Up a GUM tree? Try the Full Monte!

by
Maurice Cox,[1] Peter Harris[2]
Centre for Mathematics & Scientific Computing
National Physical Laboratory
Teddington
Middlesex TW11 0LW

The Guide to the Expression of Uncertainty in Measurement (GUM), published by ISO, is a key document used by National Measurement Institutes and industrial calibration laboratories as the basis of evaluating the uncertainty in the output of a measurement system.

The system is modelled using a functional relationship between measured quantities $\mathbf{x} = \{x_i\}$ (the inputs) and the measurement result y (the output) in the form

$$y = f(\boldsymbol{x}).$$

The (combined) standard uncertainty of y is then evaluated from

$$u^2(y) = \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial f}{\partial x_i}\frac{\partial f}{\partial x_j} u(x_i, x_j)$$

where the quantities $\partial f/\partial x_i$ are referred to as sensitivity coefficients, and $u(x_i, x_j)$ is the covariance of x_i and x_j, with $u(x_i, x_i) = u^2(x_i)$, the variance of x_i. The GUM recommends that the uncertainty in the measurement result y is expressed as a confidence interval at some probability level (typically 95%). The half-width of this interval is the expanded standard uncertainty $U(y)$ obtained as the product of $u(y)$ and a coverage factor k.

To apply the GUM approach two main assumptions must hold:
- The adequacy of the formula for $u(y)$ which is derived by propagating uncertainties in a first-order approximation to the model of the measurement system.
- The distribution of y is known, e.g., Gaussian or Student's-*t*, in order to obtain the value of k.

Some GUM users are experiencing problems in meeting these conditions, although it is important to confirm that they apply for a given application.

NPL is involved, both internally and in collaboration with other National Measurement Institutes and standards bodies, in activities to promote sound methods of uncertainty evaluation. Although the GUM is a far-reaching document, and is appropriate for many applications, NPL's work includes presenting extensions and enhancements to the GUM to cover situations in which the assumptions indicated above do not apply or are untested. It is also an intention to provide 'easy-to-use' computer-based approaches.

Sampling techniques, such as Monte Carlo simulation, provide an alternative approach to uncertainty evaluation in which the propagation of uncertainties is undertaken *numerically* rather than analytically. Such techniques are useful for validating the results returned by the application of the GUM, as well as in circumstances where the assumptions made by the GUM do not apply. In fact, these techniques are able to provide much richer information, by propagating the *distributions* (rather than just the uncertainties) of the inputs $\mathbf{x}$ through the measurement model f to provide the distribution of the output y. From the output distribution confidence intervals or confidence regions (in the multivariate case) can be produced, as can other statistical information.

The first figure shows how sampling techniques can be used to evaluate the uncertainty $u(y)$.

[1] Tel. (+44) (0) 20 8943 6096, e-mail: maurice.cox@npl.co.uk
[2] Tel. (+44) (0) 20 8943 9661, e-mail: peter.harris@npl.co.uk

STANDARD UNCERTAINTY OF THE OUTPUT

Form N samples $\mathbf{x}_k$ of the measured quantities $\mathbf{x}$.

For k = 1 to N, evaluate y_k from $y_k = f(\mathbf{x}_k)$.

Evaluate sample standard deviation for the y_k to give $u(y)$.

The input data $\{\mathbf{x}_k\}$ for this process can be provided in a number of ways. One way is to generate random samples $\{\mathbf{x}_k\}$ from the (possibly joint) probability distribution for the inputs $\mathbf{x}$. For example, $\mathbf{x}$ may be described by a multivariate Gaussian distribution with prescribed mean vector and covariance matrix. Alternatively, the components of $\mathbf{x}$ may be independent and each follow a given univariate Gaussian or uniform distribution. The application of the sampling scheme corresponds to a 'Monte Carlo' simulation (MCS). The scheme is not difficult to implement in general terms. Just the model and the input distributions (and not also its sensitivity coefficients as required by the GUM approach) are needed.

Sampling techniques can also form the basis for calculating expanded uncertainties or confidence intervals. Given the y_k as determined in the first figure, the second figure shows how to evaluate the 95% confidence interval for the measurement result y:

CONFIDENCE INTERVAL FOR THE OUTPUT

Sort the values $\{y_k\}$ into non-decreasing order.

Form the 2.5-and 97.5-percentiles in this list to define the required confidence interval.

In the use of MCS to validate the results produced by the GUM in any individual case, the GUM results can be accepted if the resulting uncertainties agree to within, say, two significant figures (which is adequate for most purposes). If such agreement is not observed then either a mistake has been made in applying the GUM (e.g., in determining the sensitivity coefficients) *or* the conditions for its application do not hold. If the latter applies, it would be appropriate to regard the MCS results as being scientifically more sound.

Both GUM and MCS work with the same model and input distributions, and the quality of the results obtained depends on that of this information. Moreover, MCS requires a value to be specified for N, the number of Monte Carlo trials. We have found N = 100,000 to be satisfactory in many (but not all) cases.

Figures 1 and 2 illustrate the use of the GUM and MCS for the simple model $y = x^2$. Figure 1 shows the sampling distribution for the input x which is chosen to be Gaussian with mean 0.5 and standard deviation 0.2. Figure 2 shows the corresponding sampling distribution for y. Each distribution is based on a sampling size of N = 10,000. We also indicate in Figure 2 the 95% confidence intervals evaluated according to the GUM (broken lines) and using MCS (solid lines) as described above. The fact that these intervals are appreciably different reflects the non-Gaussian behaviour of the output y. The 95% confidence interval evaluated according to the GUM is clearly not reliable because it includes infeasible (negative) values of the output y.

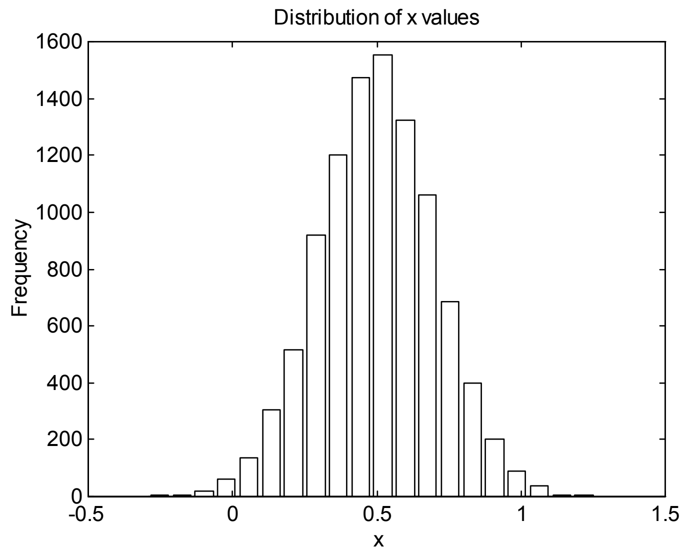

Figure1: Sampling distribution for input *x*.

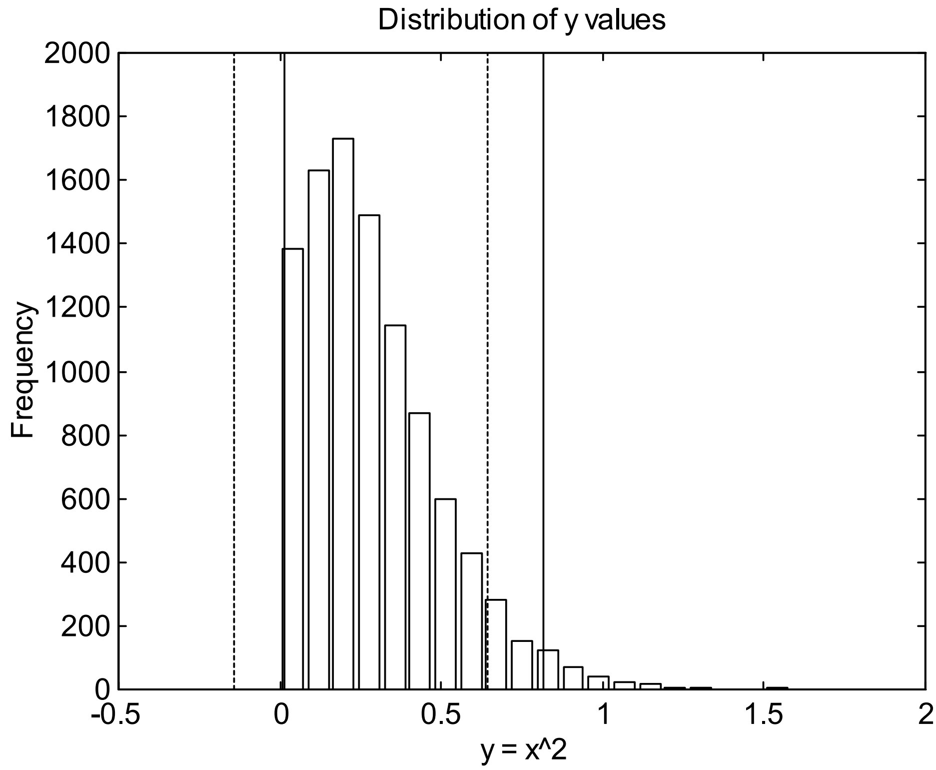

Figure 2: Sampling distribution for output $y = x^2$. The 95% confidence intervals evaluated according to the GUM (broken lines) and using MCS (solid lines) are also shown.

This article first appeared in Counting on IT Issue 8, available from NPL.

National Physical Laboratory, Teddington, Middlesex, UK, TW11 0LW