the measurement function, when expressed in terms of complex variables, is usually analytic, that is, the Cauchy-Riemann equations apply. In such a case, the replacement of real type by complex types is virtually all that needs to be done. When, rarely, the measurement function is specified in terms of functions that are not analytic (such as the absolute value function, which is not analytic at the origin), the measurement function should be re-expressed so that the Cauchy-Riemann equations are satisfied.

Jeffrey Fong : For problems involving controls based on "uncooked" models (i.e., models without a history of baseline data and experimental validation), and with small M ($\sim 10^2 - 10^3$) do you have any advice on using the approach you outlined in your talk?

Maurice Cox : The Monte Carlo method given in guidance documents [4,6] is simple and is not intended for small values of M. If the model is complicated, perhaps involving the solution of a large finite-element model, because of large computing times it may not be possible to use a sufficiently large value of M to

obtain adequate distributional knowledge of the output quantity. In such a case we give very elementary advice. It involves the use of an approximate approach in which the PDF for the output quantity is regarded as Gaussian (as in the GUM). A value of M would be selected that is economically possible. The average and covariance matrix of the resulting M values of the (vector) output quantity would be taken respectively as the best estimate and covariance matrix of that quantity. A Gaussian PDF with these parameters would be used to characterize the output quantity and a desired coverage region calculated accordingly. Although this use of a small value of M is inevitably less reliable than that of a large value in that it does not provide an approximation to the PDF for the output quantity, it does take account of model non-linearity.

Tony O'Hagan : I suggest that a variance-based sensitivity analysis would be much more useful than derivatives. This is discussed, and efficient computation with GP emulators developed in: Oakley, J. and O'Hagan, A. (2004), Probabilistic sensitivity analysis of complex models : a Bayesian approach, *J. Royal Statistical Soc., series B*, vol 66, 751–769.

Maurice Cox : We do not use derivatives alone, but in conjunction with the input standard deviations. Thus, we consider, in the notation of our paper, the contributions $u_i(y) = |c_i|u(x_i)$. These terms are the first-order contributions to the standard deviation of the output quantity, and do not account for interaction effects. They are used by metrologists as indications of the input quantities to which most attention should be paid when there is a need to meet a target uncertainty. Nevertheless, we intend to examine variance-based sensitivity analysis.

Jon Helton : Is appropriate consideration given to the problem/inconsistencies that are present when variables with bounded ranges are represented with distributions that have infinite ranges (e.g., normal distributions)?

Maurice Cox : Attention is paid to the input quantities by characterizing them by PDFs that account for knowledge of those quantities. In my opinion too much reliance is placed on the central limit theorem in representing the output quantity by a Gaussian distribution. Inconsistencies arise when it it is known that the output quantity is bounded, such as a mass concentration lying between 0% and 100%, and a coverage interval extends beyond one of these limits. The Joint Committee for Guides in Metrology, in its work on revising the GUM, will take account of such knowledge in recommending methods based on forming a suitable prior distribution and applying Bayes' theorem rather than using the propagation of distributions in such a case.

Jon Helton : Has the use of importance sampling been considered for use in the determination of the extreme tails of distributions?

Maurice Cox : The relevance of importance sampling is recognized, and it is intended that, when sufficient experience on metrology problems has been gained, appropriate guidance will be given.