NPL
National Physical Laboratory

REPORT

NPL Report MS 6

Software Support for Metrology Best Practice Guide No. 6

Uncertainty Evaluation

M G Cox and P M Harris

March 2010

National Physical Laboratory, Hampton Road, Teddington, Middlesex, United Kingdom, TW11 0LW
Switchboard 020 8977 3222 NPL Helpline 020 8943 6880 Fax 020 8614 0446 E-mail enquiry@npl.co.uk www.npl.co.uk

NPL Report MS 6

Software Support for Metrology
Best Practice Guide No. 6

Uncertainty Evaluation

M G Cox and P M Harris
Optical and Mathematical Techniques for Metrology

March 2010

ABSTRACT

This guide provides best practice in the evaluation of uncertainty within metrology, and in the support to this topic given by a treatment that is probabilistic. It is motivated by two principle considerations. The first is that although the primary guide on uncertainty evaluation, the 'Guide to the expression of uncertainty in measurement' (GUM), can be expected to be very widely applicable, the approach it predominantly endorses contains some limitations. The other is that on the basis of the authors' considerable contact with practitioners in the metrology community it is evident that important classes of problem are encountered that are subject to these limitations. A further consideration is that measurement models are encountered in practice that lie outside the scope of the model type (viz., a measurement function with a single (scalar) output quantity) that is the focus of the presentation given in the GUM.

Central to consideration is the need to carry out uncertainty evaluation in as scientific a manner as economically possible. Although several approaches to uncertainty evaluation exist, the GUM has been very widely adopted (and is strongly supported by the authors of this guide). The emphasis of this guide is on making good use of the GUM, on aspects that yield greater generality, and especially on the provision in some cases of measurement uncertainties that are more objectively based and numerically more sustainable. The guide is also concerned with validating the current usage of the GUM in circumstances where there is doubt concerning its applicability. The relationship of this guide to the work being carried out by the Joint Committee on Guides in Metrology to prepare a new edition of the GUM and to provide supporting documents to the GUM is indicated.

NPL Report MS 6

ISSN 1754–2960

National Physical Laboratory,
Hampton Road, Teddington, Middlesex, United Kingdom TW11 0LW

We gratefully acknowledge the financial support of the UK Department for Business, Innovation and Skills (National Measurement Office)

Approved on behalf of the Managing Director, NPL by Jonathan Williams,
Knowledge Leader for the Optical Technologies and Scientific Computing Division

Contents

Chapter 1

Scope

1.1 Structure of the guide

This guide provides information relating to:

1. The objective of uncertainty evaluation (Chapter 2);
2. A statement of the main problem addressed, viz., using the available information to determine the probability distribution for the quantity intended to be measured, expressed using the propagation of distributions or Bayes' theorem (Chapter 3);
3. The main stages of uncertainty evaluation (Chapter 4);
4. Approaches to the propagation of distributions for uncertainty evaluation and particularly for determining a coverage interval for the quantity intended to be measured (Chapter 5);
5. A classification of measurement models and guidance on the application to these models of the law of propagation of uncertainty (Chapter 6);
6. Details of a general numerical procedure, a Monte Carlo method, as an implementation of the propagation of distributions for uncertainty evaluation (Chapter 7);
7. A facility that enables the results of the law of propagation of uncertainty and the assumptions of the central limit theorem to be validated, thus providing assurance that that approach can legitimately continue to be used in appropriate circumstances (Chapter 8);
8. Examples to illustrate the various aspects of this guide (Chapter 9);
9. General recommendations regarding measurement uncertainty evaluation (Chapter 10);

10. Statistical concepts applied throughout the guide (Appendix A);

11. Approaches to evaluating sensitivity coefficients (Appendix B).

1.2 Summary

This guide provides best practice in the evaluation of uncertainty within metrology, and in the support to this discipline given by a treatment that is probabilistic. Central to consideration is a *measurement model*, having input quantities about which information is available, and an output quantity about which information is required. The input quantities relate to indications provided by a measuring instrument or measuring system or to information obtained from sources such as manufacturers' specifications, calibration certificates and tabulated data. The output quantity represents a well-defined physical quantity intended to be measured (sometimes known as the *measurand*).[1] The objective of uncertainty evaluation is to quantify the extent and nature of the knowledge of the output quantity given the measurement model, including knowledge of the input quantities to account for the nature of their inexactness.[2] Knowledge of the input quantities is encoded by probability density functions (PDFs) used to characterize those quantities.[3] Knowledge of the output quantity is determined by deriving a PDF for that quantity. For this reason, the problem of uncertainty evaluation can be formulated as one of *propagating distributions* through the measurement model. A main requirement is to ascribe to the output quantity a so-called *coverage interval* that contains a specified proportion, e.g., $95\,\%$, of the distribution of values that could reasonably be attributed to that quantity.[4]

The key document in the area of uncertainty evaluation is the 'Guide to the expression of uncertainty in measurement' (GUM) [10]. The GUM provides a procedure, summarized in GUM Clause 8 and Section 5.3 of this guide, for evaluating uncertainty that has been adopted by many bodies. The procedure is based on summarizing the input quantities by estimates and standard uncertainties associated with the estimates. The estimates and the associated uncertainties are 'propagated' through (a linearized version of) the measurement model to provide an estimate of the output quantity and the associated standard uncertainty. A means for obtaining a coverage interval for the output quantity is provided. The procedure also accounts for the correlation effects that arise if the input quantities are mutually dependent. The complete procedure, involving (a) the application of the *law of propagation of uncertainty* to obtain an estimate of the output quantity and the associated standard

[1] In some instances the output quantities may not individually have physical meaning. An example is the set of coefficients in a polynomial representation of a calibration function. Together, however, the set of quantities (coefficients) define a meaningful entity, the calibration function.

[2] Model validation, viz., the process of ascertaining the extent to which the model is adequate, is not treated in this guide.

[3] The encoding of knowledge by a PDF may be on the basis of a (statistical) analysis of a set of indication values for the input quantities, referred to in the GUM as a 'Type A evaluation', or by some other means, referred to in the GUM as a 'Type B'evaluation.

[4] There may be more than one output quantity, in which case a *coverage region* may be required.

uncertainty, and (b) the assumptions of the *central limit theorem* to obtain a coverage interval, is one implementation of the propagation of distributions. Henceforth, we will refer to the procedure summarized in GUM Clause 8 as the *GUM uncertainty framework*. This is consistent with the way the term is used within the first Supplement [11] to the GUM.

In order to make the GUM more immediately applicable to a wider range of problems, a classification of measurement models is provided in this guide. The classification is based on whether

- there is one or more than one output quantity,
- it is possible or not to write the measurement model as a *measurement function* in which the output quantity is expressed as a formula involving the input quantities, and
- the quantities within the measurement model are real or complex, the latter arising particularly in electrical, acoustical and optical metrology.

Guidance on uncertainty evaluation based on the law of propagation of uncertainty is provided for each measurement model type within the classification.

The measurement model employed in the GUM is an *input-output model*. For relatively simple measurements, this form can straightforwardly be obtained. In other cases, this form does not arise immediately, and must be derived. Consideration is therefore given to *statistical modelling*, a process that relates the quantities of which the measurement data are instances to the input and output quantities and other available information, such as from calibration certificates. This form of modelling can then be translated into the 'GUM model', in which the knowledge of the input quantities is summarized by estimates and associated uncertainties. Statistical modelling also covers the probabilistic and statistical analysis of the input quantities.

Although the GUM as a whole is a very rich document, there is much evidence that the GUM uncertainty framework is the approach that is adopted by most practitioners as an implementation of the propagation of distributions. It is therefore vital that the fitness for purpose of this approach (and of any other approach) is assessed, generally and in individual applications. There are some limitations and assumptions inherent in the GUM uncertainty framework and there are applications in metrology in which users of the GUM are unclear whether the limitations apply or the assumptions can be expected to hold in their circumstances. In particular, the limitations and assumptions at the basis of the 'easy-to-use' formula inherent in the law of propagation of uncertainty are highlighted.

The GUM (in Clause G.1.5) does permit 'other analytical or numerical methods' to be employed. However, if such methods are to be used they must have certain credentials in order to permit them to be applied in a sensible way. Part of this guide is concerned with such methods, their properties and their credentials.

It is natural, in examining the credentials of any alternative scientific approach, to re-visit established techniques to confirm or otherwise their appropriateness. In that sense it is appropriate to re-examine the principles of the GUM uncertainty framework to discern whether they are fit for purpose. This task is not possible as a single 'general health check'. The reason is that there are circumstances when the principles of the GUM uncertainty framework cannot be bettered by any other candidate technique, but there are others when the quality of the approach is not quantified. The circumstances in which the GUM uncertainty framework is unsurpassed are when the measurement function relating the input quantities $X_1, \ldots, X_N$ to the output quantity Y is *linear*, viz.,

$$Y = c_1 X_1 + \cdots + c_N X_N,$$

for any constants $c_1, \ldots, c_N$, any value of N, however large or small, and when the input quantities X_i are characterized by Gaussian distributions.[5] In other circumstances, the GUM uncertainty framework generally provides an approximate solution: the quality of the approximation depends on the measurement model, the estimates of its input quantities and the magnitudes of the uncertainties associated with the estimates. The approximation may in many cases be perfectly acceptable for practical application. In some circumstances this may not be so. See Clause G.6.6 of the GUM.

The concept of a *measurement model* remains central to alternative approaches to implementing the propagation of distributions. This guide advocates the use of an alternative approach in circumstances where there is doubt concerning the applicability of the GUM uncertainty framework. Guidance is provided for this approach. The approach is *numerical*, being based on a Monte Carlo method. It is thus computationally intensive, but nevertheless the calculation times taken are often only seconds or sometimes minutes on a PC, unless the measurement model is especially complicated.

It is shown how the alternative approach can also be used to *validate* the GUM uncertainty framework and thus in any specific application confirm (or otherwise) that this use of the GUM is *fit for purpose*, a central requirement of the Quality Management Systems operated by many organizations. In instances where the approach indicates that the use of the GUM uncertainty framework is invalid, the approach can itself subsequently be used for uncertainty evaluation, in place of the GUM uncertainty framework, in that it is consistent with the general principles (Clause G.1.5) of the GUM.

An overall attitude taken to uncertainty evaluation in this guide is that it consists of several stages. The first stage, *formulation*, constitutes building the measurement model and quantifying probabilistically the knowledge of its input quantities. The second stage, *propagation*, consists of using this information to quantify probabilistically the knowledge of the output quantity. The final stage, *summarizing*, involves obtaining from this information about the output quantity the required results, including an estimate of the output quantity, the associated standard uncertainty and a coverage interval containing the output quantity with a specified probability.

[5] The GUM uncertainty framework also applies when some or all of the input quantities are mutually dependent, and they are characterized by a joint (multivariate) Gaussian distribution.

The focus of this guide is on formulating the problem of uncertainty evaluation as one of propagating probability distributions through a measurement model and approaches to solving that problem. In the formulated problem it is assumed that there is no prior knowledge about the output quantity Y. However, in many cases such prior knowledge will exist and can be used to characterize Y by a prior probability distribution.[6] Such additional information can be used to provide a probability distribution for Y that can give a smaller standard deviation for Y and hence a smaller standard uncertainty associated with the estimate of Y [41]. The use of Bayes' theorem provides an approach to uncertainty evaluation that allows proper account to be taken of prior knowledge about the output quantity [47, 87].

The starting point for a Bayesian analysis is an *observation model* [80] (rather than the measurement function of the GUM), which expresses indication quantities in terms of other quantities (including the measurand) involved in the measurement. Bayes' theorem is used to combine knowledge about the indication quantities, which is in terms of indication values and sampling distributions for those quantities, with knowledge about the other quantities involved in the measurement, which is in the form of (prior) probability distributions. The analysis provides a posterior distribution for the measurand, which can be summarized in terms of an estimate, associated standard uncertainty and coverage intervals. Some information about the use and application of Bayes' theorem is given in this guide, including some comments about the relationship of the approach to the propagation of distributions [42, 43, 68].

Readers of this guide will benefit from reasonable familiarity with the GUM or the related UKAS document M3003 [100]. A companion document [34] provides specifications of relevant software for uncertainty evaluation when applying some of the principles considered here. See also [33, 45, 77, 107].

1.3 Joint Committee for Guides in Metrology

Currently, work related to the GUM is taking place under the auspices of the Joint Committee for Guides in Metrology (JCGM).[7] This work is concerned with amplifying and emphasizing key aspects of the GUM in order to make the GUM more readily usable and more widely applicable. Revision by the JCGM of the GUM itself has started in parallel with work on preparing the following supporting documents (including *Supplements*) to the

[6] Some examples of prior knowledge relate to physical constraints on values of the output quantity, e.g., a chemical concentration must take values between zero and one, mass and length are quantities that take values that are non-negative, etc. It can be argued that such knowledge *always* exists about the output quantity although it might be quite 'vague'. For example, a metre rule rather than a micrometer is used to measure a dimension of an artefact because the metrologist has some expectation about the value of the dimension, e.g., that it is $200\,\mathrm{mm}$ and not $200\,\mu\mathrm{m}$. The information is called prior knowledge because it exists before any measurements are made.

[7] The Web address of the JCGM is www.bipm.fr/en/committees/jc/jcgm/.

GUM [8]:[8]

1. JCGM 101:2008 — Supplement 1 to the 'Guide to the expression of uncertainty in measurement' — Propagation of distributions using a Monte carlo method;
2. JCGM 102 — Supplement 2 to the 'Guide to the expression of uncertainty in measurement' — Models with any number of output quantities;
3. JCGM 103 — Supplement 3 to the 'Guide to the expression of uncertainty in measurement' — Modelling;
4. JCGM 104:2009 — An introduction to the 'Guide to the expression of uncertainty in measurement' and related documents;
5. JCGM 105 — Concepts and basic principles;
6. JCGM 106 — The role of measurement uncertainty in conformity assessment;
7. JCGM 107 — Applications of the least-squares method.

The approaches to uncertainty evaluation presented here are consistent with the developments by the JCGM in this respect. This guide will be updated periodically to account for the work of this committee (Section 1.4). It will also account for the work of standards committees concerned with various aspects of measurement uncertainty, awareness of requirements in the areas indicated by workshops, etc., organized within the Software Support for Metrology (SS*f*M) programme and elsewhere, and technical developments. The authors of this guide provide input to Working Group 1, 'Expression of Uncertainty in Measurement', of the JCGM that is concerned with the GUM, and also to other relevant national or international committees, including British Standards Committee Panel SS/6/-/3, 'Measurement Uncertainty', and ISO/TC 69/SC 6, 'Measurement Methods and Results'.

1.4 Document history

The first edition of this guide was published in March 2001, having been developed during the first SS*f*M programme (April 1998 to March 2001). During that period Working Group 1, 'Expression of Uncertainty in Measurement', of the JCGM, started work, following its first meeting in March 2000, on the first Supplement [11] to the GUM concerned with a Monte Carlo method for the propagation of distributions. Material from the evolving best-practice guide was used in various parts of the Supplement and subsequently refined appropriately for consistency with the published Supplement.

[8] Those documents with associated dates have been published and may be downloaded from www.bipm.org/en/publications/guides/gum.html. The other documents are under development by Working Group 1, 'Expression of Uncertainty in Measurement', of the JCGM.

The second edition was published in March 2004, following revision during the second SS*f*M programme (April 2001 to March 2004). In this second edition, material from the drafts of the Supplement prepared during that period that had an origin in the first edition of the guide were re-used.

The third edition was published in September 2006, following revision during the third SS*f*M programme (April 2004 to March 2007). A main change from the second edition concerned a revision of Chapter 9 *Examples* to include descriptions of new examples and case studies undertaken during the third SS*f*M programme. The new examples included three examples (concerned with mass calibration, comparison loss in microwave power meter calibration and gauge block calibration) that are included as examples within the first Supplement [11] as well as three case studies (concerned with quantities subject to a normalisation constraint, area under a curve and SIR efficiency) undertaken in collaboration with other NMS programmes.

The current document (fourth edition), produced during the fourth SS*f*M programme (April 2007 to March 2010), reflects the further work of Working Group 1 of the JCGM. In particular, during that period the Working Group published an introduction to the GUM and related documents [12], and completed a draft of the second Supplement to the GUM concerned with measurement models with any number of output quantities. The Working Group also started its work to prepare an new edition of the GUM itself.

1.5 Acknowledgements

The guide has benefited from many sources of information. These include:

- The Joint Committee for Guides in Metrology;
- National and international standards committees;
- Consultative Committees of the Comitè International des Poids et Mesures (CIPM);
- National Metrology Institutes (outside the UK);
- The (UK) Royal Statistical Society;
- NPL Scientific Groups;
- The United Kingdom Accreditation Service;
- LGC Limited;
- The National Engineering Laboratory;
- The Numerical Algorithms Group Ltd;

- UK industry;
- SS*f*M workshops;
- Conferences in the Advanced Mathematical and Computational Tools in Metrology series [18, 19, 20, 21, 22, 23, 24, 79];
- Literature on uncertainty, statistics and statistical modelling;
- Many individual contacts.

Chapter 2

Introduction

2.1 Background

Measured values are not perfect. When a quantity is measured by one instrument, the value obtained will generally be different from that provided by another measuring instrument. If that quantity were to be measured a number of times by the same instrument, in the same way and in the same circumstances, a different (indication) value each time would in general be obtained.[1] These repeated indication values would form a 'cluster', the 'size' of which would depend on the nature and quality of the measurement. The 'centre' of the cluster would provide an estimate of the quantity that generally can be expected to be more reliable than individual indication values. The 'size' of the cluster would provide quantitative information relating to the quality of this central value as an estimate of the quantity. It will not furnish all the information of this type, however. The measuring instrument is likely to provide values that are influenced by one or more systematic errors.

As an illustration of a systematic error, consider domestic bathroom scales. If they are not set such that the display reads zero when there is nobody on the scales, when used to weigh a person or an object the indicated weight can be expected to be offset from what it should be. No matter how many times the person's weight is taken and a central value determined, because the scatter of values would be centred on an offset value, the effect of this offset is inherently present in the central value. See figure 2.1.

There are thus two types of error, in this example and in general. The first is a 'random' error associated with the fact that when a measurement is repeated each indication value will generally be different from the previous value. It is random in that there is no way to predict from previous indication values exactly what the next one would be.[2] The second is a

[1] This statement assumes that the recording device has sufficient resolution to distinguish between different values.

[2] If a prediction were possible, allowance for the effect could be made!

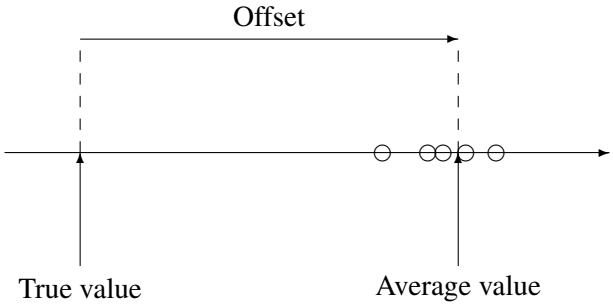

Figure 2.1: Illustration of random and systematic errors. The average of the indication values (shown as circles) is offset from the true value of the quantity. The offset relates to a systematic error, and the dispersion of the indication values about the average to random errors.

systematic error (an estimate of which is the measurement bias) associated with the fact that the indication values contain an offset. In practice there can be a number of contributions to the random error and to the systematic error, both in this situation and in many other situations. Depending on the application, the random error may dominate, the systematic error may dominate or the errors may be comparable in magnitude.

In order to make a statement concerning the measurement of a quantity it is typically required to provide an estimate of that quantity and an associated uncertainty. The estimate is (ideally) a 'best estimate' of the quantity and the uncertainty a numerical measure of the quality of the estimate.

The above discussion concerns the direct measurement of a quantity. However, the quantity *actually* measured by the device or instrument used is rarely that *intended* to be measured. For the bathroom scales the quantity measured might relate to the extension of a spring in the scales whereas the quantity intended to be measured is the weight of the person on the scales. The indication value is therefore converted or transformed into a form to provide the required (output) quantity value. For a perfect (linear) spring, the conversion is straightforward, being based on the fact that the required weight is proportional to the extension of the spring. The display on the scales constitutes a graduation or calibration of the device. For a domestic mercury thermometer, the indication value is the height of a column of mercury. The height is converted into a temperature value using another proportional relationship: a change in the height of the column is proportional to the change in temperature, again a calibration.

A relationship of types such as these constitutes a rule for converting the indication value into the output quantity value. In metrology, there are very many different types of measurement and therefore different rules. Even for one particular type of measurement there may well be more than one rule, perhaps a simple rule (e.g., a proportional rule) for everyday

domestic use, and a sophisticated rule involving more complicated calculations (a nonlinear rule, perhaps) that is capable of delivering more accurate results for industrial or laboratory purposes.

The situation is frequently more general in another way. There is often a number of different measured values that contribute to the output quantity value. Here, the concern is not simply repeated indication values, but intrinsically different measured values, e.g., some relating to temperature and some to displacement. Also, there may be more than one output quantity. For instance, by measuring the length of a bar at various temperatures it may be required to determine the coefficient of thermal expansion of the material of which the bar is made and also to determine the length of the bar at a temperature at which it may not have been measured, e.g., 27 °C, when measured values were obtained at 20 °C, 22 °C, 24 °C, 26 °C, 28 °C and 30 °C.

In addition to measured values, there is another form of data that is also frequently fed into a rule in order to provide an output quantity value. This additional data relates to a variety of 'constants', each of which is described by imperfect knowledge of the quantity concerned. An example is a material constant such as modulus of elasticity, another is a calibrated dimension of an artefact such as a length or diameter, and another is a *correction* arising from the fact that a measurement was made at, say, 22 °C rather than the stipulated 20 °C.

The complete set of quantities for which values are required by the rule to enable a value of the output quantity to be determined is known as the set of *input quantities*. The rule is usually referred to as a *measurement model* because it is the use of physical modelling (or perhaps empirical modelling or both types of modelling) [3] of a measurement, measuring system or instrument that enables the rule to be established.

This guide is concerned with the problem of determining information about the output quantity given the measurement model and information concerning the input quantities. Some advice is given on encoding the mentioned information by probability distributions for the input quantities. Because the form of the measurement model varies enormously over different metrology disciplines, it is largely assumed that a (physical) model is available (having been derived by the experts in the appropriate area). Model validity is not specifically addressed.

In particular, this guide reviews several approaches to the problem, including the widely-accepted GUM uncertainty framework. It reviews the interpretation of the GUM that is made by many organisations and practitioners concerned with measurement, the analysis of measurement data and the presentation of measurement results. The point is made that this interpretation is subject to limitations that are insufficiently widely recognized. These limitations have, however, been indicated [99] and are discussed in Chapter 5.

An approach free from these limitations, known as the *propagation of distributions*, is presented. A particular implementation of this approach is given, constituting a numerical procedure based on the use of a *Monte Carlo method*, and can be used

1. in its own right to quantify probabilistically the knowledge of the output quantity, and
2. to validate the approach based on the GUM uncertainty framework.

The described Monte Carlo method itself has deficiencies. They are of a different nature from those of the GUM uncertainty framework, and to a considerable extent controllable. They are identified in Chapter 7.

The GUM does not refer *explicitly* to the use of a Monte Carlo method. However, this option was recognized during the drafting of the GUM. The ISO/IEC/OIML/BIPM draft (First Edition) of June 1992, produced by ISO/TAG 4/WG 3, states, as Clause G.1.5:

> If the relationship between Y [the output quantity] and its input quantities is nonlinear, or if the values available for the parameters characterizing the probabilities of the X_i [the input quantities] (expectation, variance, higher moments) are only estimates and are themselves characterized by probability distributions, and a first order Taylor expansion is not an acceptable approximation, the distribution of Y cannot be expressed as a convolution. In this case, numerical methods (such as Monte Carlo calculations) will generally be required and the evaluation is computationally more difficult.

In the published version of the GUM [10], this Clause had been modified to read:

> If the functional relationship between Y and its input quantities is nonlinear and a first-order Taylor expansion is not an acceptable approximation (see 5.1.2 and 5.1.5), then the probability distribution of Y cannot be obtained by convolving the distributions of the input quantities. In such cases, other analytical or numerical methods are required.

The interpretation made here of this re-wording is that 'other analytical or numerical methods' cover any other *appropriate* approach.[3] This interpretation is consistent with that of the National Institute of Standards and Technology (NIST) of the United States [99]:

> [Clause 6.6] The NIST policy provides for exceptions as follows (see Appendix C):
>
> It is understood that any valid statistical method that is technically justified under the existing circumstances may be used to determine the equivalent of u_i [the standard deviation of the ith input quantity], u_c [the standard deviation of the output quantity], or U [the half-width of a coverage interval for the output quantity, under a Gaussian assumption]. Further, it is recognized that

[3] That this interpretation is correct has been confirmed by Working Group 1 of JCGM.

> international, national, or contractual agreements to which NIST is a party may occasionally require deviation from NIST policy. In both cases, the report of uncertainty must document what was done and why.

Further, within the context of statistical modelling in analysing the homogeneity of reference materials, it is stated [58]:

> [Clause 9.2.3] ... where lack of a normal distribution is a problem, robust or non-parametric statistical procedures may be used to obtain a valid confidence interval for the quantity of interest.

This guide adheres to these broad views. The most important aspect relates to *traceability of the results* of an uncertainty evaluation. An uncertainty evaluation should include

1. all relevant information relating to the measurement model and its input quantities,
2. an estimate of the output quantity and either or both of the standard uncertainty associated with this estimate and a coverage interval (or coverage region) for the output quantity, and
3. the manner in which these results were determined, including all assumptions made.

There would also appear to be valuable and relevant interpretations and considerations in the German standard DIN 1319 [37]. An official English-language translation of this standard would not seem to be available.

There has been massive investment in the use of the GUM. It is essential that this investment is respected and that this guide is not seen as deterring the continuation of its use, at least in circumstances where such usage can be demonstrated to be appropriate. In this respect, a recommended validation procedure for the GUM uncertainty framework is provided in this guide. The attitude taken is that if the procedure demonstrates in any particular circumstance that this usage is indeed valid, the GUM uncertainty framework can legitimately continue to be used in that circumstance. The results of the validation can be used to record the fact that fitness for purpose in this regard has been demonstrated. If the procedure indicates that there is doubt concerning the validity of the GUM uncertainty framework, then there is a case for investigation. Since in the latter case the use of a Monte Carlo method forms a constituent part (in fact the major part) of the validation procedure, the method can be used in place of the GUM uncertainty framework. Such use of an alternative method is consistent with the broader principles of the GUM (Chapter 5 of this guide and above).

There is another vital issue facing the metrologist. For a measurement it is necessary to characterize the input quantities by probability distributions and to develop the measurement model for the output quantity in terms of these quantities. Carrying out these tasks can be

far from easy. Some advice is given in this regard. However, written advice can only be general, although examples and case studies can assist. In any one circumstance, the metrologist has the responsibility, perhaps with input from a mathematician or statistician if appropriate, of characterizing the input quantities and building the measurement model. The GUM uncertainty framework and the recommended approach using a Monte Carlo method both utilize this information (but in different ways). As mentioned, the former possesses some limitations that the latter sets out to overcome.

The attitude taken here is that whatever the nature of the input quantities and the measurement model, even (and especially) if some subjective decisions are made in their derivation, the probability distribution for the output quantity should then follow objectively and without qualification from this information, rather than in a manner that is subject to limitations, in the form of effects that are difficult to quantify and beyond the control of the practitioner.

In summary, the attitude that is generally promoted in this guide is that as far as economically possible use should be made of all available knowledge. In particular, (a) the available knowledge of the input quantities should be embodied within the probability distributions used to characterize them, (b) a measurement model that relates these input quantities to the output quantity should carefully be constructed, and (c) the evaluation of uncertainty associated with the estimate of the output quantity should be carried out in terms of this information.

2.2 The objective of uncertainty evaluation

Uncertainty evaluation is the generic term used in this guide to relate to any aspect of quantifying the extent of the incomplete knowledge of the output quantity in a measurement model given incomplete knowledge of the input quantities. Also, the measurement model itself may be based on incomplete knowledge. If that is the case, the nature and extent of the incomplete knowledge of the model also need to be quantified and its influence on the output quantity established. The knowledge of the output quantity is also influenced by any algorithm or software that is used to determine its value given values of the input quantities. Such software may incorporate approximate algorithmic techniques that impart an additional uncertainty.

Example 1 *Approximate area under a curve defined by spectral data*

> Consider a measurement model necessitating the determination of an integral representing the area under a curve defined by spectral data. An algorithm might utilize the trapezoidal or some other numerical quadrature rule. Numerical approximation errors will be committed in the use of this rule. They depend on the spacing of the ordinates used and on the extent of the departure of the curve from linearity. The consequent uncertainties would need to be evaluated.

The uncertainty evaluation process could be at any level required, depending on the application. At one extreme it could involve determining the standard uncertainty associated with an estimate of the output quantity for a simple measurement model having a single output quantity. At the other extreme it might be necessary to determine the joint probability distribution for a set of output quantities of a complicated complex measurement model exhibiting non-Gaussian behaviour, and from that deduce a coverage region for the set of output quantities for a stipulated coverage probability.

The objective of uncertainty evaluation can be stated as follows:

Derive (if not already available) a measurement model relating a set of output quantities to input quantities (estimated by measured values, suppliers' specifications, etc.) that influence them. Establish probability distributions to characterize these input quantities. Calculate (in a sense required by context) estimates of the output quantities and evaluate the uncertainties associated with these estimates.

A mathematical form for this statement is given in Chapter 3.

The objective stated above may in its context be well defined or not. In a case where it is well defined there can be little dispute concerning the nature of the results, presuming they have been obtained correctly. If it is not well defined, it will be necessary to augment the information available by assumptions or assertions in order to establish a well-defined problem. It will be necessary to ensure that the assumptions and assertions made are as sensible as reasonably possible in the context of the application. It will equally be necessary to make the assumptions and assertions overt *and to record them*, so that the results can be reproduced and defended, and perhaps subsequently improved.

In very many cases the objective of uncertainty evaluation will be to determine a coverage interval (or coverage region) for the output quantity. Commonly, this coverage interval will be for a $95\,\%$ coverage probability. There is no compelling scientific reason for this choice. It almost certainly stems from the traditional use of $95\,\%$ in statistical hypothesis testing [17], although the reasons for the choice in that area are very different. The overriding reason for the use of $95\,\%$ in uncertainty evaluation is a practical one. It has become so well established that for purpose of comparison with other results its use is almost mandated. Another strong reason for the use of $95\,\%$ is the considerable influence of the Mutual Recognition Arrangement concerning the comparison of national measurement standards and of calibration and measurement certificates issued by National Metrology Institutes [9]. Such an interval will be referred to in this guide as a $95\,\%$ *coverage interval*.

It can be argued that if a coverage interval at some other level of probability is quoted, it can be 'converted' into one at some other level. Indeed, a similar operation is recommended in the GUM, when information concerning the probability distribution for an input quantity is converted into a standard deviation (standard uncertainty in GUM parlance). The standard uncertainties together with sensitivity coefficients are combined to produce the standard uncertainty associated with an estimate of the output quantity, from which a coverage interval is obtained by multiplication by a factor. The factor is selected based on the *assumption* that

the probability distribution for the output quantity is Gaussian. That this process gives rise to difficulties in some cases can be illustrated using a simple example.

Example 2 *Dominant input quantity*

> Consider the measurement function $Y = X_1 + X_2 + \ldots$, where $X_1, X_2, \ldots$ are the input quantities and Y the output quantity. Assume that all input quantities except X_1 have a small standard deviation, and X_1 is characterized by a rectangular distribution. The above-mentioned GUM procedure gives a $95\,\%$ coverage interval for Y that is longer than the $100\,\%$ coverage interval for X_1.

Instances of this type would appear to be not uncommon. For instance, an EA guide [38] gives three examples arising in the calibration area. The occurrence of such cases is recognized by the GUM:

> [GUM Clause G.6.5] ... Such cases must be dealt with on an individual basis but are often amenable to an analytic treatment (involving, for example, the convolution of a normal distribution with a rectangular distribution ...

The statement that such cases must be dealt with on an individual basis would appear to be somewhat extreme. Indeed, such a treatment is possible (see Section 5.2), but is not necessary, since a Monte Carlo method (Chapter 7) generally operates effectively in cases of this type.

The interpretation [100] of the GUM by the United Kingdom Accreditation Service recommends the inclusion of a dominant uncertainty contribution by adding the term linearly to the remaining terms combined in quadrature. This interpretation gives rise generally to a more valid result, but remains an approximation. The EA Guide [38] provides some analysis in some such cases.

It is emphasized that a result produced according to a fixed recipe that is not universally applicable, such as the GUM uncertainty framework, may well be only approximately correct, and the degree of approximation difficult to establish. The concern in this guide is with uncertainty evaluation that is reliable in the sense that the results will not exhibit inconsistent or anomalous behaviour, however simple or complicated the measurement model may be.

Appendix A reviews some relevant statistical concepts.

2.3 Standard uncertainties and coverage intervals

Arguably the most important uncertainty information to a metrologist is a coverage interval corresponding to a specified coverage probability, e.g., an interval that is expected to contain $95\,\%$ of the values that could be attributed to the output quantity. This interval is the $95\,\%$ coverage interval considered above.

There is an important distinction between the nature of the information needed to determine the standard uncertainty associated with an estimate of the output quantity and a coverage interval for the output quantity. The estimate (expectation) and associated standard uncertainty (standard deviation) can be determined knowing the probability distribution for the output quantity (Appendix A.2). The converse is not true.

Example 3 *Deducing an expectation and a standard deviation from a distribution, but not the converse*

> As an extreme example, consider a random variable X that can take only two values, a and b, with equal probability. The expectation of X is $\mu = (a + b)/2$ and the standard deviation of X is $\sigma = |b - a|/2$. However, given only the values of μ and σ, there is no way of deducing the probability distribution for X. If a Gaussian distribution were assumed, it would be concluded that the interval $\mu \pm 1.96\sigma$ contained $95\,\%$ of the possible values of X. In fact, the interval contains *all* (both) values, as does the interval $\mu \pm \sigma$, of about half that length.

Related comments are made in Clause G.6.1 of the GUM. Although knowledge of the expectation and standard deviation is valuable information, without further information it conveys nothing about the manner in which the values are distributed.[4] If, however, it is *known* that the underlying probability distribution is Gaussian, the distribution for the output quantity is completely described since just the expectation and standard deviation fully characterize a Gaussian distribution. A similar comment can be made for some other probability distributions. Some distributions require additional parameters to describe them. For instance, in addition to the expectation and standard deviation, a t–distribution requires the degrees of freedom to specify it.

Thus, if the form of the probability distribution is available generically, from analysis, empirically or from other considerations, the determination of an appropriate number of *statistical parameters* will permit it to be quantified. Once the quantified form of the distribution is available, it is possible to calculate a *percentile*, i.e., a value for the quantity of concern such that, according to the distribution, the corresponding percentage of the possible values of the quantity is smaller than that value. For instance, if the 25-percentile is determined, $25\,\%$ of the possible values can be expected to lie below it (and hence $75\,\%$ above it). Consider the determination of the 2.5-percentile and the 97.5-percentile. $2.5\,\%$ of the values

[4] See, however, the maximum entropy considerations in Appendix A.5.

will lie to the left of the 2.5-percentile and $2.5\,\%$ to the right of the 97.5-percentile. Thus, $95\,\%$ of the possible values of the quantity lie between these two percentiles. These points thus constitute the endpoints of a $95\,\%$ coverage interval for the quantity.

The 2.5-percentile of a probability distribution can be thought of as a point a certain number of standard deviations below the expectation and the 97.5-percentile as a point a certain number of standard deviations above the expectation. The numbers of standard deviations to be taken depends on the distribution. They are known as *coverage factors*. They also depend on the coverage interval required, $90\,\%$, $95\,\%$, $99.8\,\%$ or whatever.

For the Gaussian and the t–distributions, the effort involved in determining the numbers of standard deviations to be taken has been embodied in tables and software functions.[5] Since these distributions are symmetric about the expectation, the coverage factors for percentiles corresponding to probabilities $100(1-p)/2$ and $100(1+p)/2$ that are 'probabilistically symmetric',[6] such as the above 2.5- and 97.5-percentiles corresponding to $p = 0.95$, are identical. This statement is not generally true for asymmetric probability distributions. Indeed, the concept of a coverage factor is inapplicable in that case.

In order to determine percentiles in general, it is necessary to be able to evaluate the inverse G^{-1} of the distribution function G (Appendix A.3). For well-known probability distributions, such as Gaussian and t, software is available in many statistical and other libraries for this purpose. Otherwise, values of $\xi_\alpha = G^{-1}(\alpha)$ corresponding to the *α–quantile* $(0 < \alpha < 1)$ can be determined by using a zero finder to solve the equation $G(\xi_\alpha) = \alpha$ [34]. Alternatively, $G(\xi)$ can be tabulated in advance at an adequate number of values of ξ, and inverse interpolation used to determine an approximation to $\xi_\alpha = G^{-1}(\alpha)$ for any required value of α.

The coverage interval is not unique, even in the symmetric case. Suppose that a probability density function (Appendix A) $g(\xi) = G'(\xi)$ is unimodal (single-peaked), and that a value of α, $0 < \alpha < 1$, is given. Consider any interval $[a, b]$ that satisfies

$$G(b) - G(a) = \int_a^b g(\xi)\,\mathrm{d}\xi = 1 - \alpha.$$

Then [86]:

1. $[a, b]$ is a $100(1-\alpha)\,\%$ coverage interval. For instance, if a and b are such that

   $$G(b) - G(a) = 0.95,$$

   $95\,\%$ of possible values ξ lie between a and b;

[5] In most interpretations of the GUM, the output quantity is characterized by a Gaussian distribution or a t–distribution.

[6] 'Probabilistically symmetric' means that the percentage of values of the quantity to the left of the lower percentile equals the percentage of values to the right of the upper percentile.

2. The shortest such interval is given by $g(a) = g(b)$.[7] a lies to the left of the mode (the value ξ at which $g(\xi)$ is greatest) and b to the right;

3. If $g(\xi)$ is symmetric, not only is the shortest such interval given by $g(a) = g(b)$, but also a and b are equidistant from the mode, which equals the expectation in this case.

[7] A proof of this result is as follows. A $100p\,\%$ coverage interval has endpoints $a = G^{-1}(\alpha)$ and $b = G^{-1}(p + \alpha)$ for any α satisfying $0 < \alpha < 1 - p$. The length of the coverage interval is

$$L(\alpha) = b - a = G^{-1}(p + \alpha) - G^{-1}(\alpha).$$

Now, the derivative of $G^{-1}(\alpha)$ with respect to α is

$$\left[G^{-1}(\alpha)\right]' = \frac{1}{G'[G^{-1}(\alpha)]} = \frac{1}{g[G^{-1}(\alpha)]}.$$

Therefore, at a minimum of the length of the coverage interval,

$$[L(\alpha)]' = \frac{1}{g[G^{-1}(p + \alpha)]} - \frac{1}{g[G^{-1}(\alpha)]} = 0,$$

i.e.,

$$g[G^{-1}(\alpha)] = g[G^{-1}(p + \alpha)]$$

from which is obtained the result $g(a) = g(b)$.

Chapter 3

The problem formulated

3.1 Measurement models and measurement functions

As discussed in Section 2.1, regardless of the field of application, the physical quantity of concern, the output quantity or measurand, can rarely be measured directly. Rather, it is determined from a number of contributions, or input quantities, about which knowledge in the form of measured values or other information is available.

The fundamental relationship between the input quantities and the output quantity is the *measurement model*. If there is a single (scalar) output quantity Y, the measurement model takes the general form

$$h(Y, \boldsymbol{X}) = h(Y, X_1, \ldots, X_N) = 0,$$

in which $\boldsymbol{X} = (X_1, \ldots, X_N)^\top$ denote the input quantities, N, say, in number.[1]

There may be more than one output quantity, viz., $\boldsymbol{Y} = (Y_1, \ldots, Y_m)^\top$. In this case the measurement model is

$$\boldsymbol{h}(\boldsymbol{Y}, \boldsymbol{X}) = \boldsymbol{0},$$

where $\boldsymbol{h}(\boldsymbol{Y}, \boldsymbol{X}) = (h_1(\boldsymbol{Y}, \boldsymbol{X}), \ldots, h_m(\boldsymbol{Y}, \boldsymbol{X}))^\top$, a vector of measurement models. The output quantities $\boldsymbol{Y}$ would almost invariably be mutually dependent in this case, since in general each output quantity Y_j, $j = 1, \ldots, m$, would depend on several or all of the input quantities.

If the measurement model can be expressed in the equivalent form

$$Y = f(\boldsymbol{X}) = f(X_1, \ldots, X_N),$$

for a scalar output quantity, or

$$\boldsymbol{Y} = \boldsymbol{f}(\boldsymbol{X}) = \boldsymbol{f}(X_1, \ldots, X_N),$$

[1] A single input quantity (when $N = 1$) will sometimes be denoted by X (rather than X_1).

for a vector output quantity, the function f or $\boldsymbol{f}$ is the *measurement function*. The measurement function can be a mathematical formula, a step-by-step calculation procedure, computer software or other prescription, which yields a (unique) value y or $\boldsymbol{y}$ of the output quantity given values $\boldsymbol{x} = (x_1, \ldots, x_N)^\top$ of the input quantities.

A measurement model or function with a single output quantity Y is known as a *univariate* or *scalar* model or function. A measurement model or function with m (> 1) output quantities $\boldsymbol{Y}$ is known as a *multivariate* or *vector* model or function.

All input quantities X_i are regarded as *random variables* with possible values ξ_i, regardless of their source [104]. The output quantity Y is also a random variable with possible values η. Instances x_i of the X_i are *estimates* of the input quantities. The value $f(x_1, \ldots, x_n)$ provides an estimate of the output quantity. This estimate may be biased, although it is expected that the bias will be negligible in many cases. The expectation of the output quantity is unbiased. The bias associated with the estimate $f(x_1, \ldots, x_n)$ of the output quantity results from the fact that the value of Y obtained by evaluating the measurement function at the estimates $\boldsymbol{x}$ is not in general equal to the expectation of Y. These values will be equal when the measurement function is linear in $\boldsymbol{X}$, and close if the function is mildly non-linear or if the uncertainties associated with the estimates of the input quantities are small.

Example 4 *Bias associated with the estimate $f(x_1, \ldots, x_n)$ of the output quantity*

A demonstration of the bias associated with the estimate $f(x_1, \ldots, x_n)$ of the output quantity is given by the simple measurement function $Y = X^2$, where X with expectation zero and standard deviation u is characterized by a Gaussian distribution. The estimate of X is zero, and the corresponding estimate of the output quantity Y is also zero. However, the expectation of Y cannot be zero, since $Y \geq 0$, with equality occurring only when $X = 0$. (The probability distribution characterizing Y is in fact a χ^2–distribution with one degree of freedom.)

3.2 Uncertainty evaluation using the propagation of distributions

Uncertainty evaluation consists of three stages: *formulation*, *propagation* and *summarizing*. The propagation stage of uncertainty evaluation is referred to as the *propagation of distributions*.

In the formulation stage the metrologist derives the measurement model, perhaps in collaboration with a mathematician or statistician. The metrologist also characterizes the input quantities by probability distributions, e.g., rectangular (uniform), Gaussian (normal), etc., defined in terms of parameters of the probability density functions (PDFs) for these distributions, e.g., central value and semi-width for a rectangular PDF, or expectation and standard

deviation for a Gaussian PDF. These PDFs are obtained from an analysis of series of indication values [10, Clauses 2.3.2, 3.3.5] or based on scientific judgement using all the relevant information available [10, Clauses 2.3.3, 3.3.5], [99].

In the case of mutually independent input quantities and a single output quantity, the propagation stage of uncertainty evaluation can be summarized as follows. Given the measurement function $Y = f(\boldsymbol{X})$, where $\boldsymbol{X} = (X_1, \ldots, X_n)^\top$, and the PDFs $g_{X_i}(\xi_i)$ (or the distribution functions $G_{X_i}(\xi_i)$) for the input quantities X_i, for $i = 1, \ldots, N$, determine the PDF $g_Y(\eta)$ (or the distribution function $G_Y(\eta)$) for the output quantity Y. See Figure 3.1 for a graphical illustration of the propagation stage of uncertainty evaluation.

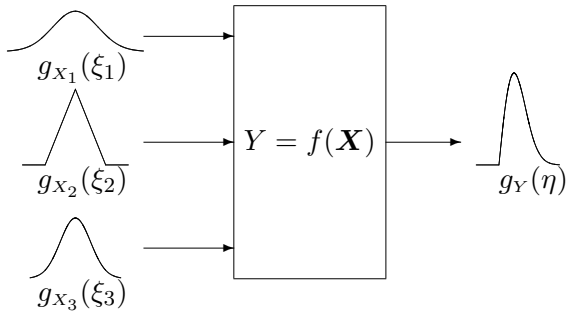

Figure 3.1: Illustration of the propagation of distributions. The measurement function has three input quantities $\boldsymbol{X} = (X_1, X_2, X_3)^\top$, where X_1 is characterized by a Gaussian PDF $g_{X_1}(\xi_1)$, X_2 by a triangular PDF $g_{X_2}(\xi_2)$ and X_3 by a (different) Gaussian PDF $g_{X_3}(\xi_3)$. The single output quantity $Y \equiv Y_1$ is illustrated as being characterized by an asymmetric PDF, as can arise for non-linear measurement functions where one or more of the input quantities has a large standard deviation.

Finally, in the summarizing stage the PDF (or corresponding distribution function) for the output quantity is used to obtain an estimate of the output quantity, the associated standard uncertainty, and a coverage interval for the output quantity for a stated coverage probability. It is reiterated that once the PDF (or distribution function) for Y has been obtained, any statistical information relating to Y can be produced from it.

When the input quantities are mutually dependent, in place of the N individual PDFs $g_{X_i}(\xi_i)$, $i = 1, \ldots, N$, there is a joint PDF $g_{\boldsymbol{X}}(\boldsymbol{\xi})$, where $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_N)^{\mathrm{T}}$. An example of a joint PDF is the multivariate Gaussian PDF (Section 4.2.7). In practice this joint PDF may be decomposable. For instance, in some branches of electrical, acoustical and optical metrology, the input quantities may be complex. The real and imaginary parts of each such quantity are generally mutually dependent but the real and imaginary parts of different quantities may be independent.

If there is more than one output quantity, $\boldsymbol{Y}$, these output quantities will almost invariably need to be described by a joint PDF $g_{\boldsymbol{Y}}(\boldsymbol{\eta})$, since each output quantity generally depends on all or several of the input quantities. See Section 9.7 for an important exception.

The propagation and summarizing stages involve the derivation of the estimate of the output quantity and the associated uncertainty, given the information provided by the formulation

stage. It is computational and requires no further information from the metrology application. The uncertainty is commonly provided as a coverage interval. A coverage interval can be determined once the distribution function $G_Y(\eta)$ (Appendix A) has been derived. The endpoints of a $95\,\%$ coverage interval[2] are given (Section 2.3) by the 0.025- and 0.975-quantiles of $G_Y(\eta)$, the α-quantile being the value of η such that $G_Y(\eta) = \alpha$.[3]

It is usually sufficient to report the uncertainty associated with the estimate of the output quantity to one or at most two significant decimal digits. In general, further digits would be spurious, because the information provided in the formulation stage is typically imprecise, involving estimates and assumptions. The attitude taken here though is that the propagation and summarizing stages should not exacerbate the consequences of the decisions made in the formulation stage.[4]

The PDF $g_Y(\eta)$ for Y cannot generally be expressed in simple or even closed mathematical form. Formally [28], if $\delta(\cdot)$ denotes the Dirac delta function,

$$g_Y(\eta) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g_{\boldsymbol{X}}(\boldsymbol{\xi}) \delta(\eta - f(\boldsymbol{\xi})) \, \mathrm{d}\xi_N \cdots \mathrm{d}\xi_1. \tag{3.1}$$

Approaches for determining $g_Y(\eta)$ or $G_Y(\eta)$ are addressed in Chapter 5. That several approaches exist is a consequence of the fact that the determination of $g_Y(\eta)$ and/or $G_Y(\eta)$ ranges from being very simple to extremely difficult, depending on the complexity of the measurement model and the PDFs for the input quantities.

3.3 Uncertainty evaluation using Bayes' theorem

Although the focus of this guide is on the propagation of distributions as a formulation of the problem of uncertainty evaluation (Section 3.2), measurement problems are encountered for which prior knowledge about the output quantity Y is available and can be used to characterize Y by a prior probability distribution. Such additional information can be used to provide a probability distribution for Y that can give a smaller standard deviation for Y and hence a smaller standard uncertainty associated with the estimate of Y. The use of Bayes' theorem (below and Appendix A.4) provides a formulation of the problem of uncertainty evaluation that allows proper account to be taken of such prior knowledge.

[2] $95\,\%$ coverage intervals are often used in this guide, but the treatment applies more generally.

[3] There are many intervals having a coverage probability of $95\,\%$, a general interval being given by the β- and $(0.95 + \beta)$-quantiles of $G_Y(\eta)$, with $0 \leq \beta \leq 0.05$. The choice $\beta = 0.025$ is natural for a $G_Y(\eta)$ corresponding to a symmetric PDF $g_Y(\eta)$. It also has the shortest length for a symmetric PDF and, in fact, for a unimodal PDF (Section 2.3).

[4] This attitude compares with that in mathematical physics where a model (e.g., a partial differential equation) is constructed and then solved. The construction involves idealizations and inexact values for dimensional quantities and material constants, for instance. The solution process involves the application of hopefully sensible and stable methods in order to make some supported statements about the quality of the solution obtained to the posed problem.

Consider a measurement in which the input quantities $\boldsymbol{X}$ in the measurement model can be partitioned as

$$\boldsymbol{X} = (W, \boldsymbol{Z}^\top)^\top,$$

in which W represents an indication quantity, about which information takes the form of the indication value w,[5] and $\boldsymbol{Z} = (Z_1, \ldots, Z_{N-1})^\top$ all other input quantities.[6] In terms of the quantities Y, W and $\boldsymbol{Z}$, suppose the measurement model takes the form[7]

$$W = \phi(Y, \boldsymbol{Z}). \tag{3.2}$$

Let $g_{Y,\boldsymbol{Z}}(\eta, \boldsymbol{\zeta})$ denote a joint *prior* PDF for Y and $\boldsymbol{Z}$ obtained on the basis of information available before any measurement is made. In addition, let $l_W(\omega|w)$ denote the *likelihood function* for W given the indication value w, which describes the probability of obtaining the value w given the value ω of W.[8]

Applying Bayes' theorem (Appendix A.4), and using the observation model (3.2),

$$g_{Y,\boldsymbol{Z}}(\eta, \boldsymbol{\zeta}|w) \propto l_W(\phi(\eta, \boldsymbol{\zeta})|w) g_{Y,\boldsymbol{Z}}(\eta, \boldsymbol{\zeta}),$$

in which $g_{Y,\boldsymbol{Z}}(\eta, \boldsymbol{\zeta}|w)$ is the joint *posterior* PDF for Y and $\boldsymbol{Z}$ obtained by 'updating' the prior PDF using the information provided by the indication value. The posterior PDF for Y *alone* is obtained by applying marginalisation (Appendix A.2) to this joint posterior PDF. A posterior estimate of Y and the associated standard uncertainty are evaluated as the expectation and standard deviation of Y characterized by its posterior PDF. The posterior PDF is also used as the basis of evaluating coverage intervals for a stipulated coverage probability.

The application of Bayes' theorem described above can be extended to include additional quantities to represent so-called 'nuisance' parameters such as might be required to specify the likelihood function. An example of such a parameter is the variance of the indication quantity W in the case that this variance is not known exactly. See Appendix A.4.

The determination of the joint posterior PDF for Y and $\boldsymbol{Z}$, and the marginal posterior PDF for Y, is generally achieved numerically by generating discrete representations of those distributions in terms of values randomly drawn from the distributions.[9] *Markov chain*

[5] There may be more than one indication value. This situation can also be handled: see Appendix A.4 and the example of gauge block calibration described therein.

[6] In the context of the GUM uncertainty framework, the uncertainty associated with an estimate of the quantity W would come from a 'Type A' evaluation of uncertainty and that for the quantity $\boldsymbol{Z}$ from a 'Type B' evaluation of uncertainty.

[7] In practice, the measurement model will often be available in this form, which is called here an *observation model* [80]. It expresses the indication quantity W (or indication quantities) for the measuring system in terms of the measurand Y, the quantity *intended* to be measured, and the (remaining) input quantities $\boldsymbol{Z}$, such as correction quantities, associated with the measuring system. In many cases, the observation model is the starting point for obtaining the measurement function, as used in the GUM, by 'solving' the observation model for the output quantity Y.

[8] The notation used for the likelihood function is such as to emphasize that the likelihood function is a function of the value ω of W and not of the indication value w, which is fixed.

[9] In the propagtion of distributions, the determination of the PDF for the output quantity given by formula (3.1) is also generally achieved numerically, and Monte Carlo methods are a powerful tool for this purpose.

Monte Carlo (MCMC) is a general computational method for generating such discrete representations, and includes the Metropolis and Metropolis-Hastings algorithms and the Gibbs sampler. Software implementations of these methods are available in packages such as WinBUGS [105] and R [82].

When there is no prior information about the measurand, the propagation of distributions can be used to provide a PDF for the output quantity that uses only the information available. Bayes' theorem can also be applied in this circumstance, but it can be difficult to know how to construct the prior PDF $g_Y(\eta)$. A requirement of the PDF might be that it has a 'minimal' influence on the posterior PDF. Such distributions are sometimes called *reference prior distributions* [50]: they are also described as *vague* or *non-informative* and are often *improper*, i.e., the distribution does not integrate to unity. For the example of gauge block calibration described in Appendix A.4 for which $Y \equiv \ell$ is the length of the gauge block being calibrated, the prior distribution $g_Y(\eta) \propto 1$ for $0 \leq \eta < \infty$, which is an improper distribution, might be chosen to express the lack of prior information about Y.

A comparison of the propagation of distributions and a Bayesian analysis in the case that no prior knowledge of the measurand is assumed is available [42]. It is shown that when the output quantity Y depends linearly on the input quantities $\boldsymbol{X} = (W, \boldsymbol{Z}^\top)^\top$, and a constant improper prior distribution is used for Y, the two approaches yield the same posterior PDF. In other cases, however, the approaches can give different results. A recommendation given in that work is to encourage the use of a *proper* PDF to express (informative) prior information about the measurand that can then be included in a Bayesian analysis.

See Section 9.8 for an example to illustrate the application of a Bayesian analysis to a problem of uncertainty evaluation concerned with mass calibration. The example illustrates the influence of the choice of prior PDF for the measurand on the results obtained.

Chapter 4

The main stages of uncertainty evaluation

4.1 Overview

In this guide, uncertainty evaluation is regarded as consisting of the main stages indicated in Section 3.2.

The *formulation stage* (Section 4.2) consists of two steps:

1. Develop a measurement model relating the input quantities to the output quantity;

2. Characterize the input quantities, regarded as random variables, by PDFs.

The *propagation stage* (Section 4.3) consists of using the information provided by the formulation stage to determine the PDF for the output quantity, also regarded as a random variable.

Finally, the *summarizing stage* (Section 4.4) consists of using the PDF for the output quantity to obtain

1. the expectation of the quantity, taken as an estimate of the quantity,

2. the standard deviation of the quantity, taken as the standard uncertainty associated with the estimate, and

3. a coverage interval for the quantity corresponding to a stated coverage probability.

4.2 Formulation stage

4.2.1 Statistical modelling

Statistical modelling can be beneficial when a measurement model is complicated, but is not always needed for simpler models. It is concerned with developing the relationships between (a) the quantities of which the measurement data obtained are instances, (b) the input and output quantities, and (c) other available information, such as from calibration certificates. It is also concerned with characterizing the input quantities by probability distributions.

Example 5 *Straight-line calibration*

A common example of statistical modelling arises when fitting a calibration function to data, representing, say, the manner in which displacement varies with temperature. The data consists, for $i = 1, \ldots, N$, say, of a measured value of a response variable X_i corresponding to a measured or assigned value of a stimulus or independent variable T_i. Suppose that the nature of the calibration is such that a straight-line calibration function is appropriate. Then, as part of the statistical modelling process [3], the equations

$$X_i = A_1 + A_2 T_i, \quad i = 1, \ldots, N, \tag{4.1}$$

relate the quantities X_i and T_i to the parameters (quantities) A_1 (intercept) and A_2 (gradient) of the calibration function.

In order to establish values for A_1 and A_2 it is necessary to make an appropriate assumption about the nature of the measurement data [3, 30]. Consider a situation in which the estimates t_i of the T_i have negligible associated uncertainties relative to those associated with the estimates x_i of the X_i. Furthermore, suppose the estimates x_i are obtained independently and the associated uncertainties are identical. Then, unbiased esimates a_1 and a_2 of A_1 and A_2 are given by *least squares*. Specifically, a_1 and a_2 are given by minimizing the sum of the squares of the residual deviations $x_i - \alpha_1 - \alpha_2 t_i$, over $i = 1, \ldots, N$, with respect to α_1 and α_2, viz.,

$$\min_{\alpha_1, \alpha_2} \sum_{i=1}^{N} (x_i - \alpha_1 - \alpha_2 t_i)^2. \tag{4.2}$$

The least-squares problem (4.2) would take an alternative form if other assumptions applied to the measured data. For example, if the estimates x_i are obtained independently and the associated uncertainties $u(x_i),\ i = 1, \ldots, N$, are generally different, then a_1 and a_2 are determined by minimizing a sum of squares of weighted residual deviations, viz.,

$$\min_{\alpha_1, \alpha_2} \sum_{i=1}^{N} w_i^2 (x_i - \alpha_1 - \alpha_2 t_i)^2, \quad w_i = 1/u(x_i),\ i = 1, \ldots, N. \tag{4.3}$$

If the only information about the quantity X_i is the estimate x_i and the associated standard uncertainty $u(x_i)$, the quantity would be characterized by a Gaussian distribution (Appendix A.5). In this case, the estimates a_1 and a_2 are maximum likelihood estimates, corresponding to the 'most likely' values of A_1 and A_2 that could have given rise to the measured data. The equations (4.1), together with the (least squares) solution criterion (4.3), and information about the distributions used to characterize the measured quantities, constitute the results of the statistical modelling process for this example.

4.2.2 Input-output modelling

Input-output modelling is the determination of the measurement model required by the GUM in its approach to uncertainty evaluation. In the GUM a measuring system is modelled, as in Section 3.1, by a functional relationship between the output quantity Y and input quantities $\boldsymbol{X} = (X_1, \ldots, X_N)^\top$ in the form of the measurement function

$$Y = f(\boldsymbol{X}). \tag{4.4}$$

In practice this functional relationship does not apply directly to all measurement systems encountered, but may instead (a) take the general form $h(Y, \boldsymbol{X}) = 0$, (b) involve a number of output quantities $\boldsymbol{Y} = (Y_1, \ldots, Y_m)^\top$, or (c) involve complex quantities. Chapter 6 is concerned with the manner in which each model type within this classification can be treated within the GUM uncertainty framework. Here, the concern is with the basic form (4.4).

Example 6 *How long is a piece of string?*

The problem of establishing a simple measurement model for the length of a piece of string, when measured with a steel tape, is considered. (An alternative treatment is available [7].) The output quantity is the length of the string. As part of the formulation stage, a measurement model for string length is established. It depends on several input quantities. This model is expressed here as the sum of four terms. Each of these terms, apart from the first, is itself expressed as a sum of terms.[1] The model takes the form[2]

String length	=	Measured string length (1)
	+	Tape length correction (2)
	+	String length correction (3)
	+	Measurement process correction (4),

where the measured string length (1) and all corrections (in (2), (3) and (4)) are regarded as quantities, and

[1] The model can therefore be viewed as a multi-stage measurement model (Section 4.2.6), although of course by substitution it can be expressed as a single model.

[2] In this formula, the correction terms are to be expressed in a way that ensures that each has the correct numerical ($\pm$) sign.

(1) Measured string length = Average of a number of repeated indications of the string length,

(2) Tape length correction = Correction due to tape calibration imperfections
+ Correction due to tape stretching (or shrinking)
+ Correction due to tape bending
+ Correction due tape thermal expansion,

(3) String length correction = Correction due to the string departing form a straight line
+ Correction due to string stretching (or shrinking),

(4) Measurement process correction = Correction due to the inability to align the end of the tape with the end of the string due to fraying of the string ends
+ Correction due to the tape and the string not being parallel
+ Correction due to assigning a numerical value to the indication on the tape.

Once this measurement model is in place statements can be made about the nature of the various quantities in the model as part of the formulation stage of uncertainty evaluation. The propagation and summarizing stages can then be carried out to evaluate the uncertainty associated with an estimate of the string length.

Each input quantity is characterized by an appropriate PDF. For instance, the measured string length (1) would be characterized by a t–distribution, based on the average and standard deviation associated with the average of the repeated indication values, with a degrees of freedom one less than the number of indication values. As another instance, the quantity describing the length deviation due to bending of the tape (2) would be characterized by a distribution related to the χ^2–distribution.[3] This quantity, characterized by such a probability distribution, does not have zero expectation, since the *minimum* effect of tape bending on the output quantity is zero.

[3]This distribution follows from a simple assumption relating to the nature of the bending. This degree of sophistication would not be warranted when measuring the length of a piece of string. It can be important in other applications.

Example 7 *Straight-line calibration (re-visited)*

For the straight-line calibration example of Section 4.2.1 (Example 5), the measurement model constitutes a formula or prescription (not in general in the form of a measurement function) derived from the statistical modelling process. Specifically, the estimates $\boldsymbol{a} = (a_1, a_2)^\top$ of the parameters of the straight-line calibration function are given in terms of the measurement data $\boldsymbol{x} = (x_1, \ldots, x_N)^\top$ by an equation of the form

$$\boldsymbol{H}\boldsymbol{a} = \boldsymbol{q}. \tag{4.5}$$

(Compare [3], [10, Clause H.3].) Here, $\boldsymbol{H}$ is a matrix of dimension 2×2 that depends on the values t_i, and $\boldsymbol{q}$ a vector of dimension 2×1 that depends on the values t_i and x_i. By expressing the solution of this equation as

$$\boldsymbol{a} = \boldsymbol{H}^{-1}\boldsymbol{q}, \tag{4.6}$$

a measurement function for the parameters of the calibration function is obtained. It is, at least superficially, an explicit expression for $\boldsymbol{a}$.[4]

4.2.3 Example of statistical and input-output modelling

Consider the measurement of two nominally identical lengths under suitably controlled conditions using a steel rule. Suppose there are two contributions to the uncertainty of measurement due to

1. imperfection in the manufacture and calibration of the rule, and
2. operator effect in positioning and reading the scale.

Let the lengths be denoted by L_1 and L_2. Let the measured values of the lengths be denoted by ℓ_1 and ℓ_2. Then the measurements may be modelled by

$$L_1 = \ell_1 + D_0 + D_1, \quad L_2 = \ell_2 + D_0 + D_2,$$

where D_0 is a quantity describing the imperfection in the steel rule when measuring lengths close to those of concern, and D_1 and D_2 are quantities describing the deviations attributable to the operator in obtaining the measurement data. Make the assumption that the quantities D_0, D_1 and D_2 have zero expectations, i.e.,

$$E(D_0) = d_0 = 0, \quad E(D_1) = d_1 = 0, \quad E(D_2) = d_2 = 0.$$

[4] The expression is termed superficially explicit, since the determination of $\boldsymbol{a}$ *via* a formal matrix inversion is *not* recommended [3, 30]. The form (4.6), or forms like it in other such applications, should not be regarded as an *implementable* formula. Rather, numerically stable matrix factorization algorithms [52] should be employed. This point is not purely academic. The instabilities introduced by inferior numerical solution algorithms can themselves be an appreciable source of *computational* uncertainty. It is not generally straightforward to quantify this effect.

Then,

$$E(L_1) = \ell_1, \quad E(L_2) = \ell_2,$$

i.e., the expectations (estimates) of the lengths are the measured values. Make the further assumption that the quantities are independent and have variances

$$V(D_0) = u^2(d_0), \quad V(D_1) = u^2(d), \quad V(D_2) = u^2(d).$$

Then, the variances of the lengths are given by

$$V(L_1) = u^2(d_0) + u^2(d), \quad V(L_2) = u^2(d_0) + u^2(d),$$

and the covariance by

$$\mathrm{Cov}(L_1, L_2) = \mathrm{Cov}(D_0, D_0) = u^2(d_0).$$

Suppose that it is required to evaluate an estimate of the difference in the lengths and the associated uncertainty. From the above equations,

$$L_1 - L_2 = (\ell_1 - \ell_2) + (D_1 - D_2).$$

Then,

$$E(L_1 - L_2) = \ell_1 - \ell_2,$$

and, since D_1 and D_2, are independent,

$$V(L_1 - L_2) = V(D_1 - D_2) = V(D_1) + V(D_2) = 2u^2(d). \tag{4.7}$$

As expected, the uncertainty associated with the imperfection in the steel rule does not enter this result.

Compare the above with the input-output modelling approach:

Input quantities: Lengths L_1 and L_2.

Output quantity: Difference in lengths Y.

Measurement model: $Y = L_1 - L_2$.

Estimates of the input quantities: Measured values ℓ_1 and ℓ_2.

Uncertainties associated with the estimates of the input quantities:

$$u(\ell_1) = u(\ell_2) = \left(u^2(d_0) + u^2(d)\right)^{1/2}, \quad \mathrm{cov}(\ell_1, \ell_2) = u^2(d_0).$$

Partial derivatives of model (evaluated at the estimates of the input quantities):

$$\partial Y/\partial L_1 = 1, \quad \partial Y/\partial L_2 = -1.$$

Estimate of the output quantity:

$$y = \ell_1 - \ell_2.$$

Uncertainty associated with the estimate of the output quantity (using GUM Formula (13)):

$$u^2(y) = (\partial Y/\partial L_1)^2\, u^2(\ell_1) + (\partial Y/\partial L_2)^2\, u^2(\ell_2) + (\partial Y/\partial L_1)\,(\partial Y/\partial L_2)\,\mathrm{cov}(\ell_1, \ell_2),$$

the partial derivatives being evaluated at $L_1 = \ell_1$ and $L_2 = \ell_2$ (here they are constants), giving

$$\begin{aligned} u^2(y) &= (1)^2(u^2(d_0) + u^2(d)) + (-1)^2(u^2(d_0) + u^2(d)) + 2(1)(-1)u^2(d_0) \\ &= 2u^2(d), \end{aligned}$$

which is the same as the result (4.7) obtained using statistical modelling.

4.2.4 Mutually dependent input quantities

In a range of circumstances some choice is possible regarding the manner in which the input quantities to the measurement model are provided. A group of input quantities can be mutually dependent in that each depends on a common quantity. It may be possible to re-express such input quantities so that the common quantity appears explicitly as a further input quantity. By doing so, this cause of correlation is *eliminated*, with the potential for a simplification of the analysis. See GUM Clause F.1.2.4. Also, an example in mass comparison [1] illustrates the principle. An example of the approach, in the context of measuring the sides of a right-angled triangle, is given in Section 9.6.

In general, the use of modelling principles, before input quantities are characterized by probability distributions or uncertainties associated with estimates of the quantities are evaluated, is often helpful in understanding correlation effects.

4.2.5 Constraints in uncertainty evaluation

Constraints in uncertainty evaluations arise as a consequence of physical limits or conditions associated with the input or output quantities. Some examples are given below.

When measuring the concentrations of the constituent parts of a chemical mixture, it is appropriate to ensure that the estimates of the concentrations sum to unity (or $100\,\%$).

In assessing the departure from perfect form in dimensional metrology, the output quantity is flatness, roundness, perpendicularity, concentricity, etc. These quantities are defined as the *unsigned* departure, assessed in an unambiguously defined way, of a real feature from an ideal feature, and are often very small, but nonzero. Any uncertainty statement associated with an estimate of such a quantity expressed as a coverage interval that embraces zero is physically unrealistic.

In triangulation, photogrammetry and similar applications, using theodolites, laser interferometers and metric cameras, redundancy of measurement ensures that smaller uncertainties are generally obtained compared with the use of a near-minimal number of measurements. The various quantities, point co-ordinates, distances, etc. are interrelated by equality conditions deducible from geometrical considerations. The extent of the improvement in uncertainty is limited by inevitable systematic errors.

Within analytical chemistry, measurement of, e.g., trace elements, is often performed near the limit of detection. At this limit the measurement uncertainty is comparable to the magnitude of the measured value. This situation has aspects in common with that in dimensional metrology above, although there are appreciable contextual differences.

The Eurachem Guide to quantifying uncertainty in analytical measurement states

> [46, Appendix F] At low concentrations, an increasing variety of effects becomes important, including, for example,
>
> - the presence of noise or unstable baselines,
> - the contribution of interferences in the (gross) signal
> - ...
>
> Because of such effects, as analyte concentrations drop, the relative uncertainty associated with the result tends to increase, first to a substantial fraction of the result and finally to the point where the (symmetric) uncertainty interval includes zero. This region is typically associated with the practical limit of detection for a given method.
>
> ...
>
> Ideally, therefore, quantitative measurements should not be made in this region. Nevertheless, so many materials are important at very low levels that it is inevitable that measurements must be made, and results reported, in this region. ... The ISO Guide to the Expression of Uncertainty in Measurement does not give explicit instructions for the estimation of uncertainty when the results are small and the uncertainties large compared to the results. Indeed, the basic form of the 'law of propagation of uncertainties' ... may cease to apply accurately in this region; one assumption on which the calculation is based is that the uncertainty is small relative to the value of the measurand. An additional, if philosophical, difficulty follows from the definition of uncertainty given by the ISO Guide: though negative observations are quite possible, and even common in this region, an implied dispersion including values below zero cannot be 'reasonably ascribed to the value of the measurand' when the measurand is a concentration, because concentrations themselves cannot be negative.
>
> ...
>
> Observations are not often constrained by the same fundamental limits that apply to real concentrations. For example, it is perfectly sensible to report an

> 'observed concentration' that is an estimate below zero. It is equally sensible to speak of a dispersion of possible observations which extends into the same region. For example, when performing an unbiased measurement on a sample with no analyte present, one *should* see about half of the observations falling below zero. In other words, reports like
>
> $$\text{observed concentration} = 2.4 \pm 8 \text{ mgl}^{-1}$$
> $$\text{observed concentration} = -4.2 \pm 8 \text{ mgl}^{-1}$$
>
> are not only possible; they should be seen as valid statements.

It is the view of the authors of this guide that these statements by Eurachem are sound. However, the current guide takes a further step, related to *modelling* the measurement and through the use of the measurement model defining and making a statement about the output quantity, as opposed to the indication quantities (observations) on which estimates of the input quantities are based. Because a (simple) measurement model is established, this step arguably exhibits even closer consistency with the GUM.

The Eurachem statements stress that indication quantities (of which the indication values or observation values are instances) are not often constrained by the same fundamental limits that apply to real concentrations. It is hence appropriate to demand that the output quantity, defined to be the *real* analyte concentration (or its counterpart in other applications) should be constrained to be non-negative. Also, the indication quantities *should not* and *cannot* be constrained, because instances of these quantities are the values delivered by the measurement method. Further, again consistent with the Eurachem considerations, the input quantity, analyte concentration, is characterized by a PDF that is symmetric about the estimate of the quantity (unless information to the contrary is available). Thus, the input quantity, X, say, is *unconstrained analyte concentration* and is characterized by a symmetric PDF, and the output quantity, Y, say, *real analyte concentration*, has a PDF to be determined.

Considerations compatible with the above reasoning gives the measurement model[5]

$$Y = \max(X, 0). \tag{4.8}$$

The rationale behind this simple choice of model is as follows. Should the average x of the indications prove to be non-negative, it would naturally be taken as an estimate of Y. Such an estimate would conventionally be used at points removed from the limit of detection. Should x prove to be negative, it cannot be used as a physically feasible estimate of Y, since by definition Y is the real analyte concentration and hence non-negative. Taking $y = 0$ in this case is the optimal compromise between the indication value and feasibility (the closest feasible value to the average of the indication values).

[5] Related considerations [65, p129] show that if an indication v is an instance of $\mathrm{N}(\theta, 1)$, i.e., drawn from a Gaussian distribution with expectation θ and standard deviation unity, but $\theta \geq 0$, the *maximum likelihood estimate* of θ is $\max(v, 0)$.

Other approaches to accounting for physical knowledge in obtaining measurement results are available [25, 31]. In particular, consideration may be given to modelling *probabilistically* (rather than *functionally* as above) knowledge of the quantities concerned (see Section 3.3). A comparison of approaches, including the use of the principle of maximum entropy, a Bayesian treatment, and the application of the propagation of distributions using the GUM uncertainty framework and a Monte Carlo method, is available [31].

4.2.6 Multi-stage measurement models

Multi-stage measurement models are widespread in metrology. Even the string example (Section 4.2.2, Example 6) can be interpreted this way. Any situation in which the output quantities from one uncertainty evaluation become the input quantities to a subsequent evaluation constitute (part of) a multi-stage measurement model. Within a measurement model there are frequently sub-models, and therefore multi-staging arises also in this context. Examples abound, especially within calibration.

In the first stage of a multi-stage measurement model, the metrologist is responsible for providing (information about) all the input quantities. In subsequent stages, the input quantities constitute some or all of the output quantities from previous stages plus, possibly, further input quantities from the metrologist.

Example 8 *Multi-stage measurement model in calibration*

An example of a multi-stage measurement model occurs regularly in calibration, when it is required to establish and use a calibration function. The following description is in the context of the GUM uncertainty framework. There would be an analogous description for circumstances where it was necessary to avoid any limitation of the GUM uncertainty framework and use instead a Monte Carlo method as an implementation of the propagation of distributions.

Stage 1, calibration, involves an indication quantity (a response quantity) that is a function of a second quantity, e.g., displacement as a function of temperature. Values of displacement, and perhaps the corresponding values of temperature, if they are not known accurately, constitute instances of the (first-stage) input quantities. The measurement model specifies the process of fitting a calibration function to the input quantities to provide the coefficients or parameters of the function. These parameters constitute the output quantities. If there is more than one parameter (the usual case), they will almost invariably be correlated, since each parameter will generally be a function of the (same) input quantities. Given estimates of the input quantities and associated uncertainties (and covariances if relevant), estimates of the output quantities can be produced, as can the associated (non-diagonal) covariance matrix.

Stage 2, prediction, involves using the output quantities from Stage 1, viz., the calibration function parameters (realized by estimates of these parameters and the associ-

ated covariance matrix), as input quantities to a measurement model that constitutes a rule for evaluating the calibration function (inversely) for appropriate values of the argument (displacement in the above instance).[6] The output quantities will be the calibration function evaluated (inversely) at these designated points (realized by estimates from the calibration function together with the associated covariance matrix). Again, because these output quantities depend in general on common input quantities, i.e., the calibration function parameters, the covariance matrix will be non-diagonal.

There may be no further stage, since the predicted values provided by the calibration function may be the primary requirement.

Otherwise, Stage 3 will be the use of the (output) quantities obtained in Stage 2 as input quantities to provide further output quantities. As an example, take the area under (a specified portion of) the calibration function. Suppose that this area is to be determined by numerical quadrature[7] because of the difficulty or impossibility of carrying out the integration analytically. This area can typically be expressed as a linear combination of the calibration function for a number of values of the stimulus quantity. As another instance, if more than one output quantity is required, such as gradients to the calibration function at various points, these again can typically be expressed as linear combinations of the calibration function for values of the stimulus quantity. Again at Stage 3 the input quantities will be estimated and associated uncertainties will be available. The output quantities will, as a result, be estimated and an associated covariance matrix be provided.

4.2.7 Characterizing the input quantities by probability distributions

The provision of PDFs for the input quantities requires the quantities are characterized by appropriate probability distributions (rectangular, Gaussian, etc.). It can be a challenging step in the formulation stage of uncertainty evaluation. Valuable guidance is given in the GUM and the first Supplement to the GUM [11] on this matter. Additional aspects are considered here.

Sometimes these PDFs will be the consequence of a previous 'uncertainty calculation' within the context of a multi-stage measurement model (Section 4.2.6).

In the above straight-line calibration example (Section 4.2.1, Example 5, and Section 4.2.2, Example 7), the PDF for each input quantity would often be taken as Gaussian. There would be other types of measurement for which quantities would be characterized by other distributions (such as Poissonian distributions for counting processes).

[6] In most situations, it is necessary, as here, to use the calibration function *inversely*. Typically, the data in Stage 1 represents a set of measurement *standards*, e.g., established controls or stimuli. At Stage 2, it is required to use the calibration function to determine an estimate of the stimulus corresponding to a measured response value.

[7] A quadrature rule is a numerical integration procedure. Examples are the trapezoidal rule and Simpson's rule.

Information concerning the underlying PDF should be deduced in any one instance from all the knowledge that can economically be brought to bear (Appendix A.5).

There is an important class of metrology problems, viz., calibration as described in Section 4.2.6, Example 8, or generally the analysis of experimental data. Suppose that there is a large number of measurement data of comparable size, such as in the straight-line calibration example in Section 4.2.1. Suppose also that the corresponding quantities can be taken as mutually independent. For a calibration function that can be expressed as a linear combination of calibration parameters, the estimates of these parameters can formally be written as a linear combination of the measured values. For the large number of measurement data envisaged, the statistics of the situation are such that almost regardless of the nature of the underlying probability distribution, the quantity realized by a linear combination of the measurement data, as here, can be expected to have essentially a Gaussian distribution, as a consequence of the central limit theorem [84, p165]. When there are several such parameters (output quantities) they will almost invariably be mutually dependent, since each is a linear combination of the input quantities. These parameters would be characterized by a multivariate (joint) Gaussian distribution: see Section 4.2.7. The straight line in Section 4.2.1, Example 5, would have an intercept and a gradient that are generally mutually dependent.[8]

Even if the calibration function depends non-linearly on its parameters, by linearizing this function about the estimates of the parameters, to a first order approximation similar considerations apply as in the linear case [3]. In cases of doubt the validation procedures of Chapter 8 should be undertaken to determine whether linearization is justified.

The GUM discriminates between Type A evaluation of uncertainty—that based on statistical means—and Type B evaluation of uncertainty—that based on non-statistical means. Although this terminology is sometimes used in this guide for compatibility with the GUM, no great distinction is made here, since all types of uncertainties can be classified by appealing to unifying principles such as Bayes' theorem (Appendix A.4) and the Principle of Maximum Entropy (Appendix A.5). It is sometimes more useful to examine the distinction between error quantities that can be regarded as random and those that can be regarded as systematic. The subdivision into Type A and Type B evaluations of uncertainty will correspond in some instances to random and systematic error quantities, respectively, but not in all circumstances. In some instances a systematic error quantity can be treated as an offset

[8] It is possible in some applications such as this one to re-express (re-parametrize) the straight line such that its parameters are mutually independent [3]. Also see Section 4.2.4. The example in Clause H.3 of the GUM illustrates this point. Such re-parametrization is not *always* a practical proposition, however, because of the conflict between a numerically or statistically convenient representation and the requirements of the application. However, the possibility of re-parametrization should always be considered carefully for at least two reasons. One reason is that the result corresponding to a sound parametrization can be obtained in a numerically stable manner [3], whereas a poor parametrization can lead to numerically suspect results. Another reason is that a poor parametrization leads to correlations associated with the estimates of the output quantities that are large compared with the correlations associated with parameter estimates using a good parametrization. Decisions about the natural correlation present in the results cannot readily be made in terms of these *induced* correlation effects.

and handled as part of statistical modelling (Sections 4.2.1 and 9.6).

Instances of the PDF used to characterize an input quantity X for some common circumstances are given in Table 4.1 [11].

All available information concerning quantity X	PDF for X
Estimate x and the associated standard uncertainty $u(x)$	Gaussian distribution $\mathrm{N}(x, u^2(x))$
Estimate $\boldsymbol{x}$ of a multivariate quantity $\boldsymbol{X}$ and the associated covariance matrix $\boldsymbol{U}_{\boldsymbol{x}}$	Multivariate Gaussian distribution $\mathrm{N}(\boldsymbol{x}, \boldsymbol{U}_{\boldsymbol{x}})$
Endpoints a and b of an interval containing X	Rectangular distribution $\mathrm{R}(a, b)$ with endpoints a and b
Estimate x (> 0) and X is known to be nonnegative	Exponential distribution $\mathrm{Ex}(\lambda)$ with parameter $\lambda = 1/x$ for which the PDF is $\exp(-\xi/x)/x$ for $\xi \geq 0$, and zero otherwise
Indication values regarded as random draws made independently from a Gaussian distribution with unknown expectation (equal to X) and unknown variance. From a sample of size n, an average $\overline{x}$ and a standard deviation s have been calculated	t-distribution $\mathrm{t}_\nu(\overline{x}, s^2/n)$ with $\nu = n - 1$ degrees of freedom
Lower and upper limits a and b of an interval within which X is known to cycle sinusoidally	U-shaped (arcsine) distribution $\mathrm{U}(a, b)$ with endpoints a and b for which the PDF is $(2/\pi)/\{(b-a)^2 - (2\xi - b - a)^2\}^{-1/2}$ for $a < \xi < b$, and zero otherwise [67, Section 3.5]

Table 4.1: PDF used to characterize an input quantity X based on available information for some common circumstances.

Univariate Gaussian distribution

There are many circumstances where measured quantities are influenced by a large number of error quantities and no one error quantity dominates. In these situations it is reasonable

to characterize the measured quantities by Gaussian distributions. One common instance is a parameter arising from least-squares fitting to a large number of measured points.

The (univariate) *Gaussian* or *normal* distribution $\mathrm{N}(\mu, \sigma^2)$ for the quantity X with expectation μ and standard deviation σ, has the PDF

$$g_X(\xi) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-(\xi-\mu)^2/(2\sigma^2)\right\}, \quad -\infty < \xi < \infty.$$

The *standardized Gaussian* distribution for the quantity X with zero expectation and unit standard deviation is

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \exp(-z^2/2), \quad -\infty < z < \infty.$$

Its distribution function, denoted by $\Phi(z)$, is

$$\Phi(z) = \int_{-\infty}^{z} \phi(\xi)\,\mathrm{d}\xi.$$

The probability that X lies between c and d, where $c < d$, is

$$\begin{aligned}\frac{1}{\sigma\sqrt{2\pi}} \int_c^d \exp\left\{-(\xi-\mu)^2/(2\sigma^2)\right\}\,\mathrm{d}\xi &= \int_{(c-\mu)/\sigma}^{(d-\mu)/\sigma} \exp(-z^2/2)\,\mathrm{d}z \\ &= \Phi((d-\mu)/\sigma) - \Phi((c-\mu)/\sigma).\end{aligned}$$

The inverse function $\Phi^{-1}(p)$ gives the value of z such that $\Phi(z) = p$, a stated probability. Tables and software for Φ and its inverse are widely available.

Multivariate Gaussian distribution

In general, multivariate distributions are defined in terms of joint PDFs $g_{\boldsymbol{X}}(\boldsymbol{\xi})$. The *multivariate Gaussian* distribution (or multinormal distribution) $\mathrm{N}(\boldsymbol{\mu}, \boldsymbol{V})$ for the quantities $\boldsymbol{X} = (X_1, \ldots, X_N)^\top$ with expectation $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_N)^\mathrm{T}$ and covariance matrix $\boldsymbol{V}$ of dimension $N \times N$ has PDF

$$g_{\boldsymbol{X}}(\boldsymbol{\xi}) = \frac{1}{((2\pi)^N \det \boldsymbol{V})^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{\xi}-\boldsymbol{\mu})^\top \boldsymbol{V}^{-1}(\boldsymbol{\xi}-\boldsymbol{\mu})\right\}.$$

The *set* of parameters arising from least-squares fitting can often be described by such a distribution.

Rectangular distribution

It is often assumed that when the value of an input quantity is given in a manufacturer's specification in the form of a 'plus/minus accuracy statement', the corresponding PDF should be

taken as rectangular with limits given by the accuracy statement. If there is no other information available, this attitude is consistent with the Principle of Maximum Entropy (PME). See Appendix A.5.

The *rectangular* or *uniform* distribution $\mathrm{R}(a, b)$ with endpoints a and b has PDF

$$g_X(\xi) = \begin{cases} 1/(b-a), & a \le \xi \le b, \\ 0, & \text{otherwise.} \end{cases}$$

It states that any value of X in the interval $[a, b]$ is equally probable and that the probability of a value of X outside this interval is zero.

The probability that X lies between c and d, where $c < d$, is

$$\int_c^d g_X(\xi)\,\mathrm{d}\xi = \begin{cases} 0, & d \le a, \\ (d-a)/(b-a), & c \le a \le d \le b, \\ (d-c)/(b-a), & a \le c < d \le b, \\ (b-c)/(b-a), & a \le c \le b \le d, \\ 0, & b \le c. \end{cases}$$

Can taking a rectangular PDF be better in general than using, say, a Gaussian? There are indeed genuine instances for the use of a rectangular PDF. An example is the digital resolution of an instrument, in which the deviation can be regarded as being equally likely anywhere within plus or minus half a unit in the last displayed digit.[9]

The quantization error in analogue to digital conversion also falls (with some exceptions) into this category. There would appear to be few other genuine instances. It would be desirable, especially in a competitive environment or when particularly reliable uncertainty statements are required, to approach suppliers to relate the provided accuracy statement to the context in which it was made. The supplier might, for example, be speaking loosely, e.g., to imply a $99\,\%$ coverage interval, say, with the previously unmentioned information that an underlying Gaussian PDF was reasonable. The contextual information might relate, for example, to reject rates in a production process.

Information is available [16] on a method for reducing the uncertainty associated with instrument resolution when a series of indication values is obtained. It involves randomizing

[9] This statement is correct for a single indication value. There are additional considerations for a sequence of indication values corresponding to a slowly varying signal. The deviation values in the resolved sequence have *serial correlation* associated with them as a consequence of the resolution of the instrument. Figure 4.1 shows successive deviation values displayed by a simulated instrument having a resolution of two decimal places. The values shown are the differences between the values of $\sin t$ that would be displayed by the instrument and the actual values of $\sin t$, for $t = 1.00, 1.01, \ldots, 1.10$ radians. Any analysis of such data that did not take account of the very obvious serial correlation would yield a flawed result. The effects of serial correlation depend on the relative sizes of the uncertainties associated with values of the signal, the instrument resolution and the magnitudes of the changes in successive values of the signal (the last-mentioned item depending on the sampling rate). In hopefully many cases they will be negligible, but it is appropriate to establish when this is indeed the case. In the context of calibration it is stated [38], but the point is more general, that the measurement uncertainty associated with the calibration of all low-resolution indicating instruments is dominated by the finite resolution provided this resolution is the only dominant source in the uncertainty budget.

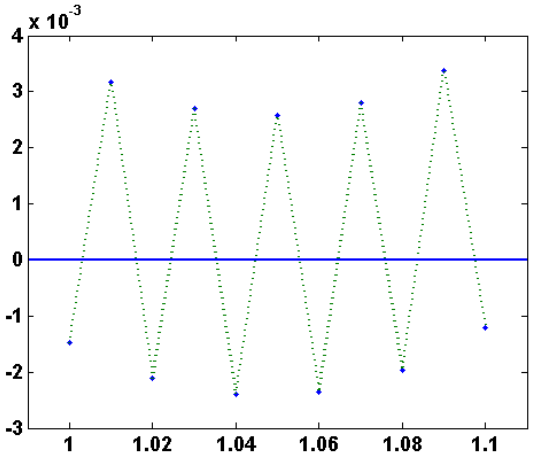

Figure 4.1: Successive deviation values displayed by a simulated instrument having a resolution of two decimal places. The values shown are the differences between the values of $\sin t$ that would be displayed by the instrument and the actual values of $\sin t$, for $t = 1.00, 1.01, \ldots, 1.10$ radians.

the zero setting, where this is possible, before obtaining each indication value. The average of a set of q indication values so obtained can be expected to have an associated uncertainty that is smaller than that of an individual indication value by a factor of $\sqrt{q}$. This result is to be compared with conventional repeated indication values in situations where the uncertainties are dominated by those of the instrument resolution: the average of the indication values has no better property than the individual indication values.

Inexactly specified rectangular distributions

Consider a random variable X, having nominally a rectangular PDF, specified in terms of its lower limit A and upper limit B. The knowledge about these endpoints may be incomplete. For instance, suppose the estimates $a = -1$ and $b = 1$ are given, only the quoted figures are reliable, and no other information is available. Then, it can be concluded that A lies between -1.5 and -0.5 and B between 0.5 and 1.5.[10] Thus, X in fact lies in the broader interval $[-1.5, 1.5]$ rather than $[-1, 1]$. See Figure 4.2. How important is this consideration in practice? In what manner is X distributed over this interval?

These considerations are a direct counterpart of those in the GUM in which the standard uncertainty associated with the estimate of an input quantity is obtained from a Type B evaluation and cannot be treated as exactly known. See GUM Clause G.4.2. There the inexactness is manifested as an effective degrees of freedom.

Suppose that the left endpoint is regarded as lying in the interval $[a - d, a + d]$ and that of

[10] If instead the estimates $a = -1.0$ and $b = 1.0$ were quoted, it would be concluded that the left endpoint were between -1.05 and -0.95 and the right endpoint between 0.95 and 1.05.

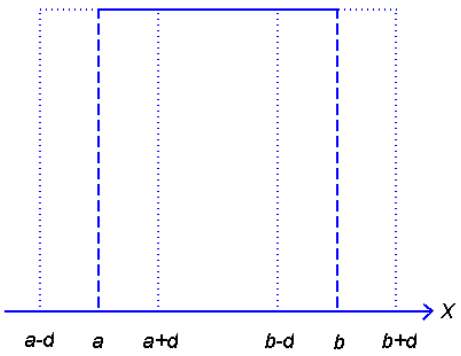

Figure 4.2: Rectangular PDF with inexact endpoints. The diagram is conceptual: the 'height' of the PDF would in fact vary with the endpoints in order to maintain unit area.

the right endpoint in $[b-d, b+d]$. It is assumed that 'the d' is the same for each endpoint. The treatment can be generalized if needed. It is henceforth assumed that the left endpoint is equally likely to lie *anywhere* in $[a-d, a+d]$, with a similar statement for the right endpoint.[11] Thus, the left and right endpoints are taken as rectangular random variables, A and B, say. It follows that

$$X = A + (B - A)R,$$

where A is rectangular over $[a-d, a+d]$, B is rectangular over $[b-d, b+d]$ and R is rectangular over $[0, 1]$.

An application of a Monte Carlo method with $a = -1.0$, $b = 1.0$ and $d = 0.5$ gave the approximation in Figure 4.3 to the PDF for X.

Note the 'shape' of the PDF. It is rectangular over the interval between the inner extremities of the inexact endpoints, i.e., where there is no doubt concerning their location. Between the inner and outer extremities it reduces from the rectangular height to zero in what is approximately a quadratic manner. Beyond the outer extremities the PDF is zero. The piecewise nature of the PDF is comparable to that for the sum of quantities characterized by rectangular PDFs, where the pieces form polynomial segments [36].

The standard deviation of X, characterized by a rectangular distribution over $[-1, 1]$ and assuming the exactness of the endpoints (equivalent to taking $d = 0$), is $1/\sqrt{3} = 0.577$. That for the above finite value of d is 0.625. As might be expected, the inexactness of the endpoints increases the value. The extent to which this increase ($8\,\%$) is important depends

[11] An alternative approach can be used. It could be assumed, for instance, that each endpoint can be regarded as a Gaussian (rather than a rectangular) variable, centred on that endpoint, with a stated standard deviation. The analysis and the result would differ from that here. The choice of approach would be made using expert judgement.

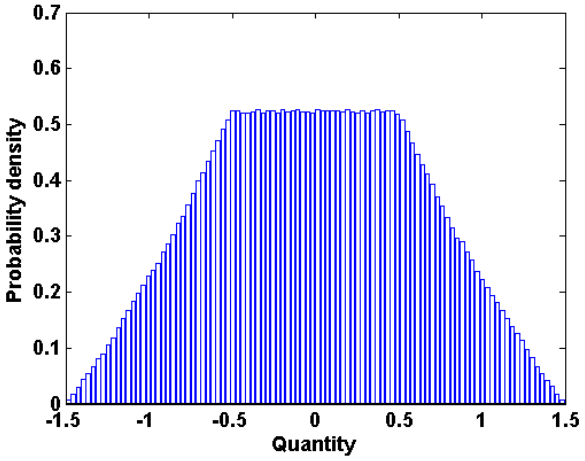

Figure 4.3: Histogram produced using an application of a Monte Carlo method for the model $X = A + (B - A)R$, where A is rectangular over $[a - d, a + d]$, B is rectangular over $[b - d, b + d]$ and R is rectangular over $[0, 1]$, with $a = -1.0$, $b = 1.0$ and $d = 0.5$. Compare with Figure 4.2. The histogram provides a scaled approximation to the PDF for X. It corresponds to an input quantity that is characterized by a rectangular distribution with inexact limits, each being represented by a rectangular distribution.

on circumstances.

There will be situations where the inexact endpoints would be expected to 'move together', i.e., the knowledge of one of them would imply the other. In this circumstance the PDF for X is slightly different. See Figure 4.4. The standard deviation of X is now 0.600 (a $4\,\%$ increase over that for $d = 0$), roughly halfway between that for the PDF illustrated in Figure 4.3 and the pure rectangular PDF. The flanks of the PDF now have greater curvature.

The final statement to be made here concerning rectangular distributions with inexactly defined endpoints is that the effects of such endpoints on the evaluation of uncertainty increase with the relative amount of inexactness. This point is qualitatively consistent with the use of an effective degrees of freedom, as above, in the GUM. Increased inexactness will give rise to a smaller degrees of freedom and yield greater uncertainty through a larger coverage factor from the t–distribution.

Taking account of the available information

It is beyond the scope, in this edition of the best-practice guide, to state how all information available can properly be taken into account. Some remarks are made, however, indicating how Bayes' theorem and the Principle of Maximum Entropy can be used to advantage [31].

If only a lower and an upper limit were available the Principle of Maximum Entropy would support the choice of a rectangular PDF (Appendix A.5).

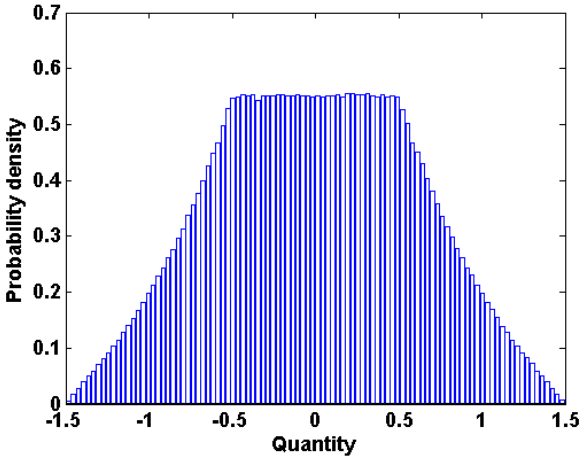

Figure 4.4: As Figure 4.3 except that the endpoints are related as described in the text.

Suppose a prior rectangular PDF were available, perhaps from sound knowledge of limits a and b, and that one measured value x was available. From the Principle of Maximum Entropy, the rectangular PDF that would be inferred from the limits alone would be modified by the measured value. The GUM provides (in GUM Clause 4.3.8) the PDF in this case. It is given by

$$g_X(\xi) = \begin{cases} Ae^{-\lambda(\xi - x)}, & a \leq \xi \leq b, \\ 0, & \text{otherwise}, \end{cases}$$

where A and λ are parameters of the distribution that are functions of a, b and x. See reference [31] for further information.

Suppose a prior PDF, say a Gaussian, were available, perhaps from historical information such as that obtained in previous calibrations. Suppose that further measured values were available. The use of Bayes' Theorem (Appendix A.4) would permit both sources of information to be combined to deliver a t–distribution that could be expected to be more reliable than the PDF from either source alone.

Other cases can be handled, and give superior results in general than if treated without taking account of the available information. It is relevant to note that in the context of the GUM uncertainty framework, which works only with the standard deviations (and the expectations) of the input quantities, the GUM states

> [GUM Clause E.4.2] When the standard uncertainty of an input quantity cannot be evaluated by analysis of the results of an adequate number of repeated observations, a probability distribution must be adopted based on knowledge that is much less extensive than might be desirable. That does not, however, make the distribution invalid or unreal; like all probability distributions it is an expression of what knowledge exists.

This attitude is consistent with a Bayesian view [104].

4.3 Propagation stage

The PDF for the output quantity is completely defined by the measurement model together with the PDFs for the input quantities. Appropriate analysis or calculation is needed, however, to determine it. Chapter 5 covers candidate approaches for forming the PDF for the output quantity in the univariate case, and indicates its counterpart in the multivariate case.

4.4 Summarizing stage

The provision of a coverage interval involves the use of the PDF for the output quantity to determine a lower limit and an upper limit of an interval that can be expected to contain $95\,\%$ (or some other specified proportion) of the values that can reasonably be attributed to the output quantity. See Chapter 5 for methods for determining the PDF for the output quantity. See Section 2.3 and Appendix A.3 for information on coverage intervals. Coverage intervals can be obtained objectively from a PDF. They can also be obtained from *coverage factors* and an *assumption* concerning the PDF.

4.4.1 Coverage intervals from distribution functions

If the distribution function is known, a coverage interval can be obtained as indicated in Section 2.3 and Appendix A.3.

4.4.2 Coverage intervals from coverage factors and an assumed form for the distribution function

The approach based on the GUM uncertainty framework (see GUM Clause G.1.1) to determining a coverage interval is as follows. The aim (using the notation of the GUM) is to provide, using the estimate y of the output quantity Y and the standard uncertainty $u(y)$ associated with the estimate, an expanded uncertainty $U_p = k_p u(y)$. With the estimate y, this value U_p defines an interval $[y - U_p, y + U_p]$ corresponding to a specified coverage probability p.

In summarizing its recommendations for determining this coverage interval, the GUM states:

> [GUM Clause G.6.1] The coverage factor k_p that provides an interval having a level of confidence p close to a specified level can only be found if there

> is extensive knowledge of the probability distribution of each input quantity and if these distributions are combined to obtain the distribution of the output quantity. The input estimates x_i and their standard uncertainties $u(x_i)$ by themselves are inadequate for this purpose.

Further,

> [GUM Clause G.6.2] Because the extensive computations required to combine probability distributions are seldom justified by the extent and reliability of the available information, an approximation to the distribution of the output quantity is acceptable. Because of the Central Limit Theorem, it is usually sufficient to assume that the probability distribution of $(y-Y)/u_{\rm c}(y)$ is the t-distribution and take $k_p = t_p(\nu_{\rm eff})$, with the t-factor based on an effective degrees of freedom $\nu_{\rm eff}$ of $u_{\rm c}(y)$ obtained from the Welch-Satterthwaite formula ...

The statement[12] concerning the extensive computation to combine probability distributions is no longer tenable. Unless the measurement model is complicated, the determination of the PDF for the output quantity and hence the required coverage interval to the required number of decimal digits, can, with today's PCs, be carried out in computation times of seconds.

4.4.3 Acceptability of an approximation?

The statement from Clause G.6.2 of the GUM reproduced in Section 4.4.2 concerning the central limit theorem demands investigation. It is accepted that it is *usually* sufficient to assume that the PDF for $(y - Y)/u(y)$ is a t–distribution. The difficulty lies in deciding when this assumption can be made. The GUM offers no *specific* guidance in this regard. This document supports that approach when it can be justified, but recommends that in any case of doubt the validation approach of Chapter 8 should be employed.

However, the GUM does provide some advice regarding the circumstances when the GUM uncertainty framework can be expected to hold:

> [GUM Clause G.6.6] For many practical measurements in a broad range of fields, the following conditions prevail:
>
> — the estimate y of the measurand Y is obtained from estimates x_i of a significant number of input quantities X_i that are describable by well-behaved probability distributions, such as the normal and rectangular distributions;

[12] The GUM uses the notation $u_{\rm c}(y)$ for *combined standard uncertainty*, i.e., that associated with y. This guide simply uses $u(y)$.

— the standard uncertainties $u(x_i)$ of these estimates, which may be obtained from either Type A or Type B evaluations, contribute comparable amounts to the combined standard uncertainty $u_c(y)$ of the measurement result y;

— the linear approximation implied by the law of propagation of uncertainty is adequate (see 5.1.2 and E.3.1);

— the uncertainty of $u_c(y)$ is reasonably small because its effective degrees of freedom ν_{eff} has a significant magnitude, say greater than 10.

Under these circumstances, the probability distribution characterized by the measurement result and its combined standard uncertainty can be assumed to be normal because of the Central Limit Theorem; and $u_c(y)$ can be taken as a reasonably reliable estimate of the standard deviation of the normal distribution because of the significant size of ν_{eff}.

This advice is sound in a qualitative sense but, again, it is unclear when the circumstances hold. The problem is that the distinction between the formulation and calculation (propagation and summarizing) stages of uncertainty evaluation, as indicated in Section 3.2, becomes blurred. The intention of the subdivision into the stages is to permit all decisions to be made in the formulation stage and the calculations to be made in the propagation and summarizing stages.

In terms of the set of conditions in GUM Clause G.6.6, listed above, it is unclear what is meant by

- 'a significant number of input quantities',
- 'well-behaved probability distributions',
- the standard uncertainties of the x_i contributing comparable amounts,[13]
- the adequacy of linear approximation, and
- the output uncertainty being reasonably small.

The concern is that because none of these considerations is explicitly quantified, different practitioners might adopt different interpretations of the same situation, thus causing divergence of results.

4.4.4 When the worst comes to the worst

Consider a situation in which no assumption is to be made concerning the PDF for the output quantity other than an estimate y of its expectation and the standard uncertainty $u(y)$

[13] This statement is taken here to mean that the standard uncertainties associated with the estimates of the input quantities, when scaled by the magnitudes of the corresponding sensitivity coefficients, contribute comparable amounts.

associated with this estimate as its standard deviation. One reason for wishing to make no assumption is that it may be difficult or impossible to obtain distributional information about some of the input quantities and it is deemed inappropriate to invoke the Principle of Maximum Entropy. In such a circumstance, a *conservative* estimate of a coverage interval can be obtained using some traditional results from the statistical literature.[14] Two results are possible. One result is general, applying to all distributions. The other relates to instances in which one is prepared to make a single assumption, viz., that the distribution is symmetric.

General distributions

Suppose that it is required to quote a coverage interval for the output quantity Y corresponding to a coverage probability of $95\,\%$, and that *nothing* is known about the distribution.

The coverage interval $y \pm ku(y)$, where $k = 4.47$, contains at least $95\,\%$ of the values of Y.

This result is derived from *Chebyshev's inequality* which states that the probability that Y lies in the interval $y \pm ku(y)$ is at least $1 - k^{-2}$. The value of k for which $1 - k^{-2} = 0.95$ is 4.47. It is stressed that this result applies *regardless of the distribution*. By its nature it cannot be as sharp as an interval derived from knowledge of the PDF for Y, e.g.,

- if Y is characterized by a rectangular distribution this interval is $y \pm 1.65u(y)$;
- if Y is characterized by a Gaussian distribution it is $y \pm 1.96u(y)$.

The length of the interval derived from Chebyshev's inequality is 2.7 times the length of that for a rectangular distribution for Y and 2.3 times that for a Gaussian distribution for Y.

Note. These results apply only if the degrees of freedom is infinite, or in practice large. Otherwise, the k-factor becomes inflated, as in the case of the t–distribution [89].

Symmetric distributions

If it is known that the distribution is *symmetric* and *unimodal* (single-peaked), tighter results based on *Gauss's inequality* are possible.

The coverage interval $y \pm ku(y)$, where $k = 2.98$, contains at least $95\,\%$ of the values of Y.

[14] Such an estimate is inconsistent with the intention of the GUM which promotes the use of a *realistic* coverage interval:

> [GUM, Clause 0.4] ... the ideal method for evaluating and expressing uncertainty in measurement should be capable of readily providing such an interval, in particular, one with a coverage probability or level of probability that corresponds in a realistic way with that required.

There may, however, be special situations where a conservative estimate is useful.

Gauss's inequality states that the probability that Y lies in the interval $y \pm ku(y)$ is at least $1 - \frac{4}{9}k^{-2}$. The value of k for which $1 - \frac{4}{9}k^{-2} = 0.95$ is 2.98.

It is noted that this interval is only approximately $50\,\%$ longer than that when Y is characterized by a Gaussian distribution.

Note. Again, these results apply only if the degrees of freedom is infinite, or in practice large. Otherwise, the k-factor becomes inflated, as in the case of the t–distribution.

Chapter 5

Approaches to the propagation of distributions

5.1 Overview

This chapter covers candidate solution procedures for the propagation of distributions in the propagation stage of the uncertainty evaluation problem formulated in Section 3.2. The starting point is (a) the availability of a measurement function f or $\boldsymbol{f}$ that relates the input quantities $\boldsymbol{X} = (X_1, \ldots, X_N)^\top$ to the scalar output quantity Y or vector output quantity $\boldsymbol{Y} = (Y_1, \ldots, Y_m)^\top$ through $Y = f(\boldsymbol{X})$ or $\boldsymbol{Y} = \boldsymbol{f}(\boldsymbol{X})$, and (b) PDFs $g_{X_1}(\xi_1), \ldots, g_{X_N}(\xi_N)$ for the input quantities. If the input quantities are mutually dependent, they are characterized by a joint PDF $g_{\boldsymbol{X}}(\boldsymbol{\xi})$.

It is required to determine the PDF $g_Y(\eta)$ for the output quantity Y or the (joint) PDF $g_{\boldsymbol{Y}}(\boldsymbol{\eta})$ for $\boldsymbol{Y}$. Once $g_Y(\eta)$ has been obtained a $95\,\%$ coverage *interval* for the (scalar) output quantity Y can be derived. Once $g_{\boldsymbol{Y}}(\boldsymbol{\eta})$ has been obtained, a $95\,\%$ coverage *region* for the (vector) output quantity $\boldsymbol{Y}$ can be derived.

Three approaches to the determination of the PDF for Y or $\boldsymbol{Y}$ are considered and contrasted:

1. Analytical methods;
2. The GUM uncertainty framework;
3. Numerical methods.

All three approaches are consistent with the GUM. The GUM uncertainty framework is the procedure that is widely used and summarized in GUM Clause 8. Analytical methods and numerical methods fall in the category of 'other analytical and numerical methods' (GUM

Clause G.1.5). Under the heading of 'Analytical methods' below, mention is also made of 'Approximate analytical methods'.

5.2 Analytical methods

Analytical methods to obtain the PDF for Y or $\boldsymbol{Y}$ are preferable in that they do not introduce any approximation, but can be applied in relatively simple cases only. A treatment of such methods, based essentially on the use of formula (3.1) is available [36]. Instances that can be so handled include linear measurement functions, $Y = c_1X_1 + \ldots + c_NX_N$, where all X_i are Gaussian or all are rectangular. In the latter case, unless N is small, the multipliers c_i must be equal and the semi-widths of the rectangular PDFs identical[1] to avoid formidable algebra.

5.2.1 Scalar input quantity

The case of a single (scalar) input quantity ($N = 1$) is amenable to analytic treatment [84, pp57-61]. If the measurement function $f(X)$ is differentiable and strictly monotonic, Y has the PDF

$$g_Y(\eta) = \left|\frac{\mathrm{d}\xi}{\mathrm{d}\eta}\right| g_X(\xi), \quad \eta = f(\xi). \tag{5.1}$$

Example 9 *Logarithmic transformation*

If the measurement function is $Y = \ln X$ with $X > 0$ and X characterized by a rectangular PDF with limits a and b ($0 < a < b$), the application of formula (5.1) gives

$$g_Y(\eta) = \begin{cases} 0, & \eta < \ln a, \\ \exp(\eta)/(b-a), & \ln a < \eta < \ln b, \\ 0, & \ln b < \eta. \end{cases}$$

Figure 5.1 depicts the rectangular PDF for X and (right) the corresponding PDF for Y in the case $a = 1$, $b = 3$.

This case is important in, say, electromagnetic compatibility measurement, where conversions are often carried out between quantities expressed in linear and decibel units using exponential or logarithmic transformations [101].

[1] In this case Y has a PDF that is a B-spline with uniform knots (Example 12) [26].

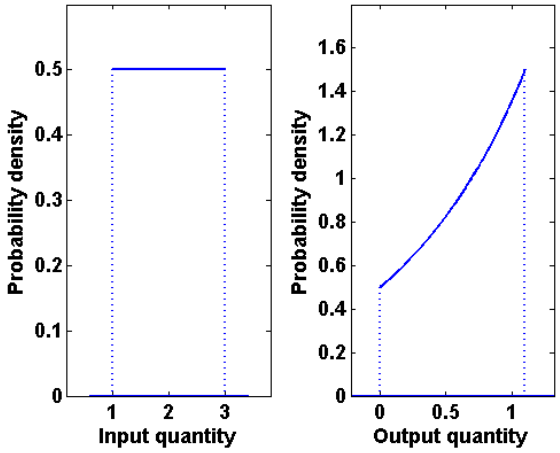

Figure 5.1: Rectangular PDF (left) for the input quantity X and the corresponding PDF for the output quantity Y, where Y is related to X by the measurement function $Y = \ln X$.

Example 10 *Linear combination of Gaussian distributions*

Suppose the measurement function is

$$Y = c_1 X_1 + \cdots + c_N X_N,$$

where $c_1, \ldots, c_N$ are specified constants, and, for $i = 1, \ldots, N$, X_i is characterized by the Gaussian distribution $\mathrm{N}(\mu_i, \sigma_i^2)$. Then, Y is described by the Gaussian distribution $\mathrm{N}(\mu, \sigma^2)$, where $\mu = c_1\mu_1 + \cdots + c_N\mu_N$ and $\sigma^2 = c_1^2\sigma_1^2 + \cdots + c_N^2\sigma_N^2$.

Example 11 *Sum of two rectangular distributions with the same semi-widths*

Suppose the measurement function is

$$Y = X_1 + X_2$$

and, for $i = 1, 2$, X_i with expectation μ_i and standard deviation $a/\sqrt{3}$ is characterized by a rectangular distribution (with semi-width a). Then, Y has expectation $\mu = \mu_1 + \mu_2$, standard deviation $a\sqrt{2/3}$ and is described by a symmetric triangular PDF $g_Y(\eta)$ with semi-width $2a$. For the case $\mu_1 + \mu_2 = 0$, this PDF takes the form

$$g_Y(\eta) = \begin{cases} 0, & \eta \leq -2a, \\ (2a + \eta)/(4a^2), & -2a \leq \eta \leq 0, \\ (2a - \eta)/(4a^2), & 0 \leq \eta \leq 2a, \\ 0, & 2a \leq \eta. \end{cases}$$

For general μ_1 and μ_2, the PDF is the same, but centred on $\mu_1 + \mu_2$ rather than zero. Geometrically, this PDF takes the form indicated in Figure 5.2.

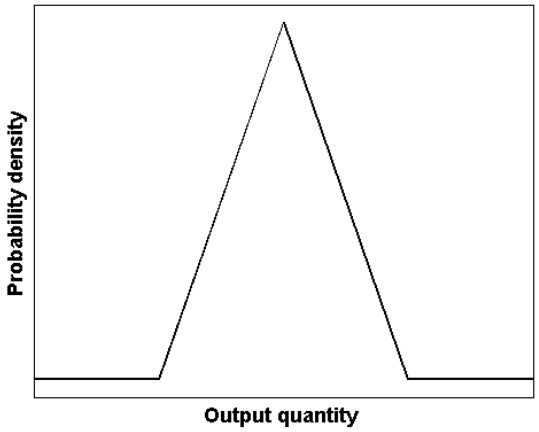

Figure 5.2: PDF for the output quantity Y in the measurement function $Y = X_1 + X_2$, where the input quantities X_1 and X_2 are characterized by rectangular distributions with identical semi-widths.

Example 12 *Sum of N rectangular distributions of the same semi-width*

Suppose the measurement function is

$$Y = X_1 + \cdots + X_N$$

and, for $i = 1, \ldots, N$, X_i with expectation μ_i and standard deviation $a/\sqrt{3}$ is characterized by a rectangular distribution (with semi-width a). Then, Y has expectation $\mu = \mu_1 + \cdots + \mu_N$ and standard deviation $a\sqrt{N/3}$ and is described by a PDF that is a B-spline of order N (degree $N - 1$) with $N - 1$ uniformly spaced knots. This PDF is symmetric about $\eta = \mu$ and has semi-width $Na/2$.

Example 13 *Sum of two rectangular distributions of arbitrary semi-widths*

Suppose the measurement function is

$$Y = c_1 X_1 + c_2 X_2$$

and, for $i = 1, 2$, X_i with expectation μ_i and standard deviation $a_i/\sqrt{3}$ is characterized by a rectangular distribution (with semi-width a_i). Then, Y has expectation $\mu = c_1\mu_1 + c_2\mu_2$, standard deviation $[(c_1^2 a_1^2 + c_2^2 a_2^2)/3]^{1/2}$ and is described by a symmetric trapezoidal PDF $g_Y(\eta)$ with semi-width $c_1 a_1 + c_2 a_2$. Geometrically, this PDF takes the form indicated in Figure 5.3.

Analytical solutions in some other simple cases are available [36, 38].

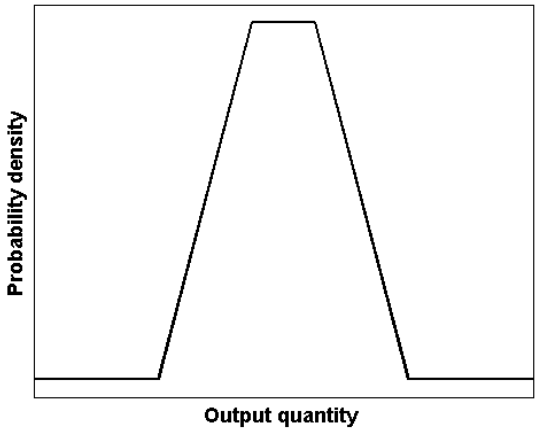

Figure 5.3: PDF for the output quantity in the measurement function $Y = c_1X_1 + c_2X_2$, where X_1 and X_2 are characterized by rectangular distributions with arbitrary semi-widths. It is symmetric about its midpoint.

5.2.2 Approximate analytical methods

Approximate analytical methods are approaches that fall part-way between the analytical methods of this section and the numerical methods of Section 5.4. They are related to the GUM uncertainty framework (Section 5.3), but take the analysis further in order to provide approximate analytical expressions for the PDF for the output quantity in cases where a Gaussian distribution (or t–distribution) obtained in the conventional way would be invalid.

A treatment [38] of some calibration examples using approximate analytic methods provides PDFs for the output quantity in the form of

1. a *rectangular* PDF in the calibration of a hand-held digital multimeter,

2. a *symmetric trapezoidal* PDF in the calibration of a vernier caliper, and

3. a further *trapezoidal* PDF in the calibration of a temperature block calibrator.

The first of these examples is used subsequently in this guide (Section 9.5) in the context of a Monte Carlo approach to uncertainty evaluation and the results compared with those of [38] and the GUM uncertainty framework.

5.3 The GUM uncertainty framework

The GUM makes the following statement about the PDFs for the input quantities:[2]

> [GUM Clause 4.1.6] Each input estimate x_i and its associated uncertainty $u(x_i)$ are obtained from a distribution of possible values of the input quantity X_i. This probability distribution may be frequency based, that is, based on a series of indications $x_{i,k}$ of X_i, or it may be an *a priori* distribution. Type A evaluations of standard uncertainty components are founded on frequency distributions while Type B evaluations are founded on *a priori* distributions. It must be recognized that in both cases the distributions are models that are used to represent the state of our knowledge.

Given the PDFs for the input quantities X_i of the model, the intent of the GUM uncertainty framework is to derive an estimate y of the output quantity Y, the associated standard uncertainty $u(y)$, and the effective degrees of freedom ν, and to characterize $(Y - y)/u(y)$ by a standard Gaussian distribution ($\nu = \infty$) or a t–distribution ($\nu < \infty$).

For the approach based on the GUM uncertainty framework, the following steps constitute the propagation and summarizing stages:

1. Obtain from the PDFs for the input quantities $X_1, \ldots, X_N$, respectively, the expectation $\boldsymbol{x} = (x_1, \ldots, x_N)^{\mathrm{T}}$ and the standard deviations $\boldsymbol{u}(\boldsymbol{x}) = (u(x_1), \ldots, u(x_N))^{\top}$. Use the joint PDF for $\boldsymbol{X}$ instead if the X_i are mutually dependent;
2. Take the *covariances* (mutual uncertainties) $\mathrm{cov}(x_i, x_j)$ as $\mathrm{Cov}(X_i, X_j)$, the covariances of mutually dependent pairs (X_i, X_j) of input quantities;
3. Form the partial derivatives of first order of f with respect to the input quantities. See Section 5.6;
4. Calculate the estimate y of the output quantity by evaluating the model at $\boldsymbol{x}$;
5. Calculate the sensitivity coefficients (GUM Clause 5.1) as the above partial derivatives evaluated at $\boldsymbol{x}$. See Section 5.6;
6. Determine the standard uncertainty $u(y)$ by combining $\boldsymbol{u}(\boldsymbol{x})$, the $\mathrm{cov}(x_i, x_j)$ and the sensitivity coefficients (GUM Formulae (10) and (13)). See Chapter 6;
7. Calculate ν, the effective degrees of freedom associated with $u(y)$, using the Welch-Satterthwaite formula (GUM Formula (G.2b));[3]

[2] To this statement, the comment must be added that some PDFs may be based on *both* types of information, viz., prior knowledge *and* repeated indication values. Evaluations of standard uncertainty in this setting are not purely Type A or Type B. The GUM gives one such instance (GUM Clause 4.3.8, Note 2.)

[3] The approach based on the GUM uncertainty framework does not state how ν is to be calculated when the input quantities are mutually dependent.

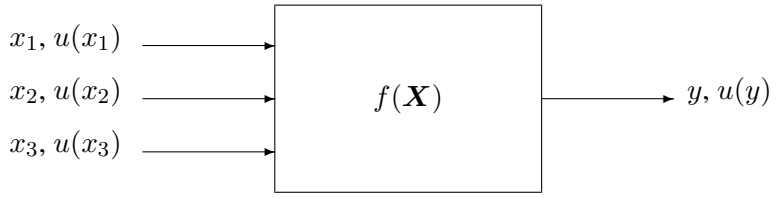

Figure 5.4: Illustration of the propagation of uncertainty in the GUM uncertainty framework. The measurement function $f(\boldsymbol{X})$ has three input quantities $\boldsymbol{X} = (X_1, X_2, X_3)^\top$, estimated by x_i with associated standard uncertainty $u(x_i)$, for $i = 1, 2, 3$. There is a single output quantity $Y \equiv Y_1$, estimated by y with associated standard uncertainty $u(y)$.

8. Compute the expanded uncertainty U_p, and hence a coverage interval for the output quantity (having a stipulated coverage probability p), by forming the appropriate multiple of $u(y)$ through taking the probability distribution of $(y-Y)/u(y)$ as a standard Gaussian distribution ($\nu = \infty$) or t–distribution ($\nu < \infty$).

Figure 5.4 illustrate the 'propagation of uncertainty' [10] for a measurement function $f(\boldsymbol{X})$ with three input quantities $\boldsymbol{X} = (X_1, X_2, X_3)^\top$, where X_i is estimated by x_i with associated standard uncertainty $u(x_i)$, and a single output quantity $Y \equiv Y_1$, estimated by $y = y_1$ with associated standard uncertainty $u(y) = u(y_1)$. The figure is the counterpart of Figure 3.1, which illustrates the propagation of distributions for a similar problem. In a more complicated circumstance, the input quantities would be mutually dependent, i.e., correlated, and additional information would be needed to quantify the correlations.

Figure 5.5 illustrates the GUM uncertainy framework diagrammatically. This figure applies in the case of a univariate, real measurement function with input quantities that are mutually independent. Other measurement model types would give diagrams that constitute a variant of Figure 5.5.

A review of the approach based on the GUM uncertainty framework is given in Section 5.5. Details, procedures and examples are given in Chapter 6.

5.4 Numerical methods

It would rarely be a practical proposition to use the integral expression (3.1) in Chapter 3 as the basis for the numerical determination of $g_Y(\eta)$, the PDF for the output quantity. A multivariate quadrature rule[4] would need to be devised that was capable of delivering $g_Y(\eta)$ to a prescribed numerical accuracy for each choice of η. Further, the quadrature rule would

[4] A quadrature rule is a numerical integration procedure. Examples in the univariate case are the trapezoidal rule and Simpson's rule.

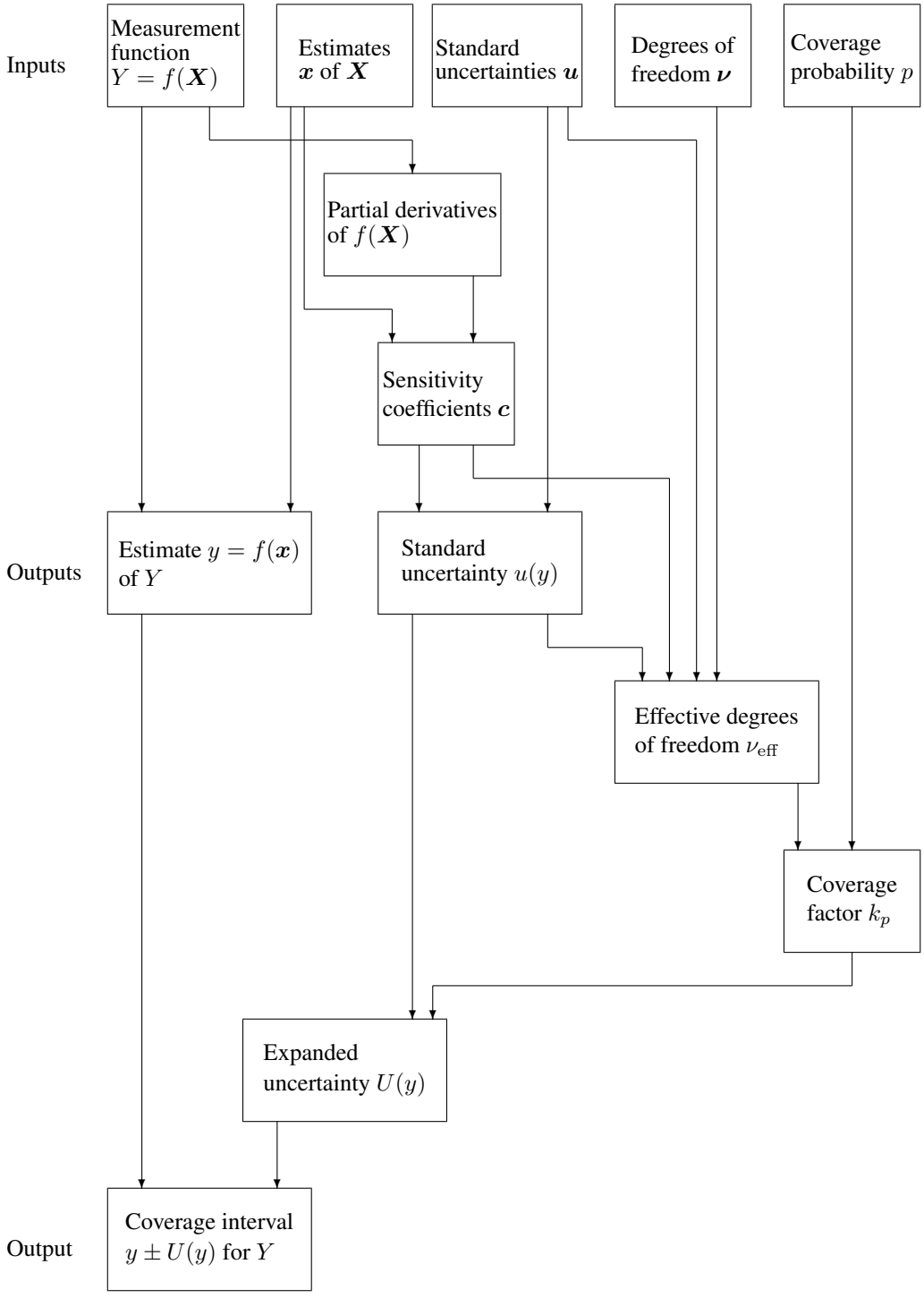

Figure 5.5: Uncertainty evaluation using the GUM uncertainty framework for a univariate, real measurement function with mutually independent input quantities.

have to be applied at a sufficiently fine set of η-values to provide $g_Y(\eta)$ adequately.

Convolution principles, implemented numerically using the Fast Fourier Transform and its inverse, provide an approach for the class of linear measurement functions with mutually independent input quantities. For example, for the measurement function $Y = X_1 + X_2$ and mutually independent input quantities X_1 and X_2, the integral expression (3.1) takes the form of the convolution integral [14]

$$g_Y(\eta) = \int_{-\infty}^{\infty} g_{X_1}(\xi_1) g_{X_2}(\eta - \xi_1)\, \mathrm{d}\xi_1,$$

where, for $i = 1, 2$, $g_{X_i}(\xi_i)$ is the PDF for X_i. A numerical method to obtain $g_Y(\eta)$ is based on replacing the convolution integral by a convolution sum evaluated using the Fast Fourier Transform. A discussion of the approach, illustrated with examples, is available [64].

A Monte Carlo method (Section 5.4.1 and Chapter 7) provides a generally applicable numerical implementation of the propagation of distributions, and is a focus of this guide.

5.4.1 A Monte Carlo method

Rather than attempting to evaluate the integral expression (3.1), an application of a Monte Carlo method [5, 28, 29, 32, 90, 102] encompasses an entirely different approach, based on the following considerations. The expected value of the output quantity Y is conventionally obtained by evaluating the measurement function for the estimated (expected) values $x_1, \ldots, x_N$ of the input quantities to give the value y. However, since each input quantity is described by a PDF rather than a single number, a value can alternatively be obtained by drawing a value at random from this PDF.

A Monte Carlo method operates in the following manner,[5] based on this consideration. Draw a value at random from the PDF for each input quantity and form the corresponding value of the output quantity by evaluating the measurement function for these values of the input quantities. Repeat this process many times, to obtain in all M, say, values of the output quantity. According to the central limit theorem [84, p169], the average y of the values of the output quantity obtained in this manner converges as $1/M^{1/2}$ to the expectation of Y, if the standard deviation of Y exists. Irrespective of the dimensionality of the problem, i.e., the number N of input quantities, it is (only) necessary to quadruple M in order to expect to improve the numerical accuracy of y by one binary digit. In contrast, standard numerical quadrature would require a factor of $2^{M/2}$ for this purpose. Thus, the Monte Carlo calculation has reasonable convergence properties. It is straightforward to implement for simple or even moderately complicated problems. Its *general* implementation requires more effort: see Chapter 7. A broad introduction to Monte Carlo methods is available [55], as is a discussion on uncertainty propagation in Monte Carlo calculations [93].

[5] This description applies to a measurement function with a single output quantity when the input quantities are mutually independent. Chapter 7 gives a general treatment.

Details, procedures and examples are given in Chapter 7.

5.5 Discussion of approaches

The approach used for any particular problem needs to be chosen with care. As indicated, an approach based on the GUM uncertainty framework is the 'method of choice' for many organizations. Analytical methods are in a sense ideal when applicable. Numerical methods offer flexibility. The described Monte Carlo method is increasingly used by laboratories and industrial organizations.

5.5.1 Conditions for the GUM uncertainty framework

The GUM uncertainty framework requires

1. the non-linearity of the measurement function to be insignificant (Note to GUM Clause 5.1.2),[6]
2. the central limit theorem (GUM Clauses G.2.1 and G.6.6) to apply, implying the representativeness of the PDF for the output quantity by a Gaussian distribution or a t–distribution, and
3. the adequacy of the Welch-Satterthwaite formula for calculating the effective degrees of freedom (GUM Clause G.4.2, [54]).[7]

5.5.2 When the conditions do or may not hold

In practice, the GUM uncertainty framework is sometimes used in violation of the conditions listed in Section 5.5.1, and the results thus produced can only be regarded as approximate (with an unquantified degree of approximation). Or, more frequently, it is used without knowing whether these conditions hold (again with an unquantified degree of approximation). As indicated in Chapter 7, a *basic* form of a Monte Carlo method is readily implemented, requiring only evaluation of the measurement function and random-number generation.[8] Because control can be exercised over the number of digits delivered (see

[6] If the linearization of the measurement function is not sufficiently accurate, the quality of the evaluated uncertainty is affected, as is the estimate of the output quantity. The latter point may be less well appreciated in some quarters.

[7] The Welch-Satterthwaite formula is an approximation and applies when the input quantities are mutually independent.

[8] Implementations made by the authors have been applied to measurement functions and general measurement models (where Y can and cannot be expressed directly in terms of $\boldsymbol{X}$), and complex models (for electrical metrology), with univariate and multivariate output quantities.

Section 7.2.5), the described Monte Carlo method can also be used to validate (a) the results provided by the GUM uncertainty framework, and (b) software implementations of the GUM uncertainty framework. Although many uncertainty evaluations based on the GUM uncertainty framework may be sound, it is important to demonstrate that this is so. If (a legitimate implementation of) the described Monte Carlo method indicated that certain results obtained using the GUM uncertainty framework were invalid, it is recommended that consideration be given to using the Monte Carlo method instead.

5.5.3 Probability density functions or not?

The application of the GUM uncertainty framework might not appear to require the specification of the PDFs for the input quantities *per se*. It operates in terms of the expectations and standard deviations of the input quantities characterized by these PDFs (Section 5.3). The GUM uncertainty framework therefore has the apparent advantage that it is not necessary to provide the PDFs for the input quantities, i.e., just expectations and standard deviations would 'suffice'.

The Type A evaluations of the uncertainties associated with estimates of the input quantities are obtained by analysing 'repeated indication values', from which expectations and standard deviations (but not PDFs) are obtained. Conversely, for Type B evaluations, the expectations and standard deviations are determined from known PDFS for the input quantities (Section 4.2.7). These PDFs are then used no further.

Thus, for some of the input quantities, the PDFs are not required and for the others they are not used. This attitude is seen as being incompatible with the Bayesian view that is increasingly used as a consistent basis for uncertainty evaluation. With a Bayesian approach, each input quantity would be characterized by a PDF, based on whatever information, however meagre, is available.

As indicated in Section 5.3, the GUM in fact states (in Clause 4.1.6) that the estimate of each input quantity and the associated standard uncertainty are obtained from a distribution of possible values of the input quantity. Thus, a distribution is at least *implied*, although many practitioners would not obtain or even postulate it, simply computing, for a Type A evaluation of uncertainty, an estimate and a standard deviation from repeated indication values.

This guide encourages the characterization of each input quantity by a PDF (see Section 4.2.7). By so doing any of the candidate solution approaches considered in this section can be applied. They can also be contrasted, if required.

5.6 Obtaining sensitivity coefficients

Sensitivity coefficients are the partial derivatives of first order of the measurement function with respect to the input quantities, evaluated at the estimates of the input quantities. Their determination can present an algebraically difficult task. There are two stages:

1. Form algebraically the N first-order partial derivatives;[9]
2. Evaluate these derivatives at the estimates of the input quantities.

These stages constitute Steps 3 and 5, respectively, of the procedure (as outlined in Section 5.3) based on the GUM uncertainty framework.

If the effort of determining these derivatives manually is considerable, there are two alternative approaches [13]:

- Computer-assisted algebraic methods;
- Finite-difference methods.

Some advice on the use of finite-difference methods is given in Appendices B.1 and B.2. The (complex-step) method described in Appendix B.2 can deliver considerably greater accuracy than conventional finite differences.

In the context of the propagation and summarizing stages of the uncertainty evaluation problem there is no *essential* concept of sensitivity coefficients. They are of course required in an implementation of the GUM uncertainty framework (Section 5.3). Independently of the approach used, they also convey valuable quantitative information about the influences of the various input quantities on the output quantity (at least in cases where the input quantities are mutually independent and linearization of the measurement function is justified). If an approach is used that does not require these coefficients for its operation, approximations to them can additionally be calculated if needed. Within the context of the described Monte Carlo method, it is also possible to apply, at least approximately, the concept of sensitivity coefficients. Some comments are given in Appendix B.3.

[9] Expert advice may be required if the measurement function is not continuously differentiable with respect to some or all of the input quantities.

Chapter 6

The law of propagation of uncertainty

6.1 Overview

In the GUM a measurement is modelled by a functional relationship between the input quantities $\boldsymbol{X} = (X_1, \ldots, X_N)^\top$ and the output quantity Y in the form of the measurement function

$$Y = f(\boldsymbol{X}). \tag{6.1}$$

In practice, however, this functional relationship does not apply *directly* to all measurements encountered, but may instead (a) take the form of a general measurement model, $h(Y, \boldsymbol{X}) = 0$, (b) involve a number of output quantities $\boldsymbol{Y} = (Y_1, \ldots, Y_m)^\top$, or (c) involve complex quantities. Although measurement models other than the form (6.1) are not directly considered in the GUM, the same underlying principles may be used to propagate uncertainties associated with estimates of the input quantities to those associated with the output quantities.

In Section 6.2 a classification of measurement models is described. The classification is motivated by actual measurements, examples of which are given. For each measurement model it is indicated how the uncertainty associated with the estimate of the output quantity is evaluated. Mathematical expressions for the uncertainty are stated using matrix-vector notation, rather than the subscripted summations given in the GUM, because generally such expressions are more compact and more naturally implemented within modern software packages and computer languages. The law of propagation of uncertainty based on a first order Taylor series expansion of the measurement model is used throughout this section. Any doubt concerning its applicability should be addressed as appropriate, for instance by using the concepts of Chapter 8.

The GUM states that whenever the non-linearity of the measurement model is significant,

higher order terms in the Taylor series expansion of the model *must* be included. The manner in which this can be achieved is indicated in Section 6.3. A detailed derivation for the case of a measurement function with a single input quantity is given. The general result for more than one input quantity is conceptually straightforward, but algebraically complicated to derive. The law of propagation of uncertainty with higher order terms is applied to the measurement problems described in Sections 9.8 and 9.9, and the results compared with those obtained from (a) the law of propagation of uncertainty based on a first order Taylor series expansion, and (b) a Monte Carlo method.

A Supplement to the GUM, related to the material in this chapter, is under development by JCGM/WG1.

6.2 Classification of measurement models

A classification of measurement models is presented that depends on whether

- there is one or more than one output quantity, denoted by the scalar Y or vector $\boldsymbol{Y} = (Y_1, \ldots, Y_m)^\top$, respectively,
- the relationship between the input and output quantities is in the form of a general measurement model, viz. $h(Y, \boldsymbol{X}) = 0$ or $\boldsymbol{h}(\boldsymbol{Y}, \boldsymbol{X}) = \boldsymbol{0}$, or can be expressed as a measurement function, viz. $Y = f(\boldsymbol{X})$ or $\boldsymbol{Y} = \boldsymbol{f}(\boldsymbol{X})$, and
- the input and output quantities are real or complex.

The following information is assumed to be available:

1. Estimates $\boldsymbol{x} = (x_1, \ldots, x_N)^\top$ of the input quantities $\boldsymbol{X} = (X_1, \ldots, X_N)^\top$;
2. For $i = 1, \ldots, N$, either
   (a) the standard uncertainty $u(x_i)$ associated with x_i, for mutually independent input quantities, or
   (b) for $j = 1, \ldots, N$, the covariance $\mathrm{cov}(x_i, x_j)$ associated with x_i and x_j, for mutually dependent input quantities.[1] Note that $\mathrm{cov}(x_i, x_i) = u^2(x_i)$, the variance associated with x_i.

The following eight sub-sections provide matrix expressions for the uncertainty associated with the estimate of the output quantity, in the form of the variance $u^2(y)$ associated with y in the scalar (univariate) case or the covariance matrix $\boldsymbol{U}_{\boldsymbol{y}}$ containing the covariances

[1] Some or many of these covariance values may be zero.

$\mathrm{cov}(y_i, y_j)$ associated with y_i and y_j in the vector (multivariate) case. Derivation of the formulae and equations is not given here. It is straightforward using basic statistical concepts and matrix manipulation.

The concentration is on providing information on the various types of measurement model that arise in practice, and for each of these types giving relevant advice. The guidance is especially relevant when *software* is to be used to help provide uncertainty evaluations [34]. For the first two model types (univariate, real and multivariate, real measurement functions), the detail of the manner in which the matrices used are formed is provided through an example. The remaining measurement model types are treated analogously.

6.2.1 Univariate, real measurement function

In a univariate, real measurement function, a single real output quantity Y is related to a number of real input quantities $\boldsymbol{X} = (X_1, \ldots, X_N)^\top$ by a functional relationship f in the form of expression (6.1). This is the form of measurement model directly considered in the GUM.

The estimate of the output quantity is $y = f(\boldsymbol{x})$.

The standard uncertainty $u(y)$ associated with y is evaluated from

$$u^2(y) = \sum_{i=1}^{N} \sum_{j=1}^{N} c_i \mathrm{cov}(x_i, x_j) c_j, \tag{6.2}$$

where c_i is the partial derivative $\partial f / \partial X_i$ evaluated at $\boldsymbol{X} = \boldsymbol{x}$. The $c_i,\ i = 1, \ldots, N$, are referred to as *sensitivity coefficients*.

Define the covariance matrix

$$\boldsymbol{U}_{\boldsymbol{x}} = \begin{bmatrix} \mathrm{cov}(x_1, x_1) & \ldots & \mathrm{cov}(x_1, x_N) \\ \vdots & \ddots & \vdots \\ \mathrm{cov}(x_N, x_1) & \ldots & \mathrm{cov}(x_N, x_N) \end{bmatrix} \tag{6.3}$$

of dimension $N \times N$, containing the covariances $\mathrm{cov}(x_i, x_j)$, and the (row) vector

$$\boldsymbol{c}^\top = [c_1, \ldots, c_N]$$

of dimension $1 \times N$, containing the sensitivity coefficients. Then, a compact way of writing the sum (6.2), which avoids the use of doubly-scripted summations, is

$$u^2(y) = \boldsymbol{c}^\top \boldsymbol{U}_{\boldsymbol{x}} \boldsymbol{c}. \tag{6.4}$$

Example 14 *End-gauge calibration*

(GUM Example H.1) The length of a nominally $50\,\mathrm{mm}$ gauge block is determined by comparing it with a known gauge block standard of the same nominal length. An expression for the direct output of the comparison of the two gauge blocks is the difference

$$d = \ell\,\{1 + (\alpha_\mathrm{s} + \delta\alpha)\theta\} - \ell_\mathrm{s}\,\{1 + \alpha_\mathrm{s}(\theta - \delta\theta)\} \tag{6.5}$$

in their lengths, where[2]

- ℓ is the output quantity, viz., the length at $20\,^\circ\mathrm{C}$ of the gauge block being calibrated,
- ℓ_s is the length of the standard at $20\,^\circ\mathrm{C}$ as given on its calibration certificate,
- α_s is the coefficient of thermal expansion of the gauge block standard,
- $\delta\alpha = \alpha - \alpha_\mathrm{s}$, where α is the coefficient of thermal expansion of the gauge block being calibrated,
- θ is the deviation in temperature from the $20\,^\circ\mathrm{C}$ reference temperature of the gauge block being calibrated, and
- $\delta\theta = \theta - \theta_\mathrm{s}$, where θ_s is the deviation in temperature from the $20\,^\circ\mathrm{C}$ reference temperature of the gauge block standard.

From expression (6.5) the output quantity ℓ can be expressed in terms of the quantities d, ℓ_s, α_s, $\delta\alpha$, θ and $\delta\theta$ as the measurement function

$$\ell = \frac{\ell_\mathrm{s}\,[1 + \alpha_\mathrm{s}(\theta - \delta\theta)] + d}{1 + (\alpha_\mathrm{s} + \delta\alpha)\theta}.$$

In terms of the general formulation above, the input quantities are

$$\boldsymbol{X} \equiv (d, \ell_\mathrm{s}, \alpha_\mathrm{s}, \delta\alpha, \theta, \delta\theta)^\top$$

and the output quantity is

$$Y \equiv \ell.$$

The estimates of the input quantities are denoted by

$$\boldsymbol{x} \equiv (\widehat{d}, \widehat{\ell}_\mathrm{s}, \widehat{\alpha}_\mathrm{s}, \widehat{\delta\alpha}, \widehat{\theta}, \widehat{\delta\theta})^\top, \tag{6.6}$$

and the estimate

$$y \equiv \widehat{\ell}$$

of the output quantity ℓ is

$$\widehat{\ell} = \frac{\widehat{\ell}_\mathrm{s}\big[1 + \widehat{\alpha}_\mathrm{s}(\widehat{\theta} - \widehat{\delta\theta})\big] + \widehat{d}}{1 + (\widehat{\alpha}_\mathrm{s} + \widehat{\delta\alpha})\widehat{\theta}}.$$

[2]This choice of input quantities is made for consistency with GUM, Example H.1. Other choices are possible.

The partial derivatives of first order of the measurement function with respect to the input quantities are

$$\begin{aligned}
\frac{\partial \ell}{\partial d} &= \frac{1}{1+(\alpha_{\mathrm{s}}+\delta\alpha)\theta}, \\
\frac{\partial \ell}{\partial \ell_{\mathrm{s}}} &= \frac{1+\alpha_{\mathrm{s}}(\theta-\delta\theta)}{1+(\alpha_{\mathrm{s}}+\delta\alpha)\theta}, \\
\frac{\partial \ell}{\partial \alpha_{\mathrm{s}}} &= \frac{\ell_{\mathrm{s}}(\theta^2\delta\alpha-\theta\delta\theta\delta\alpha-\delta\theta)-d\theta}{[1+(\alpha_{\mathrm{s}}+\delta\alpha)\theta]^2}, \\
\frac{\partial \ell}{\partial(\delta\alpha)} &= -\frac{\{\ell_{\mathrm{s}}\left[1+\alpha_{\mathrm{s}}(\theta-\delta\theta)\right]+d\}\,\theta}{[1+(\alpha_{\mathrm{s}}+\delta\alpha)\theta]^2}, \\
\frac{\partial \ell}{\partial \theta} &= \frac{(\alpha_{\mathrm{s}}+\delta\alpha)(\ell_{\mathrm{s}}\alpha_{\mathrm{s}}\delta\theta-d)-\ell_{\mathrm{s}}\delta\alpha}{[1+(\alpha_{\mathrm{s}}+\delta\alpha)\theta]^2}, \\
\frac{\partial \ell}{\partial(\delta\theta)} &= \frac{-\alpha_{\mathrm{s}}\ell_{\mathrm{s}}}{1+(\alpha_{\mathrm{s}}+\delta\alpha)\theta}.
\end{aligned}$$

The substitution ($\widehat{d}$ for d, $\widehat{\ell_{\mathrm{s}}}$ for ℓ_{s}, etc.) of the numerical values of the estimates (6.6) into these expressions for the partial derivatives yields the values of the sensitivity coefficients.

The set of six sensitivity coefficients, arranged as a row vector, constitutes the row vector $\boldsymbol{c}^{\top}$ in expression (6.4). The variances given by the squares of the standard uncertainties associated with the estimates of the six input quantities constitute the diagonal elements of the covariance matrix $\boldsymbol{U_x}$ in expression (6.4). The remaining elements of $\boldsymbol{U_x}$ are taken as zero, since the input quantities are regarded as mutually independent [GUM Example H.1]. Thus, $u^2(y)$ and hence $u(y)$ can be formed from expression (6.4).

6.2.2 Multivariate, real measurement function

Although not directly considered in the body of the GUM, instances of measurements are included in that guide for which there is more than one output quantity. This form of measurement occurs commonly in metrology, e.g., in calibration, instrument design and experimental data analysis. Formally, the measurement function takes the form

$$\boldsymbol{Y} = \boldsymbol{f}(\boldsymbol{X}), \tag{6.7}$$

where $\boldsymbol{Y} = (Y_1, \ldots, Y_m)^{\top}$ is a vector of m output quantities.

The estimates of the output quantities are given by $\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x})$.

The uncertainty associated with $\boldsymbol{y}$ is expressed using a covariance matrix $\boldsymbol{U_y}$ that contains the covariances $\mathrm{cov}(y_i, y_j)$, and is evaluated by matrix multiplication from

$$\boldsymbol{U_y} = \boldsymbol{C}\boldsymbol{U_x}\boldsymbol{C}^{\top}, \tag{6.8}$$

where $\boldsymbol{C}$ is the matrix of dimension $m \times N$ containing the partial derivatives $\partial f_i/\partial X_j$, for $i = 1, \ldots, m$, $j = 1, \ldots, N$, evaluated at $\boldsymbol{X} = \boldsymbol{x}$.

Example 15 *Resistance and reactance of a circuit element*

(GUM Example H.2) The resistance R and reactance[3] X of a circuit element are determined by measuring the amplitude U of a sinusoidal alternating potential difference applied to it, the amplitude I of the alternating current passed through it, and the phase shift angle ϕ between the two from

$$R = \frac{U}{I}\cos\phi, \qquad X = \frac{U}{I}\sin\phi.$$

In terms of the above notation, $\boldsymbol{X} \equiv (U, I, \phi)^\top$ and $\boldsymbol{Y} \equiv (R, X)^\top$.

The matrix $\boldsymbol{C}$, of dimension 2×3, is

$$\begin{bmatrix} \dfrac{\partial R}{\partial U} & \dfrac{\partial R}{\partial I} & \dfrac{\partial R}{\partial \phi} \\ \dfrac{\partial T}{\partial U} & \dfrac{\partial T}{\partial I} & \dfrac{\partial T}{\partial \phi} \end{bmatrix} = \begin{bmatrix} \dfrac{1}{I}\cos\phi & -\dfrac{U}{I^2}\cos\phi & -\dfrac{U}{I}\sin\phi \\ \dfrac{1}{I}\sin\phi & -\dfrac{U}{I^2}\sin\phi & \dfrac{U}{I}\cos\phi \end{bmatrix}$$

evaluated at the estimates $\boldsymbol{x}$ of the input quantities. Given the covariance matrix $\boldsymbol{U}_{\boldsymbol{x}}$ of dimension 3×3 associated with these estimates (cf. Section 6.2.1), the covariance matrix $\boldsymbol{U}_{\boldsymbol{y}}$ of dimension 2×2 associated with the estimates $\boldsymbol{y}$ of the output quantities is given by matrix multiplication using expression (6.8).

6.2.3 Univariate, real measurement model

In a univariate, real measurement model, a single real output quantity Y is related to real input quantities $\boldsymbol{X}$ in a way that cannot readily or stably be represented in the form of a measurement function. Instead, a model for the measurement takes the general form

$$h(Y, \boldsymbol{X}) = 0. \tag{6.9}$$

The estimate y of Y is the value of η that solves the equation $h(\eta, \boldsymbol{x}) = 0$. This equation is solved numerically for y using a zero-finding algorithm [35, 51], such as the bisection algorithm in a case where suitable lower and upper bounds for y are known.

The standard uncertainty $u(y)$ associated with y is evaluated from

$$u^2(y)c_y^2 = \boldsymbol{c_x}^\top \boldsymbol{U_x} \boldsymbol{c_x}, \tag{6.10}$$

where $\boldsymbol{c_x}^\top$ is the (row) vector of dimension $1 \times N$ of partial derivatives $\partial h/\partial X_i$, and c_y is the partial derivative $\partial h/\partial Y$, with all derivatives evaluated at $\boldsymbol{X} = \boldsymbol{x}$ and $Y = y$.

[3] In the GUM, reactance is denoted by X, which is the notation used here. The reactance X, a (scalar) component of the output quantity $\boldsymbol{Y}$, is not to be confused with $\boldsymbol{X}$, the (vector) input quantity.

Example 16 *Pressure generated by a pressure balance*

The pressure p generated by a pressure balance is defined implicitly by the measurement model[4]

$$p = \frac{M(1 - \rho_{\mathrm{a}}/\rho_{\mathrm{w}})g}{A_0(1 + \lambda p)\left[1 + \alpha(T - 20)\right]}, \tag{6.11}$$

where M is the total applied mass, ρ_{a} and ρ_{w} are the densities of air and the applied masses, g is the local acceleration due to gravity, A_0 is the effective cross-sectional area of the balance at zero pressure, λ is the distortion coefficient of the piston-cylinder assembly, α is the temperature coefficient, and T is temperature [66].

There are eight input quantities, $\boldsymbol{X} \equiv (A_0, \lambda, \alpha, T, M, \rho_{\mathrm{a}}, \rho_{\mathrm{w}}, g)^{\top}$ and a single output quantity $Y \equiv p$ related by the measurement model[5]

$$h(Y, \boldsymbol{X}) = A_0 p(1 + \lambda p)\left[1 + \alpha(T - 20)\right] - M(1 - \rho_{\mathrm{a}}/\rho_{\mathrm{w}})g = 0. \tag{6.12}$$

Given estimates $\boldsymbol{x}$ of $\boldsymbol{X}$, equation (6.12) is solved for y. The first-order partial derivatives of h, in equation (6.12), with respect to $\boldsymbol{X}$, evaluated at $\boldsymbol{X} = \boldsymbol{x}$ and $Y = y$, provide the elements of the row vector $\boldsymbol{c_x}^{\top}$ in expression (6.10). Together with the covariance matrix $\boldsymbol{U_x}$ associated with the estimates $\boldsymbol{x}$ and the partial derivative $\partial h/\partial Y$ evaluated at $\boldsymbol{X} = \boldsymbol{x}$ and $Y = y$, this information permits $u(y) \equiv u(p)$ to be formed using expression (6.10).

6.2.4 Multivariate, real measurement model

A multivariate, real measurement model is identical to equation (6.9), but the output quantity is now in the form of a vector output quantity $\boldsymbol{Y}$:

$$\boldsymbol{h}(\boldsymbol{Y}, \boldsymbol{X}) = \boldsymbol{0}. \tag{6.13}$$

The estimate $\boldsymbol{y}$ of $\boldsymbol{Y}$ is the value of $\boldsymbol{\eta}$ that solves the system of equations $\boldsymbol{h}(\boldsymbol{\eta}, \boldsymbol{x}) = \boldsymbol{0}$. This system is solved numerically for $\boldsymbol{y}$ using an iterative algorithm such as Newton's method [51], starting from an approximate solution $\boldsymbol{y}^{(0)}$.

The covariance matrix $\boldsymbol{U_y}$ associated with $\boldsymbol{y}$ is related to that, $\boldsymbol{U_x}$, associated with $\boldsymbol{x}$ by

$$\boldsymbol{C_y U_y C_y}^{\top} = \boldsymbol{C_x U_x C_x}^{\top}, \tag{6.14}$$

where $\boldsymbol{C_x}$ is the matrix of dimension $m \times N$ containing the values of the derivatives $\partial h_i/\partial X_j$, for $i = 1, \ldots, m$, $j = 1, \ldots, N$, at $\boldsymbol{X} = \boldsymbol{x}$ and $\boldsymbol{Y} = \boldsymbol{y}$, and $\boldsymbol{C_y}$ is the matrix of

[4] More complete measurement models can also be considered [66] that include, for example, a correction to account for surface tension effects.

[5] There is not a unique way to write the measurement model. For instance, in place of equation (6.12) the measurement model given by the difference between the left- and right-hand sides of expression (6.11) could be used. The efficiency and stability of the solution of the measurement model equation depends on the choice made.

dimension $m \times m$ containing the values of $\partial h_i/\partial Y_j$, for $i = 1, \ldots, m$, $j = 1, \ldots, m$, at $\boldsymbol{X} = \boldsymbol{x}$ and $\boldsymbol{Y} = \boldsymbol{y}$. Expression (6.14) defines a system of linear equations that is solved for $\boldsymbol{U_y}$.[6]

Example 17 *Pressures, generated by a pressure balance, having associated correlation*

In the example of Section 6.2.3, let p_i, $i = 1, \ldots, m$, denote the generated pressures for applied masses M_i and temperatures T_i, with A_0, λ, α, ρ_{a}, ρ_{w} and g as before. An estimate of each p_i is obtained by solving an equation of the form (6.12) given estimates of $A_0, \lambda, \alpha, T_i, M_i, \rho_{\mathrm{a}}, \rho_{\mathrm{w}}$ and g. However, the quantities representing the generated pressures are not mutually independent because they all depend on the quantities A_0, λ, α, ρ_{a}, ρ_{w} and g. To obtain the correlation associated with the estimates of the quantities p_i, the measurement is described using a multivariate model in which $\boldsymbol{X} \equiv (A_0, \lambda, \alpha, T_1, M_1, \ldots, T_m, M_m, \rho_{\mathrm{a}}, \rho_{\mathrm{w}}, g)^\top$ is the vector of $2m + 6$ input quantities, $\boldsymbol{Y} \equiv (p_1, \ldots, p_m)^\top$ the vector of m output quantities, and the measurement model takes the form

$$h_i(\boldsymbol{Y}, \boldsymbol{X}) = A_0 p_i(1+\lambda p_i)\,[1 + \alpha(T_i - 20)] - M_i(1-\rho_{\mathrm{a}}/\rho_{\mathrm{w}})g = 0, \quad i = 1, \ldots, m,$$

(cf. equation (6.12)). The matrix $\boldsymbol{U_x}$ of dimension $(2m+6) \times (2m+6)$, containing the covariances associated with the estimates $\boldsymbol{x}$ of the input quantities $\boldsymbol{X}$, together with the matrices $\boldsymbol{C_x}$ and $\boldsymbol{C_y}$ of partial derivatives of first order of $h_i(\boldsymbol{Y}, \boldsymbol{X})$, $i = 1, \ldots, m$, with respect to X_j, $j = 1, \ldots, N$, and Y_j, $j = 1, \ldots, m$, evaluated at $\boldsymbol{X} = \boldsymbol{x}$ and $\boldsymbol{Y} = \boldsymbol{y}$, provides the information needed to solve (6.14) for $\boldsymbol{U_y}$.

[6] Using recognised concepts from numerical linear algebra [52], a numerically stable way to form $\boldsymbol{U_y}$, that accounts for the fact that $\boldsymbol{C_x}$ is a rectangular matrix and $\boldsymbol{C_y}$ a square matrix, is as follows:

1. Form the Cholesky factor $\boldsymbol{R_x}$ of $\boldsymbol{U_x}$, i.e., the upper triangular matrix such that $\boldsymbol{R_x}^\top \boldsymbol{R_x} = \boldsymbol{U_x}$.
2. Factorize $\boldsymbol{C_x}$ as the product $\boldsymbol{C_x} = \boldsymbol{Q_x W_x}$, where $\boldsymbol{Q_x}$ is an orthogonal matrix and $\boldsymbol{W_x}$ is upper triangular.
3. Factorize $\boldsymbol{C_y}$ as the product $\boldsymbol{C_y} = \boldsymbol{L_y W_y}$, where $\boldsymbol{L_y}$ is lower triangular and $\boldsymbol{W_y}$ is upper triangular.
4. Solve the matrix equation $\boldsymbol{W_y}^\top \boldsymbol{M_1} = \boldsymbol{I}$ for $\boldsymbol{M_1}$.
5. Solve $\boldsymbol{L_y}^\top \boldsymbol{M_2} = \boldsymbol{M_1}$ for $\boldsymbol{M_2}$,
6. Form $\boldsymbol{M_3} = \boldsymbol{Q_x}^\top \boldsymbol{M_2}$.
7. Form $\boldsymbol{M_4} = \boldsymbol{W_x}^\top \boldsymbol{M_3}$.
8. Form $\boldsymbol{M} = \boldsymbol{R_x M_4}$.
9. Orthogonally triangularize $\boldsymbol{M}$ to give the upper triangular matrix $\boldsymbol{R}$.
10. Form $\boldsymbol{U_y} = \boldsymbol{R}^\top \boldsymbol{R}$.

It is straightforward to verify this procedure using elementary matrix algebra.

6.2.5 Univariate, complex measurement function

Let X be a complex quantity written in terms of its real and imaginary parts:

$$X = X_{\mathrm{R}} + \mathrm{i}X_{\mathrm{I}}, \quad \mathrm{i}^2 = -1.$$

The uncertainty associated with an estimate $x = x_{\mathrm{R}} + \mathrm{i}x_{\mathrm{I}}$ of X is expressed using a covariance matrix of dimension 2×2:

$$\boldsymbol{U}_x = \begin{bmatrix} \mathrm{cov}(x_{\mathrm{R}}, x_{\mathrm{R}}) & \mathrm{cov}(x_{\mathrm{R}}, x_{\mathrm{I}}) \\ \mathrm{cov}(x_{\mathrm{I}}, x_{\mathrm{R}}) & \mathrm{cov}(x_{\mathrm{I}}, x_{\mathrm{I}}) \end{bmatrix}.$$

This is a more complicated data structure than for the case of real X.[7] For a vector $\boldsymbol{X} = (X_1, \ldots, X_N)^\top$ of complex quantities, the uncertainty associated with an estimate $\boldsymbol{x}$ of $\boldsymbol{X}$ is expressed using a covariance matrix $\boldsymbol{U}_{\boldsymbol{x}}$ of dimension $2N \times 2N$:

$$\boldsymbol{U}_{\boldsymbol{x}} = \begin{bmatrix} \boldsymbol{U}_{1,1} & \cdots & \boldsymbol{U}_{1,n} \\ \vdots & \ddots & \vdots \\ \boldsymbol{U}_{n,1} & \cdots & \boldsymbol{U}_{n,n} \end{bmatrix}, \tag{6.15}$$

where $\boldsymbol{U}_{i,i}$ is a sub-matrix of dimension 2×2 containing the uncertainty associated with x_i, and $\boldsymbol{U}_{i,j}$, $i \neq j$, is a sub-matrix of dimension 2×2 containing the covariances associated with the real and imaginary parts of x_i and those of x_j.

In a univariate, complex measurement function, a single complex output quantity Y is related to a number of complex input quantities $\boldsymbol{X}$ by a functional relationship f in the form of (6.1). The covariance matrix $\boldsymbol{U}_y$ associated with the estimate y of Y is evaluated from

$$\boldsymbol{U}_y = \boldsymbol{C}\boldsymbol{U}_{\boldsymbol{x}}\boldsymbol{C}^\top, \tag{6.16}$$

where $\boldsymbol{C}$ is a matrix of dimension $2 \times 2N$ whose first row contains the partial derivatives of first order of the real part of f with respect to the real and imaginary parts of $\boldsymbol{X}$, and in whose second row are the partial derivatives of first order of the imaginary part of f, evaluated at $\boldsymbol{X} = \boldsymbol{x}$.

Example 18 *Reflection coefficient measured by a calibrated microwave reflectometer*

The (complex) reflection coefficient Γ measured by a calibrated microwave reflectometer, such as an automatic network analyser (ANA), is given by

$$\Gamma = \frac{aW + b}{cW + 1}, \tag{6.17}$$

where W is the indicated (complex) uncorrected reflection coefficient and a, b and c are (complex) calibration coefficients characterizing the reflectometer [44, 63, 96].

[7] The data structure is, however, like that for the vector $\boldsymbol{X} = (X_{\mathrm{R}}, X_{\mathrm{I}})^\top$.

There are four (complex) input quantities $\boldsymbol{X} \equiv (a, b, c, W)^\top$ and one (complex) output quantity $Y \equiv \Gamma$. $\boldsymbol{C}$ is a matrix of dimension 2×8 containing the partial derivatives of first order of the real and imaginary parts of Γ with respect to the real and imaginary parts of $a,\ b,\ c$ and W, evaluated at the estimates of $a,\ b,\ c$ and W. $\boldsymbol{U_y}$, a matrix of dimension 2×2 containing the covariances associated with the real and imaginary parts of an estimate y of Y, is formed using expression (6.16), in which $\boldsymbol{U_x}$ contains the covariances associated with the real and imaginary parts of estimates $\boldsymbol{x}$ of $\boldsymbol{X}$.

6.2.6 Multivariate, complex measurement function

In a multivariate, complex measurement function the relationship (6.7) applies with $\boldsymbol{X}$ and $\boldsymbol{Y}$ complex. The uncertainty associated with $\boldsymbol{y}$ is given by the covariance matrix $\boldsymbol{U_y}$ of dimension $2m \times 2m$ (see expression (6.15)) evaluated from

$$\boldsymbol{U_y} = \boldsymbol{C}\boldsymbol{U_x}\boldsymbol{C}^\top, \tag{6.18}$$

where $\boldsymbol{C}$ is a matrix of dimension $2m \times 2N$ containing the partial derivatives of first order of the real and imaginary parts of each component of $\boldsymbol{f}$ with respect to the real and imaginary parts of each component of $\boldsymbol{X}$, evaluated at $\boldsymbol{X} = \boldsymbol{x}$.

Example 19 *Calibrated microwave reflectometer used to measure mutually dependent reflection coefficients*

Let a, b and c be the calibration coefficients for an automatic network analyser (ANA) as in Example 18. Suppose W_i, $i = 1, \ldots, m$, are quantities describing m indicated uncorrected reflection coefficients and the quantities describing the corresponding 'corrected' reflection coefficients are Γ_i, $i = 1, \ldots, m$. Estimates of Γ_i are obtained by evaluating m formulae of the form of expression (6.17). However, these quantities are not mutually independent because they all depend on the calibration coefficients. To obtain the correlation associated with estimates of the quantities, the measurement is described using a multivariate measurement function in which the (complex) vector of input quantities is $\boldsymbol{X} \equiv (a, b, c, W_1, \ldots, W_m)^\top$ and the (complex) vector of output quantities is $\boldsymbol{Y} \equiv (\Gamma_1, \ldots, \Gamma_m)^\top$.

The covariance matrix $\boldsymbol{U_x}$ of dimension $(2m + 6) \times (2m + 6)$, containing the covariances associated with the real and imaginary parts of the estimates $\boldsymbol{x}$ of the input quantities $\boldsymbol{X}$, together with the matrix $\boldsymbol{C}$ of dimension $2m \times (2m + 6)$, provides the information needed to determine $\boldsymbol{U_y}$ from expression (6.18).

6.2.7 Univariate, complex measurement model

In a univariate, complex measurement model, the relationship (6.9) applies with Y and $\boldsymbol{X}$ complex. The uncertainty associated with y is evaluated from

$$\boldsymbol{C}_y \boldsymbol{U}_y \boldsymbol{C}_y^{\top} = \boldsymbol{C}_{\boldsymbol{x}} \boldsymbol{U}_{\boldsymbol{x}} \boldsymbol{C}_{\boldsymbol{x}}^{\top}, \tag{6.19}$$

where $\boldsymbol{C}_y$ is a matrix of dimension 2×2 containing the partial derivatives of first order of the real and imaginary parts of h with respect to the real and imaginary parts of Y, evaluated at $\boldsymbol{X} = \boldsymbol{x}$ and $Y = y$. Compare with Section 6.2.4.

Example 20 *Reflection coefficient measured by a calibrated microwave reflectometer*

Another approach to Example 18 is to relate the input quantities $\boldsymbol{X} \equiv (a, b, c, W)^{\top}$ and the output quantity $Y \equiv \Gamma$ using the (complex) measurement model

$$h(Y, \boldsymbol{X}) = \Gamma(cW + 1) - (aW + b) = 0.$$

An advantage of this approach is that the calculation of partial derivatives of first order and thence sensitivity coefficients is easier. The matrix $\boldsymbol{C}_y$ of dimension 2×2 contains the partial derivatives of first order of the real and imaginary parts of h with respect to the real and imaginary parts of Γ, and the matrix $\boldsymbol{C}_{\boldsymbol{x}}$ of dimension 2×8 the partial derivatives of first order with respect to the real and imaginary parts of $a,\ b,\ c$ and W, with all derivatives evaluated at the estimates of Γ and $a,\ b,\ c$ and W. This information, together with the covariance matrix $\boldsymbol{U}_{\boldsymbol{x}}$ of dimension 8×8, containing the covariances associated with the real and imaginary parts of the estimates $\boldsymbol{x}$ of the input quantities $\boldsymbol{X}$, is used to determine $\boldsymbol{U}_y$ from expression (6.19).

6.2.8 Multivariate, complex measurement model

In a multivariate, complex measurement model, the relationship (6.13) applies with $\boldsymbol{X}$ and $\boldsymbol{Y}$ complex. The covariance matrix associated with $\boldsymbol{y}$ is then evaluated from equation (6.14), which constitutes a linear system for $\boldsymbol{U}_{\boldsymbol{y}}$.

Example 21 *Calibration of a microwave reflectometer using three standards*

The calibration of a reflectometer is undertaken by measuring values W corresponding to a number of standards Γ. Typically, three standards are used, giving the three simultaneous equations

$$\Gamma_i(cW_i + 1) - (aW_i + b) = 0, \qquad i = 1, 2, 3,$$

that are solved for estimates of the three calibration coefficients a, b and c. There are six (complex) input quantities $\boldsymbol{X} \equiv (W_1, \Gamma_1, W_2, \Gamma_2, W_3, \Gamma_3)^\top$ and three (complex) output quantities $\boldsymbol{Y} \equiv (a, b, c)^\top$ related by a measurement model of the form (6.13), where

$$h_i(\boldsymbol{Y}, \boldsymbol{X}) = \Gamma_i(cW_i + 1) - (aW_i + b) = 0, \qquad i = 1, 2, 3.$$

The matrix $\boldsymbol{C_y}$ of dimension 6×6 contains the partial derivatives of first order of the real and imaginary parts of $h_i,\ i = 1, 2, 3$ with respect to the real and imaginary parts of $a,\ b$ and c, and the matrix $\boldsymbol{C_x}$ of dimension 6×12 the partial derivatives of first order with respect to the real and imaginary parts of Γ_i and $W_i,\ i = 1, 2, 3$, with all derivatives evaluated at the estimates of W_i and $\Gamma_i,\ i = 1, 2, 3$. This information, together with the covariance matrix $\boldsymbol{U_x}$ of dimension 12×12, containing the covariances associated with the real and imaginary parts of the estimates $\boldsymbol{x}$ of the input quantities $\boldsymbol{X}$, is used to determine $\boldsymbol{U_y}$ from equation (6.14).

6.3 The law of propagation of uncertainty with higher order terms

The GUM states that whenever the non-linearity of the measurement model is significant, higher order terms in the Taylor series expansion of the model *must* be included. The manner in which this can be achieved is indicated. A detailed derivation for the case of a measurement function with a single input quantity is given. The general result for more than one input quantity is conceptually straightforward, but algebraically complicated, and is not given.

Consider, therefore, a measurement modelled by a functional relationship between a single (scalar) input quantity X and a single (scalar) output quantity Y of the form

$$Y = f(X).$$

Let x denote an estimate of X (the expectation of X) and $u = u(x)$ the standard uncertainty associated with x (the standard deviation of X). Define random variables δX and δY by

$$X = x + \delta X, \qquad Y = y + \delta Y = f(x + \delta X),$$

where

$$y = f(x).$$

Now, δX has expectation and variance

$$E(\delta X) = 0, \qquad V(\delta X) = u^2.$$

Since

$$V(\delta X) = E\left((\delta X)^2\right) - \left(E(\delta X)\right)^2,$$

it follows also that

$$E\left((\delta X)^2\right) = u^2.$$

Consider, to begin with, a first order Taylor series approximation to the measurement function f about the estimate x, i.e,

$$y + \delta Y = f(x) + f'(x)\delta X.$$

Then,

$$\delta Y = f'(x)\delta X, \qquad (\delta Y)^2 = \left[f'(x)\right]^2 (\delta X)^2,$$

and, taking expectations,

$$E\left(\delta Y\right) = f'(x)E\left(\delta X\right) = 0,$$

and

$$E\left((\delta Y)^2\right) = \left[f'(x)\right]^2 E\left((\delta X)^2\right) = \left[f'(x)\right]^2 u^2.$$

It follows that

$$E\left(Y\right) = y + E\left(\delta Y\right) = y \tag{6.20}$$

and

$$V\left(Y\right) = E\left((\delta Y)^2\right) - \left[E\left(\delta Y\right)\right]^2 = \left[f'(x)\right]^2 u^2. \tag{6.21}$$

Expression (6.20) says that the expectation of Y, obtained on the basis of a linear approximation to the measurement function, is y. Expression (6.21) is a special case of formula (6.2) applied to a univariate, real measurement function with $N = 1$ (Section 6.2.1).

Now consider a higher (third) order Taylor series approximation to the measurement function f about the estimate x, i.e.,

$$y + \delta Y = f(x) + f'(x)\delta X + \frac{1}{2}f''(x)(\delta X)^2 + \frac{1}{6}f'''(x)(\delta X)^3.$$

Then, to a second order approximation,

$$\delta Y = f'(x)\delta X + \frac{1}{2}f''(x)(\delta X)^2,$$

and, to fourth order approximation,

$$(\delta Y)^2 = \left[f'(x)\right]^2 (\delta X)^2 + f'(x)f''(x)(\delta X)^3 + \left\{\frac{1}{4}\left[f''(x)\right]^2 + \frac{1}{3}f'(x)f'''(x)\right\}(\delta X)^4.$$

Taking expectations,

$$E\left(\delta Y\right) = f'(x)E\left(\delta X\right) + \frac{1}{2}f''(x)E\left((\delta X)^2\right) = \frac{1}{2}f''(x)u^2,$$

and

$$\begin{aligned} E\left((\delta Y)^2\right) = {} & (f'(x))^2 E\left((\delta X)^2\right) + f'(x)f''(x)E\left((\delta X)^3\right) \\ & + \left\{\frac{1}{4}\left[f''(x)\right]^2 + \frac{1}{3}f'(x)f'''(x)\right\} E\left((\delta X)^4\right). \end{aligned}$$

Assume that X (and hence δX) is characterized by a Gaussian distribution, so that

$$E\left((\delta X)^3\right) = 0, \qquad E\left((\delta X)^4\right) = 3u^4.$$

It follows that

$$E\left(Y\right) = y + E\left(\delta Y\right) = y + \frac{1}{2}f''(x)u^2, \tag{6.22}$$

and

$$\begin{aligned} V\left(Y\right) &= E\left((\delta Y)^2\right) - \left(E\left(\delta Y\right)\right)^2 \\ &= \left[f'(x)\right]^2 u^2 + 3\left\{\frac{1}{4}\left[f''(x)\right]^2 + \frac{1}{3}f'(x)f'''(x)\right\}u^4 - \frac{1}{4}\left[f''(x)\right]^2 u^4, \end{aligned}$$

i.e.,

$$V\left(Y\right) = \left[f'(x)\right]^2 u^2 + \left\{\frac{1}{2}\left[f''(x)\right]^2 + f'(x)f'''(x)\right\}u^4. \tag{6.23}$$

Expression (6.22) says that the expectation of Y, obtained on the basis of a higher order approximation to the measurement function, is no longer y. Expression (6.23) can be used to evaluate the standard uncertainty associated with the estimate y accounting for higher order terms in the Taylor series approximation to the measurement function. Unlike expression (6.21), which is based on a linearization of the measurement function, the derivation of expression (6.23) requires knowledge of the probability distribution used to characterize X.[8] A generalisation of the result given by expression (6.23) to more than one input quantity is available; see, e.g., [GUM Note to Clause 5.1.2]. The result requires that the input quantities are uncorrelated and are characterized by Gaussian distributions.[9]

[8] A different result is obtained if X is characterized by a rectangular distribution, since for that distribution $E\left((\delta X)^4\right) = \frac{9}{5}u^4$.

[9] In this regard, the conditions stated in the GUM are incomplete, which require only that the distributions for the input quantities are symmetric.

Chapter 7

A Monte Carlo method

7.1 Overview

The manner in which a general *numerical* approach, a Monte Carlo method, can be applied to uncertainty evaluation is described. Practical implementation considerations are given.[1]

In the context of uncertainty evaluation, a Monte Carlo method provides an implementation of the propagation of distributions: the process is undertaken numerically rather than analytically. The method is also useful for validating the results returned by the application of the GUM uncertainty framework, as well as in circumstances where the assumptions made by the GUM uncertainty framework may not apply. In fact, it provides much richer information, by propagating the PDFs for the input quantities $\boldsymbol{X}$ (rather than just the uncertanties associated with estimates of these quantities) through the measurement model to provide the PDF for the output quantity Y or the joint PDF for multivariate output quantities $\boldsymbol{Y}$. From the PDF for the output quantity, coverage intervals (in the univariate case) can then straightforwardly be produced, as can other statistical information.[2]

The Monte Carlo method enables account to be taken of any specified PDFs for the input quantities. Such PDFs include asymmetric distributions such as Poisson (counting rates) and Gamma (special cases of which are exponential and chi-squared). The PDFs for the input quantities form the necessary basis for determining the PDF for an output quantity by a Monte Carlo method. (A calculated expectation and standard deviation, as provided by the GUM uncertainty framework, do not alone form such a basis.)

If the input quantities are mutually dependent, the corresponding joint PDF would be used,

[1] A Supplement [11] to the GUM, related to the approach in this chapter, has been developed by JCGM/WG1.

[2] The determination of coverage *regions* (for the multivariate case) remains a research problem. See Section 7.4. A Supplement to the GUM, concerned with the application of a Monte Carlo method to uncertainty evaluation for multivariate measurement functions and general measurement models, is under development by JCGM/WG1.

for which a general approach is available [102].

The Monte Carlo method is a step-by-step procedure, like the approach based on the GUM uncertainty framework. The difference is that in the Monte Carlo method a small number of steps is repeated very many times, with each repetition constituting a single trial, and the results obtained aggregated. Hence, computer implementation is essential. Increasingly, software is used in applying the GUM uncertainty framework, so the use of software for a Monte Carlo calculation is seen as a practical and acceptable (and more general) alternative. Specifications for key software units are available [34]. See also [33, 45, 77, 107].

The method makes repeated random draws from the PDFs describing the input quantities. The fact that draws are made from the provided PDFs, rather than being based on approximations the quality of which is difficult to quantify, is seen as highly relevant in removing the influence of uncontrollable limitations.

Given the measurement model and the PDFs for its input quantities, the Monte Carlo method constitutes a tool to approximate the PDF for the scalar output quantity Y or vector output quantity $\boldsymbol{Y}$. The PDF for the output quantity is fundamental to determining any or all statistics associated with the output quantity.[3] From it can be obtained

- expectation[4], median, mode and other estimates of location,
- standard deviation (standard uncertainty), variance (squared standard deviation), and higher moments such as skewness and kurtosis[5],

[3] Recall that the output quantity may become the input quantity to a subsequent stage in a multi-stage measurement model (Section 4.2.6), and hence in these circumstances the Monte Carlo method provides valuable problem-specific information that would not necessarily be provided by more traditional approaches to uncertainty evaluation.

[4] There is a debate, at the time of publication, concerning whether the expectation or the value of the model corresponding to the estimates of the input quantities should be used (GUM Clause 4.1.4). In many instances it will make negligible practical difference. In some cases, however, the difference can be appreciable. Consider the simple model $Y = X^2$, where X has expectation zero and standard deviation u and is characterized by a symmetric PDF. Taking the expectation of values of Y involves averaging a set of non-negative values and hence will be positive. In contrast, the value of Y corresponding to the expectation of X is zero. In this case, the former value is more reasonable, since zero lies at an extremity of the interval of possible values for the output quantity and is hardly representative, as an expectation should be. In other, somewhat more complicated situations, the expectation of the output quantity constitutes an estimate that contains unwanted bias. In this circumstance, it can be more meaningful to take instead the value of Y corresponding to the expectation of X. In general, circumstances should dictate the choice. In one sense the degree of arbitrariness is genuine. The Monte Carlo method naturally provides *quantiles* of the distribution function for the output quantity. In particular the 0.025– and 0.975–quantiles define a $95\,\%$ coverage interval for the output quantity. Such a coverage interval is also given by any other pair of quantiles that differ by 0.95, such as 0.015 and 0.965, or 0.040 and 0.990. In this setting, it is less meaningful to quote a coverage interval in the form $y \pm U$, involving the 'expectation' y. For comparison with a 'Monte Carlo' interval, it would instead be appropriate to quote the interval $[y - U, y + U]$. See Section 7.2.4.

[5] The first moment of a PDF is the expectation, a measure of location, the second is the variance, a measure of dispersion or spread about the expectation, the third is skewness, a measure of asymmetry about the expectation, and the fourth is kurtosis, a measure of heaviness of the tails of the PDF or the peakedness in the centre [84, p143, p329].

- a coverage interval corresponding to some stated coverage probability (the generalization of 'estimate $\pm$ expanded uncertainty' in the case of an output quantity characterized by a Gaussian distribution or t–distribution), and

- any other statistical estimator or derived statistical quantity.

For multivariate output quantities, there would be higher-dimensional counterparts of these statistics.

The use of a Monte Carlo method is in principle straightforward, but a solid implementation requires (a) generators (algorithms) to make random draws from all (joint) PDFs for the input quantities that are likely to be useful in practice (some of which will be multidimensional because of mutual dependencies), and (b) consideration of the number of Monte Carlo trials needed to deliver two (say) correct digits in the standard uncertainty associated with the estimate of the output quantity. Work is needed in (a) to cater for the variety of possible PDFs. As indicated, a general approach to making such random draws is available [102]. This and other approaches need to be reviewed carefully for their suitability in the current context. For some of the commonest PDFs (univariate rectangular, univariate Gaussian and multivariate Gaussian), generators to carry out the sampling are available [11, 34]. For (b), see Section 7.2.5.

The Monte Carlo method is also valuable for validating the results returned by the application of the GUM uncertainty framework (Chapter 8), as well as in circumstances where the assumptions made by the GUM uncertainty framework do not apply. Further, the fact that the Monte Carlo method permits general PDFs rather than just estimates and uncertainties to be propagated through measurement models cannot be underestimated. As indicated, all statistical information relating to the variability of measured data, including correlation effects, can be discerned from the distributions for the output quantities. The quality of this information will depend on that of the measurement model and the input quantities and, if those are considered appropriate, is only limited by the number of Monte Carlo trials made. In particular, quantitative results relating to the GUM uncertainty framework can be obtained from the propagated PDFs. In contrast, the converse is not true: the GUM uncertainty framework cannot be used to derive the PDFs for the output quantities (unless it can be shown that it is acceptable to characterize them by Gaussian or t-distributions).

The GUM uncertainty framework is based on propagating uncertainties in a first-order approximation to the measurement model. The Monte Carlo method [39, 40] provides an alternative approach in which instead the probability distributions are propagated. Although no first-order approximation to the measurement model is made, e.g., the non-linearity of the model is taken into account, the quality of the results obtained depends on the number of trials taken (Section 7.2.5). In contrast, with the GUM uncertainty framework there is no control over the extent of the approximation introduced by linearizing the measurement model, or assuming the PDF for the output quantity to be Gaussian.

7.2 A Monte Carlo method for a univariate measurement function

Consider first the univariate measurement function[6]

$$Y = f(\boldsymbol{X}),$$

where

$$\boldsymbol{X} = (X_1, \ldots, X_N)^{\top}.$$

Let the PDF for the ith input quantity X_i be $g_{X_i}(\xi_i)$ and the PDF for Y be $g_Y(\eta)$. Let

$$G_Y(\eta) = \int_{-\infty}^{\eta} g_Y(z)\,\mathrm{d}z$$

denote the distribution function (DF) corresponding to g_Y. An adequate approximation to $G_Y(\eta)$ will permit all the required statistics associated with Y to be determined.

Advice on a simple implementation of a Monte Carlo method is given in the case of the univariate measurement function above. Its use will permit for instance the evaluation of the standard uncertainty associated with an estimate y of the output quantity Y, and a $95\,\%$ coverage interval for Y.

The procedure is as follows:

1. Select the number M of Monte Carlo trials to be made;
2. Generate M vectors $\boldsymbol{x}_r$ by making random draws from the PDFs for the (set of N) input quantities $\boldsymbol{X}$. See Section 7.2.1;
3. For each vector $\boldsymbol{x}_r$, evaluate the measurement function to give the corresponding value $y_r = f(\boldsymbol{x}_r)$ of the output quantity[7];
4. Calculate the estimate y of the output quantity and the associated standard uncertainty $u(y)$ as the average and standard deviation of the output quantity values y_r, $r = 1, \ldots, M$. See Section 7.2.2;
5. Sort the values y_r, $r = 1, \ldots, M$, into non-decreasing order, and use the sorted values to provide a discrete representation $\boldsymbol{G}$ of the distribution function $G_Y(\eta)$ for the output quantity. See Section 7.2.3;

[6] Multivariate measurement functions and general measurement models, including complex models, are considered in Section 7.4 and 7.5, respectively.

[7] The values y_r, $r = 1, \ldots, M$, when assembled into a histogram (with suitable cell widths) provide a (scaled) approximation to the PDF $g_Y(\eta)$ for Y. Most calculations will not be carried out in terms of this histogram, the 'shape' of which depends on the choice of cell size [48], but in terms of the distribution function. The histogram is, however, a useful visual aid to understanding the nature of the PDF.

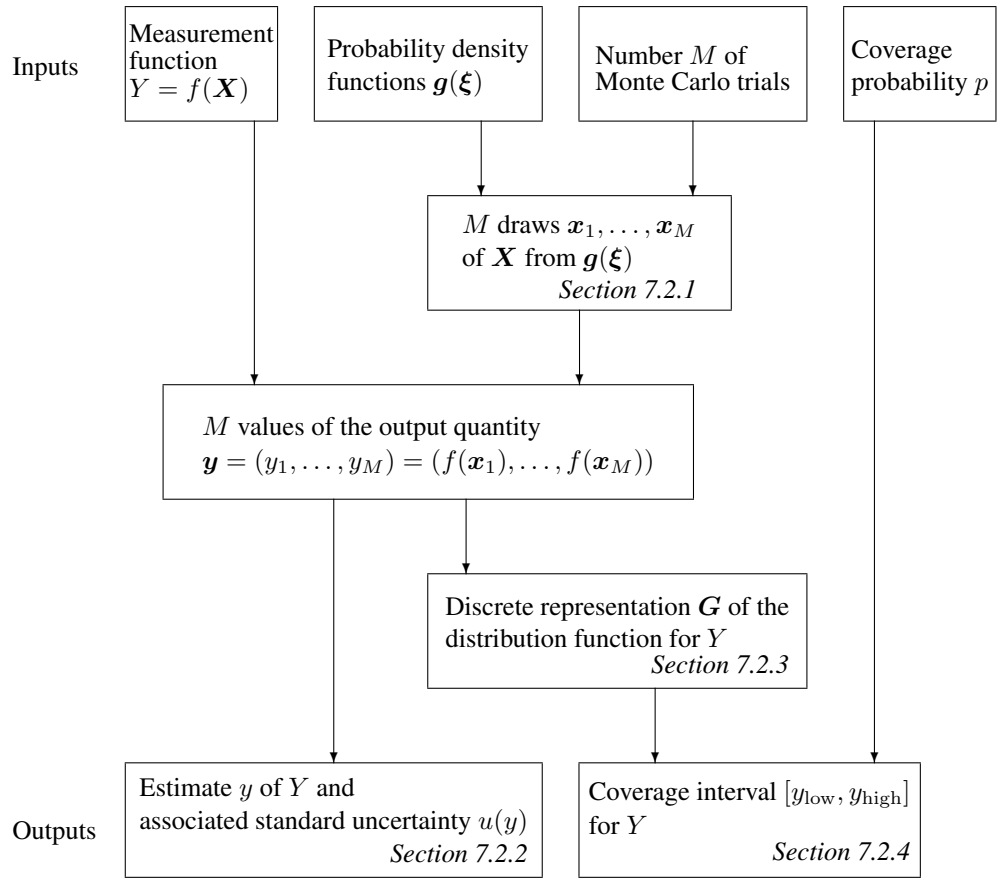

Figure 7.1: Uncertainty evaluation using a Monte Carlo method for a univariate, real measurement function.

6. Use the discrete representation of the distribution function to calculate a $95\,\%$ coverage interval for the output quantity. See Section 7.2.4.

Figure 7.1 shows the procedure diagrammatically.

The number M of Monte Carlo trials is selected initially at Step 1. Variants of the procedure, in which the number is chosen adaptively as the procedure proceeds, are available [77, 107, 108]. One such procedure is described in Section 7.2.5.

7.2.1 Making random draws from the PDFs for the input quantities

This section is concerned with the manner in which vectors $\boldsymbol{x}_r$ are drawn randomly from the PDFs for the input quantities. Consideration is given to PDFs that are univariate or

multivariate (joint).

Univariate probability density functions

Each independent input quantity is characterized by a PDF (GUM Clause G.1.4). Making random draws from the commonest distributions, e.g., rectangular, Gaussian or t, is carried out using a (pseudo-)random number generator that is appropriate for that PDF. Methods, in the form of pseudo-random number generators, for making random draws from these PDFs are available in a companion document [34].

Consider now information about an input quantity in the form of a sufficiently large number of values of the quantity characterized by an unknown PDF.[8] In such a case these values can be used to approximate the PDF for the quantity. Consider such an input quantity X, realized by n values $x_1, \ldots, x_n$. Let $x_{(1)}, \ldots, x_{(n)}$ denote these values arranged in non-decreasing order. Then, the piecewise-linear function $\widetilde{G}_X(\xi)$ joining the points $(x_{(r)}, (r - 1/2)/n)$, $r = 1, \ldots, n$, provides an approximation to the distribution function $G_X(\xi)$ for X. Making random draws from this distribution function can be carried out using a generator for the rectangular distribution and inverse linear interpolation. Specifically, a random draw from the distribution function is given by using

1. a generator to provide a value z from the rectangular PDF defined over zero to one, and
2. inverse linear interpolation to provide a value x satisfying $\widetilde{G}_X(x) = z$.

n should be large if coverage intervals corresponding to a large coverage probability are required, since the tails of the distribution would otherwise not be well characterized.

Multivariate probability density functions

Making random draws from a joint PDF that is defined continuously is largely a research topic. However, a joint (multivariate) Gaussian distribution can straightforwardly be handled [34].

Making random draws from a joint PDF that is defined by a set of values can be carried out straightforwardly. Such values are likely to have been obtained from a previous application of a Monte Carlo method to a multivariate measurement model (Section 7.4). Suppose M Monte Carlo trials have previously be carried out and m is the number of output quantities in that calculation. The information from the calculation will have been preserved as an array of values of dimension $m \times M$. It embodies fully the correlation effects present in

[8] For example, this situation would arise in the case that the values are the result of a previous Monte Carlo calculation.

those output quantities. A column chosen at random from the array will be a valid random draw accounting for the correlation effects.

7.2.2 Estimate of the output quantity and the associated standard uncertainty

The average $\widetilde{y}$ of the values $y_r,\ r = 1, \ldots, M$, of the output quantity is taken as the estimate y of the output quantity, and the standard deviation $u(\widetilde{y})$ of the values is taken as the standard uncertainty $u(y)$ associated with y. $\widetilde{y}$ is evaluated from

$$\widetilde{y} = \frac{1}{M}\sum_{r=1}^{M} y_r, \tag{7.1}$$

and the standard deviation $u(\widetilde{y})$ from

$$u^2(\widetilde{y}) = \frac{1}{M-1}\sum_{r=1}^{M}(y_r - \widetilde{y})^2. \tag{7.2}$$

Advice is available on using an 'updating' procedure to evaluate the estimate of the output quantity and the associated standard uncertainty that avoids the need to store the complete set of values y_r [34].

The value of y so obtained yields the smallest mean squared deviation over all possible choices of the estimate of the output quantity. However, the value will not in general agree with the measurement function evaluated at the estimates of the input quantities [10, Clause 4.1.4]. Agreement (in a practical sense) will be achieved for a large value of M when the measurement function is linear in the input quantities. Whether this general lack of agreement is important depends on the application.

7.2.3 Discrete representation of the distribution function

A discrete representation $\boldsymbol{G}$ of the distribution function for the output quantity is obtained by sorting the values y_r, $r = 1, \ldots, M$, into non-decreasing order. Denoting the sorted values by $y_{(r)}$, $r = 1, \ldots, M$, the discrete representation is given by $\boldsymbol{G} = (y_{(1)}, \ldots, y_{(M)})$. The discrete representation is used as the basis for calculating a coverage interval for the output quantity (Section 7.2.4). It is also used as the basis for obtaining a (continuous) approximation to the distribution function for the output quantity (below) that may used, for example, to make random draws from the distribution for the output quantity (in the manner described in Section 7.2.1).

An approximation $\widetilde{G}_Y(\eta)$ to the distribution function $G_Y(\eta)$ for the output quantity is obtained as follows. Assign uniformly spaced cumulative probabilities $p_r = (r - 1/2)/M$,

$r = 1, \ldots, M$, to the ordered values $y_{(r)}$ in the discrete representation $\boldsymbol{G}$ of the distribution function for the output quantity.[9] Form $\widetilde{G}_Y(\eta)$ as the piecewise-linear function joining the M points $(y_{(r)}, p_r),\ r = 1, \ldots, M$:

$$\widetilde{G}_Y(\eta) = \frac{r - 1/2}{M} + \frac{\eta - y_{(r)}}{M(y_{(r+1)} - y_{(r)})}, \qquad y_{(r)} \leq \eta \leq y_{(r+1)}, \tag{7.3}$$

for $r = 1, \ldots, M - 1$.

Formulae (7.1) and (7.2) for the estimate of the output quantity and the associated standard uncertainty do not in general provide values that are identical to the expectation and standard deviation of the quantity characterized by the distribution function $\widetilde{G}_Y(\eta)$. The latter values are given by

$$\widetilde{y} = \frac{1}{M} {\sum_{r=1}^{M}}'' y_{(r)} \tag{7.4}$$

and

$$u^2(\widetilde{y}) = \frac{1}{M} \left({\sum_{r=1}^{M}}'' (y_{(r)} - \widetilde{y})^2 - \frac{1}{6} \sum_{r=1}^{M-1} (y_{(r+1)} - y_{(r)})^2 \right), \tag{7.5}$$

where the double prime on the summation in expression (7.4) and on the first summation in expression (7.5) indicates that the first and the last terms are to be taken with weight one half. However, for a sufficiently large value of M, the values obtained using expressions (7.1) and (7.2) are generally indistinguishable for practical purposes from those given by expressions (7.4) and (7.5).

7.2.4 Coverage interval for the output quantity

Let α denote any value between zero and $1 - p$, where p is the required coverage probability (e.g., 0.95). The endpoints of a $100p\,\%$ coverage interval for the output quantity are $G_Y^{-1}(\alpha)$ and $G_Y^{-1}(p + \alpha)$, i.e., the α– and $(p + \alpha)$–quantiles of $G_Y(\eta)$. Here, the β–quantile is the value of η for which $G_Y(\eta) = \beta$.

The choice $\alpha = 0.025$ gives the coverage interval defined by the 0.025– and 0.975–quantiles. This choice provides a $95\,\%$ coverage interval that is probabilistically symmetric. The probability is $2.5\,\%$ that Y is smaller than the left-hand endpoint of the interval and $2.5\,\%$ that it is larger than the right-hand endpoint. If $g_Y(\eta)$ is symmetric about its expectation, this coverage interval is symmetric about the estimate y of the output quantity, and the left-hand and right-hand endpoints of the coverage interval are equidistant from y.

A value of α different from 0.025 would generally be appropriate were the PDF asymmetric. Usually the shortest coverage interval is required, because it corresponds to the best possible

[9] The values p_r, $r = 1, \ldots, M$, are the midpoints of M contiguous probability intervals of width $1/M$ between zero and one.

location of the output quantity Y for a specified coverage probability. It is given by the value of α satisfying $g_Y(G_Y^{-1}(\alpha)) = g_Y(G_Y^{-1}(p+\alpha))$, if $g_Y(\eta)$ is single-peaked,[10] and in general by the value of α such that $G_Y^{-1}(p+\alpha) - G_Y^{-1}(\alpha)$ is a minimum. If $g_Y(\eta)$ is symmetric, the shortest coverage interval is given by taking $\alpha = (1-p)/2$.

The endpoints of a coverage interval can be obtained from the discrete representation of the distribution function for the output quantity (Section 7.2.3) as follows. Let $q = pM$, if pM is an integer, or the integer part of $pM + 1/2$, otherwise. Then, $[y_{\mathrm{low}}, y_{\mathrm{high}}] = [y_{(r)}, y_{(r+q)}]$ for any $r = 1, \ldots, M-q$, is a $100p\,\%$ coverage interval. The probabilistically symmetric $100p\,\%$ coverage interval is given by $r = (M-q)/2$ if $(M-q)/2$ is an integer, or the integer part of $(M-q+1)/2$, otherwise. The shortest $100p\,\%$ coverage interval is given by determining $r = r^*$ such that, for $r = 1, \ldots, M-q$, $y_{(r^*+q)} - y_{(r^*)} \leq y_{(r+q)} - y_{(r)}$.

The endpoints of a coverage interval can also be obtained from the approximation $\widetilde{G}_Y(\eta)$ to $G_Y(\eta)$ obtained in Section 7.2.3. For a sufficiently large value of M, the coverage interval obtained using the discrete representation $\boldsymbol{G}$ of $G_Y(\eta)$ can be expected to be indistinguishable for practical purposes from those obtained using the approximation $\widetilde{G}_Y(\eta)$. To find the left-hand endpoint y_{low} such that $\alpha = \widetilde{G}_Y(y_{\mathrm{low}})$, identify the index r for which the points $(y_{(r)}, p_r)$ and $(y_{(r+1)}, p_{r+1})$ satisfy

$$p_r \leq \alpha < p_{r+1}.$$

Then, by inverse linear interpolation,

$$y_{\mathrm{low}} = y_{(r)} + \left(y_{(r+1)} - y_{(r)}\right) \frac{\alpha - p_r}{p_{r+1} - p_r}.$$

Similarly, the upper endpoint y_{high} is calculated from

$$y_{\mathrm{high}} = y_{(s)} + \left(y_{(s+1)} - y_{(s)}\right) \frac{p + \alpha - p_s}{p_{s+1} - p_s},$$

where the index s is identified to satisfy

$$p_s \leq p + \alpha < p_{s+1}.$$

The shortest coverage interval can generally be obtained computationally from $\widetilde{G}_Y(\eta)$ by determining α such that $\widetilde{G}_Y^{-1}(p+\alpha) - \widetilde{G}_Y^{-1}(\alpha)$ is a minimum. A straightforward approach to determining the minimum is to evaluate $\widetilde{G}_Y^{-1}(p+\alpha) - \widetilde{G}_Y^{-1}(\alpha)$ for a sufficient number of choices $\{\alpha_k\}$ of α between zero and $1-p$, and to choose that value α_ℓ from the set $\{\alpha_k\}$ yielding the minimum value from the set $\{\widetilde{G}_Y^{-1}(p+\alpha_k) - \widetilde{G}_Y^{-1}(\alpha_k)\}$.

7.2.5 An adaptive Monte Carlo method

A basic implementation of an adaptive Monte Carlo method is described as follows. It is based on carrying out an increasing number of Monte Carlo trials until the various quantities

[10] See Section 2.3 for a proof of this result.

of interest have stabilised in a statistical sense. A quantity is deemed to have stabilised if twice the standard deviation associated with the estimate of its value is less than a positive numerical tolerance. The numerical tolerance is chosen to reflect the number (usually one or two) of significant decimal digits regarded as meaningful in the standard uncertainty $u(y)$.

A practical approach consists of carrying out a sequence of Monte Carlo calculations, each containing a relatively small number, say $M = 10^4$, trials. For each Monte Carlo calculation in the sequence, y, $u(y)$ and the endpoints of a $95\,\%$ coverage interval are formed from the results obtained as in Sections 7.2.2 and 7.2.4. Denote by $y^{(h)}$, $u(y^{(h)})$, $y_{\rm low}^{(h)}$ and $y_{\rm high}^{(h)}$ the values of y, $u(y)$ and the left- and right-hand endpoints of the $95\,\%$ coverage interval for the hth member of the sequence.

After the hth Monte Carlo calculation (apart from the first) in the sequence, the arithmetic mean of the values $y^{(1)}, \ldots, y^{(h)}$ and the standard deviation s_y associated with this arithmetic mean are formed. The counterparts of these statistics are determined for $u(y)$, $y_{\rm low}$ and $y_{\rm high}$. If the largest of $2s_y$, $2s_{u(y)}$, $2s_{y_{\rm low}}$ and $2s_{y_{\rm high}}$ does not exceed the numerical tolerance, the overall computation is regarded as having stabilised. The results from the total number of Monte Carlo trials taken are then used to provide the estimate of the output quantity, the associated standard uncertainty and the coverage interval for the output quantity.

7.2.6 Computation time

An indication (at the time of publishing this guide) of the computation time required for Monte Carlo calculations can be obtained from the following figures.

A problem with a measurement function consisting of the sum of five terms, a cosine, a sine, an inverse tangent, an exponential and a square root was synthesised. The quantity in each term was characterized by a Gaussian PDF. $M = 10^6$ Monte Carlo trials were made. Average computation times for a $2\,\mathrm{GHz}$ Intel processor using MATLAB [72] were as follows:

- The generation of the $5M$ Gaussian pseudo-random numbers took $0.2\,\mathrm{sec}$;
- The evaluation of the M values of the output quantity took $0.2\,\mathrm{sec}$;
- To sort the M values of the output quantity into non-decreasing order to produce a discrete representation of the distribution function for the output quantity took $6\,\mathrm{sec}$.[11]

The computation time is dominated by that to sort the values of the output quantity. This information provides a simple basis for estimating the computation time for other measurement functions, other values of M and other processors.

[11] The sorting should be carried out using a sorting algorithm that takes a number of operations proportional to $M \log M$ [91]. A naive sorting algorithm would take a number of operations proportional to M^2 and that might make the computation time unacceptably long.

7.3 A Monte Carlo method applied to a simple non-linear measurement function

Consider the univariate measurement function $Y = X^2$, where the (single) input quantity X has expectation 1.2 and standard deviation 0.5 and is characterized by a Gaussian PDF.

First, the number M of Monte Carlo trials was taken as 500. Values x_r, $r = 1, \ldots, M$, were drawn randomly from the Gaussian distribution for X. The corresponding values $y_r = x_r^2$, $r = 1, \ldots, M$, were calculated according to the measurement function. An approximation to the distribution function for Y was formed, in accordance with the above procedure, as the piecewise-linear function joining the points $(y_{(r)}, p_r)$, $r = 1, \ldots, M$, where $p_r = (r - 1/2)/M$. Figure 7.2 shows the approximation to the distribution function so obtained. The figure also shows a histogram of the values y_r, which constitutes a discrete, scaled approximation to the PDF for Y.

The approximation to the distribution function is a much smoother function than the approximation to the corresponding PDF. Such a result is generally to be expected, and relates to the fact that the PDF is the derivative of the distribution function, and that numerical approximations to the derivative of a function tend to be considerably less accurate than approximations to the function itself.

The exercise was repeated for $M = 50\,000$ trials. See Figure 7.3. The enhanced smoothness of the results is evident. Statistics computed from the results corresponding to the larger number of trials would be more reliable. It can be expected that increasing the number of trials by a factor of 100, as here, would yield an additional significant digit of accuracy in the computed statistics [28].

The enhanced resolution permits a feature to be discerned in the PDF for $M = 50\,000$ (Figure 7.3) that was not evident in that for $M = 500$ (Figure 7.2). The PDF is *bimodal*, there being a narrow peak near the origin, in addition to the main peak. This is not an artefact introduced by the particular set of draws made, but a genuine effect. Its presence is due to the fact that $0.8\,\%$ of the values of X according to its PDF are negative. These values lie in the left-hand tail of the Gaussian PDF for X, i.e., that with parameters $\mu = 1.2$ and $\sigma = 0.5$. The above proportion corresponds to the area under the standardized Gaussian curve (i.e., that for a quantity with expectation zero and standard deviation unity) to the left of the value $z = (0 - \mu)/\sigma = -2.4$. These values when squared, through the measurement function $Y = X^2$, are aggregated with those arising from small positive values of X. Even for such a superficially innocent example, there can be a 'surprise' such as this!

The estimate of the output quantity and the associated standard uncertainty as provided by the law of propagation of uncertainty are 1.44 and 1.20. Those provided by the described Monte Carlo method were 1.70 and 1.26. The standard uncertainty in this example is reasonably estimated by the law of propagation of uncertainty, but the expectation is estimated to be lower than the correct value, which could in fact be calculated explicitly.

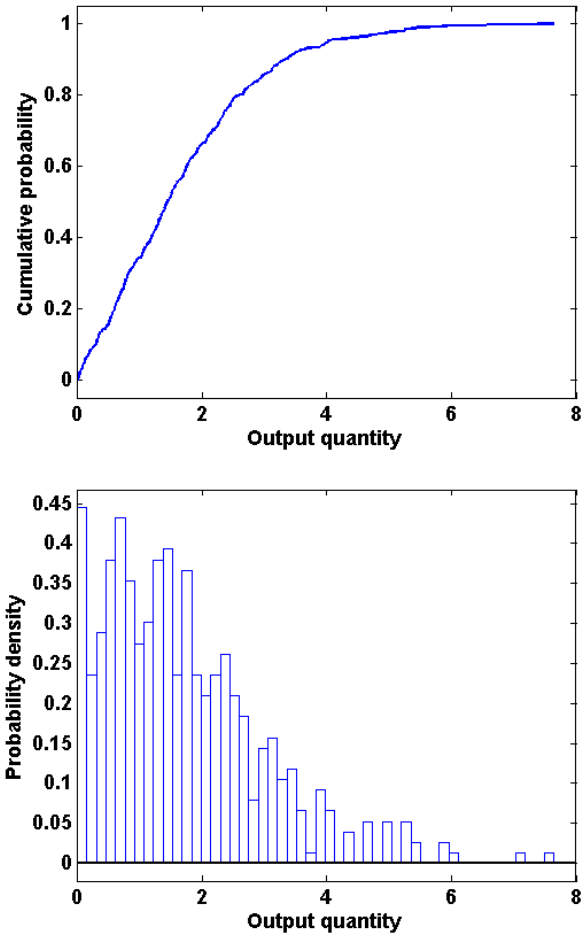

Figure 7.2: Approximation (top) obtained using a Monte Carlo method, with 500 trials, to the distribution function for the output quantity Y, where $Y = X^2$ and X has expectation 1.2 and standard deviation 0.5 and is characterized by a Gaussian distribution. A histogram (below) of the values used to produce the (approximate) distribution function. It constitutes a discrete, scaled approximation to the PDF for Y.

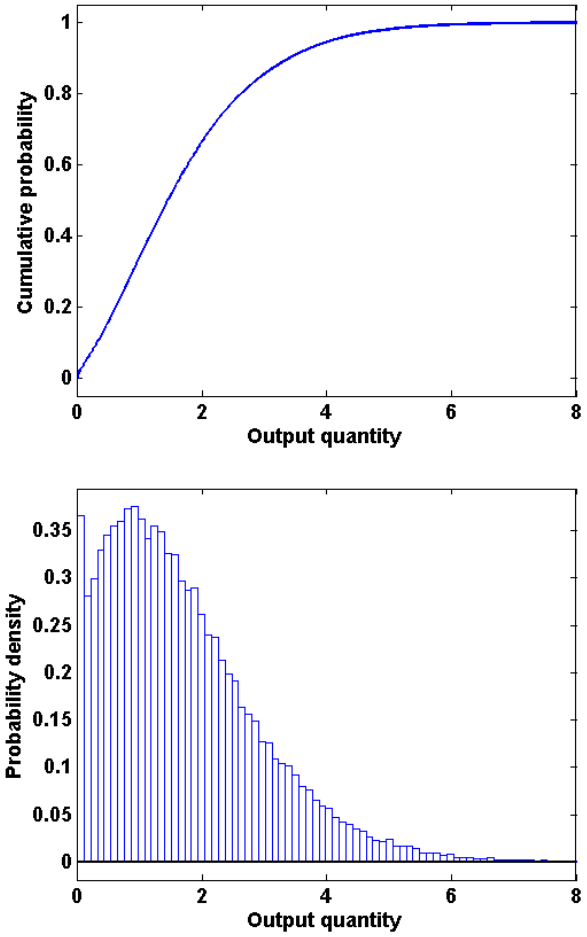

Figure 7.3: As Figure 7.2 but based on 50 000 Monte Carlo trials.

A further noteworthy feature arises from this example. In the case $M = 50\,000$, a $95\,\%$ coverage interval for Y, determined from the 0.025- and 0.975-quantiles of the (approximate) distribution function was $[0.1, 4.8]$. That provided using the GUM uncertainty framework is $[-1.1, 3.9]$ or, equivalently, 1.4 ± 2.5. The lengths of the coverage interval are similar. However, the interval provided by the GUM uncertainty framework is shifted to the left relative to that delivered by the described Monte Carlo method. In fact, the portion of the coverage interval provided by the GUM uncertainty framework from -0.8 to zero is infeasible, since, from its definition, Y cannot take negative values.

Coverage intervals at other levels of probability were also obtained using the described Monte Carlo method and by applying the GUM uncertainty framework. Appreciable differences were again observed. For instance, for a $99.8\,\%$ coverage probability (corresponding to a coverage factor of 3 under the Gaussian assumption), the coverage interval provided by the described Monte Carlo method was $[0.0, 7.5]$ and that provided by the GUM uncertainty framework is $[-2.3, 5.2]$.

Although this example might seem extreme, situations with large uncertainties arise in metrology areas such as EMC measurement. Instances where the standard uncertainty and the estimate are of similar size also arise, e.g., in dimensional metrology and in photometry and radiometry.

The effect of bias in the evaluated endpoints of the coverage interval constructed in this way, resulting from the use of a finite number of Monte Carlo trials, can be reduced using so-called bias-corrected methods [40].[12]

7.4 A Monte Carlo method for a multivariate measurement function

Consider the counterpart of Section 7.2 for a multivariate measurement function

$$\boldsymbol{Y} = \boldsymbol{f}(\boldsymbol{X}).$$

M vectors $\boldsymbol{x}_r$ are drawn randomly from the PDFs for the input quantities $\boldsymbol{X}$ as before. For each $\boldsymbol{x}_r$, evaluate the measurement function as previously, except now the output values $\boldsymbol{y}_r = \boldsymbol{f}(\boldsymbol{x}_r)$ are vectors of dimension $m \times 1$.

Assemble these vectors into a matrix of dimension $m \times M$:[13]

$$\boldsymbol{\Psi} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_M).$$

[12] In the authors' experience these corrections typically have a small effect. This aspect, however, merits further study.

[13] The symbol $\boldsymbol{\Psi}$ is (reluctantly) used to denote the matrix of vectors $\boldsymbol{y}_r$, since Y is used to denote a scalar output quantity and $\boldsymbol{Y}$ a vector output quantity.

From this matrix the covariance matrix $\boldsymbol{U}_{\boldsymbol{y}}$ associated with estimates $\boldsymbol{y}$ of the output quantities $\boldsymbol{Y}$ is calculated from

$$\boldsymbol{U}_{\boldsymbol{y}} = \frac{1}{M-1}\boldsymbol{\Psi}'(\boldsymbol{\Psi}')^{\top},$$

where $\boldsymbol{\Psi}'$ is $\boldsymbol{\Psi}$ corrected for the sample averages over all M trials, i.e., with the average of the values in the jth row subtracted from all elements in that row, for $j = 1, \ldots, M$.

This covariance matrix contains (generally a more reliable estimate of) the information that would be delivered by a linear analysis such as based on the law of propagation of uncertainty. (In fact, it provides more than the law of propagation of uncertainty, since that procedure does not in general cover multivariate measurement functions.) The matrix $\boldsymbol{\Psi}$ provides much richer information, however, in the following sense. Any column of $\boldsymbol{\Psi}$ corresponds to the values of the output quantities for one choice (random draw)) of the input quantities. *Any* (scalar) derived quantity can be determined from this single set of output values. This quantity can be calculated for all columns, the resulting (row) vector of dimension $1 \times M$ being used to provide a discrete representation of the distribution function for that quantity (as in Section 7.2.3). In particular, the discrete representation can be used to provide a coverage interval for the derived quantity (as in Section 7.2.4). Another quantity could be so introduced and the two row vectors used to compute any statistics required (expectation, median, etc.) and the pair of vectors used to approximate the covariance (or correlation) associated with the pair of derived quantities. Thus, the matrix $\boldsymbol{\Psi}$ is a very valuable array, being the basis of almost unlimited statistical information.

The extension of the approach to the evaluation of coverage regions for multivariate output quantities is not straightforward, because the operation of sorting multivariate data is generally not well-defined . Some approaches have been proposed [4], including the ranking of multivariate data using the metric

$$(\boldsymbol{y}_r - \boldsymbol{a})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_r - \boldsymbol{a}), \tag{7.6}$$

where $\boldsymbol{a}$ is a location statistic, such as the expectation or (spatial) median [94], for the set $\boldsymbol{y}_r$ and $\boldsymbol{\Sigma}$ is a dispersion statistic, such as the covariance matrix $\boldsymbol{U}_{\boldsymbol{y}}$, for the set.

The provision of coverage regions in general is currently a research topic. A simple, practical approach is therefore proposed for current purposes. As indicated in Section 3, even in the univariate case the coverage interval is not unique. There is far greater freedom of choice in the multivariate case, where any domain containing $95\,\%$ of the distribution of possible values constitutes a $95\,\%$ coverage region. Moreover, a coverage interval can be expressed in terms of just two quantities, such as the interval endpoints or the interval midpoint and the semi-width. In more than one dimension, there is an infinite number of possibilities for the *shape* of the region.

A working approach is as follows. For linear or linearized problems the covariance matrix associated with an estimate of the multivariate output quantity defines a one-standard-deviation ellipsoid [84] centred on the point denoting the estimate of the output quantity. Ellipsoids concentric with this one contain various fractions of the distribution of values

attributed to the output quantity. For a given coverage probability, $95\,\%$, say, the size of the ellipsoid from this set can be found (using the theory of multidimensional Gaussian distributions) that contains $95\,\%$ of the possible values of the output quantity. Such an ellipsoid can be constructed from the above covariance matrix, but its *size* would be dictated by the Gaussian assumption and not depend on the actual distribution of $\boldsymbol{Y}$. *An* ellipsoid is required that contains $95\,\%$ of the actual distribution. In the univariate case, it is more valid, as considered in Section 7.2, to derive the coverage interval from an approximation to the probability distribution for the output quantity. Similarly, in the multivariate case the points $\boldsymbol{y}_r$ define a cluster of points centered on $\boldsymbol{y}$. These $\boldsymbol{y}_r$ can be expected to reflect faithfully the actual distribution, as a consequence of the sampling approach used. Therefore, it is recommended to define *a* coverage region by the (first-order) ellipsoid that (just) contains $95\,\%$ of these $\boldsymbol{y}_r$.

It is emphasized that this approach will provide *a* $95\,\%$ coverage region. The extent to which it is appropriate depends on the context. It may be highly inappropriate if the actual distribution of points $\boldsymbol{y}_r$ is, say, star-shaped. However, the approach is consistent with the use of the metric (7.6) with $\boldsymbol{\Sigma} = \boldsymbol{U}_{\boldsymbol{y}}$.

7.5 Extensions to general measurement models

The extension of the above concepts to general measurement models is conceptually straightforward. Instead of forming values $y_r = f(\boldsymbol{x}_r)$ of the output quantity Y in the univariate case or $\boldsymbol{y}_r = \boldsymbol{f}(\boldsymbol{x}_r)$ of $\boldsymbol{Y}$ in the multivariate case, by evaluating a formula or formulae, it is necessary to solve, respectively, the equation $h(y_r, \boldsymbol{x}_r) = 0$ for y_r or the equations $\boldsymbol{h}(\boldsymbol{y}_r, \boldsymbol{x}_r) = \boldsymbol{0}$ for $\boldsymbol{y}_r$.

A measurement model for which the output quantity Y or $\boldsymbol{Y}$ is complex can be treated as a multivariate measurement model (as in Section 7.4) in which the output quantities are the real and imaginary parts of Y or $\boldsymbol{Y}$.

7.6 Properties of the Monte Carlo method

The attributes of the approach, and implementation issues for the approach, are briefly reviewed.

1. *Availability of pseudo-random number generators*. Pseudo-random number generators are required that are appropriate for the PDFs and joint PDFs for the input quantities that are likely to arise in metrology.

2. *Quality of pseudo-random number generators*. Some pseudo-random number generators are known to yield sequences of values that fail to satisfy standard tests for

randomness.

3. *Repeatability properties*. The results may not be repeatable, thus making the testing of software for the described Monte Carlo method harder. The same random number generator, using the same seed, must be employed to provide exactly repeatable results.

4. *Complicated measurement models*. The computational time required to carry out a sufficient number of Monte Carlo trials may be prohibitive if the measurement model is complicated. See Section 7.7.

5. *Calculation of values of the output quantity*. In the described Monte Carlo method a value of the output quantity is calculated corresponding to each set of values of the input quantities randomly drawn from the PDFs for those quantities, and hence for some values that may be a number of 'standard deviations' away from the estimates of the input quantities. This is in contrast to the procedure based on the law of propagation of uncertainty in which a value of the output quantity is calculated only at the estimates of the input quantities and, if finite difference approximations are used, also at points perturbed from these estimates by $\pm$ one standard uncertainty for each quantity in turn. For this reason some issues may arise regarding the numerical procedure used to calculate a value of the output quantity, e.g., ensuring its convergence (where iterative schemes are used) and numerical stability.

6. *Straightforward use*. Software can be implemented such that the user provides information concerning just the measurement model and the parameters defining the PDFs for the input quantities.

7. *A discrete representation of the distribution function for the output quantity (for univariate problems) is provided* (rather than a single statistic such as the standard deviation). Any required statistic (standard deviation, higher-order moments, etc.), coverage intervals and derived statistics such as the uncertainties associated with an estimate of any function of the output quantity can be calculated from this representation.

8. *A discrete representation of the (joint) PDF for the output quantities for multivariate problems is provided*. This takes the form of a set of (M) values (points) of the output quantities. This information is valuable in the context of multi-stage measurement models in which the output from one stage becomes the input to the next. Drawing randomly from these points embodies all the distributional information (including correlation) that is present.

9. *Applicability to a wide range of measurement models*. The described Monte Carlo method is broadly applicable regardless of the nature of the measurement model:

   (a) The measurement model may be linear, mildly non-linear or strongly non-linear. No initial analysis of the measurement model is required to decide, for instance, how many terms in the Taylor-series expansion are required to approximate the

model adequately for purposes of determining unbiased estimates of statistics associated with the estimate of the output quantity.

(b) The uncertainties associated with estimates of the input quantities may be arbitrarily large.

(c) No assumption is made concerning the PDF for the output quantity Y. Thus, distributions for quantities that cannot be negative for instance, such as a distribution for distance, can be sensibly approximated.

10. *Symmetry is not assumed.* No assumption is made in using the described Monte Carlo method concerning the symmetry of the PDFs for the input quantities or of the PDF for the output quantity. Thus, there is no need to 'symmetrize' any PDF, or indeed any advantage gained from doing so.[14]

11. *Sensitivity coefficients are not required.* There is no need for algebraic expressions for the first-order partial derivatives of the measurement function with respect to the input quantities and for the evaluation of these expressions at the estimates of the input quantities.

12. *Avoidance of the concept of effective degrees of freedom.* The described Monte Carlo method avoids the concept of effective degrees of freedom: an experimental mean and a standard deviation of a quantity, for which a Gaussian prior has been assumed, are described by a posterior density, viz., a t–distribution with the mean as the location parameter and the standard deviation as the scale parameter.[15]

13. *Linear computing time.* The computing time is dominated by the product of the number of trials and the time to evaluate the measurement model. Over and above this, it is independent of the number N of input quantities. (This is not the case for the numerical evaluation of the multivariate integral (3.1) that defines Y, where the computation time is essentially proportional to C^N, for some constant C.)

14. *Sensitivities can approximately be calculated.* The described Monte Carlo method does not automatically provide sensitivity coefficients, for two reasons. First, they are not required for purposes of its operation. Second, for a non-linear model, sensitivity coefficients are in general approximate, the quality of the approximations worsening with increased standard uncertainties associated with estimates of the input quantities. However, simply by holding all input quantities but one fixed at their expected values the described Monte Carlo method can be used to provide the PDF for the

[14] The (UK) Royal Society of Chemistry states [2] that 'Such an assumption [the use of a single parameter–often taken to be the half-range] is appropriate only if the 'dispersion of values' is nearly symmetric about the measured result. It is easy to think of reasons why this may be false, including the effects of contamination and of range shifting on instruments'.

[15] A discussion [53] of this issue suggests that in the case of finite degrees of freedom the standard deviation u of the Gaussian distribution should also be regarded as a random variable. In this regard, a PDF should also be attached to u. Because of the Gaussian assumption this PDF is distributed as χ^2. An example of pseudo-code is available [53] that accounts for this effect.

output quantity for the measurement model having just that input quantity. See Appendix B.3.

15. *Multi-stage measurement models*. The described Monte Carlo method can take the output matrix $\boldsymbol{\Psi}$ of vectors $\boldsymbol{y}_r$ from one stage of a multi-stage measurement model, and draw randomly from $\boldsymbol{\Psi}$ at the input to the next.

7.7 Summary remarks on the described Monte Carlo method

The described Monte Carlo method is a tool that is consistent with general GUM philosophy (GUM Clause G.1.5) and also with its interpretation [99] for scientists at the National Institute for Standards and Technology (NIST) in the United States. The major difference is that rather than propagating uncertainty through a linearized measurement model, the PDFs for the input quantities are propagated through the model *per se* to calculate (an approximation to) the PDF for output quantity. From the PDF for the output quantity a coverage interval is obtained without making a Gaussian or any other assumption concerning the form of this PDF.

The described Monte Carlo method can straightforwardly be applied to a range of uncertainty evaluation problems. For the most general such problem, it is emphasized that it would be necessary to provide:

1. Pseudo-random number generators for the univariate and joint PDFs needed in the application;

2. A mechanism for determining coverage regions for multivariate output quantities.

Recommendations [34] are intended to assist in this regard.

The degree of belief in the PDFs for the input quantities can be considered by repeating a Monte Carlo calculation after having varied these functions. The sensitivity of the results to such critical information can thus be investigated.

For simple measurement models the number of Monte Carlo trials can be chosen to be substantially large, e.g., 10^6 (Section 7.2.6). A complete uncertainty calculation might take six seconds on a 2 GHz Intel processor (Section 7.2.6). The majority of this time is taken with *sorting* the Monte Carlo values of the output quantity. For measurement models of modest complexity, taking say 100 times longer to evaluate, to achieve a comparable quality of result would take of the order of 26 seconds. The figure is not 600 seconds because the sorting time remains about 6 seconds. Nevertheless the computation time is now noticeable, particularly if many similar calculations have to be carried out. In such cases it would be desirable to consider the use of an automatic stopping rule (Section 7.2.5), rather than fixing the number of Monte Carlo trials in advance.

For very complicated measurement models[16] it would not be economic to take more than a small number of trials (say 10). In such a case it would be impossible to provide a coverage interval reliably. Rather, an expectation and standard deviation should be obtained and a coverage interval obtained assuming a Gaussian distribution.[17] For multivariate output quantities a covariance matrix can be calculated from the results and a multivariate Gaussian distribution assumed. As always, the results obtained should be accompanied by a statement that indicates clearly how they were obtained and what assumptions were made.

The computational task undertaken in applying a Monte Carlo method is suitable for parallel computing and, in particular, for a distributed computing system in which the spare processing capacity of a large number of desktop computers is coordinated to undertake the computation. Parallel computing provides another approach to treating complicated measurement models. See reference [45], which considers some of the issues in using a distributed computing system, such as the way the pseudo-random number generators on the different computers in the distributed system are seeded, and describes an application.

[16] An instance is a model defined by a partial differential equation.

[17] Appeal can be made to the Principle of Maximum Entropy: see Section A.5.

Chapter 8

Validation of the GUM uncertainty framework

The GUM uncertainty framework has some limitations [10]. Although the procedure can be expected to work well in many circumstances, it is generally difficult to quantify the effects of the approximations involved, viz., linearization, the Welch-Satterthwaite formula for the effective degrees of freedom and the assumption that the output quantity is Gaussian (i.e., that the central limit theorem is applicable). Indeed, the degree of difficulty of doing so would typically be considerably greater than that required to apply a Monte Carlo method. Therefore, since these circumstances cannot readily be tested, any cases of doubt should be validated. To this end, it is recommended that both the GUM uncertainty framework and the described Monte Carlo method are applied and the results compared. If the comparison is favourable, the GUM uncertainty framework can be used on this occasion and for sufficiently similar problems in the future. Otherwise, consideration can be given to using a Monte Carlo method instead.

Specifically, it is recommended that the two steps below and the following comparison process are carried out.

1. Apply the GUM uncertainty framework to yield a $95\,\%$ coverage interval $y \pm U(y)$ for the output quantity.

2. Apply the described Monte Carlo method to yield the standard uncertainty $u(y)$ associated with an estimate of the output quantity and the endpoints y_{low} and y_{high} of a $95\,\%$ coverage interval for the output quantity.[1]

[1] A sufficient number of Monte Carlo trials should be performed in obtaining results from the method. The adaptive procedure described in Section 7.2.5 can be applied with a numerical tolerance for deciding that the results provided by the procedure have stabilized that is smaller than that used in the validation procedure. See reference [11] for further advice.

A comparison procedure is based on the following objective: determine whether the coverage intervals obtained by the GUM uncertainty framework and a Monte Carlo method agree to a stipulated degree of approximation. This degree of approximation is assessed in terms of the endpoints of the coverage intervals and corresponds to that given by expressing the standard uncertainty $u(y)$ to what is regarded as a meaningful number of *significant decimal digits*.

The procedure is as follows:

1. Let n_{ndig} denote the number of significant digits regarded as meaningful in the numerical value of $u(y)$. Usually, $n_{\mathrm{ndig}} = 1$ or 2. Express the value of $u(y)$ in the form $a \times 10^r$, where a is an n_{ndig}–digit integer and r an integer. The comparison accuracy is
$$\delta = \frac{1}{2}10^r.$$

2. Compare the coverage intervals obtained by the GUM uncertainty framework and the described Monte Carlo method to determine whether the required number of correct digits in the coverage interval provided by the GUM uncertainty framework has been obtained. Specifically, determine the quantities
$$d_{\mathrm{low}} = |y - U(y) - y_{\mathrm{low}}| \tag{8.1}$$
and
$$d_{\mathrm{high}} = |y + U(y) - y_{\mathrm{high}}|, \tag{8.2}$$
viz., the absolute values of the differences of the respective endpoints of the two coverage intervals. Then, if both these quantities are no larger than δ the comparison is successful and the GUM uncertainty framework has been validated in this instance.

Example 22 *Setting the degree of approximation*

The estimate of the mass of a nominally 100 g standard of mass [10, Clause 7.2.2] is $y = 100.021\,47\,\mathrm{g}$ with associated standard uncertainty $u(y) = 0.000\,35\,\mathrm{g}$. Thus, $n_{\mathrm{ndig}} = 2$ and $u(y)$ is expressed as $35 \times 10^{-5}\,\mathrm{g}$, and so $a = 35$ and $r = -5$. Take $\delta = 0.5 \times 10^{-5}\,\mathrm{g} = 0.000\,005\,\mathrm{g}$.

Chapter 9

Examples

The examples in this chapter are intended to illustrate the principles contained in the body of this guide. Where appropriate, two approaches, the GUM uncertainty framework (Section 5.3 and Chapeter 6) and the Monte Carlo method of Chapter 7, are used and contrasted. Analytical solutions are also obtained in some cases for purposes of further comparison. Some examples are typical of those that arise in metrology. Others attempt to indicate the considerations that are necessary when those of 'normal circumstances' fail to apply. Perhaps, unfortunately, such adverse circumstances arise more frequently than would be wished, in areas such as electromagnetic compliance, photometry and dimensional metrology.

The following examples are presented:

- A simple summation model (Section 9.1) and a logarithmic transformation model (Section 9.2). Comparisons are made of results obtained analytically and from the applications of the GUM uncertainty framework and a Monte Carlo method.

- Flow in a channel (Section 9.3). This example illustrates a multi-stage measurement model with a sub-model that takes the form of an implicit equation. A Monte Carlo method is used to validate the GUM uncertainty framework as an approach to uncertainty evaluation.

- Electrical resistance (Section 9.4). This example illustrates an instance where it is important to take specific account of the PDFs for the input quantities in the measurement model in order to ensure that valid results are obtained from an uncertainty evaluation.

- Calibration of a digital multimeter (Section 9.5). A comparison is made of the results obtained using a semi-analytical approach and from the application of a Monte Carlo method.

- Measuring the lengths of the sides of a right-angled triangle (Section 9.6). This example illustrates the use of statistical modelling. It also illustrates the manner in which correlation effects can be removed by the introduction of additional quantities.
- Fourier transform (Section 9.7). This example illustrates the application of the GUM uncertainty framework for a multivariate measurement function and a multi-stage measurement model. It concerns the calculation and subsequent use of the discrete Fourier transform.
- Mass calibration (Section 9.8). A comparison is made of the results obtained from the application of the GUM uncertainty framework (with first and higher order terms) and a Monte Carlo method. The example illustrates the use of a Monte Carlo method to validate the results returned by the GUM uncertainty framework. The problem is also treated using a Bayesian analysis. In particular, the example illustrates the influence of the choice of prior PDF for the measurand on the results obtained using such an analysis. The example is included in the first Supplement [11] to the GUM.
- Comparison loss in microwave power meter calibration (Section 9.9). A comparison is made of the results obtained from the application of the GUM uncertainty framework (with first and higher order terms) and a Monte Carlo method. The example also illustrates the effect of mutual dependence between the input quantities in the measurement model. The example is included in the first Supplement [11] to the GUM.
- Calibration of a gauge block (Section 9.10). This example corresponds to that given in Annex H.1 of the GUM [10]. The example illustrates how information about each input quantity in a measurement model may be used to characterize those quantities by probability distributions. A comparison is made of the results obtained from the application of the GUM uncertainty framework and a Monte Carlo method. The example is included in the first Supplement [11] to the GUM.
- Quantities subject to a normalisation constraint (Section 9.11). This example illustrates how measured (indication) values of the composition of a mixture, such as a certified reference material used in gas analysis, can be analysed to furnish estimates of the concentrations of the components in the mixture that satisfy a normalisation constraint (that they sum to unity).
- Area under a curve defined by measured data (Section 9.12). This example concerns the application of the law of propagation of uncertainty to evaluate the uncertainty associated with the result delivered by a quadrature rule used to approximate the value of a definite integral. The example illustrates how to obtain the sensitivity coefficients for a linear measurement function defined by a numerical procedure. The example also illustrates how to use the results of the uncertainty evaluation in a procedure to decide the order of the quadrature rule.
- Modelling of SIR efficiency curves (Section 9.13). This example illustrates the use of least-squares to fit a curve to measured data in the case that the curve depends

on additional (reference) data for which estimates are available as tabulated values. The evaluation of the uncertainties associated with estimates of the parameters of the curve is treated using the GUM uncertainty framework and the application of a Monte Carlo method.

Many other examples are given throughout this guide, some to illustrate basic points and others more comprehensive. More than the one or two significant decimal digits recommended [10, Clause 7.2.6] are used for reporting the uncertainties in some of the examples for purposes of comparison.

9.1 Summation model

The measurement function is $Y = X_1 + X_2$, where, for $i = 1, 2$, X_i is characterized by a rectangular PDF with endpoints a_i and b_i, with $a_i < b_i$. Using convolution principles [84, p93], the PDF (Figure 9.1) for Y is

$$g_Y(\eta) = \frac{1}{\lambda_1 + \lambda_2} \min\left(\frac{1}{\lambda_2 - \lambda_1} \max\left(\lambda_2 - |\eta - \mu|, 0\right), 1\right),$$

where $\mu = (a_1 + a_2 + b_1 + b_2)/2$, $\lambda_1 = |a_2 - a_1 + b_1 - b_2|/2$ and $\lambda_2 = (b_1 + b_2 - a_1 - a_2)/2$. From this PDF the expectation of Y is taken as y and the variance of Y as $u^2(y)$, where

$$y = \int_{a_1+a_2}^{b_1+b_2} \eta g_Y(\eta)\, \mathrm{d}\eta = \mu,$$

and

$$u^2(y) = \int_{a_1+a_2}^{b_1+b_2} (\eta - y)^2 g_Y(\eta)\, \mathrm{d}\eta = \frac{\lambda_1^2 + \lambda_2^2}{6}.$$

The PDF for Y is symmetric and hence a coverage interval I_p with endpoints equidistant from the expectation μ is the shortest such interval (Section 2.3). Hence, for a coverage probability p,

$$I_p = \mu \pm \omega, \qquad \int_{\mu-\omega}^{\mu+\omega} g_Y(\eta)\, \mathrm{d}\eta = p,$$

from which it follows that

$$I_p = \mu \pm \begin{cases} (\lambda_1 + \lambda_2)p/2, & p < 2\lambda_1/(\lambda_1 + \lambda_2), \\ \left(\lambda_2 - \left[(\lambda_2^2 - \lambda_1^2)(1-p)\right]^{1/2}\right), & \text{otherwise.} \end{cases}$$

A value of p satisfying $p < 2\lambda_1/(\lambda_1 + \lambda_2)$ corresponds to endpoints of I_p lying in the interval $\mu \pm \lambda_1$. Otherwise, the endpoints lie outside this interval.

Applying the GUM uncertainty framework, for $i = 1, 2$, the estimate x_i of X_i is taken as the expectation of X_i, viz., $x_i = (a_i + b_i)/2$, and the associated standard uncertainty $u(x_i)$

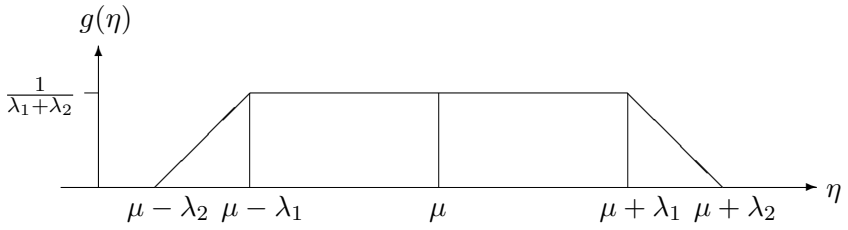

Figure 9.1: The (trapezoidal) PDF for $Y = X_1 + X_2$, where, for $i = 1, 2$, the PDF for X_i is rectangular.

as the standard deviation of X_i, viz., $(b_i - a_i)/\sqrt{12}$. The estimate y of Y is Y evaluated at $X_1 = x_1$ and $X_2 = x_2$, i.e., $y = \mu$. The sensitivity coefficients are $c_1 = c_2 = 1$. The law of propagation of uncertainty gives the uncertainty $u(y)$ associated with y from $u^2(y) = u^2(x_1) + u^2(x_2) = [(b_1 - a_1)^2 + (b_2 - a_2)^2]/12$. The quantity Y is characterized by $\mathrm{N}(y, u^2(y))$ and a coverage interval for this value corresponding to a $95\,\%$ coverage probability is $y \pm 2u(y)$.

A Monte Carlo method was carried out with $M = 10^6$ trials, five times in all in this example, in order to indicate the dispersion of results obtained. The results obtained when $a_1 = 0$, $b_1 = 1$, $a_2 = 0$ and $b_2 = 10$, from which $\mu = 5.5$, $\lambda_1 = 4.5$ and $\lambda_2 = 5.5$, are shown in Table 9.1. The distribution functions and PDFs obtained using the GUM uncertainty framework and a Monte Carlo method are shown in Figure 9.2, in which broken and continuous vertical lines indicate, respectively, the endpoints of the $95\,\%$ coverage intervals determined using the two approaches. The (trapezoidal) PDF obtained from the analytical solution is also shown in the figure, and is seen to match well with the approximation to the PDF obtained using a Monte Carlo method (displayed as a scaled frequency distribution). In contrast the PDFs obtained from the analytical solution and the GUM uncertainty framework are very different. Furthermore, the coverage interval provided by the GUM uncertainty framework is about $10\,\%$ longer than that provided by a Monte Carlo method, and includes infeasible values for Y, i.e., values outside the interval $[0, 11]$ that are taken by Y with zero probability.

9.2 Logarithmic transformation model

The measurement function is $Y = \ln X$, i.e., having a single input quantity $X \equiv X_1$, where X is characterized by a rectangular PDF with endpoints a and b, with $0 < a < b$. Such a measurement function arises, e.g., when converting electrical quantities from natural to decibel units [101]. The PDF for Y is (Section 5.2)

$$g_Y(\eta) = \begin{cases} \exp(\eta)/(b-a), & \ln a < \eta < \ln b, \\ 0, & \text{otherwise.} \end{cases}$$

Method	y	$u(y)$	Endpoints of 95 % coverage interval	
Analytical	5.50	2.90	0.71	10.29
GUF	5.50	2.90	−0.19	11.19
MCM 1	5.50	2.90	0.73	10.32
MCM 2	5.50	2.90	0.71	10.30
MCM 3	5.50	2.90	0.70	10.28
MCM 4	5.50	2.90	0.73	10.32
MCM 5	5.50	2.90	0.72	10.30

Table 9.1: Results for the summation model from the analytical solution, the GUM uncertainty framework (GUF) and five runs of a Monte Carlo method (MCM 1–5), each with $M = 10^6$ trials.

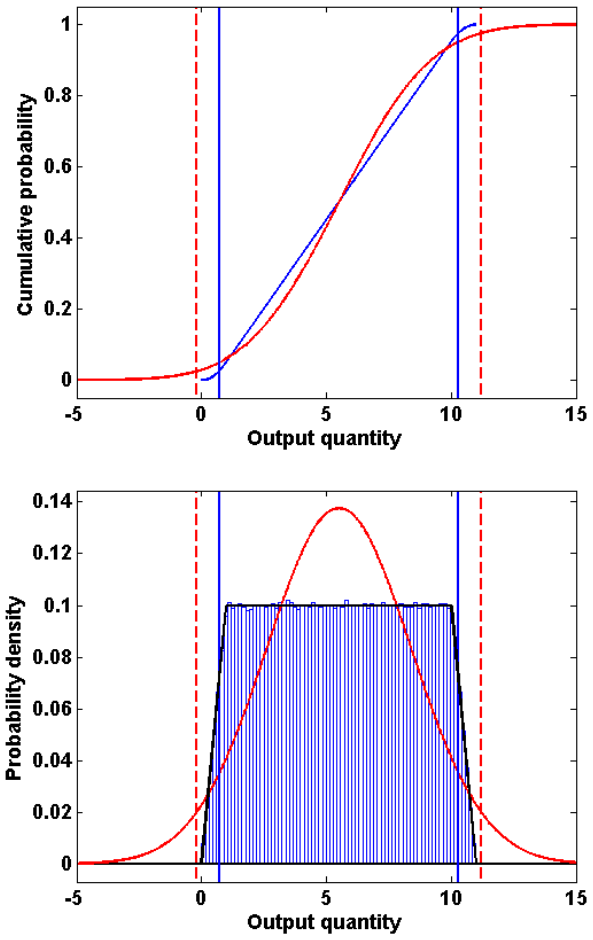

Figure 9.2: Distribution functions and (below) PDFs for the summation model obtained using the GUM uncertainty framework and a Monte Carlo method. The (trapezoidal) PDF obtained from the analytical solution is shown in the lower graph.

From this PDF the expectation of Y is taken as y and the variance of Y as $u^2(y)$, where

$$y = \int_{\ln a}^{\ln b} \frac{\eta e^{\eta}}{b-a}\,\mathrm{d}\eta = \frac{b(\ln b - 1) - a(\ln a - 1)}{b-a},$$

and

$$u^2(y) = \int_{\ln a}^{\ln b} \frac{(\eta - y)^2 e^{\eta}}{b-a}\,\mathrm{d}\eta = \frac{b(\ln b - y - 1)^2 - a(\ln a - y - 1)^2}{b-a} + 1.$$

The PDF for Y is a monotonically increasing function over the interval $[\ln a, \ln b]$, and hence the shortest coverage interval for Y has $y_{\mathrm{high}} = \ln b$ as its right-hand endpoint. For a coverage probability p, the left-hand endpoint y_{low} is such that

$$\int_{y_{\mathrm{low}}}^{\ln b} \frac{e^{\eta}}{b-a}\,\mathrm{d}\eta = p,$$

giving

$$y_{\mathrm{low}} = \ln[pa + (1-p)b].$$

Applying the GUM uncertainty framework, the estimate x of X is taken as the expectation of X, viz., $x = (a+b)/2$, and the associated standard uncertainty $u(x)$ as the standard deviation of X, viz., $(b-a)/\sqrt{12}$. The estimate y of Y is Y evaluated at $X = x$, i.e., $y = \ln x = \ln((a+b)/2)$. The (single) sensitivity coefficient is $c = \partial \ln X/\partial X$ evaluated at $X = x$, i.e., $c = 1/x = 2/(a+b)$. The law of propagation of uncertainty gives the uncertainty associated with y as $u(y) = |c|u(x) = (b-a)/[(a+b)\sqrt{3}]$. The quantity Y is characterized by $\mathrm{N}(y, u^2(y))$ and a coverage interval for this value corresponding to a $95\,\%$ coverage probability is $y \pm 2u(y)$.

Table 9.2 gives the results obtained when $a = 0.1$ and $b = 1.1$. The distribution functions and PDFs obtained using the GUM uncertainty framework and a Monte Carlo method with $M = 10^6$ trials are shown in Figure 9.3, in which broken and continuous vertical lines indicate, respectively, the endpoints of the $95\,\%$ coverage interval determined using the two approaches. The (exponential) PDF obtained from the analytical solution is also shown in the figure, and is seen to match well with the approximation to the PDF obtained using a Monte Carlo method (displayed as a scaled frequency distribution). In contrast the PDFs obtained from the analytical solution and the GUM uncertainty framework are very different. Furthermore, the coverage interval provided by the GUM uncertainty framework includes infeasible values for Y, i.e., values outside the interval $[\ln a, \ln b]$ that are taken by Y with zero probability.

9.3 Flow in a channel

This example was provided by the National Engineering Laboratory. It concerns a multi-stage measurement model arising in channel flow with a sub-model that takes the form of an implicit equation.

Method	y	$u(y)$	Endpoints of 95 % coverage interval	
Analytical	-0.665	0.606	-1.897	0.095
GUF	-0.511	0.481	-1.454	0.432
MCM	-0.664	0.606	-1.895	0.095

Table 9.2: Results for the logarithmic transformation model from the analytical solution, the GUM uncertainty framework (GUF) and a Monte Carlo method (MCM) with $M = 10^6$ trials.

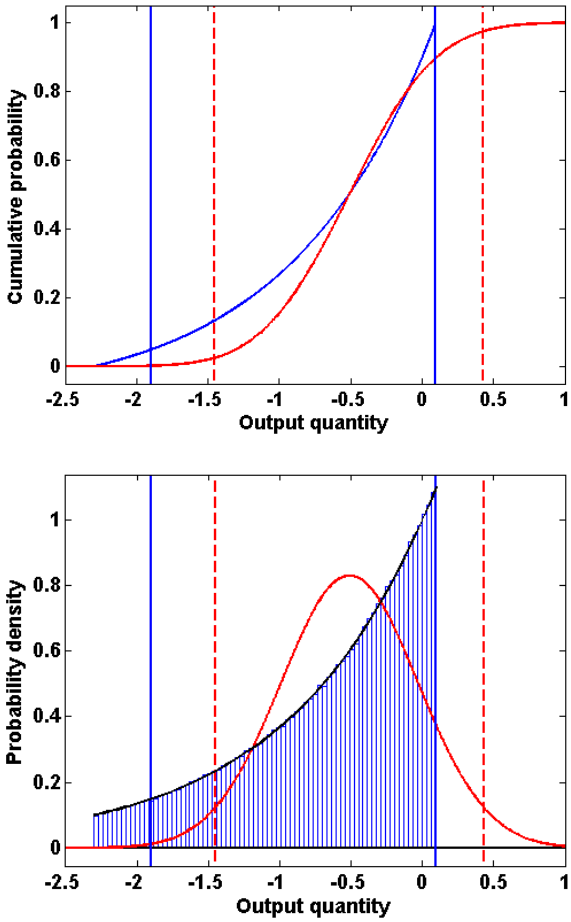

Figure 9.3: Distribution functions and (below) PDFs for the logarithmic transformation model obtained using the GUM uncertainty framework and a Monte Carlo method. The (exponential) PDF obtained from the analytical solution is shown in the lower graph.

Open channel flows are common in the water and hydroelectric power industries and where river extraction provides cooling water for industrial processes. Such a flow can be measured by introducing a specially constructed restriction in the flow channel. The flow is then a function of the geometry of the restriction (width upstream, throat width and length, height of the hump in the floor of the restriction) and the depth of water passing through the restriction.

The input quantities (and estimates of them) are:

Approach channel width	B	$(2.0\,\mathrm{m})$,
Hump height	p	$(0.25\,\mathrm{m})$,
Nominal head	h	$(1.0\,\mathrm{m})$,
Throat width	b	$(1.0\,\mathrm{m})$,
Throat length	L	$(3.0\,\mathrm{m})$.

The output quantity is the flow rate Q. The measurement model relating Q to the input quantities is

$$Q = (2/3)^{3/2} g^{1/2} C_v C_D b h^{3/2}, \tag{9.1}$$

with $g = 9.812\,\mathrm{m\,s^{-2}}$, the acceleration due to gravity,

$$C_D = (1 - 0.006L/b)(1 - 0.003L/h)^{3/2} \tag{9.2}$$

and

$$4b^2h^2C_v^2 - 27B^2(h+p)^2(C_v^{2/3} - 1) = 0. \tag{9.3}$$

To calculate the value of Q for values of the input quantities, it is first necessary to form C_D from the explicit formula (9.2) and C_v from the implicit equation (9.3), which may be regarded as sub-models (in a multi-stage measurement model) to that defined by (9.1). The equation (9.3) is in fact a cubic equation in the variable $C_v^{2/3}$ and, as a consequence, C_v can be expressed explicitly in terms of B, h and p. Doing so is unwise because of the possible numerical instability due to subtractive cancellation in the resulting form. Rather, the cubic equation can be solved using a recognised stable numerical method.

The first four input quantities are geometric dimensions obtained by a series of measurements made with a steel rule at various locations across the flume. There are uncertainties associated with the estimates of these quantities due to location, rule reading errors and rule calibration. Head height is measured with an ultrasonic detector, with uncertainties arising from fluctuations in the water surface and instrument calibration.

All uncertainty sources were quantified and the corresponding input quantities characterized by appropriate probability distributions. All PDFs were based on Gaussian or rectangular distributions. The standard deviations of the input quantities (standard uncertainties associated with estimates of the input quantities), characterized by these PDFs, were all less than $0.3\,\%$ relative to the corresponding expectations (estimates of the input quantities).

Both the GUM uncertainty framework and a Monte Carlo method were applied. The results obtained from the GUM uncertainty framework were validated using a Monte Carlo procedure, under the requirement that results to two significant digits were required. In fact, the coverage interval for Q as produced by the GUM uncertainty framework was confirmed correct (by carrying out further Monte Carlo trials) to *three* significant digits. To give the comparison in a relative sense, the quotient of (a) the half-length of the $95\,\%$ coverage interval for Q and (b) the standard uncertainty associated with the estimate of the output quantity, was 1.96, which agrees to three significant digits with the value obtained from the GUM uncertainty framework, viz., the (Gaussian) coverage factor for $95\,\%$ coverage probability. For further comparison, the corresponding quotients corresponding to $92.5\,\%$ and $97.5\,\%$ coverage intervals were 1.78 and 2.24, also in three-digit agreement with results obtained from the GUM uncertainty framework. It is concluded that the use of the GUM uncertainty framework is validated for this example for the coverage probabilities indicated.

9.4 Graded resistors

This example is intended to cover an instance where it would be important to take specific account of the PDF for an input quantity to help ensure that valid results are obtained from an uncertainty evaluation.

The uncertainties associated with a mass-produced electrical circuit are to be evaluated. The circuit contains electrical components of various types. One of these component types, a resistor, is considered here.

Nominally $1\,\Omega$ resistors are graded according to their specification. A-grade resistors are those that lie within $1\,\%$ of nominal, B-grade within $5\,\%$ and C-grade within $10\,\%$. The allocation of resistors to the various grades is decided by measurement. For the purposes of this example, the uncertainty associated with this measurement is taken as negligible. As each resistor is measured it is allocated to an A-grade, a B-grade, a C-grade or an unclassified 'bin'. The allocation is made in the following sequential manner. If a resistor has resistance in the interval $(1.00 \pm 0.01)\,\Omega$, it is allocated to the A-grade bin. If not, and it has resistance in the interval $(1.00 \pm 0.05)\,\Omega$, it is allocated to the B-grade bin. If not, and it has resistance in the interval $(1.00 \pm 0.10)\,\Omega$, it is allocated to the C-grade bin. Otherwise, it is allocated to the unclassified bin.

For the circuit application, C-grade resistors are selected. All such resistors have a resistance in the interval [0.90, 0.95] Ω or the interval [1.05, 1.10] Ω. From the knowledge of the manufacturing process, the expectation of the resistance of a resistor before the allocation process is carried out is $1.00\,\Omega$, the standard deviation is $0.04\,\Omega$ and the PDF for the resistance can be taken as Gaussian.

Consider the use of three such resistors in series within the circuit to form a (nominally)

No. N of resistors	MCM /Ω		GUF /Ω	
1	0.91	1.09	0.87	1.13
3	2.78	3.22	2.77	3.23
6	5.69	6.31	5.68	6.32
10	9.58	10.42	9.58	10.42
20	19.41	20.59	19.41	20.59

Table 9.3: The endpoints of the probabilistically symmetric $95\,\%$ coverage intervals for N Grade-C $1\,\Omega$ resistors in series evaluated using a Monte Carlo method (MCM) and the GUM uncertainty framework (GUF).

$3\,\Omega$ resistance. The measurement function for the $3\,\Omega$ resistance R is

$$R = R_1 + R_2 + R_3,$$

where R_i denotes the resistance of resistor i. Each R_i is characterized by a PDF as above. What is the PDF for R and what is a $95\,\%$ coverage interval for R?

The following figures show diagrammatically approximations to the distribution functions and PDFs obtained using a Monte Carlo method. An analytic or semi-analytic treatment is possible, but a Monte Carlo method enables results to be provided rapidly. All results are based on the use of $M = 10^5$ Monte Carlo trials.

Figure 9.4 shows the distribution function and PDF for R_i obtained using a Monte Carlo method. The PDF is basically Gaussian, with the central and tail regions removed as a consequence of the grading process. The endpoints of the probabilistically symnmetric $95\,\%$ coverage interval obtained from the distribution function are indicated by vertical continuous lines and the corresponding endpoints under the Gaussian assumption by vertical broken lines.

Figure 9.5 shows the distribution function and PDF for R, three Grade-C resistors in series. The PDF is multimodal, possessing four maxima. The expectation of R, characterised by this PDF, is $3.00\,\Omega$, the sum of the expectations of the resistances R_i. This value is, however, unrepresentative, an 'expectation' that could rarely occur. The PDF could be perceived as an overall bell-shape, with strong structure within it. Indeed, the counterpart of these results for six resistors in series, as illustrated in Figure 9.6, lies even more in that direction.

Table 9.3 summarises the numerical results, and also includes the results for $N = 10$ and 20 resistors. It is reassuring that, considering the appreciable departure from normality, the coverage interval 'converges' rapidly to that obtained under the assumption that the PDF for the output quantity is Gaussian (as in the GUM uncertainty framework). There are no grounds for complacency, however: there will be situations where the use of the GUM uncertainty framework is not so favourable.

As stated, an analytical treatment would be possible for this problem. It might be difficult to

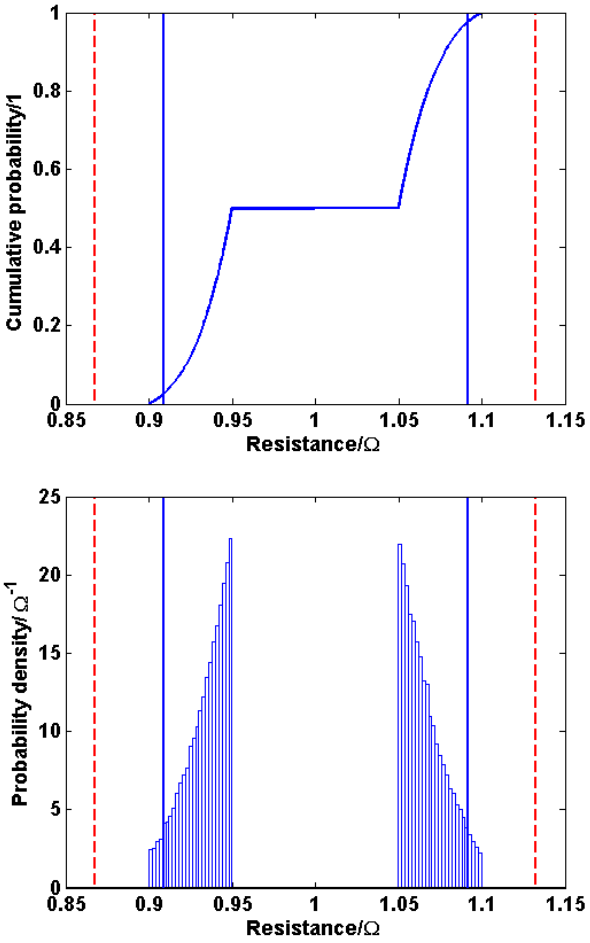

Figure 9.4: Distribution function and (below) PDF for the resistance of a single Grade-C resistor.

justify the effort required, however, unless the analysis provided some general insight that would give added value to the application. Using existing software implementing a Monte Carlo method, it required approximately one hour to enter the problem and produce the numerical and graphical results. The computation time itself was negligible, being a few seconds in all.

9.5 Calibration of a digital multimeter

A hand-held digital multimeter (DMM) is calibrated at an input of 100 V DC using a multifunction calibrator as a working standard. A measurement function for the error of indica-

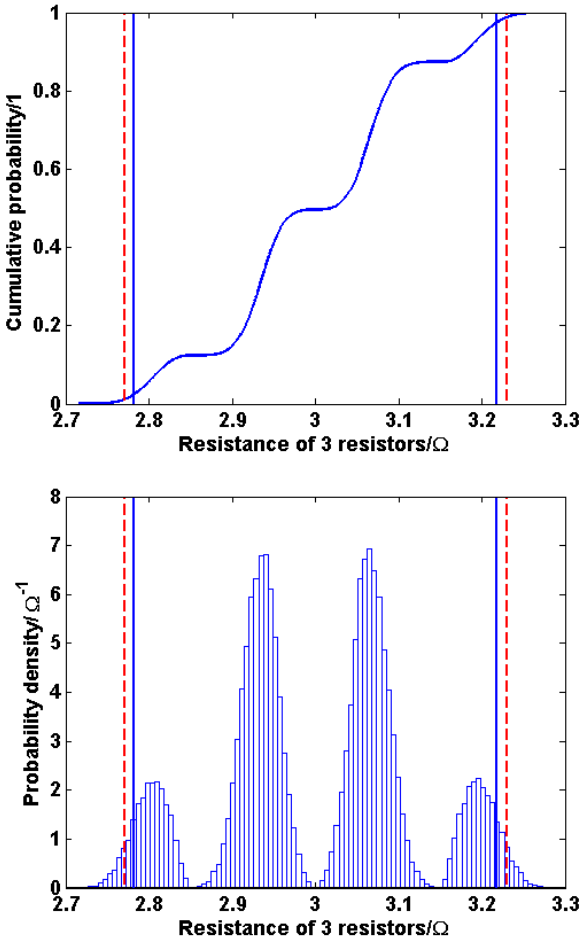

Figure 9.5: Distribution function and (below) PDF for the resistance R of three Grade-C resistors in series.

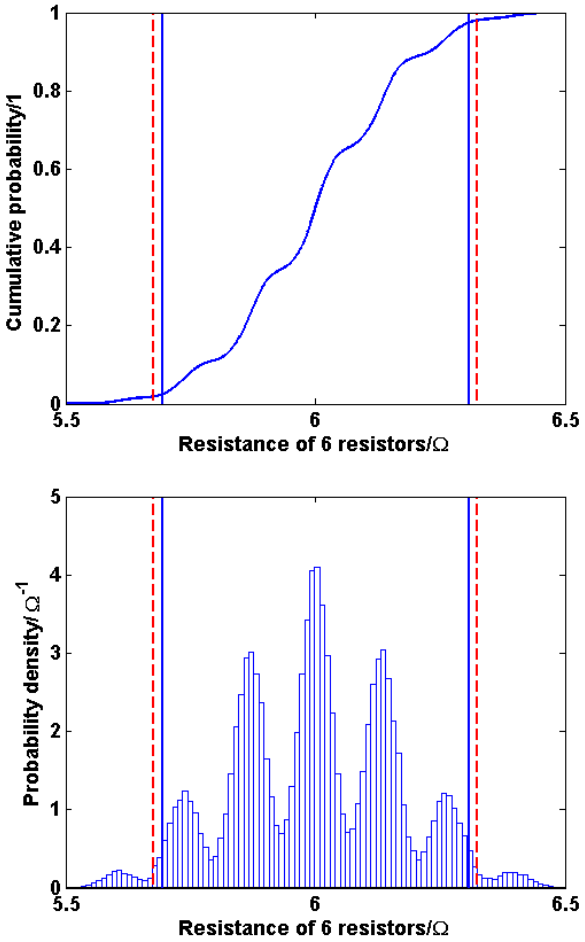

Figure 9.6: Distribution function and (below) PDF for the resistance of six Grade-C resistors in series.

tion $E_{\rm X}$ of the DMM [38] is

$$E_{\rm X} = V_{\rm iX} - V_{\rm S} + \delta V_{\rm iX} - \delta V_{\rm S},$$

where the input quantities and the PDFs used to characterize them are defined as follows:

DMM reading $V_{\rm iX}$. The voltage indicated by the DMM (the index 'i' meaning 'indication'). Because of the limited resolution of the device, no scatter is observed in the indicated values. Therefore, the indicated voltage, $100.1\,\mathrm{V}$, at the calibrator setting of $100\,\mathrm{V}$, is taken as exact.

Voltage $V_{\rm S}$ generated by the calibrator. The calibration certificate for the calibrator states that the voltage generated is the value indicated by the calibrator setting and that the expanded uncertainty associated with the $100\,\mathrm{V}$ setting is $U = 0.002\,\mathrm{V}$ with a coverage factor of $k = 2$. In the absence of other knowledge, this input quantity with expectation $100\,\mathrm{V}$ and standard deviation $0.001\,\mathrm{V}$ (obtained from $U/k = 0.002/2$) is characterized by a Gaussian distribution.

Correction $\delta V_{\rm iX}$ of the indicated voltage of the DMM. The least significant digit of the DMM display corresponds to $0.1\,\mathrm{V}$ as a consequence of the finite resolution of the instrument. The correction therefore lies in the interval $\pm 0.05\,\mathrm{V}$, with best estimate $0\,\mathrm{V}$. In the absence of other knowledge, this input quantity with expectation $0\,\mathrm{V}$ and standard deviation $0.029\,\mathrm{V}$ (obtained from $0.05/\sqrt{3}$) is characterized by a rectangular distribution.

Correction $\delta V_{\rm S}$ of the calibrator voltage. The calibrator voltage is in principle corrected for a number of effects including drift, mains power deviations and loading. An analysis [38] states that the correction lies in the interval $\pm 0.011\,\mathrm{V}$. In the absence of other knowledge, this input quantity with expectation $0\,\mathrm{V}$ and standard deviation $0.006\,4\,\mathrm{V}$ (obtained from $0.011/\sqrt{3}$) is characterized by a rectangular distribution. (Since this correction is based on a number of effects, it does not seem reasonable to regard the correction as equally likely to take any value in this interval. However, since the effect of this input quantity on the output quantity is relatively small, and since an intention is to compare the EA approach with that of a Monte Carlo method, the rectangular form is taken.)

This measurement function was analyzed using a Monte Carlo method (employing 10^5 trials). The error of indication of the DMM was found to be $0.100\,\mathrm{V}$ with a $95\,\%$ coverage interval of [0.050, 0.151] V. The corresponding result obtained by an approximate analytical treatment [38] was $(0.10 \pm 0.05)\,\mathrm{V}$, i.e., in agreement to the digits quoted.

Figure 9.7 shows the distribution function and PDF for the error of indication of the DMM obtained using a Monte Carlo method. The PDF is essentially trapezoidal in shape, to be compared with the statement [38], made following an approximate analytical treatment, that the distribution is essentially rectangular. The endpoints of the $95\,\%$ coverage interval, defined by the 2.5- and 97.5-percentiles of the distribution, are indicated in this figure by vertical lines.

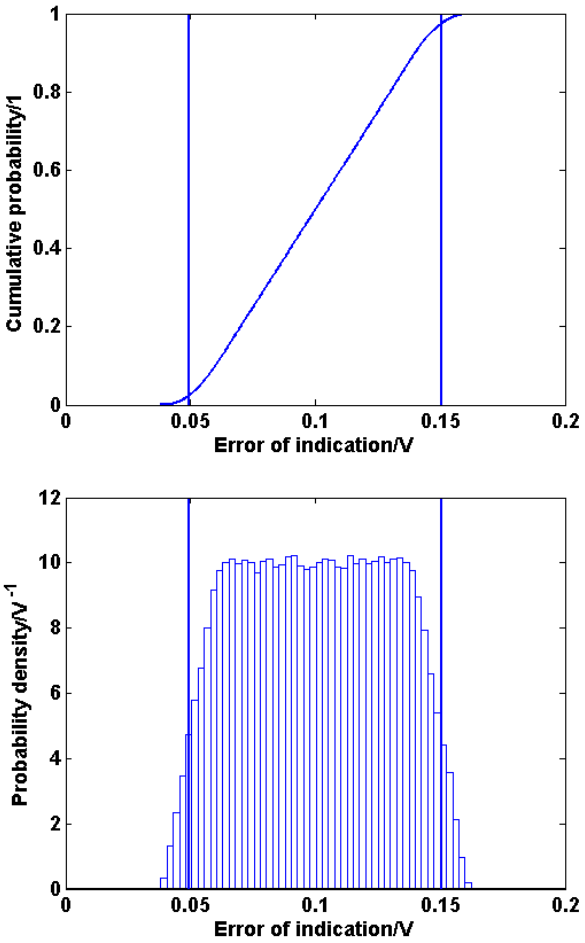

Figure 9.7: Distribution function and (below) PDF for the error of indication of a DMM.

9.6 Sides of a right-angled triangle

The sides of a right-angled triangle are repeatedly measured with a length-measuring instrument. Measurement is influenced by random and systematic errors. Use all the measured values to estimate the sides of the triangle and evaluate the associated uncertainties.

Denote the shorter sides of the triangle by A and B and the hypotenuse by H. Let there be n_A measurements of A, n_B of B and n_H of H. Let the quantity representing the ith measured value of A be A_i, with deviation ΔA_i from A, and similarly for B and H. According to Pythagoras' theorem, the sides are *physically* related by

$$A^2 + B^2 = H^2. \tag{9.4}$$

For consistency, the solution values (estimates) of A, B and H are to satisfy this condition. The deviations $\Delta A_i = A_i - A$, etc. are *statistically* related because the instrumental systematic error will manifest itself in all these values. Its 'presence' means that the quantities of which these deviations are realizations are correlated. In order to quantify this correlation, it is conventionally necessary to know the standard uncertainty $u(\Delta L)$ associated with the instrumental systematic error and the standard uncertainty $u_{\rm rep}$ associated with measurement repeatability.

A covariance matrix based on this information can be established and solution values obtained by solving a least-squares problem taking account of the covariance matrix. The covariance matrix, of dimension $(n_A + n_B + n_H) \times (n_A + n_B + n_H)$, is built from

1. $V(\Delta A_i) = V(\Delta B_i) = V(\Delta H_i) = u^2_{\rm rep} + u^2(\Delta L)$
2. All covariances are equal to $u^2(\Delta L)$.

The standard uncertainty $u(\Delta L)$ associated with the instrumental systematic error may not be available explicitly from the calibration certificate of the instrument, but should be part of the detailed uncertainty budget for the calibration. Generic details of the approach, *Gauss-Markov estimation*, are available [3]. Formally, the measurement model is in the form of a measurement function

$$\boldsymbol{Y} \equiv (A,\ B)^\top = \boldsymbol{f}(A_1, \ldots, A_{n_A}, B_1, \ldots, B_{n_B}, H_1, \ldots, H_{n_H}).$$

The quantities A_i, B_i and H_i that are measured are the input quantities ($n_A + n_B + n_H$ in number) and A and B are the (two) output quantities. The third side, H, the hypotenuse of the triangle, is not included as an output quantity, since it can be formed from A and B using equation (9.4). The measurement function $\boldsymbol{f}$ cannot, at least conveniently, be written down mathematically, but is defined by the computational procedure that implements the least-squares solution process.

Propagation of the covariance matrix associated with the measured values through the measurement function to provide the covariance matrix associated with estimates of $(A,\ B)^\top$

can be carried out as discussed in Chapter 6. The use of equation (9.4) as a 'next-stage' model (cf. Section 4.2.6), providing the output quantity H in terms of (input quantities) A and B, can then be used to evaluate the uncertainty associated with an estimate of H. The results can be combined to provide the covariance matrix associated with estimates of $(A,\ B,\ H)^\top$.

Statistical modelling can alternatively be used to provide the required estimates and the associated standard uncertainties without having to work with mutually dependent quantities and, in this instance, without prior knowledge of the above standard uncertainties. Regard the error effect as an unknown deviation ΔL and write

$$\Delta A_i = \Delta L + \delta A_i,$$

etc., where δA_i is the random deviation associated with A_i, etc. The δA_i, etc. are mutually independent, the associated covariance matrix being diagonal with all entries equal to u_{rep}^2. Best estimates of the sides (and ΔL) are then given by an ordinary least-squares problem (Gauss estimation). First, it is necessary to incorporate the condition (9.4). There are various ways to do so in general, but here it is simplest to use the condition to eliminate a variable. One possibility is to replace H by $(A^2 + B^2)^{1/2}$ or, letting θ denote the angle between sides A and H, set

$$A = H\cos\theta \tag{9.5}$$

and

$$B = H\sin\theta. \tag{9.6}$$

The latter choice gives the least-squares formulation

$$\begin{aligned}\min_{H,\theta,\Delta L} S = &\sum_{i=1}^{n_A}\left(\frac{a_i - H\cos\theta - \Delta L}{u_{\text{rep}}}\right)^2 \\ &+ \sum_{i=1}^{n_B}\left(\frac{b_i - H\sin\theta - \Delta L}{u_{\text{rep}}}\right)^2 \\ &+ \sum_{i=1}^{n_H}\left(\frac{h_i - H - \Delta L}{u_{\text{rep}}}\right)^2\end{aligned}$$

in terms of measured values (estimates) a_i, b_i and h_i, respectively, of A_i, B_i and H_i. Its solution could be found using the Gauss-Newton algorithm or one of its variants [3]. However, advantage can be taken as follows of the fact that the problem is linear in two of the unknowns, H and ΔL. Equate to zero the partial derivatives of S, with respect to H, θ and ΔL, to give three algebraic equations. Eliminate H and ΔL to give a single nonlinear equation in θ, and solve this equation using a suitable 'zero-finder'. Finally, determine H and ΔL by substitution.

Many such problems would be solved in this manner. In this particular case, by defining transformed parameters

$$V_1 = H\cos\theta + \Delta L, \quad V_2 = H\sin\theta + \Delta L, \quad V_3 = H + \Delta L \tag{9.7}$$

and $\boldsymbol{V} = (V_1,\ V_2,\ V_3)^\top$, the problem becomes

$$\min_{\boldsymbol{V}} \left[\sum_{i=1}^{n_A} (a_i - V_1)^2 + \sum_{i=1}^{n_B} (b_i - V_2)^2 + \sum_{i=1}^{n_H} (h_i - V_3)^2 \right].$$

The problem separates into three trivial independent minimization problems, giving the solution

$$v_1 = \overline{a} = \frac{1}{n_A} \sum_{i=1}^{n_A} a_i,$$

and similarly $v_2 = \overline{b}$ and $v_3 = \overline{h}$.

In terms of the estimate $\boldsymbol{v}$ of $\boldsymbol{V}$, the equations (9.7) can then straightforwardly be solved for estimates of H, ΔL and θ, from which estimates of A and B are determined from formulae (9.5) and (9.6). The associated standard uncertainties and covariance matrices are readily obtained using the principles of Chapter 6.

Since the value of u_{rep}^2 is common to all terms in the sum, the minimizing values of H, θ and ΔL do not depend on it. The term may therefore be replaced by unity (or any other constant). Thus, the solution can be obtained without knowledge of the uncertainty associated with random repeatability or that associated with the systematic instrumental effect.

9.7 Fourier transform

Consider the measurement of a periodic phenomenon. Such a measurement is commonplace in many branches of metrology. Suppose a complete period is measured, with N quantities $\boldsymbol{X} = (X_1, \ldots, X_N)^\top$ available. These quantities correspond to the uniformly spaced angles $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_N)^\top$, where $\theta_i = 2\pi(i-1)/N$.

A Fourier transform of such data provides information concerning the frequency content of the data. Each Fourier coefficient depends on all (or most of) the X_i, regarded as the input quantities, and is a linear combination of them.

Suppose that measured values are obtained independently with associated standard uncertainty σ. The covariance matrix associated with estimates x_i of the input quantities X_i is therefore given by

$$\boldsymbol{U}_{\boldsymbol{x}} = \sigma^2 \boldsymbol{I}, \tag{9.8}$$

where $\boldsymbol{I}$ is the identity matrix of dimension $N \times N$. It is required to evaluate the uncertainties associated with the Fourier transform of this data, i.e., associated with estimates of the coefficients of the Fourier representation of the data. The coefficients constitute the (vector) output quantity.

The Fourier *representation* of the data is

$$h(\theta) = a_0 + a_1 \cos\theta + b_1 \sin\theta + \cdots + a_r \cos r\theta + b_r \sin r\theta,$$

where $r = \lfloor N/2 \rfloor$. (When N is even, the coefficient b_r of $\sin r\theta$ is in fact zero.) Let $\boldsymbol{Y} = (Y_1, \ldots, Y_{2r+1})^\top \equiv (a_0, a_1, b_1, \ldots, a_r, b_r)^\top$, denote the output quantities. The Fourier transform $\boldsymbol{Y}$ of $\boldsymbol{X}$ is then given implicitly by

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{Y}, \tag{9.9}$$

where

$$\boldsymbol{A} = \begin{bmatrix} 1 & \cos\theta_1 & \sin\theta_1 & \ldots & \cos r\theta_1 & \sin r\theta_1 \\ 1 & \cos\theta_2 & \sin\theta_2 & \ldots & \cos r\theta_2 & \sin r\theta_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \cos\theta_N & \sin\theta_N & \ldots & \cos r\theta_N & \sin r\theta_N \end{bmatrix}$$

is the matrix of dimension $N \times N$ of Fourier basis-function values. Formally, the Fourier coefficients are given in terms of the data using

$$\boldsymbol{Y} = \boldsymbol{A}^{-1}\boldsymbol{X} \tag{9.10}$$

or, equivalently, from a formula that expresses the Y_i as linear combinations of the X_i, where the multipliers are sine and cosine terms. In practice, $\boldsymbol{Y}$ would be computed from $\boldsymbol{X}$ using the fast Fourier transform (FFT) [14]. The FFT gives far greater efficiency than would be obtained from the application of general-purpose linear-algebra techniques, and also greater numerical accuracy. In exact arithmetic, the FFT and the expression (9.10) give identical results, since mathematically they are equivalent ways of expressing the solution.

Denote the covariance matrix associated with estimates $\boldsymbol{y}$ of $\boldsymbol{Y}$ by $\boldsymbol{U}_{\boldsymbol{y}}$. The application of the law of propagation of uncertainty (Chapter 6) to the relationship (9.9) gives

$$\boldsymbol{U}_{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{U}_{\boldsymbol{y}}\boldsymbol{A}^\top.$$

This result is *exact* since the output quantity $\boldsymbol{Y}$ and the input quantity $\boldsymbol{X}$ are related linearly through the relationship (9.9), and linearization introduces no error in this case. Since $\boldsymbol{A}$ is invertible,

$$\boldsymbol{U}_{\boldsymbol{y}} = \boldsymbol{A}^{-1}\boldsymbol{U}_{\boldsymbol{x}}\boldsymbol{A}^{-\top}.$$

This expression enables in general the uncertainty associated with the Fourier transform to be computed from that associated with the data. As a consequence of expression (9.8),

$$\boldsymbol{U}_{\boldsymbol{y}} = \sigma^2\boldsymbol{A}^{-1}\boldsymbol{A}^{-\top} = \sigma^2\left(\boldsymbol{A}^\top\boldsymbol{A}\right)^{-1}.$$

Now, using the fact that the elements of $\boldsymbol{\theta}$ are equiangular and the fundamental properties of the trigonometric functions, it is straightforward to show that

$$\boldsymbol{A}^\top\boldsymbol{A} = \frac{N}{2}\text{diag}\left\{\ 2,\quad 1,\quad \ldots,\quad 1\ \right\},$$

giving

$$(\boldsymbol{A}^\top\boldsymbol{A})^{-1} = \frac{2}{N}\text{diag}\left\{\ \tfrac{1}{2},\quad 1,\quad \ldots,\quad 1\ \right\}.$$

Consequently,

$$\boldsymbol{U_y} = \frac{2}{N}\sigma^2 \text{diag}\left\{ \begin{array}{cccc} \frac{1}{2}, & 1, & \ldots, & 1 \end{array} \right\}.$$

This result states that for measured values that are obtained independently with associated standard uncertainty σ, the Fourier coefficients are (also) realizations of independent quantities, with associated standard uncertainty equal to σ scaled by the factor $\sqrt{2/N}$, where N is the number of measured values (with the exception of the constant term for which the factor is $\sqrt{1/N}$). Moreover, each Fourier coefficient is a linear combination of N measured values, the multipliers being the products of a constant value and that of values of cosines and sines (and thus lying between -1 and $+1$). Consequently, if N is large, regardless of the statistical distributions of the quantites of which the data are realizations, the Fourier coefficients can be expected to be very close to realizations of normally distributed quantities. This result is an immediate consequence of the central limit theorem when using the Fourier transform to analyse large numbers of measured values obtained independently. Thus, it is valid to regard the resulting Fourier coefficients as if they were realizations of *independent Gaussian-distributed quantities*.

The output quantities, the Fourier coefficients, from this process become the input quantities to a subsequent stage, viz., the evaluation of the Fourier series $h(\theta)$ for any value of θ. Now, since, as shown, the Fourier coefficients are mutually independent,

$$u^2(h(\theta)) = u^2(a_0) + u^2(a_1)\cos^2\theta + u^2(b_1)\sin^2\theta \cdots + u^2(a_r)\cos^2 r\theta + u^2(b_r)\sin^2 r\theta.$$

Using the results above,

$$u^2(h(\theta)) = \frac{\sigma^2}{N} + \frac{2\sigma^2}{N}\cos^2\theta + \frac{2\sigma^2}{N}\sin^2\theta + \ldots + \frac{2\sigma^2}{N}\cos^2 r\theta + \frac{2\sigma^2}{N}\sin^2 r\theta, \quad (9.11)$$

which simplifies to σ^2. Thus,

$$u(h(\theta)) = \sigma,$$

i.e., the standard uncertainty associated with the Fourier representation of a data set when evaluated at any point is identical to the standard uncertainty associated with the data itself. This property is remarkable in that the (interpolatory) replacement of data by other functions usually gives an amplification of the raw data uncertainty, at least in some regions of the data.

9.8 Mass calibration

Consider the calibration of a weight W of mass density ρ_{W} against a reference weight R of mass density ρ_{R} having nominally the same mass, using a balance operating in air of mass density ρ_{a} [78]. Since ρ_{W} and ρ_{R} are generally different, it is necessary to account for buoyancy effects. Applying Archimedes' principle, the measurement model takes the form

$$m_{\mathrm{W}}(1 - \rho_{\mathrm{a}}/\rho_{\mathrm{W}}) = (m_{\mathrm{R}} + \delta m_{\mathrm{R}})(1 - \rho_{\mathrm{a}}/\rho_{\mathrm{R}}), \quad (9.12)$$

where δm_{R} is the mass of a small weight of density ρ_{R} added to R to balance it with W.

It is usual to work in terms of conventional masses. The conventional mass $m_{\mathrm{W,c}}$ of W is the mass of a (hypothetical) weight of density $\rho_0 = 8\,000\,\mathrm{kg/m^3}$ that balances W in air at density $\rho_{\mathrm{a}_0} = 1.2\,\mathrm{kg/m^3}$. Thus,

$$m_{\mathrm{W}}(1 - \rho_{\mathrm{a}_0}/\rho_{\mathrm{W}}) = m_{\mathrm{W,c}}(1 - \rho_{\mathrm{a}_0}/\rho_0).$$

In terms of conventional masses $m_{\mathrm{W,c}}$, $m_{\mathrm{R,c}}$ and $\delta m_{\mathrm{R,c}}$, the measurement model (9.12) becomes

$$m_{\mathrm{W,c}}(1-\rho_{\mathrm{a}}/\rho_{\mathrm{W}})(1-\rho_{\mathrm{a}_0}/\rho_{\mathrm{W}})^{-1} = (m_{\mathrm{R,c}}+\delta m_{\mathrm{R,c}})(1-\rho_{\mathrm{a}}/\rho_{\mathrm{R}})(1-\rho_{\mathrm{a}_0}/\rho_{\mathrm{R}})^{-1}, \quad (9.13)$$

from which, to an approximation adequate for most practical purposes,

$$m_{\mathrm{W,c}} = (m_{\mathrm{R,c}} + \delta m_{\mathrm{R,c}}) \left[1 + (\rho_{\mathrm{a}} - \rho_{\mathrm{a}_0}) \left(\frac{1}{\rho_{\mathrm{W}}} - \frac{1}{\rho_{\mathrm{R}}}\right)\right].$$

Let

$$\delta m = m_{\mathrm{W,c}} - m_{\mathrm{nom}}$$

be the deviation of $m_{\mathrm{W,c}}$ from the nominal mass

$$m_{\mathrm{nom}} = 100\,\mathrm{g}.$$

The measurement function used in this example is given by

$$\delta m = (m_{\mathrm{R,c}} + \delta m_{\mathrm{R,c}}) \left[1 + (\rho_{\mathrm{a}} - \rho_{\mathrm{a}_0}) \left(\frac{1}{\rho_{\mathrm{W}}} - \frac{1}{\rho_{\mathrm{R}}}\right)\right] - m_{\mathrm{nom}}. \quad (9.14)$$

The only information available concerning $m_{\mathrm{R,c}}$ and $\delta m_{\mathrm{R,c}}$ is an estimate and an associated standard uncertainty for each of these quantities. Accordingly, each of these quantities is characterized by a Gaussian distribution, with these estimates used as the expectations of the corresponding quantities and the associated standard uncertainties as the standard deviations [11, Clause 6.4.7.1]. The only information available concerning ρ_{a}, ρ_{W} and ρ_{R} is lower and upper limits for each of these quantities. Accordingly, each of these quantities is characterized by a rectangular distribution, with these limits used as the endpoints of the distribution [11, Clause 6.4.2.1]. The quantity ρ_{a_0} in the measurement function (9.14) is assigned the value $1.2\,\mathrm{kg/m^3}$ with no associated uncertainty.

Table 9.4 summarizes the input quantities and the PDFs used to characterize them. In the table, a Gaussian distribution $\mathrm{N}(\mu, \sigma^2)$ is described in terms of expectation μ and standard deviation σ, and a rectangular distribution $\mathrm{R}(a, b)$ with endpoints a and b ($a < b$) in terms of expectation $(a + b)/2$ and semi-width $(b - a)/2$.

X_i	Distribution	Parameters			
		Expectation μ /mg	Standard deviation σ /mg	Expectation $x = (a+b)/2$ /kg/m^3	Semi-width $(b-a)/2$ /kg/m^3
$m_{\mathrm{R,c}}$	$\mathrm{N}(\mu, \sigma^2)$	100 000.000	0.050		
$\delta m_{\mathrm{R,c}}$	$\mathrm{N}(\mu, \sigma^2)$	1.234	0.020		
ρ_{a}	$\mathrm{R}(a, b)$			1.20	0.10
ρ_{W}	$\mathrm{R}(a, b)$			8×10^3	1×10^3
ρ_{R}	$\mathrm{R}(a, b)$			8.00×10^3	0.05×10^3

Table 9.4: The input quantities X_i and the PDFs used to characterize them for the measurement function (9.14).

9.8.1 Propagation of distributions

The GUM uncertainty framework and the adaptive Monte Carlo procedure (Section 7.2.5) were each used to obtain an estimate $\widehat{\delta m}$ of δm, the associated standard uncertainty $u(\widehat{\delta m})$, and the shortest $95\,\%$ coverage interval for δm. The results obtained are shown in Table 9.5, in which GUF_1 denotes the GUM uncertainty framework with first-order terms, MCM the adaptive Monte Carlo procedure, and GUF_2 the GUM uncertainty framework with higher-order terms.

0.72×10^6 trials were taken by the adaptive Monte Carlo procedure for a degree of approximation of 0.001 required in $u(\widehat{\delta m})$ (Section 7.2.5). The chosen numerical tolerance corresponds to a value of $\delta/5$ with δ set for the case where one significant decimal digit in $u(\widehat{\delta m})$ is regarded as meaningful (Chapter 8 and below).

Figure 9.8 shows the approximations to the distribution function and the PDF for δm obtained from the GUM uncertainty framework with first-order terms and a Monte Carlo method. The continuous curve represents a Gaussian PDF with parameters given by the GUM uncertainty framework. The inner pair of (broken) vertical lines indicates the shortest $95\,\%$ coverage interval for δm based on this PDF. The histogram is the scaled frequency distribution obtained using a Monte Carlo method as an approximation to the PDF. The outer pair of (continuous) vertical lines indicates the shortest $95\,\%$ coverage interval for δm based on the discrete representation of the distribution function provided by a Monte Carlo method.

The results show that, although the GUM uncertainty framework (first order) and a Monte Carlo method give estimates of δm in good agreement, the numerical values for the associated standard uncertainty are noticeably different. The value ($0.075\,4\,\mathrm{mg}$) of $u(\widehat{\delta m})$ returned by a Monte Carlo method is $40\,\%$ larger than that ($0.053\,9\,\mathrm{mg}$) returned by the GUM uncertainty framework (first order). The latter is thus optimistic in this respect. There is good agreement between $u(\widehat{\delta m})$ determined by a Monte Carlo method and that ($0.075\,0\,\mathrm{mg}$) provided by the GUM uncertainty framework with higher-order terms.

Method	$\widehat{\delta m}$ /mg	$u(\widehat{\delta m})$ /mg	Shortest 95 % coverage interval/mg	d_{low} /mg	d_{high} /mg	GUF validated ($\delta = 0.005$)?
GUF_1	1.234 0	0.053 9	[1.128 4, 1.339 6]	0.045 3	0.042 6	No
MCM	1.234 1	0.075 4	[1.083 1, 1.382 2]			
GUF_2	1.234 0	0.075 0	[1.087 0, 1.381 0]	0.003 9	0.001 2	Yes

Table 9.5: Results of the calculation stage for the measurement function (9.14).

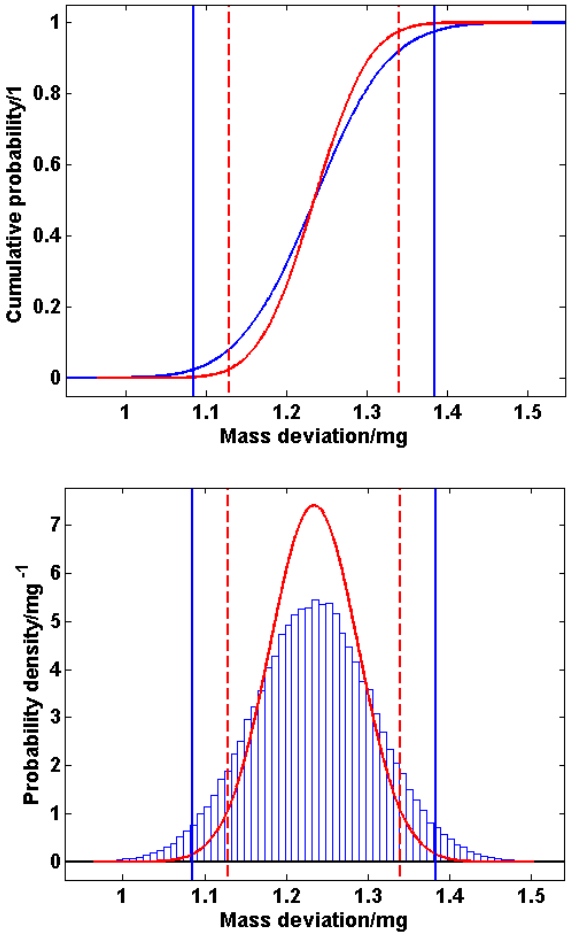

Figure 9.8: Approximations to the distribution function and (below) PDF for the output quantity δm in the measurement function (9.14) obtained using the GUM uncertainty framework (first order) and a Monte Carlo method.

X_i	Partial derivative	Sensitivity coefficient
$m_{\mathrm{R,c}}$	$1 + (\rho_{\mathrm{a}} - \rho_{\mathrm{a}_0})(1/\rho_{\mathrm{W}} - 1/\rho_{\mathrm{R}})$	1
$\delta m_{\mathrm{R,c}}$	$1 + (\rho_{\mathrm{a}} - \rho_{\mathrm{a}_0})(1/\rho_{\mathrm{W}} - 1/\rho_{\mathrm{R}})$	1
ρ_{a}	$(m_{\mathrm{R,c}} + \delta m_{\mathrm{R,c}})(1/\rho_{\mathrm{W}} - 1/\rho_{\mathrm{R}})$	0
ρ_{W}	$-(m_{\mathrm{R,c}} + \delta m_{\mathrm{R,c}})(\rho_{\mathrm{a}} - \rho_{\mathrm{a}_0})/\rho_{\mathrm{W}}^2$	0
ρ_{R}	$(m_{\mathrm{R,c}} + \delta m_{\mathrm{R,c}})(\rho_{\mathrm{a}} - \rho_{\mathrm{a}_0})/\rho_{\mathrm{R}}^2$	0

Table 9.6: Sensitivity coefficients for the measurement function (9.14).

Table 9.6 contains the partial derivatives of first order for the measurement function (9.14) with respect to the input quantities together with the sensitivity coefficients, viz. these derivatives evaluated at the best estimates of the input quantities. These derivatives indicate that, for the purposes of the GUM uncertainty framework with first-order terms, the measurement function for this example can be considered as being replaced by the additive measurement function

$$\delta m = m_{\mathrm{R,c}} + \delta m_{\mathrm{R,c}} - m_{\mathrm{nom}}.$$

A Monte Carlo method makes no such (implied) approximation to the measurement function.

Table 9.5 also shows in the right-most three columns the results of applying the validation procedure of Chapter 8 in the case where one significant decimal digit in $u(\widehat{\delta m})$ is regarded as meaningful. Using the terminology of that Chapter, $n_{\mathrm{dig}} = 1$, since a numerical tolerance of one significant decimal digit in $u(\widehat{\delta m})$ is required. Hence, $u(\widehat{\delta m}) = 0.08 = 8 \times 10^{-2}$, and so $a = 8$ and $r = -2$. Thus $\delta = 1/2 \times 10^{-2} = 0.005$. d_{low} and d_{high} denote the magnitudes of the endpoint differences (8.1) and (8.2), where y there corresponds to $\widehat{\delta m}$. Whether the results were validated to one significant decimal digit in $u(\widehat{\delta m})$ is indicated in the final column of the table. If only first-order terms are accounted for, the application of the GUM uncertainty framework is not validated. If higher-order terms are accounted for [10, Clause 5.1.2 note], the GUM uncertainty framework is validated. Thus, the non-linearity of the measurement function is such that accounting for first-order terms only is inadequate.

9.8.2 Bayesian analysis

To apply a Bayesian analysis to this problem, the measurement function (9.14) is written in the equivalent form

$$\delta m_{\mathrm{R,c}} = (\delta m + m_{\mathrm{nom}})/[1 + (\rho_{\mathrm{a}} - \rho_{\mathrm{a}_0})(1/\rho_{\mathrm{W}} - 1/\rho_{\mathrm{R}})] - m_{\mathrm{R,c}}, \tag{9.15}$$

which is an observation model [80] that expresses the indication quantity $W \equiv \delta m_{\mathrm{R,c}}$ in terms of the measurand $Y \equiv \delta m$ and other (input) quantities $\boldsymbol{Z} \equiv (m_{\mathrm{R,c}}, \rho_{\mathrm{a}}, \rho_{\mathrm{W}}, \rho_{\mathrm{R}})^{\top}$ involved in the measurement (see Section 3.3).

The information given in Table 9.4 is used to define the likelihood function for $\delta m_{\mathrm{R,c}}$ given the indication value of $1.234\,\mathrm{mg}$ and prior PDFs for the components of $\boldsymbol{Z}$. It is necessary also to define a prior distribution for the measurand δm, which is taken to be the Gaussian distribution $\mathrm{N}(\delta m_0, u^2(\delta m_0))$ specified by the prior estimate $\delta m_0 = 0\,\mathrm{mg}$ of δm with associated standard uncertainty (or prior standard deviation) $u(\delta m_0)$. Three cases are considered in which $u(\delta m_0)$ is set equal to $10\,\mathrm{mg}$, $1\,\mathrm{mg}$ and $0.5\,\mathrm{mg}$, respectively. In all cases the prior standard deviation is large relative to the standard deviation of $0.020\,\mathrm{mg}$ used to define the likelihood function (Table 9.4).

In each case, an approximation to the posterior PDF for δm was determined using the WinBUGS software package [105], and an estimate, associated standard uncertainty and the shortest $95\,\%$ coverage interval for δm evaluated in terms of that approximation. The results obtained are shown in Table 9.7 and Figures 9.9 to 9.11. In each figure is shown as a scaled frequency distribution (histogram) the approximation to the posterior PDF for δm obtained using a Bayesian analysis. The broken and continuous curves represent the Gaussian PDFs obtained by applying the GUM uncertainty framework (with first and higher order terms, respectively) as an approach to the propagation of distributions.

The results show that when the prior PDF expresses 'weak' information about the measurand, in terms of a large prior standard deviation, the posterior PDF obtained is not very different to that obtained from the propagation of distributions (implemented using the GUM uncertainty framework with higher order terms or, indeed, a Monte Carlo method). However, as the prior standard deviation becomes smaller there is a noticeable influence of the prior PDF on the posterior PDF, which exhibits asymmetry and, consequently, on the coverage interval, which 'shifts' towards zero, the prior estimate of the measurand. In these cases, the posterior PDF is a compromise between the information expressed by the prior PDF and that provided by the 'data'.

When (proper) prior information is available about the measurand δm, the Bayesian analysis described here will provide a PDF for δm that is generally different from that delivered by the propagation of distributions (as described in Section 9.8.1), but will correctly take account of that information. Furthermore, since δm does not depend linearly on the (indication and input) quantities $\delta m_{\mathrm{R,c}}$ and $(m_{\mathrm{R,c}}, \rho_{\mathrm{a}}, \rho_{\mathrm{W}}, \rho_{\mathrm{R}})^{\top}$, the PDFs can also be expected to be different when, in the Bayesian analysis, a non-informative prior distribution of the form $\delta m \propto 1$ is used [42]. The propagation of distributions can be used to provide a PDF for δm in the case that no (informative) prior information is available about δm. A discussion and comparison of different approaches applied to this problem is available [97].

9.9 Comparison loss in microwave power meter calibration

During the calibration of a microwave power meter, the power meter and a standard power meter are connected in turn to a stable signal generator. The power absorbed by each meter will in general be different because their complex input voltage reflection coefficients are

Prior standard deviation /mg	$\widehat{\delta m}$ /mg	$u(\widehat{\delta m})$ /mg	Shortest 95 % coverage interval/mg
10	1.233 9	0.076 1	[1.082 0, 1.382 0]
1	1.225 2	0.075 0	[1.080 0, 1.376 0]
0.5	1.207 1	0.075 4	[1.058 5, 1.355 0]

Table 9.7: Results of a Bayesian analysis applied to a problem of mass calibration for different choice of prior PDF for the measurand.

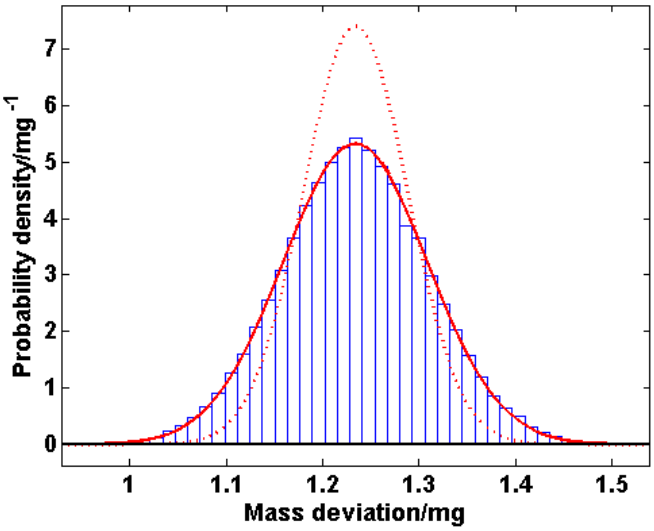

Figure 9.9: Approximation to the posterior PDF for the measurand δm in the observation model (9.15) obtained using a Bayesian analysis with prior PDF $\mathrm{N}(100\,000\,\mathrm{mg}, (10\,\mathrm{mg})^2)$. The approximations to the PDF for δm obtained using the GUM uncertainty framework (first and higher order) as an approach to the propagation of distributions are also shown.

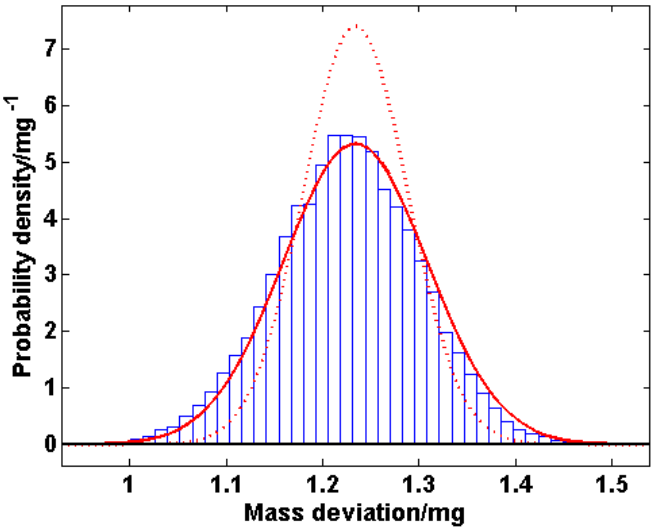

Figure 9.10: As Figure 9.9 except with prior PDF $\mathrm{N}(100\,000\,\mathrm{mg}, (1\,\mathrm{mg})^2)$.

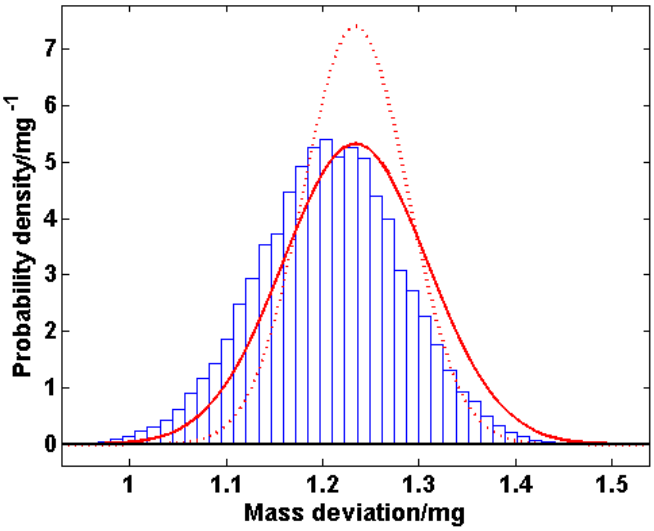

Figure 9.11: As Figure 9.9 except with prior PDF $\mathrm{N}(100\,000\,\mathrm{mg}, (0.5\,\mathrm{mg})^2)$.

not identical. The ratio Y of the power $P_{\rm M}$ absorbed by the meter being calibrated and that, $P_{\rm S}$, by the standard meter is [85]

$$Y = \frac{P_{\rm M}}{P_{\rm S}} = \frac{1-|\Gamma_{\rm M}|^2}{1-|\Gamma_{\rm S}|^2} \times \frac{|1-\Gamma_{\rm S}\Gamma_{\rm G}|^2}{|1-\Gamma_{\rm M}\Gamma_{\rm G}|^2}, \tag{9.16}$$

where $\Gamma_{\rm G}$ is the voltage reflection coefficient of the signal generator, $\Gamma_{\rm M}$ that of the meter being calibrated and $\Gamma_{\rm S}$ that of the standard meter. This power ratio is an instance of 'comparison loss' [6, 63].

Consider the case where the standard and the signal generator are reflectionless, so that $\Gamma_{\rm S} = \Gamma_{\rm G} = 0$, and measured values are obtained of the real and imaginary parts X_1 and X_2 of $\Gamma_{\rm M} = X_1 + {\rm j}X_2$, where ${\rm j}^2 = -1$. Since $|\Gamma_{\rm M}|^2 = X_1^2 + X_2^2$, formula (9.16) becomes

$$Y = 1 - X_1^2 - X_2^2. \tag{9.17}$$

Given are best estimates x_1 and x_2 of the quantities X_1 and X_2 from measurement and the associated standard uncertainties $u(x_1)$ and $u(x_2)$. X_1 and X_2 are often not independent. Denote by $\mathrm{cov}(x_1, x_2)$ the covariance associated with x_1 and x_2. Equivalently [10, Clause 5.2.2], $\mathrm{cov}(x_1, x_2) = r(x_1, x_2)u(x_1)u(x_2)$, where $r = r(x_1, x_2)$ denotes the associated correlation coefficient. $\boldsymbol{X} = (X_1, X_2)^\top$, with expectation and covariance matrix

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \qquad \begin{bmatrix} u^2(x_1) & ru(x_1)u(x_2) \\ ru(x_1)u(x_2) & u^2(x_2) \end{bmatrix}, \tag{9.18}$$

is characterized by a bivariate Gaussian PDF [11, Clause 6.4.8.1]. Because the magnitudes of X_1 and X_2 in expression (9.17) are in practice small compared with unity, the resulting Y is close to unity. Results are accordingly expressed in terms of the quantity

$$\delta Y = 1 - Y = X_1^2 + X_2^2, \tag{9.19}$$

taken as the measurement function. For physical reasons, $0 \le Y \le 1$, and so $0 \le \delta Y \le 1$.[1]

The determination of an estimate δy of δY, the associated standard uncertainty $u(\delta y)$, and a coverage interval for δY are considered for choices of x_1, x_2, $u(x_1)$, $u(x_2)$ and $r(x_1, x_2)$. All quantities have dimension one. Six cases are considered, in all of which x_2 is taken as zero and $u(x_1) = u(x_2) = 0.005$. The first three of these cases correspond to taking $x_1 = 0$, 0.010, and 0.050, each with $r(x_1, x_2) = 0$. The other three cases correspond to taking the same x_1, but with $r(x_1, x_2) = 0.9$. The various numerical values of x_1 (comparable to those occurring in practice) are used to investigate the extent to which the results obtained using the considered approaches differ. For the cases in which $r = r(x_1, x_2) = 0$, the covariance matrix given in Formula (9.18) reduces to $\mathrm{diag}(u^2(x_1), u^2(x_2))$ and the corresponding joint distribution for X_1 and X_2 to the product of two univariate Gaussian distributions for X_i, for $i = 1, 2$, with expectation x_i and standard deviation $u(x_i)$.

[1] None of the approaches considered constrain the PDF for δY to be no greater than unity. However, for sufficiently small uncertainties $u(x_1)$ and $u(x_2)$, as here, the PDF for δY may adequately be approximated by a simpler PDF defined over all non-negative values of δY. A rigorous treatment, using Bayesian inference ([104] and Section 3.3), which applies regardless of the magnitudes of $u(x_1)$ and $u(x_2)$, is possible.

9.9.1 Zero covariance

The evaluation of uncertainty is treated by applying the propagation of distributions (a) analytically (for purposes of comparison), (b) using the GUM uncertainty framework, and (c) using a Monte Carlo method. δy and $u(\delta y)$ can generally be formed analytically as the expectation and standard deviation of δY, as characterized by the PDF for δY [11, Clause F.1]. The PDF for δY can be formed analytically when $x_1 = 0$ and, in particular, used to determine the endpoints of the shortest $95\,\%$ coverage interval in that case [11, Clause F.2]. The GUM uncertainty framework with first-order terms and with higher-order terms is applied for each of the three estimates x_1 in the uncorrelated case [11, Clause F.3]. An estimate δy of δY is formed in each case [11, Clause 4.1.4] from

$$\delta y = x_1^2 + x_2^2.$$

A Monte Carlo method is applied in each case with $M = 10^6$ trials.

Input estimate $x_1 = 0$

For the input estimate $x_1 = 0$, higher-order terms must be used when applying the law of propagation of uncertainty, because the partial derivatives of δY with respect to X_1 and X_2, evaluated at $X_1 = x_1$ and $X_2 = x_2$, are identically zero when $x_1 = x_2 = 0$. Thus, if the law of propagation of uncertainty with first-order terms only were applied, the resulting standard uncertainty would incorrectly be computed as zero.[2]

Figure 9.12 shows the PDFs for δY determined by applying the propagation of distributions (a) analytically (the exponentially decreasing curve for $\delta Y \geq 0$ and zero elsewhere), (b) using the GUM uncertainty framework with higher-order terms in order to characterize the output quantity by a Gaussian PDF (bell-shaped curve), and (c) using a Monte Carlo method (scaled frequency distribution). It is seen in the figure that the use of the GUM uncertainty framework with higher-order terms in order to characterize the output quantity by a Gaussian distribution yields a PDF that is very different from the analytic solution. The latter takes the form of a particular chi-squared distribution—the sum of squares of two standard Gaussian variables [11, Clause F.2]. Since the partial derivatives of the measurement function (9.19) of order higher than two are all identically zero, the solution obtained essentially corresponds to taking all Taylor-series terms, i.e. the full non-linearity of the problem, into account. Thus, the particular Gaussian distribution so determined is the best that is possible using the GUM uncertainty framework to characterize the output quantity by such a distribution. It can therefore be concluded that the reason for the departure from the analytic solution of the results of the use of the approach based on the GUM uncertainty framework is that the output quantity is characterized by a Gaussian PDF. No Gaussian PDF, however it is obtained, could adequately represent the analytic solution in this case. It is also seen in Figure 9.12 that the PDF provided by a Monte Carlo method is consistent with the analytic solution.

[2] A similar difficulty would arise for x_1 close to zero.

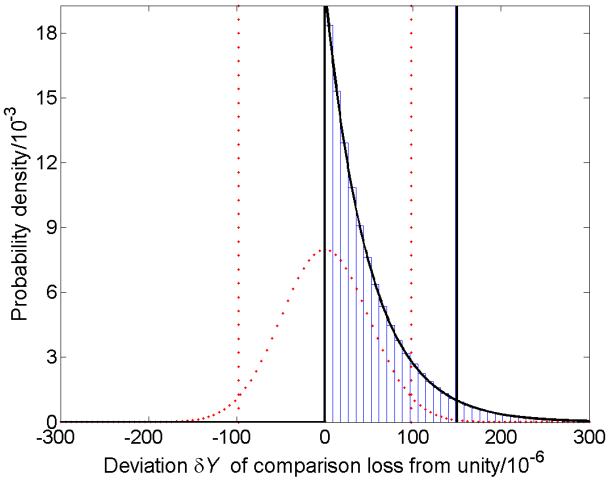

Figure 9.12: Results for the problem of comparison loss in power meter calibration in the case $x_1 = x_2 = 0$, with $u(x_1) = u(x_2) = 0.005$ and $r(x_1, x_2) = 0$.

The estimates δy determined as the expectation of δY described by the PDFs obtained (a) analytically, (b) using the GUM uncertainty framework, and (c) applying a Monte Carlo method are given in columns 2–4 of the row corresponding to $x_1 = 0.000$ in Table 9.8. Columns 5–8 contain the corresponding $u(\delta y)$, with those obtained using the GUM uncertainty framework with first-order terms (G_1) and higher-order terms (G_2).

x_1	Estimate $\delta y/10^{-6}$			Standard uncertainty $u(\delta y)/10^{-6}$			
	A	G	M	A	G_1	G_2	M
0.000	50	0	50	50	0	50	50
0.010	150	100	150	112	100	112	112
0.050	2 550	2 500	2 551	502	500	502	502

Table 9.8: Comparison loss results (estimates and associated standard uncertainties), for input estimates with associated zero covariance, obtained analytically (A), and using the GUM uncertainty framework (G) with first-order terms (G_1) and higher-order terms (G_2) and a Monte Carlo method (M).

The estimate $\delta y = 0$ obtained by evaluating the measurement function at the input estimates is invalid: the correct (analytic) $g_{\delta Y}(\eta)$ is identically zero for $\delta Y < 0$; this estimate lies on the boundary of the non-zero part of that function. The estimate provided by a Monte Carlo method agrees with that obtained analytically. The law of propagation of uncertainty based on first-order terms gives the wrong, zero, value for $u(\delta y)$ already noted. The value (50×10^{-6}) from the law of propagation of uncertainty based on higher-order

x_1	Shortest 95 % coverage interval for $\delta Y/10^{-6}$			
	A	G_1	G_2	M
0.000	[0, 150]	[0, 0]	[−98, 98]	[0, 150]
0.010	—	[−96, 296]	[−119, 319]	[0, 366]
0.050	—	[1 520, 3 480]	[1 515, 3 485]	[1 597, 3 551]

Table 9.9: Comparison loss results (endpoints of the shortest 95 % coverage interval), for input estimates with associated zero covariance, obtained analytically (A), and using the GUM uncertainty framework with first-order terms (G_1) and higher-order terms (G_2) and a Monte Carlo method (M).

terms agrees with that obtained analytically and from a Monte Carlo method.[3]

Figure 9.12 also shows the shortest 95 % coverage intervals for the corresponding approximations to the distribution function for δY. The 95 % coverage interval, indicated by dotted vertical lines, as provided by the GUM uncertainty framework is infeasible: it is symmetric about $\delta Y = 0$ and therefore erroneously implies there is a 50 % probability that δY is negative. The continuous vertical lines are the endpoints of the shortest 95 % coverage interval derived from the analytic solution [11, Clause F.2]. The endpoints of the shortest 95 % coverage interval determined using a Monte Carlo method are indistinguishable to graphical accuracy from those for the analytic solution. The endpoints of the shortest coverage intervals are given in columns 2–5 of the row corresponding to $x_1 = 0.000$ in Table 9.9.

Input estimate $x_1 = 0.010$

For the input estimate $x_1 = 0.010$, with correlation coefficient $r(x_1, x_2) = 0$, Figure 9.13 shows the PDFs obtained using the GUM uncertainty framework with first-order terms only and with higher-order terms, and using a Monte Carlo method. The PDF provided by a Monte Carlo method exhibits a modest left-hand flank, although it is truncated at zero, the smallest possible numerical value of δY. Further, compared with the results for $x_1 = 0$, it is closer in form to the Gaussian PDFs provided by the GUM uncertainty framework. These Gaussian PDFs are in turn reasonably close to each other, δY having expectation 1.0×10^{-4} and standard deviations 1.0×10^{-4} and 1.1×10^{-4}, respectively.

Figure 9.13 also shows the endpoints of the shortest 95 % coverage intervals obtained by the three approaches. The continuous vertical lines denote the endpoints of the interval provided by a Monte Carlo method, the broken vertical lines those resulting from the GUM uncertainty framework with first-order terms, and the dotted vertical lines from the GUM

[3] When the Monte Carlo method was repeated several times the results obtained were scattered about 50×10^{-6}. When it was repeated a number of times with a larger numerical value of M the results were again scattered about 50×10^{-6}, but with a reduced dispersion. Such dispersion effects are expected, and were observed for the other Monte Carlo calculations made. Reporting the results to greater numbers of significant decimal digits would be necessary to see the actual numerical differences.

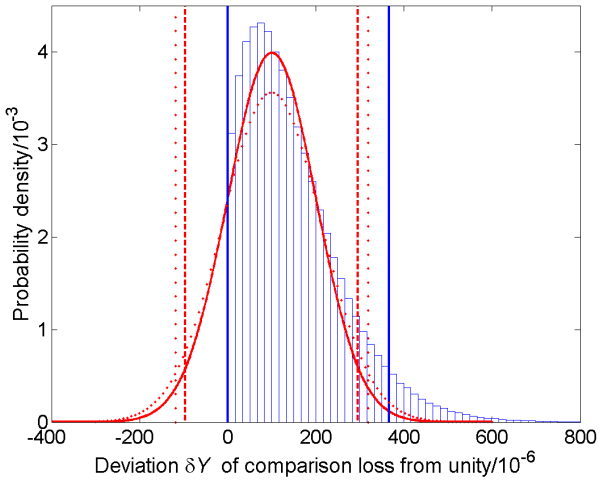

Figure 9.13: As Figure 9.12 except that $x_1 = 0.010$, and showing the PDFs resulting from the GUM uncertainty framework with first-order (higher-peaked curve) and with higher-order terms (lower-peaked curve).

uncertainty framework with higher-order terms. The intervals provided by the GUM uncertainty framework are shifted to the left compared with the shortest $95\,\%$ coverage interval provided by a Monte Carlo method. As a consequence, they again include infeasible values of δY. The shift is about $70\,\%$ of the standard uncertainty. The interval provided by a Monte Carlo method has its left-hand endpoint at zero, the smallest feasible value. The corresponding results are given in the penultimate rows of Tables 9.8 and 9.9.

Input estimate $x_1 = 0.050$

Figure 9.14 is similar to Figure 9.13, but for $x_1 = 0.050$. Now, the PDFs provided by both variants of the GUM uncertainty framework are virtually indistinguishable from each other. Further, they are now much closer to the approximation to the PDF provided by a Monte Carlo method. That PDF exhibits a slight skewness, as evidenced in the tail regions. The coverage intervals provided by the two variants of the GUM uncertainty framework are visually almost identical, but still shifted from those provided by a Monte Carlo method. The shift is now about $10\,\%$ of the standard uncertainty. The intervals provided by the GUM uncertainty framework are now feasible. The corresponding results are given in the final rows of Tables 9.8 and 9.9.

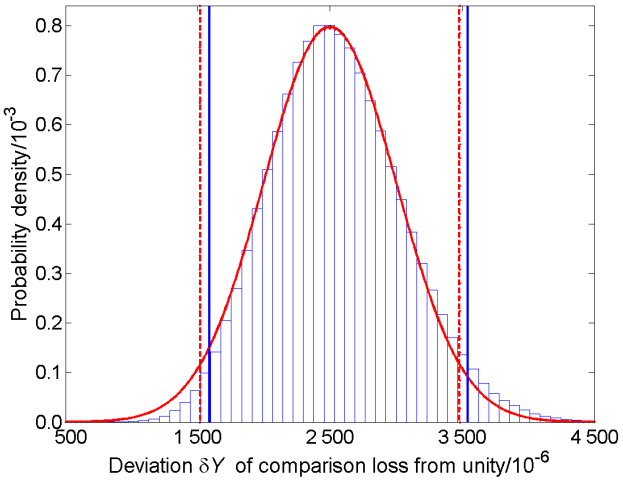

Figure 9.14: As Figure 9.13 except that $x_1 = 0.050$.

9.9.2 Non-zero covariance

The three approaches used in the cases where the X_i are uncorrelated (Section 9.9.1) are now applied for the three cases in which they are correlated, with $r(x_1, x_2) = 0.9$. However, the GUM uncertainty framework with first-order terms only is used. Unlike the cases where the X_i are uncorrelated, the GUM uncertainty framework with higher-order terms is not applied, no counterpart being provided in the GUM for the formula containing higher-order terms when the x_i have associated non-zero covariances. Other aspects match those in Section 9.9.1.

For the GUM uncertainty framework with first-order terms, $u(\delta y)$ is evaluated from [11, Clause F.3]

$$u^2(\delta y) = 4x_1^2 u^2(x_1).$$

Consequently, $u(\delta y)$ does not depend on $r(x_1, x_2)$ and the GUM uncertainty framework with first-order terms gives identical results to those presented in Section 9.9.1. In particular, for the case $x_1 = 0$, $u(\delta y)$ is (incorrectly) computed as zero, as in Section 9.9.1.

A Monte Carlo method was implemented by making random draws from a bivariate Gaussian PDF used to characterize a quantity with the given expectation and covariance matrix (expressions (9.18)) [11, Clause C.5].[4]

Tables 9.10 and 9.11 contain the results obtained. Those from a Monte Carlo method indicate that although δy is unaffected by the correlation between the X_i, $u(\delta y)$ is so influenced,

[4] Apart from the requirement to draw from a multivariate distribution, the implementation of a Monte Carlo method for input quantities that are correlated is no more complicated than when the input quantities are uncorrelated.

more so for small x_1. The 95 % coverage intervals are influenced accordingly.

Figures 9.15 and 9.16 show the PDFs provided by the GUM uncertainty framework with first-order terms (bell-shaped curves) and a Monte Carlo method (scaled frequency distributions) in the cases $x_1 = 0.010$ and $x_1 = 0.050$, respectively. The endpoints of the shortest 95 % coverage interval provided by the two approaches are also shown, as broken vertical lines for the GUM uncertainty framework and continuous vertical lines for a Monte Carlo method. In the case $x_1 = 0.010$ (Figure 9.15), the effect of the correlation has been to change noticeably the results returned by a Monte Carlo method (compare with Figure 9.13). Not only has the shape of (the approximation to) the PDF changed, but the corresponding coverage interval no longer has its left-hand endpoint at zero. In the case $x_1 = 0.050$ (Figure 9.16), the differences between the results for the cases where the input quantities are uncorrelated and correlated (compare with Figure 9.14) are less obvious.

x_1	Estimate $\delta y/10^{-6}$			Standard uncertainty $u(\delta y)/10^{-6}$		
	A	G	M	A	G	M
0.000	50	0	50	67	0	67
0.010	150	100	150	121	100	120
0.050	2 550	2 500	2 550	505	500	505

Table 9.10: Comparison loss results (estimates and associated standard uncertainties), for input estimates with associated non-zero covariance ($r(x_1, x_2) = 0.9$), obtained analytically (A), and using the GUM uncertainty framework (G) and a Monte Carlo method (M).

x_1	Shortest 95 % coverage interval for $\delta Y/10^{-6}$	
	G	M
0.000	[0, 0]	[0, 185]
0.010	[−96, 296]	[13, 397]
0.050	[1 520, 3 480]	[1 627, 3 559]

Table 9.11: Comparison loss results (endpoints of the shortest 95 % coverage interval), for input estimates with associated non-zero covariance ($r(x_1, x_2) = 0.9$), obtained using the GUM uncertainty framework (G) and a Monte Carlo method (M).

9.10 Gauge block calibration

The length of a nominally 50 mm gauge block is determined by comparing it with a known reference standard of the same nominal length. The direct output of the comparison of the two gauge blocks is the difference d in their lengths given by

$$d = \ell(1 + \alpha\theta) - \ell_{\mathrm{s}}(1 + \alpha_{\mathrm{s}}\theta_{\mathrm{s}}), \tag{9.20}$$

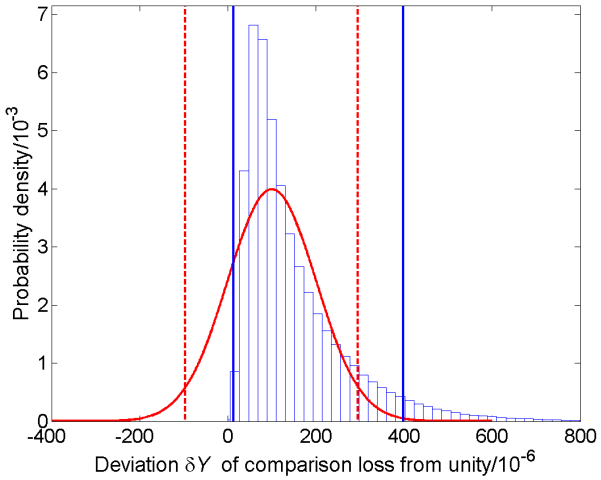

Figure 9.15: Results for the problem of comparison loss in power meter calibration in the case $x_1 = 0.010$, $x_2 = 0$, with $u(x_1) = u(x_2) = 0.005$ and $r(x_1, x_2) = 0.9$.

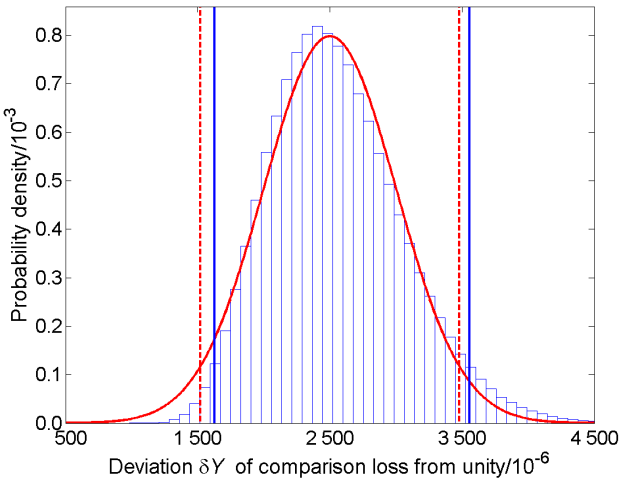

Figure 9.16: As Figure 9.15 except that $x_1 = 0.050$.

where ℓ is the length at 20 °C of the gauge block being calibrated, ℓ_s is the length of the reference standard at 20 °C as given in its calibration certificate, α and α_s are the coefficients of thermal expansion, respectively, of the gauge being calibrated and the reference standard, and θ and θ_s are the deviations in temperature from the 20 °C reference temperature, respectively, of the gauge block being calibrated and the reference standard.

From expression (9.20), the output quantity ℓ is given by

$$\ell = \frac{\ell_s(1 + \alpha_s\theta_s) + d}{1 + \alpha\theta}, \tag{9.21}$$

from which, to an approximation adequate for most practical purposes,

$$\ell = \ell_s + d + \ell_s(\alpha_s\theta_s - \alpha\theta). \tag{9.22}$$

If the difference in temperature between the gauge block being calibrated and the reference standard is written as $\delta\theta = \theta - \theta_s$, and the difference in their thermal expansion coefficients as $\delta\alpha = \alpha - \alpha_s$, expressions (9.21) and (9.22) become, respectively,

$$\ell = \frac{\ell_s[1 + \alpha_s(\theta - \delta\theta)] + d}{1 + (\alpha_s + \delta\alpha)\theta} \tag{9.23}$$

and

$$\ell = \ell_s + d - \ell_s(\delta\alpha\theta + \alpha_s\delta\theta). \tag{9.24}$$

The difference d in the lengths of the gauge block being calibrated and the reference standard is determined as the average of a series of five indication values, obtained independently, of the difference using a calibrated comparator. d can be expressed as

$$d = D + d_1 + d_2, \tag{9.25}$$

where D is a quantity of which the average of the five indication values is a realization, and d_1 and d_2 are quantities describing, respectively, the random and systematic errors associated with using the comparator.

The quantity θ, representing deviation of the temperature from 20 °C of the gauge block being calibrated, can be expressed as

$$\theta = \theta_0 + \Delta, \tag{9.26}$$

where θ_0 is a quantity representing the average temperature deviation of the gauge block from 20 °C and Δ a quantity describing a cyclic variation of the temperature deviation from θ_0.

Substituting expressions (9.25) and (9.26) into expressions (9.23) and (9.24), and working with the quantity $\delta\ell$ representing the deviation of ℓ from the nominal length

$$\ell_{\mathrm{nom}} = 50\,\mathrm{mm}$$

of the gauge block, gives

$$\delta\ell = \frac{\ell_{\rm s}[1 + \alpha_{\rm s}(\theta_0 + \Delta - \delta\theta)] + D + d_1 + d_2}{1 + (\alpha_{\rm s} + \delta\alpha)(\theta_0 + \Delta)} - \ell_{\rm nom} \tag{9.27}$$

and

$$\delta\ell = \ell_{\rm s} + D + d_1 + d_2 - \ell_{\rm s}[\delta\alpha(\theta_0 + \Delta) + \alpha_{\rm s}\delta\theta] - \ell_{\rm nom} \tag{9.28}$$

as measurement functions for the measurement problem.

The treatment here of the measurement problem is in terms of the measurement functions (9.27) and (9.28) with output quantity $\delta\ell$ and input quantities $\ell_{\rm s}$, D, d_1, d_2, $\alpha_{\rm s}$, θ_0, Δ, $\delta\alpha$ and $\delta\theta$. It differs from that given in GUM example H.1 in that in the GUM the measurement functions (9.25) and (9.26) are treated as sub-models to measurement functions (9.23) and (9.24), i.e. the GUM uncertainty framework is applied to each measurement function (9.25) and (9.26), with the results obtained used to provide information about the input quantities d and θ in measurement functions (9.23) and (9.24). The treatment here avoids having to use the results obtained from a Monte Carlo method applied to the measurement functions (9.25) and (9.26) to provide information about the distributions for the input quantities d and θ in expressions (9.23) and (9.24).

In the following the available information about each input quantity in the measurement functions (9.27) and (9.28) is provided. This information is extracted from the description given in the GUM, and for each item of information the GUM subclause from which the item is extracted is identified. Also provided is an interpretation of the information in terms of characterizing the quantity by a probability distribution [11, Clause 6.4].

Length $\ell_{\rm s}$ of the reference standard The calibration certificate for the reference standard gives $\widehat{\ell_{\rm s}} = 50.000\,623\,{\rm mm}$ as its length at $20\,^{\circ}{\rm C}$ [10, Clause H.1.5]. It also gives $U_p = 0.075\,\mu{\rm m}$ as the expanded uncertainty of the reference standard and states that U_p was obtained using a coverage factor of $k_p = 3$ [10, Clause H.1.3.1]. The certificate states that the effective degrees of freedom associated with the combined standard uncertainty, from which the quoted expanded uncertainty was obtained, is $\nu_{\rm eff}(u(\widehat{\ell_{\rm s}})) = 18$ [10, Clause H.1.6].

The quantity $\ell_{\rm s}$ is characterized by a scaled and shifted t-distribution ${\rm t}_\nu(\mu, \sigma^2)$[11, Clause 6.4.9.7], with

$$\mu = 50\,000\,623\,{\rm nm}, \qquad \sigma = \frac{U_p}{k_p} = \frac{75}{3}\,{\rm nm} = 25\,{\rm nm}, \qquad \nu = 18.$$

Average length difference D The average $\widehat{D}$ of the five indication values of the difference in lengths between the gauge block being calibrated and the reference standard is $215\,{\rm nm}$ [10, Clause H.1.5]. The pooled experimental standard deviation characterizing the comparison of ℓ and $\ell_{\rm s}$ was determined from 25 indication values, obtained independently, of the difference in lengths of two standard gauge blocks, and equalled $13\,{\rm nm}$ [10, Clause H.1.3.2].

The quantity D is characterized by a scaled and shifted t-distribution $\mathrm{t}_\nu(\mu, \sigma^2)$ [11, Clause 6.4.9], with

$$\mu = 215\,\mathrm{nm}, \qquad \sigma = \frac{13}{\sqrt{5}}\,\mathrm{nm} = 6\,\mathrm{nm}, \qquad \nu = 24.$$

Random effect d_1 of comparator According to the calibration certificate of the comparator used to compare ℓ with ℓ_s, the associated uncertainty due to random errors is $0.01\,\mu\mathrm{m}$ for a coverage probability of $95\,\%$ and is obtained from six indication values, obtained independently [10, Clause H.1.3.2].

The quantity d_1 is characterized by a scaled and shifted t-distribution $\mathrm{t}_\nu(\mu, \sigma^2)$ [11, Clause 6.4.9.7], with

$$\mu = 0\,\mathrm{nm}, \qquad \sigma = \frac{U_{0.95}}{k_{0.95}} = \frac{10}{2.57}\,\mathrm{nm} = 4\,\mathrm{nm}, \qquad \nu = 5.$$

Here, $k_{0.95}$ is obtained from Table G.2 of the GUM with $\nu = 5$ degrees of freedom and $p = 0.95$.

Systematic effect d_2 of comparator The uncertainty of the comparator due to systematic errors is given in the calibration certificate as $0.02\,\mu\mathrm{m}$ at the 'three sigma level' [10, Clause H.1.3.2]. This uncertainty may be assumed to be reliable to $25\,\%$, and thus the degrees of freedom is $\nu_\mathrm{eff}(u(\widehat{d}_2)) = 8$ [10, Clause H.1.6].

The quantity d_2 is characterized by a scaled and shifted t-distribution $\mathrm{t}_\nu(\mu, \sigma^2)$ [11, Clause 6.4.9], with

$$\mu = 0\,\mathrm{nm}, \qquad \sigma = \frac{U_p}{k_p} = \frac{20}{3}\,\mathrm{nm} = 7\,\mathrm{nm}, \qquad \nu = 8.$$

Thermal expansion coefficient α_s The coefficient of thermal expansion of the reference standard is given as $\widehat{\alpha}_\mathrm{s} = 11.5 \times 10^{-6}\,^\circ\mathrm{C}^{-1}$ with possible values of this quantity represented by a rectangular distribution with limits $\widehat{\alpha}_\mathrm{s} \pm 2 \times 10^{-6}\,^\circ\mathrm{C}^{-1}$ [10, Clause H.1.3.3].

The quantity α_s is characterized by a rectangular distribution $\mathrm{R}(a, b)$ [11, Clause 6.4.2], with limits

$$a = 9.5 \times 10^{-6}\,^\circ\mathrm{C}^{-1}, \qquad b = 13.5 \times 10^{-6}\,^\circ\mathrm{C}^{-1}.$$

There is no information about the reliability of the limits and so the quantity is characterized by a rectangular distribution with exactly known limits. Such information may have been omitted from the description in the GUM because the corresponding sensitivity coefficient is zero, and so the quantity makes no contribution in an application of the GUM uncertainty framework based on first order terms only.

Average temperature deviation θ_0 The temperature of the test bed is reported as $(19.9 \pm 0.5)\,^\circ\mathrm{C}$. The average temperature deviation $\widehat{\theta}_0 = -0.1\,^\circ\mathrm{C}$ is reported as

having an associated standard uncertainty due to the uncertainty associated with the average temperature of the test bed of $u(\widehat{\theta_0}) = 0.2\,^{\circ}\mathrm{C}$ [10, Clause H.1.3.4].

The quantity θ_0 is characterized by a Gaussian distribution $\mathrm{N}(\mu, \sigma^2)$ [11, Clause 6.4.7], with

$$\mu = -0.1\,^{\circ}\mathrm{C}, \qquad \sigma = 0.2\,^{\circ}\mathrm{C}.$$

There is no information about the source of the evaluation of the uncertainty and so the quantity is characterized by a Gaussian distribution.

Effect Δ of cyclic temperature variation The temperature of the test bed is reported as $(19.9 \pm 0.5)\,^{\circ}\mathrm{C}$. The stated maximum offset of $0.5\,^{\circ}\mathrm{C}$ for Δ is said to represent the amplitude of an approximately cyclical variation of temperature under a thermostatic system. The cyclic variation of temperature results in a U-shaped (arc sine) distribution [10, Clause H.1.3.4].

The quantity Δ is characterized by an arc sine distribution $\mathrm{U}(a, b)$ [11, Clause 6.4.6], with limits

$$a = -0.5\,^{\circ}\mathrm{C}, \qquad b = 0.5\,^{\circ}\mathrm{C}.$$

There is no information about the reliability of the limits and so the quantity is characterized by a U-shaped distribution with exactly known limits.

Difference $\delta\alpha$ in expansion coefficients The estimated bounds on the variability of $\delta\alpha$ are $\pm 1 \times 10^{-6}\,^{\circ}\mathrm{C}^{-1}$, with an equal probability of $\delta\alpha$ having any value within those bounds [10, Clause H.1.3.5]. These bounds are deemed to be reliable to $10\,\%$, giving $\nu(u(\delta\widehat{\alpha})) = 50$ [10, Clause H.1.6].

The quantity $\delta\alpha$ is characterized by a rectangular distribution with inexactly prescribed limits [11, Clause 6.4.3], with

$$a = -1.0 \times 10^{-6}\,^{\circ}\mathrm{C}^{-1}, \qquad b = 1.0 \times 10^{-6}\,^{\circ}\mathrm{C}^{-1}, \qquad d = 0.1 \times 10^{-6}\,^{\circ}\mathrm{C}^{-1}.$$

The stated reliability of $10\,\%$ on the estimated bounds provides the basis for this value of d [11, Clause 6.4.3].

Difference $\delta\theta$ in temperatures The reference standard and the gauge block being calibrated are expected to be at the same temperature, but the temperature difference $\delta\theta$ could lie with equal probability anywhere in the estimated interval $-0.05\,^{\circ}\mathrm{C}$ to $0.05\,^{\circ}\mathrm{C}$ [10, Clause H.1.3.6]. This difference is believed to be reliable only to $50\,\%$, giving $\nu(u(\delta\widehat{\theta})) = 2$ [10, Clause H.1.6].

The quantity $\delta\theta$ is characterized by a rectangular distribution with inexactly prescribed limits [11, Clause 6.4.3], with

$$a = -0.050\,^{\circ}\mathrm{C}, \qquad b = 0.050\,^{\circ}\mathrm{C}, \qquad d = 0.025\,^{\circ}\mathrm{C}.$$

The stated reliability of $50\,\%$ provides the basis for this value of d [11, Clause 6.4.3].

The application of the GUM uncertainty framework is based on (a) a first-order Taylor series approximation to the measurement function (9.27) or (9.28), (b) use of the Welch-Satterthwaite formula to evaluate an effective degrees of freedom (rounded towards zero) associated with the uncertainty obtained from the law of propagation of uncertainty, and (c) characterizing the output quantity by a scaled and shifted t-distribution with the above degrees of freedom. The application of a Monte Carlo method requires (a) making random draws from a rectangular distribution [11, Clause 6.4.2.4], Gaussian distribution [11, Clause 6.4.7.4], t-distribution [11, Clause 6.4.9.5], U-shaped distribution [11, Clause 6.4.6.4], and rectangular distribution with inexactly prescribed limits [11, Clause 6.4.3.4], and (b) implements an adaptive Monte Carlo procedure (Section 7.2.5) with a numerical tolerance ($\delta = 0.005$) set to deliver $n_{\mathrm{dig}} = 2$ significant decimal digits in the standard uncertainty.

Table 9.12 gives the results obtained for the approximate measurement function (9.28) using the information above about each input quantity. Figure 9.17 shows the distribution functions and PDFs for $\delta\ell$ obtained from the application of the GUM uncertainty framework (solid curve) and a Monte Carlo method (scaled frequency distribution). The distribution obtained from the GUM uncertainty framework is a t-distribution with $\nu = 16$ degrees of freedom. The endpoints of the shortest $99\,\%$ coverage intervals for $\delta\ell$ obtained from the PDFs are indicated as (continuous) vertical lines (obtained from the GUM uncertainty framework) and (broken) vertical lines (obtained from a Monte Carlo method).

1.36×10^6 trials were taken by the adaptive Monte Carlo procedure. The calculations were also carried out for a coverage probability of $95\,\%$, for which 0.52×10^6 trials were taken.

Method	$\widehat{\delta\ell}$ /nm	$u(\widehat{\delta\ell})$ /nm	Shortest $99\,\%$ coverage interval for $\delta\ell$/nm
GUF	838	32	[746, 930]
MCM	838	36	[745, 931]

Table 9.12: Results obtained for the approximate measurement function (9.28). GUF denotes the GUM uncertainty framework and MCM a Monte Carlo method.

Results obtained for the measurement function (9.27) are identical to the results in Table 9.12 to the number of decimal digits given there.

There are some modest differences in the results obtained. $u(\widehat{\delta\ell})$ was $4\,\mathrm{nm}$ greater for the application of a Monte Carlo method than for the GUM uncertainty framework. The length of the $99\,\%$ coverage interval for $\delta\ell$ was $2\,\mathrm{nm}$ greater. These results apply equally to both measurement functions considered. Whether such differences are important has to be judged in terms of the manner in which the results are to be used.

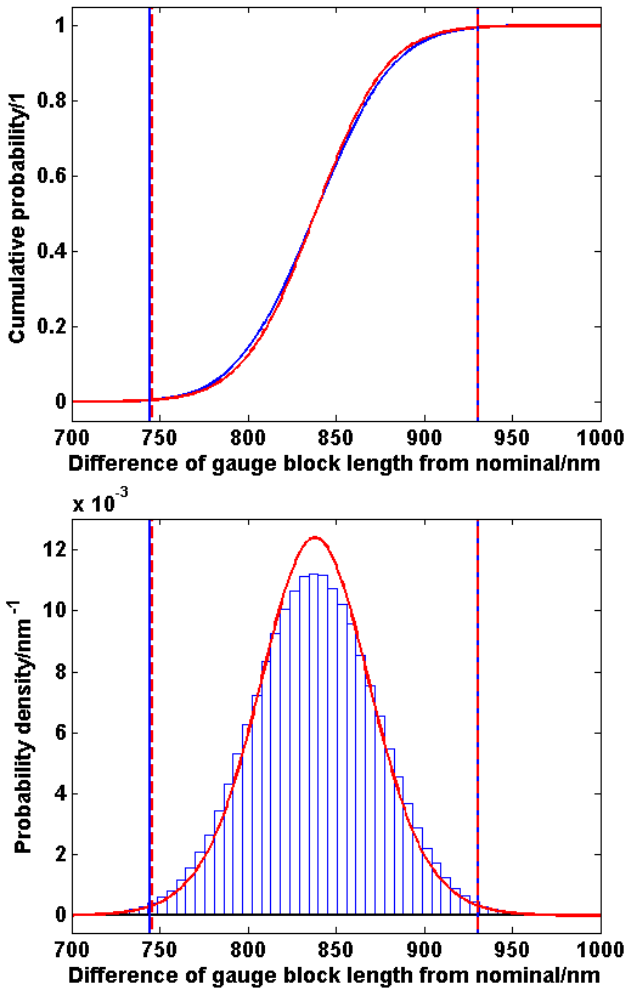

Figure 9.17: Distribution functions and (below) PDFs for $\delta\ell$ using the GUM uncertainty framework and a Monte Carlo method.

9.11 Quantities subject to a normalisation constraint

The composition of natural gas extracted from subterranean reservoirs varies widely. Since the composition determines the energy content as well as the combustion and condensation characteristics of the gas, it has a strong influence on its value as a traded commodity. Consequently, there is a requirement for the analysis of the composition of natural gas and strong economic pressures to reduce the uncertainty associated with the measured composition to facilitate efficient and safe trading.

The most widely used method for the analysis of the composition of natural gas is gas chromatography. The processing of the raw data from gas chromatographic analysis presents a number of mathematical challenges, particularly when the detailed statistical structure of the measurement data is taken fully into account. The task is made more complex by the presence of a normalisation constraint that requires the sum of all component fractions is unity.

Let $x_i,\ i = 1, \ldots, N$, denote the measured value of amount fraction of component i and $\mathrm{cov}(x_i, x_j)$ the covariance associated with measured values x_i and x_j. The problem is to obtain from the x_i estimates y_i, with associated uncertainties, of the amount fractions that satisfy the normalisation constraint

$$\sum_{i=1}^{N} y_i = 1. \tag{9.29}$$

This constraint expresses the fact that the sum of the y_i is unity because they are defined to be fractions of the whole mixture. The physical mechanisms that prevent the measured values meeting the normalisation constraint have been discussed elsewhere [15, 103]. A fuller treatment of the problem, which compares a number of measurement models and applies these models to real data, is available [56, 57, 75, 76].

The problem addressed is described by the model equations

$$\boldsymbol{Y} = \boldsymbol{X} \qquad \text{subject to} \qquad \mathbf{1}^\top \boldsymbol{Y} = 1, \tag{9.30}$$

where $\boldsymbol{X}$ is the vector of input quantities of which the measured values $\boldsymbol{x}$ are estimates with associated covariance matrix $\boldsymbol{U}_{\boldsymbol{x}}$, $\boldsymbol{Y}$ is the vector of output quantities representing the amount fractions constrained to satisfy the normalisation constraint, and $\mathbf{1} = (1, \ldots, 1)^\top$.

The generalised least-squares solution [70] to the model equations (9.30) is the vector $\boldsymbol{z} = \boldsymbol{y}$ that solves

$$\min_{\boldsymbol{z}} (\boldsymbol{x} - \boldsymbol{z})^\top \boldsymbol{U}_{\boldsymbol{x}}{}^{-1} (\boldsymbol{x} - \boldsymbol{z}) \qquad \text{subject to} \qquad \mathbf{1}^\top \boldsymbol{z} = 1. \tag{9.31}$$

Problem (9.31) is a linear equality constrained minimization problem, for which (optimality) conditions for a solution take the form [51]

$$\mathbf{1}^\top \boldsymbol{y} = 1,$$

and

$$-2\boldsymbol{U_x}^{-1}(\boldsymbol{x}-\boldsymbol{y}) = \mathbf{1}\lambda,$$

in which λ denotes a Lagrange multiplier. From the above optimality conditions is then obtained the solution

$$\boldsymbol{y} = \frac{\boldsymbol{U_x}\mathbf{1}}{\mathbf{1}^\top\boldsymbol{U_x}\mathbf{1}} - \left(\boldsymbol{I} - \frac{\boldsymbol{U_x}\mathbf{1}\mathbf{1}^\top}{\mathbf{1}^\top\boldsymbol{U_x}\mathbf{1}}\right)\boldsymbol{x}, \tag{9.32}$$

where $\boldsymbol{I}$ is the identity matrix. Then, applying the law of propagation of uncertainty for *multivariate, real, measurement functions* (Section 6.2.2),

$$\boldsymbol{U_y} = \left(\boldsymbol{I} - \frac{\boldsymbol{U_x}\mathbf{1}\mathbf{1}^\top}{\mathbf{1}^\top\boldsymbol{U_x}\mathbf{1}}\right)\boldsymbol{U_x}\left(\boldsymbol{I} - \frac{\boldsymbol{U_x}\mathbf{1}\mathbf{1}^\top}{\mathbf{1}^\top\boldsymbol{U_x}\mathbf{1}}\right)^\top,$$

from which is obtained

$$\boldsymbol{U_y} = \boldsymbol{U_x} - \frac{\boldsymbol{U_x}\mathbf{1}\mathbf{1}^\top\boldsymbol{U_x}}{\mathbf{1}^\top\boldsymbol{U_x}\mathbf{1}}. \tag{9.33}$$

In order to help with the interpretation of this result, write

$$\boldsymbol{w} = \boldsymbol{U_x}\mathbf{1}.$$

Then, from expression (9.32),

$$\boldsymbol{y} = \boldsymbol{x} + \frac{\boldsymbol{w}}{\mathbf{1}^\top\boldsymbol{w}}\left(1 - \mathbf{1}^\top\boldsymbol{x}\right), \tag{9.34}$$

and, from expression (9.33),

$$\boldsymbol{U_y} = \boldsymbol{U_x} - \frac{\boldsymbol{w}\boldsymbol{w}^\top}{\mathbf{1}^\top\boldsymbol{w}}. \tag{9.35}$$

It follows from expression (9.34) that the correction applied to each measured amount fraction is a weighted proportion of the amount by which the measured values fail to satisfy the normalisation constraint. Furthermore, since the weights sum to unity, the total correction is the amount by which the measured values fail to satisfy that constraint. It is interesting to note that when the measured values $\boldsymbol{x}$ satisfy the normalisation constraint (9.29), then expression (9.34) simplifies to $\boldsymbol{y} = \boldsymbol{x}$, but expression (9.35) implies $\boldsymbol{U_y} \neq \boldsymbol{U_x}$. This demonstrates that the normalisation constraint itself carries information about the amount fractions that is *additional* to that provided by the measured values. This information serves to 'update' the uncertainties associated with the estimates although, in this special case, it does not update the estimates themselves.

9.12 Area under a curve

This example is concerned with the problem of evaluating the uncertainty associated with an estimate of the definite integral

$$I = \int_a^b x(t)\,\mathrm{d}t$$

given finite values of a and b with $a < b$, and inexact data x_i regarded as measured values of the smooth unknown function $x = x(t)$ at exact abscissa values $t_i,\ i = 1, \ldots, N$, with $a = t_1 < t_2 < \ldots < t_N = b$. Such integrals arise in several branches of metrology, and most notably in the field of radiometry as part of the determination of (a) photometric quantities such as illuminance from spectral irradiance measurements, (b) filtered-detector response, or (c) colorimetric quantities of a source [49].

For $i = 1, \ldots, N$, let X_i denote the value of $x(t)$ at $t = t_i$. Consider the determination of an approximation Y to I given by applying a linear *quadrature rule* of the form

$$Y = \sum_{i=1}^{N} w_i X_i, \tag{9.36}$$

where the *quadrature rule weights* w_i depend only on the t_i [27]. Regard the given x_i as particular realizations, obtained by measurement, of the quantities X_i. An estimate y of Y is

$$y = \sum_{i=1}^{N} w_i x_i.$$

For quadrature rules of the form (9.36) that are *linear* in the measured quantities X_i, the law of propagation of uncertainty based on a first-order Taylor series expansion can be applied, making no further approximation, to evaluate the uncertainty associated with the estimate y. Such application gives

$$u^2(y) = \boldsymbol{w}^\top \boldsymbol{U}_{\boldsymbol{x}} \boldsymbol{w},$$

where $\boldsymbol{w} = (w_1, w_2, \ldots, w_N)^\top$ and $\boldsymbol{U}_{\boldsymbol{x}}$ is the covariance matrix associated with the estimates $\boldsymbol{x}$ (Section 6.2.3). In the case of mutually independent X_i, the result reduces to

$$u^2(y) = \sum_{i=1}^{N} w_i^2 u^2(x_i). \tag{9.37}$$

The quadrature rule most commonly used in metrology is the *trapezoidal rule*. Consider the $x_i,\ i = 1, \ldots, N$, at a uniform spacing $h = (b - a)/(N - 1)$ in the interval $[a, b]$, i.e., at abscissae $t_i = a + (i - 1)h$. The trapezoidal rule is given by the formula

$$y = h \sum_{i=1}^{N-1} \frac{x_i + x_{i+1}}{2} = h \left(\frac{1}{2} x_1 + \sum_{i=2}^{N-1} x_i + \frac{1}{2} x_N \right),$$

i.e., a rule of the form (9.36) with weights

$$w_i = \begin{cases} h/2, & i = 1, \\ h, & i = 2, \ldots, N-1, \\ h/2, & i = N. \end{cases}$$

By applying (9.37), we obtain for the trapezoidal rule and mutually independent X_i,

$$u^2(y) = h^2 \left(\frac{1}{4} u^2(x_1) + \sum_{i=2}^{N-1} u^2(x_i) + \frac{1}{4} u^2(x_N) \right),$$

which, in the case where the uncertainties $u(x_i)$ are identical, becomes

$$u^2(y) = h^2(N - 3/2)u^2(x_1).$$

Substituting $h = (b - a)/(N - 1)$ gives

$$u(y) = (b - a)\frac{(N - 3/2)^{1/2}}{N - 1}u(x_1),$$

which, for large N, is approximately

$$u(y) = \frac{b - a}{\sqrt{N}}u(x_1).$$

Since the length $b - a$ of the interval of integration is fixed, the uncertainty essentially decreases with the number N of measurements as $1/\sqrt{N}$. Such behaviour is similar to that for the average of a set of N measured values, of uncorrelated quantities, having the same associated uncertainty.

The trapezoidal rule can be derived by determining the piecewise-linear function joining the points $(t_1, x_1), \ldots, (t_N, x_N)$ and integrating it between a and b. By integrating each piece of the piecewise-linear function separately, the rule can be expressed as

$$y = \sum_{i=1}^{N-1} y_i, \qquad y_i = h\frac{x_i + x_{i+1}}{2},$$

where y_i denotes the area under the linear piece over the sub-interval (t_i, t_{i+1}). More sophisticated rules are given by approximating the underlying function over each interval by a polynomial of degree two or higher, rather than a straight line (polynomial of degree one), where each polynomial piece is obtained by interpolating (a subset of) the data. Once these polynomial pieces have been formed, their integration over each interval (to give the values $y_i,\ i = 1, \ldots, N - 1$) and their summation (to give the value y) provides the required approximation to the definite integral of the function. A treatment of such quadrature rules is available [27].

Because of the nature of the implementation of these quadrature rules, in terms of calculations for each consecutive sub-interval (t_i, t_{i+1}), the explicit evaluation of the quadrature rule weights w_i in (9.36) is not generally undertaken. However, for the purposes of the evaluation of the uncertainty associated with the result, the quadrature rule weights are required. An approach to determining the weights is presented below that can readily be applied based on exploiting the linearity of the quadrature rule as a function of the input quantities X_i. It assumes a numerical procedure implementing the quadrature rule is available.

Define, as in expression (9.36),

$$Y(X_1, \ldots, X_N) = \sum_{i=1}^{N} w_i X_i,$$

and, for $r = 1, \ldots, N$,

$$Y_r(X_r) = Y(x_1, \ldots, x_{r-1}, X_r, x_{r+1}, \ldots, x_n) = \sum_{i \neq r} w_i x_i + w_r X_r.$$

Then,

$$Y_r(x_r) = \sum_{i \neq r} w_i x_i + w_r x_r, \qquad Y_r(x_r + \delta x_r) = \sum_{i \neq r} w_i x_i + w_r (x_r + \delta x_r),$$

and, for $\delta x_r \neq 0$,

$$w_r = \frac{Y_r(x_r + \delta x_r) - Y_r(x_r)}{\delta x_r}.$$

The sensitivity coefficient corresponding to the input quantity X_r is, therefore, calculated in terms of two applications of the quadrature rule: one for the original measured data and the other for this data with the measured value for X_r perturbed by $\delta x_r \neq 0$.

Consider now approximations $y^{(n)}$ and $y^{(n+1)}$ to the value of the integral I obtained by applying quadrature rules based on interpolating polynomials of, respectively, degrees n and $n + 1$. The approximations can be expressed as

$$y^{(n)} = \sum_{i=1}^{N} w_i^{(n)} x_i,$$

and

$$y^{(n+1)} = \sum_{i=1}^{N} w_i^{(n+1)} x_i,$$

or, equivalently, as the values of the measurement functions

$$Y^{(n)} = \sum_{i=1}^{N} w_i^{(n)} X_i,$$

and

$$Y^{(n+1)} = \sum_{i=1}^{N} w_i^{(n+1)} X_i,$$

for the estimates x_i of the input quantities X_i. The manner in which the quadrature rule weights $w_i^{(n)}$ and $w_i^{(n+1)}$ can be obtained is indicated above. The approximation $y^{(n+1)}$ can be regarded as statistically no better than $y^{(n)}$ if the magnitude of the numerical difference between the approximations is no greater than the expanded uncertainty associated with the difference. The difference is

$$\Delta y^{(n)} = y^{(n+1)} - y^{(n)},$$

and, applying the law of propagation of uncertainty, the standard uncertainty associated with the difference is obtained from

$$u^2(\Delta y^{(n)}) = u^2(y^{(n+1)}) + u^2(y^{(n)}) - 2\mathrm{Cov}(y^{(n)}, y^{(n+1)}).$$

However, rather than calculating the covariance term in the expression above, it is simpler to express $\Delta Y^{(n)} = Y^{(n+1)} - Y^{(n)}$ (of which $\Delta y^{(n)}$ is an estimate) directly in terms of the quantities X_i:

$$\Delta Y^{(n)} = \sum_{i=1}^{N} \Delta w_i^{(n)} X_i, \qquad \Delta w_i^{(n)} = w_i^{(n+1)} - w_i^{(n)}.$$

Then, it follows from the law of propagation of uncertainty, that

$$u^2(\Delta y_n) = (\boldsymbol{\Delta w}^{(n)})^\top \boldsymbol{U_x} \boldsymbol{\Delta w}^{(n)},$$

where $\boldsymbol{\Delta w}^{(n)} = (\Delta w_1^{(n)}, \Delta w_2^{(n)}, \ldots, \Delta w_N^{(n)})^\top$.

Finally, in the application of the GUM uncertainty framework, $\Delta Y^{(n)}$ is characterized by a Gaussian distribution giving $U(\Delta y^{(n)}) = 2u(\Delta y^{(n)})$ for the expanded uncertainty associated with the estimate $\Delta y^{(n)}$ corresponding to a $95\,\%$ coverage probability. If the input quantities are characterized by probability distributions, a Monte Carlo method may be used to validate this choice of distribution for $\Delta Y^{(n)}$.

9.13 SIR efficiency

The SIR (International Reference System for activity measurements of γ-ray emitting radionuclides) was established in 1976 at the Bureau International des Poids et Mesures (BIPM) to complement the international comparisons of activity measurements organized by Section II of the CCRI (Comité Consultatif des Rayonnements Ionizants). Participating laboratories submit SIR glass ampoules containing their standardized solutions of radionuclides to the BIPM, where the current produced by these samples in an ionization chamber is compared with the current obtained with a $^{226}\mathrm{Ra}$ reference source [88]. The simplicity and rapidity of the measurement as well as the long-term stability of the ionization chamber has allowed the comparison over 25 years of hundreds of radioactive solutions for a total of about 60 radionuclides.

Efficiency curves (detection efficiency versus photon energy and detection efficiency versus beta energy) are required for the calculation of the response of the ionization chamber for radionuclides not previously measured in the SIR. They are needed to evaluate the correction for a photon-emitting impurity present in an ampoule when the efficiency for this impurity is not known experimentally [73], or to give a point of comparison when a radionuclide is measured in the SIR for the first time. Each SIR measurement may be considered as a determination of the efficiency of the ionization chamber for a given radionuclide [83]. In consequence the whole set of SIR measurement data may be used, in principle, to establish the required efficiency curves.

A model-based approach is proposed [74] to obtain estimates of the efficiency curves from the set of SIR measurements. The approach uses (a) a least-squares formulation that allows

for the presence of radioactive impurities [73] and aims to account for available physical information and measurement uncertainties, and (b) families of empirical functions, viz., polynomials in the logarithm of photon or beta energy, to represent the efficiency curves.

Let $\boldsymbol{D}$ denote the measured quantities, viz., decay-corrected activity, for which $\boldsymbol{d}$ are the available measured values (estimates) with associated covariance matrix $\boldsymbol{U_d}$. Furthermore, let $\boldsymbol{F} = \boldsymbol{F}(\boldsymbol{B}, \boldsymbol{N})$ denote quantities provided by a model for the measured quantities $\boldsymbol{D}$ expressed in terms of quantities $\boldsymbol{B}$, representing the adjustable parameters of the efficiency curves, and $\boldsymbol{N}$, representing nuclear reference data including photon energies and photon emission probabilities, and also including decay-corrected impurity activities. Tabulated values $\boldsymbol{n}$ of $\boldsymbol{N}$ with the associated covariance matrix $\boldsymbol{U_n}$ are available. Estimates $\boldsymbol{b}$ of $\boldsymbol{B}$ with the associated covariance matrix $\boldsymbol{U_b}$ are to be evaluated. In this application, the model quantities $\boldsymbol{F}$ are complicated (non-linear) functions of the parameters $\boldsymbol{B}$ and reference data $\boldsymbol{N}$ [74], and are not reproduced here. However, the application provides an instance of a general problem in which a model is fitted to observed data (using least-squares) and the model depends on additional (reference) data for which estimates are available, either from observation or as tabulated values.

A formulation of the generalized least-squares problem to be solved is

$$\min_{\boldsymbol{B},\boldsymbol{N}} \boldsymbol{s}^{\top}(\boldsymbol{B}, \boldsymbol{d}, \boldsymbol{N}) \boldsymbol{U_d}^{-1} \boldsymbol{s}(\boldsymbol{B}, \boldsymbol{d}, \boldsymbol{N}) + (\boldsymbol{n} - \boldsymbol{N})^{\top} \boldsymbol{U_n}^{-1} (\boldsymbol{n} - \boldsymbol{N}) \tag{9.38}$$

where

$$\boldsymbol{s}(\boldsymbol{B}, \boldsymbol{D}, \boldsymbol{N}) = \boldsymbol{D} - \boldsymbol{F}(\boldsymbol{B}, \boldsymbol{N})$$

are the residual deviations associated with the model. The function to be minimized in expression (9.38) is purely additive, since the (measured) quantities ($\boldsymbol{D}$ and $\boldsymbol{N}$, respectively) represented by the two terms in the sum are regarded as mutually independent.

The approach formulates *a priori* the problem in a manner that respects the knowledge of the uncertainties associated with *all* relevant effects. The result of the computation would be new (adjusted) estimates $\boldsymbol{n}^*$, say, of $\boldsymbol{N}$, which would generally be different from the tabulated estimates $\boldsymbol{n}$, in addition to the provision of the efficiency curve parameter estimates $\boldsymbol{b}$. In the case of consistency between model and data, $\boldsymbol{n}^*$ 'should' arguably be used as 'replacements' for the tabulated values $\boldsymbol{n}$. Politically and logistically, however, difficulties might be encountered. It would be unreasonable to expect, for example, that the tables of the transition energies and probabilities would be updated on every occasion a (statistically consistent) model fit was made. The formulation also generates a demanding problem computationally. It constitutes a non-linear least-squares problem with a number of adjustable parameters equal not just to the dimension of $\boldsymbol{B}$ (which is typically of order 10), but to the dimension of $\boldsymbol{B}$ *and* $\boldsymbol{N}$ (of order 1000).

In the case of consistency between model and data, however, the general formulation (9.38) can be expected to provide estimates $\boldsymbol{b}$ of the efficiency curve parameters that differ only slightly from those provided by the (reduced) least-squares problem

$$\min_{\boldsymbol{B}} \boldsymbol{s}^{\top}(\boldsymbol{B}, \boldsymbol{d}, \boldsymbol{n}) \boldsymbol{U_d}^{-1} \boldsymbol{s}(\boldsymbol{B}, \boldsymbol{d}, \boldsymbol{n}), \tag{9.39}$$

in which $\boldsymbol{N}$ is regarded as fixed and equal to the estimates $\boldsymbol{n}$ of these quantities. The formulation (9.39) also constitutes a non-linear least-squares problem, but one that is computationally less demanding than that defined in formulation (9.38).

The uncertainties associated with the estimates $\boldsymbol{b}$ of the model parameters $\boldsymbol{B}$ can be formed from the information provided by the algorithm used to solve the non-linear least-squares problem (9.39). This information takes the form of a covariance matrix [3]

$$\boldsymbol{U_b} = \left(\boldsymbol{J_s}^{\top}(\boldsymbol{b}) \boldsymbol{U_d}^{-1} \boldsymbol{J_s}(\boldsymbol{b}) \right)^{-1},$$

where $\boldsymbol{J_s}(\boldsymbol{B})$ is the (Jacobian) matrix containing the partial derivatives of first order of the residual deviations $\boldsymbol{s}$ with respect to the model parameters $\boldsymbol{B}$, i.e.,

$$\boldsymbol{J_s}(\boldsymbol{B}) = \frac{\partial \boldsymbol{s}(\boldsymbol{B}, \boldsymbol{d}, \boldsymbol{n})}{\partial \boldsymbol{B}}.$$

However, there are further sources of uncertainty that are to be taken into account and that influence the uncertainties associated with the estimates of the model parameters. These sources relate to the nuclear reference data $\boldsymbol{N}$. We describe below how these further uncertainties can be taken into account using a Monte Carlo method and the GUM uncertainty framework.

The solution obtained as described corresponds to the use of the tabular estimates $\boldsymbol{n}$ of $\boldsymbol{N}$. These estimates are regarded as the expectation values of the quantities $\boldsymbol{N}$ characterized by certain probability distributions. Such distributions could be, for example, Gaussian with expectations equal to the estimates and standard deviations equal to the standard uncertainties associated with these estimates. In the application of a Monte Carlo method, other valid realizations of these quantities would be given by making random draws from these distributions. Solutions corresponding to such realizations could be constructed in the same way as for the estimates, i.e., by solving a problem of the form (9.39) in which $\boldsymbol{n}$ is replaced by a realization of $\boldsymbol{N}$. A large number of such solutions would provide approximate distributions (including joint distributions) for the model parameters $\boldsymbol{B}$, from which the associated uncertainties could be deduced (Section 7.4). This approach would be computer intensive since each Monte Carlo trial would require the numerical solution of a non-linear least-squares problem, and a large number, say 10^5, trials would be required to a give a reasonable assurance of a valid result.

At a solution to the minimization problem (9.39), the partial derivatives with respect to the adjustable quantities $\boldsymbol{B}$are zero. Thus, the solution to formulation (9.39) satisfies

$$\boldsymbol{h}(\boldsymbol{B}, \boldsymbol{d}, \boldsymbol{n}) \equiv \left(\frac{\partial \boldsymbol{F}(\boldsymbol{B}, \boldsymbol{d}, \boldsymbol{n})}{\partial \boldsymbol{B}} \right)^{\top} \boldsymbol{U_d}^{-1} \left[\boldsymbol{d} - \boldsymbol{F}(\boldsymbol{B}, \boldsymbol{n}) \right] = \boldsymbol{0}. \tag{9.40}$$

Consider the counterpart of formulation (9.39) for the quantities $\boldsymbol{D}$ and $\boldsymbol{N}$ rather than the estimates $\boldsymbol{d}$ and $\boldsymbol{n}$, i.e,

$$\min_{\boldsymbol{B}} \boldsymbol{s}^{\top}(\boldsymbol{B}, \boldsymbol{D}, \boldsymbol{N}) \boldsymbol{U_d}^{-1} \boldsymbol{s}(\boldsymbol{B}, \boldsymbol{D}, \boldsymbol{N}).$$

The expression corresponding to expression (9.40) is then

$$\boldsymbol{h}(\boldsymbol{B},\boldsymbol{D},\boldsymbol{N}) \equiv \left(\frac{\partial \boldsymbol{F}(\boldsymbol{B},\boldsymbol{D},\boldsymbol{N})}{\partial \boldsymbol{B}}\right)^{\top} \boldsymbol{U_d}^{-1}(\boldsymbol{D}-\boldsymbol{F}(\boldsymbol{B},\boldsymbol{N})) = \boldsymbol{0}.$$

This expression provides a 'measurement model' relating 'input quantities' $\boldsymbol{D}$ and $\boldsymbol{N}$ to 'output quantities' $\boldsymbol{B}$ that may be used as the basis for applying the GUM uncertainty framework. The measurement model is categorized as a *multivariate, real measurement model* (Section 6.2.4), and it follows from the law of propagation of uncertainty applied to such models that $\boldsymbol{U_b}$ satisfies the linear vector equation

$$\boldsymbol{J_h}(\boldsymbol{b})\boldsymbol{U_b}\boldsymbol{J_h}^{\top}(\boldsymbol{b}) = \boldsymbol{J_h}(\boldsymbol{d})\boldsymbol{U_d}\boldsymbol{J_h}^{\top}(\boldsymbol{d}) + \boldsymbol{J_h}(\boldsymbol{n})\boldsymbol{U_n}\boldsymbol{J_h}^{\top}(\boldsymbol{n}),$$

in which $\boldsymbol{J_h}(\boldsymbol{d})$, $\boldsymbol{J_h}(\boldsymbol{n})$ and $\boldsymbol{J_h}(\boldsymbol{b})$ are Jacobian matrices given by

$$\boldsymbol{J_h}(\boldsymbol{D}) = \frac{\partial \boldsymbol{h}(\boldsymbol{B},\boldsymbol{D},\boldsymbol{N})}{\partial \boldsymbol{D}}, \quad \boldsymbol{J_h}(\boldsymbol{N}) = \frac{\partial \boldsymbol{h}(\boldsymbol{B},\boldsymbol{D},\boldsymbol{N})}{\partial \boldsymbol{N}}, \quad \boldsymbol{J_h}(\boldsymbol{B}) = \frac{\partial \boldsymbol{h}(\boldsymbol{B},\boldsymbol{D},\boldsymbol{N})}{\partial \boldsymbol{B}},$$

evaluated at $\boldsymbol{D} = \boldsymbol{d}$, $\boldsymbol{N} = \boldsymbol{n}$ and $\boldsymbol{B} = \boldsymbol{b}$. Finally, $\boldsymbol{B}$ with expectation $\boldsymbol{b}$ and covariance matrix $\boldsymbol{U_b}$ is characterized by a multivariate Gaussian distribution.

Chapter 10

Recommendations

The 'Guide to the expression of uncertainty in measurement' (GUM) [10] provides recommendations that have been internationally agreed for the evaluation of measurement uncertainties. Central to the GUM is a measurement model with input quantities, characterized by probability distributions, and an output quantity, also characterized by a probability distribution. The use of the GUM permits the uncertainty associated with an estimate of the output quantity to be evaluated. In particular, an interval (termed here a coverage interval) that can be expected to encompass a specified fraction of the distribution of values that could reasonably be attributed to the output quantity can be obtained.

It will always be necessary to make some assertions about the measurement uncertainties associated with the estimates of the input quantities in the measurement model. That is the metrologist's task. The metrologist needs to make statements, using expert judgement if necessary, about what he believes, and those statements provide the basis for the analysis, until better statements become available. After all, he is best-placed to do this. If everything is recorded, the reported uncertainty can be defended in that light.

Arguably, the worst-case scenario is when the metrologist genuinely feels he knows nothing about an input quantity other than lower and upper limits on the deviation of the quantity from the estimate. (If he cannot even quote that, the uncertainty evaluation cannot be progressed at all!) In this situation the Principle of Maximum Entropy would imply that a rectangular PDF should be assigned to the quantity, based on the limits.

In general, it is recommended that all input quantities are characterized in terms of probability density functions (PDFs). By doing so the metrologist is able to incorporate to the maximum his degree of belief in his knowledge of the various input quantities. In particular, if little or very little information is available, appeal to the Principle of Maximum Entropy and Bayes' theorem permits a defensible PDF to be provided.

Once the measurement model and the PDFs for the input quantities are in place the PDF for the output quantity is completely defined. It is possible to use a number of approaches for

determining the PDF for the output quantity and thence an estimate of that quantity and an associated standard uncertainty, and a coverage interval or coverage region for the quantity. The attributes of the various approaches considered, all in a sense covered or implied by the GUM, are to be taken into account when selecting whether to apply the GUM uncertainty framework, a Monte Carlo method or other analytical or numerical methods.[1]

Validation of the approach used is important in cases of doubt. The use of the described Monte Carlo method to validate the results obtained from the GUM uncertainty framework is urged when it is unclear whether the latter is applicable in a certain situation. The described Monte Carlo method can also be seen as a widely applicable tool for uncertainty evaluation.

In the case that prior knowledge of the output quantity is also available as a proper PDF, such additional information can also be used in the determination of the PDF for the output quantity. The use of Bayes' theorem provides an approach to uncertainty evaluation that allows proper account to be taken of such prior knowledge.

> [GUM, Clause 0.4] The actual quantity used to express uncertainty should be:
>
> 1. Internally consistent: it should be directly derivable from the components that contribute to it, as well as independent of how these components are grouped and of the decomposition of the components into subcomponents;
> 2. Transferable: it should be possible to use directly the uncertainty evaluated for one result as a component in evaluating the uncertainty of another measurement in which the first result is used.
> 3. ... it is often necessary to provide an interval about the measurement result that may be expected to encompass a large fraction of the distribution of values that could reasonably be attributed to the quantity subject to measurement. Thus the ideal method for evaluating and expressing uncertainty in measurement should be capable of readily providing such an interval, in particular, one with a coverage probability or level of probability that corresponds in a realistic way with that required.

These are laudable properties and objectives. It is reasonable to summarize them and to infer further aims as follows:

1. All information used to evaluate uncertainties is to be recorded.
2. The sources of the information are to be recorded.
3. Any assumptions or assertions made are to be stated.

[1] The Monte Carlo method given in Supplement 1 [11] to the GUM is, like the GUM itself, a consequence of international agreement on the evaluation of measurement uncertainty.

4. The measurement model and its input quantities are to be provided in a manner that maximizes the use of this information consistent with the assumptions made.

5. Uncertainties are to be evaluated in a manner that is consistent with quality management systems and, in particular, the results of the evaluation are to be fit for purpose.

6. If the same information is provided to different bodies, the uncertainties these bodies calculate for the required results should agree to within a stipulated numerical accuracy.

7. Difficulties in handling sparse or scarce information are to be addressed by making alternative, equally plausible assumptions, and re-evaluating the uncertainties for the required results to obtain knowledge of the variability due to this source.

The intention of this guide has been to address these aims as far as reasonably possible. Further, the three-stage approach advocated in Section 3.2 and followed in the rest of this guide supports the first point from GUM, Clause 0.4. The approach to multi-stage measurement models recommended here supports the second point. Finally, the mathematical formulation and the attitude of this guide supports the third point, through the solution approaches of Chapter 5.

Bibliography

[1] AFNOR. Aid in the procedure for estimating and using uncertainly in measurements and test results. Technical Report FD X 07-021, AFNOR, 2000. Informal English translation.

[2] Analytical Methods Committee. Uncertainty of measurement: implications of its use in analytical science. *Analyst*, 120:2303–2308, 1995.

[3] R. M. Barker, M. G. Cox, A. B., Forbes, and P. M. Harris. Best Practice Guide No. 4. Discrete modelling and experimental data analysis. Technical report DEM-ES 018, National Physical Laboratory, Teddington, UK, 2007.

[4] V. Barnett. The ordering of multivariate data. *J. R. Statist. Soc. A*, 139:318, 1976.

[5] M. Basil and A. W. Jamieson. Uncertainty of complex systems by Monte Carlo simulation. In *North Sea Flow Measurement Workshop, Gleneagles, 26-29 October 1998*, 1998.

[6] R. W. Beatty. Insertion loss concepts. *Proc. IEEE*, 52:663–671, 1964.

[7] Stephanie Bell. Measurement Good Practice Guide No. 11. A Beginner's Guide to Uncertainty of Measurement. Technical report, National Physical Laboratory, 1999.

[8] W. Bich, M. G. Cox, and P. M. Harris. Evolution of the "Guide to the Expression of Uncertainty in Measurement". *Metrologia*, 43(4):S161–S166, 2006.

[9] BIPM. Mutual recognition of national measurement standards and of calibration and measurement certificates issued by national metrology institutes. Technical report, Bureau International des Poids et Mesures, Sèvres, France, 1999.

[10] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP and OIML. *Guide to the expression of uncertainty in measurement* (GUM:1995 with minor corrections). Bureau International des Poids et Mesures, JCGM 100:2008.

[11] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP and 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*. Bureau International des Poids et Mesures, JCGM 101:2008.

[12] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. *Evaluation of measurement data—An introduction to the "Guide to the expression of uncertainty in measurement" and related documents*. Bureau International des Poids et Mesures, JCGM 104:2009.

[13] R. Boudjemaa, M. G. Cox, A. B. Forbes, and P. M. Harris. Automatic differentiation techniques and their application in metrology. Technical Report CMSC 26/03, National Physical Laboratory, Teddington, UK, 2003.

[14] E. O. Brigham. *The Fast Fourier Transform*. Prentice-Hall, New York, 1974.

[15] A. S. Brown, M. J. T. Milton, C. J. Cowper, G. D. Squire, W. Bremser, and R. W. Branch. Analysis of natural gas by gas chromatography reduction of correlated uncertainties by normalisation. *J. Chromatogr. A*, 1040:215–225, 2004.

[16] P. J. Campion, J. E. Burns, and A. Williams. A code of practice for the detailed statement of accuracy. Technical report, National Physical Laboratory, Teddington, UK, 1980.

[17] G. Casella and R. L. Berger. *Statistical Inference*. Wadsworth and Brooks, Pacific Grove, Ca., USA, 1990.

[18] P. Ciarlini, M. G. Cox, E. Filipe, F. Pavese, and D. Richter, editors. *Advanced Mathematical and Computational Tools in Metrology V. Series on Advances in Mathematics for Applied Sciences Vol. 57*, Singapore, 2001. World Scientific.

[19] P. Ciarlini, M. G. Cox, R. Monaco, and F. Pavese, editors. *Advanced Mathematical and Computational Tools in Metrology. Series on Advances in Mathematics for Applied Sciences Vol. 16*, Singapore, 1994. World Scientific.

[20] P. Ciarlini, M. G. Cox, F. Pavese, and D. Richter, editors. *Advanced Mathematical and Computational Tools in Metrology II. Series on Advances in Mathematics for Applied Sciences Vol. 40*, Singapore, 1996. World Scientific.

[21] P. Ciarlini, M. G. Cox, F. Pavese, and D. Richter, editors. *Advanced Mathematical and Computational Tools in Metrology III. Series on Advances in Mathematics for Applied Sciences Vol. 45*, Singapore, 1997. World Scientific.

[22] P. Ciarlini, M. G. Cox, F. Pavese, and G. B. Rossi, editors. *Advanced Mathematical and Computational Tools in Metrology VI. Series on Advances in Mathematics for Applied Sciences Vol. 66*, Singapore, 2004. World Scientific.

[23] P. Ciarlini, E. Filipe, A. B. Forbes, F. Pavese, C. Perruchet, and B. R. L. Siebert, editors. *Advanced Mathematical and Computational Tools in Metrology VII. Series on Advances in Mathematics for Applied Sciences Vol. 72*, Singapore, 2006. World Scientific.

[24] P. Ciarlini, A. B. Forbes, F. Pavese, and D. Richter, editors. *Advanced Mathematical and Computational Tools in Metrology IV. Series on Advances in Mathematics for Applied Sciences Vol. 53*, Singapore, 2000. World Scientific.

[25] S. Cowen and S. L. R. Ellison. Reporting measurement uncertainty and coverage intervals near natural limits. *Analyst*, 131:710–717, 2006.

[26] M. G. Cox. The numerical evaluation of B-splines. *J. Inst. Math. Appl.*, 10:134–149, 1972.

[27] M. G. Cox. The area under a curve specified by measured values *Metrologia*, 44:365–378, 2007.

[28] M. G. Cox, M. P. Dainton, A. B. Forbes, P. M. Harris, P. M. Schwenke, B. R. L Siebert, and W. Wöger. Use of Monte Carlo simulation for uncertainty evaluation in metrology. In P. Ciarlini, M. G. Cox, E. Filipe, F. Pavese, and D. Richter, editors, *Advanced Mathematical Tools in Metrology V. Series on Advances in Mathematics for Applied Sciences Vol. 57*, pages 93–104, Singapore, 2001. World Scientific.

[29] M. G. Cox, M. P. Dainton, P. M. Harris, and N. M. Ridler. The evaluation of uncertainties in the analysis of calibration data. In *Proceedings of the 16th IEEE Instrumentation and Measurement Technology Conference*, pages 1093–1098, Venice, 24-26 May 1999, 1999.

[30] M. G. Cox, A. B. Forbes, P. M. Harris, and I. M. Smith. The classification and solution of regression problems for calibration. Technical Report CMSC 24/03, National Physical Laboratory, Teddington, UK, 2003.

[31] M. G. Cox, A. B. Forbes, P. M. Harris, and J. Sousa. Accounting for physical knowledge in obtaining measurement results and associated uncertainties. In *Proceedings of XVIII IMEKO World Congress, September 17–22, 2006, Rio de Janeiro, Brazil*, 2006.

[32] M. G. Cox and P. M. Harris. Design and use of reference data sets for testing scientific software. *Analytica Chimica Acta*, 380:339–351, 1999.

[33] M. G. Cox, P. M. Harris, and I. M. Smith. GUM Supplement 1 Software: User Manual. Technical Report MS-1, National Physical Laboratory, Teddington, UK, 2008.

[34] M. G. Cox, P. M. Harris, and I. M. Smith. Software specifications for uncertainty evaluation. Technical Report MS 7, National Physical Laboratory, Teddington, UK, 2010.

[35] T. J. Dekker. Finding a zero by means of successive linear interpolation. In B. Dejon and P. Henrici, editors, *Constructive Aspects of the Fundamental Theorem of Algebra*, London, 1969. Wiley Interscience.

[36] C. F. Dietrich. *Uncertainty, Calibration and Probability*. Adam Hilger, Bristol, UK, 1991.

[37] DIN. DIN 1319 Teil 4 1985 Grundbegriffe der Meßtechnik; Behandlung von Unsicherheiten bei der Auswertung von Messungen, 1985.

[38] EA. Expression of the uncertainty of measurement in calibration. Technical Report EA-4/02, European Co-operation for Accreditation, 1999.

[39] B. Efron. *The Jackknife, the Bootstrap and Other Resampling Plans*. Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, 1982.

[40] B. Efron and R. Tibshirani. *An Introduction to the Bootstrap*. Monographs on Statistics and Applied Probability 57. Chapman and Hall, New York, 1993.

[41] C. Elster. Calculation of uncertainty in the presence of prior knowledge. *Metrologia*, 44:111–116, 2007.

[42] C. Elster and B. Toman. Bayesian uncertainty analysis under prior ignorance of the measurand versus analysis using Supplement 1 to the *Guide*: a comparison. *Metrologia*, 46:161–166, 2009.

[43] C. Elster, W. Wöger, and M. G. Cox. Draft GUM Supplement 1 and Bayesian analysis. *Metrologia*, 44:L31–32, 2007.

[44] G. F. Engen. *Microwave Circuit Theory and Foundations of Microwave Metrology*. Peter Peregrinus, London, 1992.

[45] T. J. Esward, A. de Ginestous, P. M. Harris, I. D. Hill, I. M. Smith, B. A. Wichmann, and E. R. Woolliams. A Monte Carlo method for uncertainty evaluation implemented on a distributed computing system. *Metrologia*, 44:319–326, 2007.

[46] EURACHEM/CITAC. Quantifying uncertainty in analytical measurement. Technical Report Guide CG4, EURACHEM/CITEC, 2000. Second edition.

[47] A. B. Forbes, M. G. Cox, G. B. Rossi, and P.M. Harris. A probabilistic approach to the analysis of measurement processes. *Metrologia*, 45:493–502, 2008.

[48] M. J. Fryer. A review of some non-parametric methods of density estimation. *J. Inst. Maths. Applics.*., 20:335–354, 1977.

[49] J. L. Gardner. Correlations in primary spectral standards. *Metrologia*, 40:S167–S171, 2003.

[50] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian Data Analysis*. Chapman and Hall, 2004.

[51] P. E. Gill, W. Murray, and M. H. Wright. *Practical Optimization*. Academic Press, London, 1981.

[52] G. H. Golub and C. F. Van Loan. *Matrix Computations*. John Hopkins University Press, Baltimore, MD, USA, 1996. Third edition.

[53] B. D. Hall. Computer modelling of uncertainty calculations with finite degrees of freedom. *Meas. Sci. Technol.*, 11:1335–1341, 2000.

[54] B. D. Hall and R. Willink. Does Welch-Satterthwaite" make a good uncertainty estimate? *Metrologia*, 38:9–15, 2001.

[55] J. M. Hammersley and D. C. Handscomb. *Monte Carlo Methods*. Wiley, New York, 1964.

[56] W. Hässelbarth and W. Bremser. Correlation in natural gas composition data. In *Proc. Int. Gas Research Conf. (Amsterdam)*, 2001.

[57] W. Hässelbarth and W. Bremser. Measurement uncertainty for multiple measurands: characterization and comparison of uncertainty matrices. *Metrologia*, 44:128–145, 2007.

[58] ISO. ISO Guide 35 - Certification of reference materials - general and statistical principles, 1989. International Standards Organization, Geneva.

[59] ISO. ISO 3534-1:1993. Statistics - Vocabulary and symbols - Part 1: Probability and general statistical terms, 1993. International Standards Organization, Geneva.

[60] E. T. Jaynes. Information theory and statistical mechanics. *Phys. Rev*, 106:620–630, 1957.

[61] E. T. Jaynes. Where do we stand on maximum entropy? In R. D. Rosenkrantz, editor, *Papers on Probability, Statistics, and Statistical Physics*, pages 210–314, Dordrecht, The Netherlands, 1989. Kluwer Academic.

[62] R. Kacker. An interpretation of the Guide to the Expression of Uncertainty in Measurement. In *NCSL-2000 proceedings, Toronto*, 2000.

[63] D. M. Kerns and R. W. Beatty. *Basic Theory of Waveguide Junctions and Introductory Microwave Network Analysis*. Pergamon Press, London, 1967.

[64] M. J. Korczynski, M. G. Cox, and P. M. Harris. Convolution and uncertainty evaluation. In P. Ciarlini, E. Filipe, A. B. Forbes, F. Pavese, C. Perruchet, and B. R. L. Siebert, editors, *Advanced Mathematical and Computational Tools in Metrology VII. Series on Advances in Mathematics for Applied Sciences Vol. 72*, pages 188–195, Singapore, 2006. World Scientific.

[65] T. Leonard and J. S. J. Hsu. *Bayesian Methods*. Cambridge University Press, Cambridge, 1999.

[66] S. Lewis and G. Peggs. *The Pressure Balance: A Practical Guide to its Use. Second edition*. HMSO, London, 1991.

[67] I. Lira. *Evaluating the Uncertainty of Measurement. Fundamentals and Practical Guidance*. Institute of Physics, Bristol, UK, 2002.

[68] I. Lira and W. Wöger. Comparison between the conventional and Bayesian approaches to evaluate measurement data. *Metrologia*, 43(4):S249–S259, 2006.

[69] J. N. Lyness and C. B. Moler. Numerical differentiation of analytic functions. *SIAM J. Numer. Anal.*, 4:202–210, 1967.

[70] K. V. Mardia, J. T. Kent, and J. M. Bibby. *Multivariate Analysis*. Academic Press, London, 1979.

[71] J. R. R. A. Martins, P. Sturdza, and J. J. Alonso. The complex-step derivative approximation. *ACM Trans. Math. Software*, 29:245–262, 2003.

[72] MATLAB. The Mathworks, Inc. www.mathworks.co.uk/products/matlab/.

[73] C. Michotte. Influence of radioactive impurities on SIR measurements. *Appl. Rad. Isot.*, 52:319–323, 2000.

[74] C. Michotte, A. K. Pearce, M. G. Cox, and J.-J. Gostely. An approach based on the SIR measurement model for determining ionization chamber efficiency curves, and a study of ^{65}Zn and ^{201}Tl photon emission intensities. *Applied Radiation and Isotopes*, 64:1147–1155, 2006.

[75] M. J. T. Milton, P. M. Harris, A. S. Brown, and C. J. Cowper. Normalization of natural gas composition data measured by gas chromatography. *Meas. Sci. Tech.*, 20, 2009.

[76] M. J. T. Milton, P. M. Harris, A. S. Brown, and C. J. Cowper. Normalization of natural gas composition data measured by gas chromatography (Corrigendum). *Meas. Sci. Tech.*, 20, 2009.

[77] M. E. Müller. *Computational Aspects of Measurement Uncertainty Calculation*. Dissertation ETH No. 18763, Swiss Federal Institute of Technology, Zürich, 2009.

[78] OIML. Conventional value of the result of weighing in air. Technical Report OIML D 28, Organisation International de Métrologie Légale, Paris, 2004.

[79] F. Pavese, M. Bär, A.B. Forbes, J.M. Linares, C. Perruchet, and N.F. Zhang, editors. *Advanced Mathematical and Computational Tools in Metrology and Testing VIII. Series on Advances in Mathematics for Applied Sciences Vol. 78*, Singapore, 2009. World Scientific.

[80] A. Possolo and B. Toman. Assessment of measurement uncertainty via observation equations. *Metrologia*, 44:464–475, 2007.

[81] W. Press and S. A. Teukolsky. Numerical calculation of derivatives. *Computers in Physics*, 5, 1991.

[82] R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. www.r-project.org/.

[83] D. F. G. Reher, M. J. Woods, B. R. S. Simpson, and G. Ratel. Portability of the calibration of SIR of BIPM to other ionization chambers for radioactivity measurements. *Appl. Rad. Isot.*, 499:1417–1419, 1998.

[84] J. R. Rice. *Mathematical Statistics and Data Analysis*. Duxbury Press, Belmont, Ca., USA, second edition, 1995.

[85] N. M. Ridler and M. J. Salter. Propagating S–parameter uncertainties to other measurement quantities. In *58th ARFTG (Automatic RF Techniques Group) Conference Digest*, 2001.

[86] C. P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer-Verlag, New York, 1999.

[87] G. B. Rossi, F. Crenna, M. G. Cox, and P. M. Harris. Combining direct calculation and the Monte Carlo method for the probabilistic expression of measurement results. In P. Ciarlini, E. Filipe, A. B. Forbes, F. Pavese, C. Perruchet, and B. R. L. Siebert, editors, *Advanced Mathematical and Computational Tools in Metrology VII. Series on Advances in Mathematics for Applied Sciences Vol. 72*, pages 221–228, Singapore, 2006. World Scientific.

[88] A. Rytz. The international reference system for activity measurements of γ-ray emitting nuclides. *Appl. Rad. Isot*, 34:1047–1056, 1983.

[89] J. G. Saw, M. C. K. Tang, and T. C. Mo. Chebychev's inequality with estimated mean and variance. *American Statistician*, 38:130–132, 1984.

[90] H. Schwenke, B. R. L. Siebert, F. Wäldele, and H. Kunzmann. Assessment of uncertainties in dimensional metrology by Monte Carlo simulation: proposal of a modular and visual software. *Annals CIRP*, 49:395–398, 2000.

[91] R. S. Scowen. QUICKERSORT, Algorithm 271. *Comm. ACM*, 8:669, 1965.

[92] C. E. Shannon and W. Weaver. *The Mathematical Theory of Communication*. University of Illinois, Urbana-Champagne, 1949.

[93] B. R. L. Siebert. Discussion of methods for the assessment of uncertainties in Monte Carlo particle transport calculations. In P. Ciarlini, A. B. Forbes, F. Pavese, and D. Richter, editors, *Advanced Mathematical and Computational Tools in Metrology IV. Series on Advances in Mathematics for Applied Sciences Vol. 53*, pages 220–229, Singapore, 2000. World Scientific.

[94] C. G. Small. A survey of multidimensional medians. *International Statistical Review*, 58:263, 1990.

[95] C. R. Smith and W. T. Grandy, editors. *Maximum-Entropy and Bayesian methods in inverse problems*, Dordrecht, 1985. Reidel.

[96] P. I. Somlo and J. D. Hunter. *Microwave Impedance Measurement*. Peter Peregrinus, London, 1985.

[97] J. A. Sousa, A. S. Ribeiro, A. B. Forbes, P. M. Harris, F. Carvalho, and L. Bacelar. The relevance of using a Monte Carlo method to evaluate uncertainty in mass calibration. In IMEKO TC3, 16 and 22, International Conference *Cultivating metrological knowledge*, November 2630, 2007, Merica, Mexico, 2007.

[98] W. Squire and G. Trapp. Using complex variables to estimate derivatives of real functions. *SIAM Rev.*, 40:110–112, 1998.

[99] B. N. Taylor and C. E. Kuyatt. Guidelines for evaluating and expressing the uncertainty of NIST measurement results. Technical Report TN1297, National Institute of Standards and Technology, USA, 1994.

[100] UKAS. The expression of uncertainty and confidence in measurement. Technical Report M 3003, United Kingdom Accreditation Service, 1997.

[101] UKAS. The expression of uncertainty in EMC testing. Technical Report LAB34, United Kingdom Accreditation Service, 2002.

[102] P. A. G. van der Geest. An algorithm to generate samples of multi-variate distributions with correlated marginals. *Comput. Stat. Data Anal.*, 27:271–289, 1998.

[103] G. Vargha, M. J. T. Milton, M. G. Cox, and S. Kamvissis. Harmonisation of coupled calibration curves to reduce correlated effects in the analysis of natural gas by gas chromatography. *J. Chromatogr. A*, 1062:239–245, 2005.

[104] K. Weise and W. Wöger. A Bayesian theory of measurement uncertainty. *Meas. Sci. Technol.*, 3:1–11, 1992.

[105] WinBUGS. Imperial College and Medical Research Council, UK. www.mrc-bsu.cam.ac.uk/bugs.

[106] W. Wöger. Probability assignment to systematic deviations by the Principle of Maximum Entropy. *IEEE Trans. Instr. Measurement*, IM-36:655–658, 1987.

[107] M. Wolf. *A Modelling Language for Measurement Uncertainty Evaluation*. Dissertation ETH No. 18676, Swiss Federal Institute of Technology, Zürich, 2009.

[108] G. Wübbeler, P. M. Harris, M. G. Cox, and C. Elster. A two-stage procedure for determining the number of trials in the application of a monte carlo method for uncertainty evaluation. *Metrologia*, 2010. Submitted.

Appendix A

Some statistical concepts

Some statistical concepts used in this guide are reviewed [59]. The concepts of random variables and probability distributions are especially important because quantities in a measurement model are treated as random variables and information about the quantities encoded as probability distributions.

A.1 Discrete random variables

A *discrete random variable* X is a variable that can take only a finite number of possible values. If X is the number of heads in an experiment consisting of tossing three coins, X can take (only) the value 0, 1, 2 or 3.

The *frequency function* $p_X(\xi)$ states the probabilities of occurrence of the possible outcomes:

$$p_X(\xi) = \Pr(X = \xi),$$

the probability that the outcome is ξ. For the coin tossing experiment,

$$p_X(0) = \frac{1}{8}, \quad p_X(1) = \frac{3}{8}, \quad p_X(2) = \frac{3}{8}, \quad p_X(3) = \frac{1}{8}.$$

The probabilities can be deduced by enumerating all $2 \times 2 \times 2 = 8$ possible outcomes arising from the fact that each coin can only land in one of two equally likely ways, or by using the binomial distribution [84, p36] that applies to the analysis of such probability problems.

The *distribution function* $G_X(\xi)$ gives the probability that a random variable takes a value no greater than a specified value:

$$G_X(\xi) = \Pr(X \leq \xi), \quad -\infty < \xi < \infty.$$

For the coin-tossing experiment,

$$\begin{aligned} G_X(\xi < 0) &= 0, \\ G_X(0 \leq \xi < 1) &= \frac{1}{8}, \\ G_X(1 \leq \xi < 2) &= \frac{1}{2}, \\ G_X(2 \leq \xi < 3) &= \frac{7}{8}, \\ G_X(3 \leq \xi) &= 1. \end{aligned}$$

The distribution function varies from zero to one throughout its range, never decreasing.

The probability that X lies in an interval $[a,\ b]$ is

$$\Pr(a < X \leq b) = G_X(b) - G_X(a).$$

Two important statistics associated with a discrete random variable are its expectation and standard deviation.

Let $\xi_1, \xi_2, \ldots$ denote the possible values of X and $p_X(\xi)$ the frequency function for X. The *expectation* μ of X is

$$\mu = \sum_i \xi_i p_X(\xi_i).$$

It is a measure of the location of X.

The *standard deviation* σ of X with frequency function $p_X(\xi)$ and expectation μ is the square root of the *variance*

$$\sigma^2 = \sum_i (\xi_i - \mu)^2 p_X(\xi_i).$$

It is a measure of the spread or dispersion of X.

A.2 Continuous random variables

A.2.1 Scalar continuous random variables

A *continuous random variable* X is a variable that can take any value within a certain interval. For example, for a machine capable of weighing any person up to $150\,\mathrm{kg}$, the indicated weight can take any value in the interval $0\,\mathrm{kg}$ to $150\,\mathrm{kg}$.

For a continuous random variable X, the counterpart of the frequency function (for a discrete random variable) is the *probability density function* (PDF) $g_X(\xi)$, where ξ denotes possible values of X. This function has the property that the probability that X lies between a and b is

$$\Pr(a < X < b) = \int_a^b g_X(\xi)\,\mathrm{d}\xi.$$

Since a random variable X must take *some* value, $g_X(\xi)$ has the property that

$$\int_{-\infty}^{\infty} g_X(\xi)\,\mathrm{d}\xi = 1.$$

The rectangular probability distribution has a PDF that describes the fact that the value of X is equally likely to lie anywhere in an interval $[a,\ b]$:

$$g_X(\xi) = \begin{cases} \frac{1}{b-a}, & a \le \xi \le b, \\ 0, & \text{otherwise.} \end{cases}$$

The *distribution function* $G_X(\xi)$ gives the probability that a random variable takes a value no greater than a specified value, and is defined as for a discrete random variable:

$$G_X(\xi) = \Pr(X \le \xi), \quad -\infty < \xi < \infty.$$

The distribution function can be expressed in terms of the probability density function as

$$G_X(\xi) = \int_{-\infty}^{\xi} g_X(v)\,\mathrm{d}v.$$

The expectation of a continuous random variable X with PDF $g_X(\xi)$ is

$$E(X) = \int_{-\infty}^{\infty} \xi g_X(\xi)\,\mathrm{d}\xi.$$

It is often denoted by μ.

The variance of a continuous random variable X with PDF $g_X(\xi)$ and expectation $\mu = E(X)$ is

$$V(X) = \int_{-\infty}^{\infty} (\xi - \mu)^2 g_X(\xi)\,\mathrm{d}\xi.$$

The variance is often denoted by σ^2 and its square root is the standard deviation σ of X.

A.2.2 Vector continuous random variables

A *vector random variable* $\boldsymbol{X}$ is a set of scalar random variables arranged as a column vector:

$$\boldsymbol{X} = (X_1, \ldots, X_N)^\top.$$

For a vector random variable $\boldsymbol{X}$, the counterparts of the PDF and distribution function (for a scalar random variable) are the *joint* PDF $g_{\boldsymbol{X}}(\boldsymbol{\xi})$ and *joint* distribution function $G_{\boldsymbol{X}}(\boldsymbol{\xi})$ where $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_N)^\top$ denotes possible values of $\boldsymbol{X}$. The joint distribution function is a function giving, for every value $\boldsymbol{\xi}$, the probability that $\boldsymbol{X}$ is less than or equal to $\boldsymbol{\xi}$:

$$G_{\boldsymbol{X}}(\boldsymbol{\xi}) = \Pr(X_1 \le \xi_1, \ldots, X_N \le \xi_N).$$

The joint distribution function and joint PDF are related by

$$G_{\boldsymbol{X}}(\boldsymbol{\xi}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g_{\boldsymbol{X}}(\boldsymbol{z})\, \mathrm{d}z_N \cdots \mathrm{d}z_1.$$

The expectation of a random variable X_i, a component of $\boldsymbol{X}$ having joint PDF $g_{\boldsymbol{X}}(\boldsymbol{\xi})$, is

$$\begin{aligned} E(X_i) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \xi_i g_{\boldsymbol{X}}(\boldsymbol{\xi})\, \mathrm{d}\xi_N \cdots \mathrm{d}\xi_1 \\ &= \int_{-\infty}^{\infty} \xi_i g_{X_i}(\xi_i)\, \mathrm{d}\xi_i, \end{aligned}$$

where $g_{X_i}(\xi_i)$ is the *marginal* PDF for X_i (alone) given by

$$g_{X_i}(\xi_i) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g_{\boldsymbol{X}}(\boldsymbol{z})\, \mathrm{d}z_N \cdots \mathrm{d}z_{i+1} \mathrm{d}z_{i-1} \cdots \mathrm{d}z_1.$$

The variance of X_i, a component of $\boldsymbol{X}$ having joint PDF $g_{\boldsymbol{X}}(\boldsymbol{\xi})$, is

$$\begin{aligned} V(X_i) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} [\xi_i - E(X_i)]^2 g_{\boldsymbol{X}}(\boldsymbol{\xi})\, \mathrm{d}\xi_N \cdots \mathrm{d}\xi_1 \\ &= \int_{-\infty}^{\infty} [\xi_i - E(X_i)]^2 g_{X_i}(\xi_i)\, \mathrm{d}\xi_i. \end{aligned}$$

The *covariance* is a property of a pair of random variables X_i and X_j, components of $\boldsymbol{X}$, given by

$$\begin{aligned} \mathrm{Cov}(X_i, X_j) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} [\xi_i - E(X_i)][\xi_j - E(X_j)] g_{\boldsymbol{X}}(\boldsymbol{\xi})\, \mathrm{d}\xi_N \cdots \mathrm{d}\xi_1 \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [\xi_i - E(X_i)][\xi_j - E(X_j)] g_{X_i, X_j}(\xi_i, \xi_j)\, \mathrm{d}\xi_i \mathrm{d}\xi_j. \end{aligned}$$

The covariance measures the 'strength' of (linear) dependence between the two random variables.

The expectation of the vector random variable $\boldsymbol{X}$ is

$$E(\boldsymbol{X}) = (E(X_1), \ldots, E(X_N))^{\top},$$

the vector of expectations of the components X_i of $\boldsymbol{X}$.

The *covariance matrix* of $\boldsymbol{X}$ is the symmetric matrix $V(\boldsymbol{X})$ of dimension $N \times N$ containing the covariances $\mathrm{Cov}(X_i, X_j)$:

$$V(\boldsymbol{X}) = \begin{bmatrix} \mathrm{Cov}(X_1, X_1) & \cdots & \mathrm{Cov}(X_1, X_N) \\ \vdots & \ddots & \vdots \\ \mathrm{Cov}(X_N, X_1) & \cdots & \mathrm{Cov}(X_N, X_N) \end{bmatrix},$$

where $\mathrm{Cov}(X_i, X_i) = V(X_i)$ is the variance of X_i and $\mathrm{Cov}(X_i, X_j) = \mathrm{Cov}(X_j, X_i)$.

A.3 Coverage intervals

Given the PDF $g_X(\xi)$ for a random variable X and coverage probability p, a *coverage interval* for X is an interval such that the proportion of $g_X(\xi)$ between its endpoints is equal to p.

Given the PDF $g_X(\xi)$ and corresponding distribution function $G_X(\xi)$ for X, the α–quantile is the value ξ_α such that

$$G_X(\xi_\alpha) = \int_{-\infty}^{\xi_\alpha} g_X(\xi)\mathrm{d}\xi = \alpha,$$

i.e., the proportion of $g_X(\xi)$ to the left of ξ_α is equal to α. A $100p\,\%$ coverage interval is therefore $[\xi_\alpha,\ \xi_{\alpha+p}]$ for any value α between zero and $1-p$.

The inverse distribution function $G_X^{-1}(p)$ permits the value of X corresponding to a specified quantile to be obtained:

$$\xi_\alpha = G_X^{-1}(\alpha).$$

Example 23 *A coverage interval for a random variable characterized by a Gaussian probability distribution*

A $95\,\%$ coverage interval for a random variable with zero mean and unit standard deviation that is characterized by a Gaussian distribution is $[-2.0, 2.0]$.

Example 24 *A coverage interval for a random variable characterized by a rectangular probability distribution*

A $95\,\%$ coverage interval for a random variable with zero mean and unit standard deviation characterized by a rectangular distribution is $[-1.6, 1.6]$.

A.4 Bayes' theorem

The probability that the random variable Y takes the value η and the random variable X takes the value ξ can be written as

$$g_{Y,X}(\eta, \xi) = g_Y(\eta|X = \xi)g_X(\xi).$$

This expression is to be interpreted as the product of the probabilities that X takes the value ξ and that Y takes the value η *given* that X takes the value ξ. Thus,

$$g_Y(\eta|X = \xi) = \frac{g_{Y,X}(\eta, \xi)}{g_X(\xi)},$$

which is a statement of Bayes' theorem. Since the roles of Y and X can be interchanged,

$$g_X(\xi|Y=\eta) = \frac{g_{Y,X}(\eta,\xi)}{g_Y(\eta)}.$$

It then follows that

$$g_Y(\eta|X=\xi) = \frac{g_X(\xi|Y=\eta)g_Y(\eta)}{g_X(\xi)} \propto g_X(\xi|Y=\eta)g_Y(\eta).$$

In the above expression, $g_Y(\eta)$ is the *prior* PDF for Y and $g_Y(\eta|X=\xi)$ is the *posterior* PDF for Y given $X=\xi$. Often, the quantity X will describe an indication quantity, ξ an indication value ('data') for X, and $g_X(\xi|Y=\eta)$ is interpreted as the probability of obtaining the indication value ξ given that Y takes the value η. Then, $g_X(\xi|Y=\eta)$ is called the *likelihood function* (for Y) and is written as $l_Y(\eta|\xi)$ (regarded as a function of η). Often, it will correspond to the *sampling distribution* for (indication quantity) X. In this context, Bayes' theorem is used to update information about the quantity Y available prior to a measurement with the information provided by the indication value to obtain posterior information about Y.

The statement of Bayes' theorem given above can straightforwardly be extended to include vector random variables $\boldsymbol{Y}$ and $\boldsymbol{X}$, and repeated indication values $\boldsymbol{\xi}$.

Example 25 *Expectation of a Gaussian distribution with known variance*

Suppose X with unknown expectation Y and known variance σ^2 is characterized by a Gaussian distribution, and let ξ be a value of X drawn randomly from this distribution. It follows that the likelihood function is

$$g_X(\xi|Y=\eta) = l_Y(\eta|\xi) \propto \exp\left\{\frac{-(\xi-\eta)^2}{2\sigma^2}\right\}.$$

Furthermore, let the prior PDF for Y be the Gaussian PDF with expectation μ_0 and variance σ_0^2,[1] i.e.,

$$g_Y(\eta) \propto \exp\left\{\frac{-(\eta-\mu_0)^2}{2\sigma_0^2}\right\}.$$

Then, applying Bayes' theorem, the posterior PDF for Y is

$$\begin{aligned} g_Y(\eta|X=\xi) &\propto \exp\left\{\frac{-(\xi-\eta)^2}{2\sigma^2}\right\}\exp\left\{\frac{-(\eta-\mu_0)^2}{2\sigma_0^2}\right\} \\ &\propto \exp\left\{\frac{-(\eta-\mu_1)^2}{2\sigma_1^2}\right\}, \end{aligned}$$

[1] This prior PDF is used to characterize Y in the case that the information about Y takes the form of an estimate μ_0 with associated standard uncertainty σ_0^2.

where

$$\mu_1 = \frac{\frac{1}{\sigma_0^2}\mu_0 + \frac{1}{\sigma^2}\xi}{\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2}}$$

and

$$\frac{1}{\sigma_1^2} = \frac{1}{\sigma_0^2} + \frac{1}{\sigma^2}.$$

It follows that the posterior expectation μ_1 of Y is a weighted average of the prior expectation μ_0 and the indicated value ξ.

Example 26 *Analyte concentration*

Suppose X, representing the indication of analyte concentration provided by a measuring system, with unknown expectation Y, representing the 'real' analyte concentration, and known variance σ^2 is characterized by a Gaussian distribution. Let ξ, drawn randomly from this distribution, provide an indication value for X. It follows that the likelihood function is

$$g_X(\xi|Y=\eta) = l_Y(\eta|\xi) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{\frac{-(\xi-\eta)^2}{2\sigma^2}\right\}.$$

In particular, at or near the limit of detection, i.e., for values η of Y close to zero, the value ξ can be expected, with some non-zero probability, to be negative. Furthermore, if the value relates to an analytical blank [46, Clause F2.3], i.e, when $\eta = 0$, used subsequently to correct other results, on average as many negative as positive values ξ would be expected from repeated measurement of the blank. On the other hand, the 'real' analyte concentration is known to lie in the interval $[0,\ 1]$, which provides prior information about Y encoded as the rectangular distribution on that interval.

Then, applying Bayes' theorem, the posterior PDF for Y is

$$g_Y(\eta|X=\xi) \propto l_Y(\eta|\xi) g_Y(\eta),$$

in which

$$g_Y(\eta) = \begin{cases} 1, & 0 \le \eta \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

It follows that the posterior PDF for Y is the Gaussian distribution with parameters ξ and σ^2, which is truncated to the interval $[0,\ 1]$ and scaled to integrate to unity.

Figure A.1 illustrates (a) the prior PDF for Y, (b) the likelihood function, and (c) the posterior distribution, in the case that $\xi = 0.1$ and $\sigma = 0.1$.

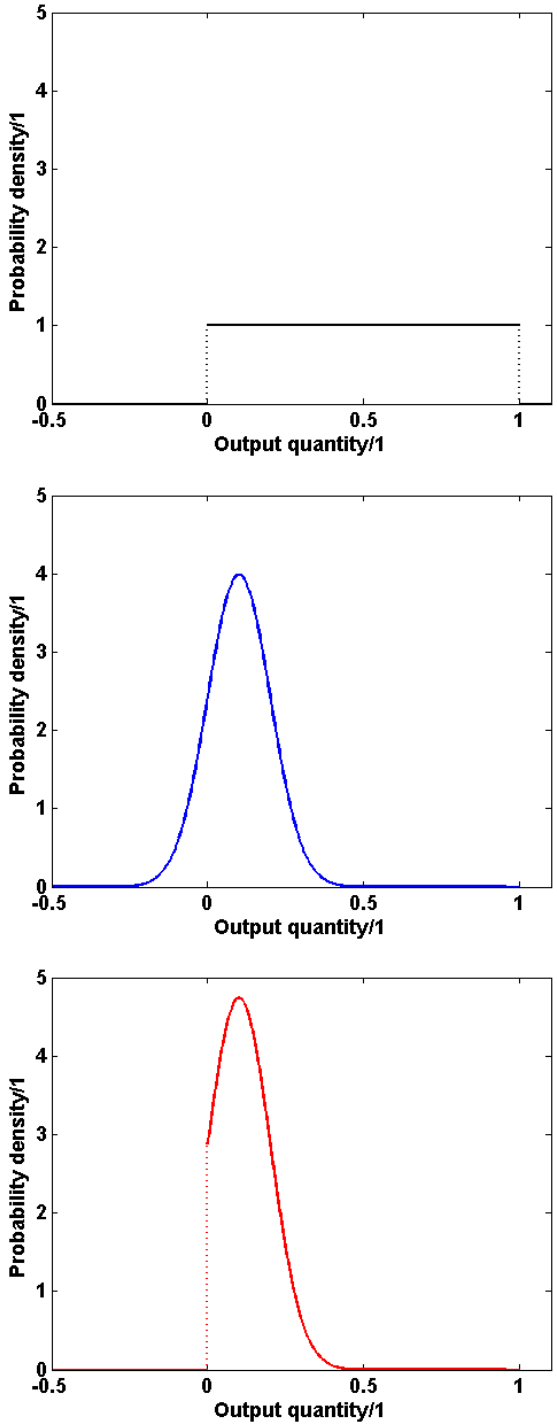

Figure A.1: Bayesian analysis for a problem of estimating analyte concentration near the limit of detection: the prior PDF for 'real' analyte concentration Y (top), the likelihood function (middle), and the posterior PDF for Y (bottom), in the case that $\xi = 0.1$ and $\sigma = 0.1$.

Example 27 *Gauge block calibration*

The length of a nominally $50\,\mathrm{mm}$ gauge block is determined by comparing it with a known reference standard of the same nominal length. The direct output of the comparison of the two gauge blocks is the difference D in their lengths given by

$$D = \ell(1 + \alpha\theta) - \ell_\mathrm{s}(1 + \alpha_\mathrm{s}\theta_\mathrm{s}) \equiv \phi(\ell, \ell_\mathrm{s}, \alpha, \alpha_\mathrm{s}, \theta, \theta_\mathrm{s}), \tag{A.1}$$

where ℓ is the length at $20\,^\circ\mathrm{C}$ of the gauge block being calibrated, ℓ_s is the length of the reference standard at $20\,^\circ\mathrm{C}$ as given on its calibration certificate, α and α_s are the coefficients of thermal expansion, respectively, of the gauge being calibrated and the reference standard, and θ and θ_s are the deviations in temperature from the $20\,^\circ\mathrm{C}$ reference temperature, respectively, of the gauge block being calibrated and the reference standard.

In the observation model (A.1), $Y \equiv \ell$ is the measurand, $\boldsymbol{Z} \equiv (\ell_\mathrm{s}, \alpha, \alpha_\mathrm{s}, \theta, \theta_\mathrm{s})^\top$, and $W \equiv D = \phi(Y, \boldsymbol{Z})$ is the indication quantity,

Suppose indication values $d_1, \ldots, d_n$ are assumed to be drawn independently from a Gaussian distribution with unknown expectation D and unknown variance S^2. In this case, an application of Bayes' theorem (Appendix A.4) gives

$$g_{Y,\boldsymbol{Z},S^2}(\eta, \boldsymbol{\zeta}, s^2|d_1, \ldots, d_n) \propto L_{D,S^2}(d, s^2|d_1, \ldots, d_N) g_{Y,\boldsymbol{Z},S^2}(\eta, \boldsymbol{\zeta}, s^2),$$

in which the likelihood function, a function of values $d = \phi(\eta, \boldsymbol{\zeta})$ of D and s^2 of S^2, takes the form

$$\begin{aligned} L_{D,S^2}(d, s^2|d_1, \ldots, d_n) &= \prod_{i=1}^{n} \frac{1}{s\sqrt{2\pi}} \exp\left\{\frac{-(d_i - d)^2}{2s^2}\right\} \\ &= \prod_{i=1}^{n} \frac{1}{s\sqrt{2\pi}} \exp\left\{\frac{-(d_i - \phi(\eta, \boldsymbol{\zeta}))^2}{2s^2}\right\}, \end{aligned}$$

and $g_{Y,\boldsymbol{Z},S^2}(\eta, \boldsymbol{\zeta}, s^2)$ is the prior PDF for Y, $\boldsymbol{Z}$ and S^2.

In practice, the quantities Y, the components of $\boldsymbol{Z}$ and S^2 are assumed *a priori* to be independent. Information about Y, the components of $\boldsymbol{Z}$ and S^2 can come from a variety of sources, and is used to characterize the quantities by prior PDFs (see, e.g., Section 4.2.7 and Appendix A.5). For example, the calibration certificate for the reference standard will provide information about the quantity ℓ_s representing its length in terms of an estimate of the quantity, the standard uncertainty associated with the estimate and a degrees of freedom attached to the standard uncertainty. Furthermore, information about S^2 may come from historical data, e.g., in the form of the (pooled) standard deviation determined from indication values, obtained independently, of the difference in lengths of two reference gauge blocks of nominally the same length. The use of Bayes' theorem serves not only to provide a posterior PDF for the measurand ℓ, but also to update the information about the other quantities, such as the length ℓ_s of the reference standard and the variance S^2 of the likelihood function.

A.5 The Principle of Maximum Entropy

> '... the virtue of wise decisions by taking into account all possibilities, i.e., by not presuming more information than we possess.' [61]

The Principle of Maximum Entropy (PME) [60] is a concept that can valuably be employed to enable maximum use to be made of available information, whilst at the same time avoiding the introduction of unacceptable bias in the result obtained. Two internationally respected experts in measurement uncertainty [104] state that predictions based on the results of Bayesian statistics and this principle turned out to be so successful in many fields of science [95], particularly in thermodynamics and quantum mechanics, that experience dictates no reason for not also using the principle in a theory of measurement uncertainty.

Bayesian statistics have been labelled 'subjective', but that is the intended nature of the approach. One builds in knowledge based on experience and other information to obtain an improved solution. However, if the same knowledge is available to more than one person, it would be entirely reasonable to ask that they drew the same conclusion. The application of PME was proposed [106] in the field of uncertainty evaluation in order to achieve this objective.

To illustrate the principle, consider a problem [106] in which a single unknown systematic effect X is present in a measurement. Suppose that all possible values for this effect lie within an interval $[-L,\ L]$, after the measured value has been corrected as carefully as possible for a known constant value. The value supplied for L is a subjective estimate based on known properties of the measurement, including the input quantities in the measurement model. In principle, the value of L could be improved by aggregating in a suitable manner the estimates of several experienced people. Let $g_X(\xi)$ denote the PDF for X. Although it is unknown, $g_X(\xi)$ will of course satisfy the normalizing condition

$$\int_{-L}^{L} g_X(\xi)\,\mathrm{d}\xi = 1. \tag{A.2}$$

Suppose that from the properties of the measurement it can be asserted that, because of the above careful correction process, the systematic effect is expected to be zero. A second condition on $g_X(\xi)$ is therefore

$$\int_{-L}^{L} \xi g_X(\xi)\,\mathrm{d}\xi = 0. \tag{A.3}$$

Suppose a value u of the standard deviation of the possible values for the systematic effect is available. Then,

$$\int_{-L}^{L} \xi^2 g_X(\xi)\,\mathrm{d}\xi = u^2. \tag{A.4}$$

Of course, $u^2 \leq L^2$. Suppose that no further information is available.

There are of course infinitely many PDFs for which the above three conditions hold. However, PME can be used to select a PDF from these. The PDF so obtained will have the property that it will be unbiased in that nothing more is implicitly assumed.

The use [60] of Shannon's theory of information [92] achieves this goal. Any given PDF represents some lack of knowledge of the quantity under consideration. This lack of knowledge can be quantified by a number,

$$S = -\int g_X(\xi) \log g_X(\xi) \, \mathrm{d}\xi,$$

called the (information) entropy [92] of that PDF. The least-biased 'probability assignment' is that which maximizes S subject to the conditions (A.2)–(A.4). Any other PDF that satisfies these conditions has a smaller value of S, thus introducing a prejudice that might contradict the missing information.

This formulation can fully be treated mathematically [106] and provides the required PDF. The 'shape' of the PDF depends on the quotient u/L. If u/L is smaller than $1/\sqrt{3}$, the PDF is bell-shaped. If u/L is larger than $1/\sqrt{3}$, the PDF is U-shaped. Between these possibilities, when $u/L = 1/\sqrt{3}$, the PDF is the rectangular PDF.

It can also be determined that if no information is available about the standard deviation u, and S is maximized with respect to conditions (A.2) and (A.3) only, the resulting PDF is the rectangular PDF.

It is evident that as more information is obtained the PDF that characterizes it becomes narrower, in that its standard deviation becomes smaller. Although obtaining such information might be time-consuming and expensive, in a competitive environment, or when it is required to state the most realistic (and hopefully the smallest) and defensible measurement uncertainty, such a treatment might be justified.

One approach to a range of such problems might be to categorize the commonest types of problem, such as those above, viz., when

1. Conditions (A.2) and (A.3) hold,

2. Conditions (A.2)–(A.4) hold.

There would be other such conditions in other circumstances. 'Solutions' to this range of problems could be determined, almost certainly in the form of algorithms that took as input the defining parameters (such as L and u above) and returned the corresponding quantified PDF.

In summary, it is regarded as scientifically flawed to discard credible information, unless it can be shown that doing so will have no influence on the results required to the accuracy needed, or the costs of doing so are prohibitive.

In particular, if knowledge of the PDFs for the input quantities is available, perhaps deduced as above using PME, these PDFs, which can be regarded as providing prior information in a Bayesian sense, should not simply be replaced by an expectation and standard deviation, unless doing so can be shown to have the mentioned negligible effect. If other information is available, such as above, or conditions on the quantities of interest or on nuisance parameters,[2] this information should be incorporated in order to render the solution physically more meaningful, and the uncertainties more realistic.

Clause 4.3.8 of the GUM provides the PDF when limits plus a single measured value are available (see Example 29). The treatment of a problem concerned with the limit of detection by the application of PME is also available [31].

Example 28 *The application of the PME to determining the PDF when lower and upper bounds only are available*

Consider that lower and upper bounds a and b for the input quantity X are available. If no further information is available the PME would yield the PDF for X as the rectangular distribution with limits a and b. It would also yield $(a + b)/2$ as an estimate of X (obtained as the expectation of X) with associated standard uncertainty $[(b - a)/12]^{1/2}$ (obtained as the standard deviation).

Example 29 *The application of the PME to determining the PDF when lower and upper bounds and a single measured value are available*

Suppose that as well as limits a and b, an estimate x of X is available. Unless $x = (a + b)/2$, i.e., it lies at the centre of the interval $(a,\ b)$, the PDF for X would not be rectangular as before. Let λ be the root of the equation

$$(e^{-\lambda a} - e^{-\lambda b})C(\lambda) - \lambda = 0,$$

where

$$C(\lambda)^{-1} = (x - a)e^{-\lambda a} + (b - x)e^{-\lambda b}.$$

The PME yields [GUM Clause 4.3.8] the PDF for X as

$$C(\lambda)e^{-\lambda X_i}.$$

All circumstances should be treated on their merits. Consider a *large* number of indication values. Suppose that the manner in which they are distributed (as seen by a histogram of their values, e.g.) indicates clearly that their behaviour is non-Gaussian, e.g., a strong

[2] Nuisance parameters are additional variables introduced as part of the modelling process to help build a realistic model. They would not by their nature constitute measurement results, but their estimates and associated uncertainties might be of interest as part of model development or enhancement.

asymmetry, 'long tails' or bi-modality. Then, the data itself, if judged to be representative, is indicating that the blind application of the PME is inappropriate in this circumstance. Since, for large samples, the principle of the bootstrap is appropriate, it can legitimately be applied here.

A.6 Discussion on Frequentist and Bayesian approaches

There are strongly-held views concerning whether statistical analysis in general or uncertainty evaluation in particular should be carried out using Frequentist or Bayesian approaches.

The Frequentist would assume that the output quantity is an unknown constant and that the result of measurement is a random variable. The Bayesian would regard the output quantity as a random variable having a PDF derived from available knowledge and the result of measurement as a known quantity [62].

These views can result in such divergent opinions that their discussion, although of considerable philosophical interest, must not be allowed to obscure the provision of clear practical guidance on uncertainty evaluation.

The attitude of this guide is predominantly Bayesian. A Bayesian would use available knowledge to make judgements, often subjective to some extent, about the PDFs for the output quantities (and in some cases also the input quantities) in a measurement model. The practitioner, with support from the expert metrologist as necessary, would also wish to employ previous information, e.g., calibration information or measurements of similar artefacts. Where (some of) the information seems suspect, as a result of common sense, experience or statistical analysis, further information should be sought if it is economical to do so.

In several instances the results that would finally be obtained by Frequentists and Bayesians would be identical or at least similar. Consider the repeated measurement of items manufactured under nominally identical conditions. The Frequentist would analyse the sample of measured values to estimate the 'population' mean and standard deviation of the manufactured items, and perhaps other parameters. The Bayesian would devise a prior distribution, based on his knowledge of the manufacturing process. He would 'update' the information contained within it in order to provide hopefully more reliable estimates of the parameters. In a case where there was no usable information available initially, the Bayesian would employ the so-called 'non-informative prior'. This prior effectively corresponds to the minimal knowledge that in the absence of information any measured value is equally possible. The parameter values so obtained can be identical in this case to those of the Frequentist. Any additional knowledge would help to give a better prior and hopefully a more valid result in that it would depend on available application-dependent information.

A valuable comparison of the Frequentist and Bayesian approaches is available [68].

Appendix B

Sensitivity coefficients

Sensitivity coefficients can be difficult to determine by hand for measurement functions that are complicated. The process by which they are conventionally determined is described in Section 5.6. If the effort of determining sensitivity coefficients manually is considerable, there are alternative approaches, including computer-assisted algebraic methods [13], finite-difference methods, and as part of an implementation of a Monte Carlo method. Some of these alternative approaches are described below.

B.1 Finite-difference methods

Finite-difference methods are the most commonly used techniques for approximating derivatives for scientific and engineering applications [81]. Typical of finite-difference formulæ are the *forward-difference* formula [51, p339]

$$f'(x) \approx \frac{f(x+h) - f(x)}{h} \tag{B.1}$$

and the *central-difference formula* [51, p340]

$$f'(x) \approx \frac{f(x+h) - f(x-h)}{2h}. \tag{B.2}$$

Here h is a 'step' selected in an appropriate way. It is typically chosen to balance truncation error (the error given by ignoring higher-order terms in the Taylor-series expansion from which the formula is derived) and subtractive-cancellation error (the error incurred when forming differences between function values at closely-neighbouring points). A Taylor expansion gives, for the error in the forward-difference form (B.1),

$$\frac{h}{2}f''(x) + \frac{h^2}{3!}f'''(x) + \frac{h^3}{4!}f^{\mathrm{iv}}(x) + \cdots$$

and, for the error in the central-difference form (B.2),

$$\frac{h^2}{3!}f^{(3)}(x) + \frac{h^4}{5!}f^{(5)}(x) + \cdots$$

The truncation error in expression (B.1) is $h|f''(\xi_1)|/2$, where $\xi_1 \in [x,\ x+h]$, and that in expression (B.2) is $h^2|f'''(\xi_2)|/6$, where $\xi_2 \in [x-h,\ x+h]$. For example, if f is a quadratic function

$$f(x) = ax^2 + bx + c,$$

then, in exact arithmetic, expression (B.1) approximates the derivative by $(2ax+b)+ah$ while expression (B.2) gives the correct value $2ax+b$. For appropriately scaled problems [51, p341], an optimal choice of h would be proportional to $\eta^{1/2}$ for expression (B.1) and $\eta^{1/3}$ for expression (B.2). Here, η is the relative machine precision, i.e., the smallest machine representable positive value η such that the value of $1+\eta$, computed in floating-point arithmetic, exceeds 1. For IEEE arithmetic, $\eta = 2.22 \times 10^{-16}$. For appropriately scaled problems, the use of expression (B.1) can be expected at best to lose 'half a wordlength of accuracy' and the use of expression (B.2) to lose one-third of the available digits. From problems involving a number of parameters each with different scaling, determining a good value of h in order to achieve the above 'best numerical accuracies' is not trivial. The GUM, in Clause 5.1.3, suggests the use of expression (B.2) with $h = u(x)$.

We note that if the function $f(x)$ has already been evaluated at x, then the forward-difference formula requires one additional function evaluation, and the central-difference two, to obtain approximate values for the derivative. The most important advantage of the finite difference approach is that it does not require access to the source code for the function evaluation component and therefore, from this point of view, it is an ideal method for library software.

B.2 Complex-step method

The complex-step method is similar to finite differences, but uses complex arithmetic to provide accurate estimates of sensitivity coefficients. Recent interest (for example, [71]) in the approach was generated by Squire and Trapp [98] who revisited earlier work by Lyness and Moler [69]. A complex number Z takes the form $Z = X + \mathrm{i}Y$, where X and Y are real and $\mathrm{i}^2 = -1$. All the arithmetic operations for real numbers also apply for complex numbers. Most real-valued functions $f(X)$ occurring in science also have complex-valued counterparts. The Taylor expansion for a complex function takes the form

$$f(Z+W) = f(Z) + Wf'(Z) + \frac{W^2}{2}f''(Z) + \frac{W^3}{3!}f'''(Z) + \frac{W^4}{4!}f^{\mathrm{iv}}(Z) + \cdots, \quad \text{(B.3)}$$

where Z and W are complex numbers. (The Taylor expansion holds for what are termed analytical functions that have continuous derivatives of all orders, but these include almost all the function of interest to science.) Taking $Z = x$ and $W = \mathrm{i}h$ where x is real and h is

real and small,

$$f(x+\mathrm{i}h) = f(x) + \mathrm{i}hf'(x) - \frac{h^2}{2}f''(x) - \mathrm{i}\frac{h^3}{3!}f'''(x) + \frac{h^4}{4!}f^{\mathrm{iv}}(x) + \cdots . \qquad \text{(B.4)}$$

Taking real and imaginary parts,

$$\Re f(x+\mathrm{i}h) = f(x) - \frac{h^2}{2}f''(x) + \frac{h^4}{4!}f^{\mathrm{iv}}(x) + \cdots ,$$

and

$$\Im f(x+\mathrm{i}h) = hf'(x) - \frac{h^3}{3!}f'''(x) + \frac{h^5}{5!}f^{\mathrm{v}}(x) + \cdots .$$

From these last equations, for h small,

$$f(x) \approx \Re f(x+\mathrm{i}h), \qquad f'(x) \approx \frac{\Im f(x+\mathrm{i}h)}{h}, \qquad \text{(B.5)}$$

both with a truncation error of order h^2. Thus, both $f(x)$ and $f'(x)$ are obtained for one complex function evaluation. Unlike the use of a finite-difference formula, h can be chosen to be *very* small with no concern about the loss of significant figures through subtractive cancellation since no subtraction is involved. The only practical restriction is that h must not be chosen so small that it underflows, i.e., is replaced by zero in floating-point arithmetic. The value $h = 10^{-100}$ should be suitable for all but pathologically-scaled problems.

The success of this approach, i.e., the use of the approximation (B.5), depends on the availability of complex arithmetic and on the integrity of the inbuilt complex-valued functions. The complex step method is particularly suitable for implementation in software such as MATLAB [72] since the default data type is complex, and all the main intrinsic functions can be evaluated for complex arguments. Most functions can be worked with satisfactorily, but care must be taken to ensure that the intrinsic functions used in the function evaluation component behave in the appropriate way in complex arithmetic.

B.3 Nonlinear sensitivity coefficients

With a Monte Carlo calculation there is no immediate counterpart of a sensitivity coefficient since such calculation operates in terms of the actual non-linear measurement function rather than a linearized counterpart. Recall that with a linear measurement function the sensitivity coefficients 'reproduce' linear effects, and for a non-linear function the sensitivity coefficients provide first-order information. Therefore, those practitioners accustomed to using the approach based on the GUM uncertainty framework may find the absence of sensitivity coefficients disconcerting.

It is possible and very straightforward, however, to adapt the Monte Carlo procedure such that it provides information that in a sense constitutes a non-linear counterpart of a sensitivity coefficient. Consider holding all input quantities but one, say X_k, at their estimates. In

this setting the measurement function effectively becomes one having a single input quantity, viz., X_k. Make random draws from the PDF for this input quantity and determine an approximation to the probability distribution of the output quantity with respect to X_k. The standard deviation $\widetilde{u}_k(y)$ determined in terms of this distribution is taken as an approximation to the component of the (combined) 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

$$\widetilde{c}_k = \frac{\widetilde{u}_k(y)}{u(x_k)}.$$

It will be equal to the magnitude $|c_k|$ of the 'linear' sensitivity coefficient c_k when the measurement function is linear in X_k, and be close to its value when the non-linearity with respect to X_k is negligible. When $\widetilde{c}_k$ is appreciably different from $|c_k|$, the non-linearity effect may noticeably influence the standard uncertainty $u(y)$. Thus, the deviation of $\widetilde{u}_k(y)$ from $u_k(y) = |c_k|u(x_k)$ can be used as an approximate measure of the influence of the non-linearity of the measurement function with regards to X_k alone.

The sensitivity coefficients so obtained are not generally to be taken in conjunction with the standard uncertainties associated with the estimates of the input quantities as the *only* contributions to the standard uncertainty associated with the estimate of the output quantity. There will be further contributions arising from any interaction (i.e., non-additive) terms in the measurement function.

In the case of complicated measurement functions the above approach is already used by some metrologists as a practical alternative to the tedious analysis required to provide (linear) sensitivity coefficients [53].