A2LA Guide for the Estimation of Measurement Uncertainty In Testing

July 2002

By
Thomas M. Adams

Table of Contents

Acknowledgements

I would like to thank A2LA's Measurement Advisory Committee and calibration assessors for their efforts in reviewing the manuscript and providing commentary and corrections.

In particular, I gratefully acknowledge the following individuals for their insightful comments and stimulating uncertainty discussions over the years: Mr. Ralph Veale, Dr. Richard Turner, Dr. Henrik Nielsen, Mr. Phil Stein, Dr. Ted Doiron, and Mr. Dan Tholen.

Finally, this guidance could not have been written without the support and encouragement of A2LA, its Advisory Committees, and management.

Thom Adams
Frederick, MD

July 2002

1.0 Introduction

The purpose of measurement is to determine a value of a quantity of interest (the measurand**)**. Examples of measurands include the boiling point of water under 1 atmosphere of pressure, or the Rockwell hardness of a metal specimen, or the tensile strength of a plastic compound, or the length of a metal bar at 20°C. Notice that the objective of a measurement is to determine a value of the measurand, in other words, to sample one value out of a universe of possible values, since, in general, when one repeats a measurement many times, one will obtain many different answers.

This observed variability in the results of repeated measurements arises because influence quantities that can affect the measurement result are not held constant. In general, there are many- if not infinitely many- influence quantities affecting a measurement result. Although it is impossible to identify all of them, the most significant can be identified and the magnitude of their effects on the measurement result can be estimated. Further, the way they impact the measurement result can, in many cases, be *mathematically modeled*.

Given the inherent variability of measurement, a statement of a measurement result is incomplete (perhaps even meaningless) without an accompanying statement of the estimated uncertainty of measurement (a parameter characterizing the range of values within which the value of the measurand can be said to lie within a specified level of confidence). The incompleteness of measurement results unqualified by uncertainty estimates is evidenced by the common situations when two technicians in the same lab determine different measurement results, or different labs determine different results, or when disagreements arise between customer and supplier. Even with Youden plots, commonly taken as a measure of the agreement among different test labs in an interlaboratory comparison, outliers are identified even if the difference between the lab's result and the grand mean is smaller than the actual uncertainty of the measurement (were it to be determined!).

However, even granting the importance of uncertainty estimates, it is just as important that the method used to estimate uncertainties be generally agreed upon, at least between interested parties. Given the nature of today's global economy, however, the method should be *universally adopted, understood, and equitably applied*. The ISO *Guide to the Expression of Uncertainty in Measurement* (GUM) and the corresponding American National Standard ANSI/NCSL Z540-2-1997 provide the current international consensus method for estimating measurement uncertainty. It is equally applicable to calibration and test results and it forms the basis for accreditation requirements relating to measurement uncertainty estimation.

The GUM method supposes that a mathematical model is available or can be derived that describes the functional relationship between the measurand and the influence quantities. In the absence of this model, the GUM method does not apply very well. Further, it must be admitted that, for many people, the mathematics and concepts in the GUM are somewhat removed from daily experience.

There are, however, measures of precision that can be used as the basis for estimating uncertainty that are well known in the testing community. These measures include reproducibility and repeatability and what ISO 5725 refers to as "intermediate measures of precision". These measures are standard deviations derived from the analysis of experimental data, and if the reproducibility experiment is designed in such a way that variability due to all of the major sources of uncertainty is sampled (as required by ISO/IEC 17025 section 5.4.6.3), then reliable estimates of uncertainty can be based entirely on experiment without having to resort to the mathematics and theory found in the GUM that so many people find daunting.

Therefore, this document also provides guidance on estimation of uncertainty based on reproducibility estimates and control charting. This guidance is applicable to all fields of testing. The method assumes that all significant systematic effects have been identified and either eliminated or else compensated for by the application of suitable corrections.

2.0 Repeatability, reproducibility, trueness estimates, and control charts in measurement uncertainty estimation

Although the GUM method presented below (c.f. Section 3) is generally regarded as the complete, rigorous method for uncertainty estimation, in practice many people find it to be too complicated and, as mentioned in the Introduction, it does not apply well in the absence of a mathematical model for the measurement. For example, in the absence of a mathematical model, it would be necessary to devise experiments in order to estimate the sensitivity coefficients (c.f. Section 3.4) of each of the identified uncertainty contributors. This could be an arduous task indeed.

In this section we will focus on two measures of precision, repeatability and reproducibility, and measures of trueness to show a simple method for basing uncertainty estimates on these measures. This method has the great advantages that most testing labs are already acquainted with repeatability and reproducibility experiments and that a suitably devised reproducibility experiment will include the effects of all of the major uncertainty contributors. Thus, adequate uncertainty estimates can be derived without recourse to the mathematics of the GUM. For complete details, the reader is urged to consult ISO 5725, in particular part 2, "Accuracy (trueness and precision) of measurement methods and results – Basic method for the determination of repeatability and reproducibility of a standard measurement method". In addition, a draft technical standard is being prepared by ISO designated as ISO/DTS 21748. This document will be an excellent guide to the use of repeatability, reproducibility and trueness estimates in estimation of measurement uncertainty. Also, this section briefly discusses the use of control charts as the basis for uncertainty estimates. For details on control charts, the reader is referred to, for example, ISO 8258, "Shewhart control charts". In what follows, we assume that the reader is already familiar with control charting.

2.1 Control charts

Recalling that:

1) *measurement uncertainty* is defined as a "parameter, associated with the result of a measurement, that characterizes the dispersion of the values that could reasonably be attributed to the measurand";

2) this parameter is a standard deviation, or a multiple of a standard deviation, that can be derived from the statistical analysis of experimental data;

3) upper and lower action limits are established on control charts so that approximately 997 measurements out of 1000 are within the action limits for a measurement process in statistical control;

4) upper and lower warning limits are established on control charts so that approximately 950 measurements out of 1000 are within the warning limits for a measurement process in statistical control,

we can see immediately that the action limits provide an estimate of measurement uncertainty at approximately the 99.7% level of confidence ("3 sigma") and that the warning limits will provide estimates of uncertainty at approximately the 95% level of confidence ("2 sigma").

There are a few caveats associated with estimating measurement uncertainty based on control chart data:

1) The control test sample should have a certified or otherwise known or accepted value. This way, bias in the measurement process may be identified and corrected for in the calculation of measurement results, or else eliminated. There will be some uncertainty associated with bias corrections so it may be necessary to identify and quantify this uncertainty and root-sum-square it with the standard deviation associated with the control limits.

2) The value of the measurand represented by the control sample should be close to the value of the measurand actually obtained during routine testing since, in general, the uncertainty of a measurement will be some function of the "level of the test", or value of the measurand. For example, the uncertainty associated with melt

flow rate determinations will depend on the value of the melt flow rate: uncertainties associated with very low flow rates will in general be different from uncertainties associated with very high melt flow rates. Consequently, it may be necessary to track several control samples at different levels of the measurand in order to properly assess the measurement uncertainty for the various levels of the measurand encountered in a testing laboratory.

3) The measurement process for control samples should be the same as for routine samples, including subsampling and sample preparation. If it is not, then additional uncertainty components may have to be considered (c.f. section 3 for the general method of doing this).

4) The measurement process must be in statistical control as demonstrated by the control chart. This means that a sufficient number of data points must be collected before a judgment can be made as to whether or not the process is in control and also to ensure that the estimate of the population standard deviation is reasonably accurate. There are no universally applicable rules here, but normally 20 – 25 subgroups of 4 or 5 are considered adequate for providing preliminary estimates. Measurement processes that are not in statistical control must be brought into control before the control chart can be properly constructed.

Even with those caveats, control charting is probably the simplest, most direct way of estimating measurement uncertainty. However, if, for example, certified reference materials are unavailable or very expensive or if a test is very expensive or time-consuming to perform, control charting may not be a viable option and the methods of the GUM presented below may have to be used.

2.2 Repeatability and reproducibility

For the purposes of this section, the following definitions are taken from ISO 3534-1 (note the differences between these definitions and the definitions from the VIM given in Appendix 1):

Precision is the closeness of agreement between independent test results obtained under stipulated conditions. Precision depends upon the distribution of random errors and does not relate to the true value or the specified value of the measurand. The measure of precision is usually computed as a standard deviation of test results. Less precision is indicated by a higher standard deviation. “Independent test results” means results obtained in a manner not influenced by any previous result on the same or similar test item. Quantitative measures of precision depend critically on the stipulated conditions. Repeatability and reproducibility conditions are particular examples of extreme stipulated conditions.

Repeatability is precision under repeatability conditions, i.e. conditions where independent test results are obtained with the same method on identical test items in the same laboratory by the same operator using the same equipment within short intervals of time. *Repeatability standard deviation* is the standard deviation of test results obtained under repeatability conditions.

Reproducibility is precision under reproducibility conditions, i.e. conditions where independent test results are obtained with the same method on identical test items in different laboratories by different operators using different equipment. *Reproducibility standard deviation* is the standard deviation of test results obtained under reproducibility conditions and is a measure of the dispersion of test results under reproducibility conditions. A valid statement of reproducibility requires specification of the conditions changed. For example, ISO 5725 considers four factors subject to change: operator, equipment, calibration, and time. A statement of reproducibility will indicate which of these factors have been varied in the course of the experiment.

Bias is the difference between the expectation of the test results and an accepted reference value. Bias is the total systematic error as contrasted to random error. There may be one or more systematic error components contributing to the bias. A larger systematic difference from the accepted reference value is reflected by a larger bias value.

Trueness is the closeness of agreement between the average value obtained from a large set of test results and an accepted reference value. The measure of trueness is normally expressed in terms of bias.

It will usually be the case that if the laboratory is applying a standard, validated test method, the test method will be accompanied by estimates of precision and bias obtained by interlaboratory comparison during the course of method validation. For example, ASTM test methods are required to be accompanied by statements of precision and, where, applicable, a statement of bias. In the case of ASTM test methods, the statement of precision will include an expression of the within-laboratory standard deviation (S_r) and the between-laboratory standard deviation (S_R).

With all of this in mind, we can see that a laboratory will be able to establish reasonable uncertainty estimates if

1) it can establish that its bias when applying the measurement method is within the bias stated in the test method, and

2) it can establish that the precision (i.e., reproducibility with several factors varied) obtained during application of the test method is within the between-laboratory standard deviation stated in the test method, and

3) it can identify any influence quantities that may not have been adequately studied during the interlaboratory comparison, and quantify the influence quantities, associated uncertainties and sensitivity coefficients (c.f. section 3.4).

If these conditions are satisfied, then the laboratory can estimate its measurement uncertainty by combining the uncertainty associated with bias corrections, its reproducibility, and the uncertainties of any influence quantities identified in (3) above via the "root-sum-square" method given below in section 3.5. To establish that these conditions are fulfilled, the laboratory may proceed as follows:

2.2.1 Control of bias

The laboratory may demonstrate that its measurement bias is within the limits expected by proper application of the test method by either of the following methods:

1) Using a certified reference material

Using an appropriate reference standard or material, the laboratory should perform replicate measurements to form an estimate of its bias, which is the difference between the mean of its test results and the certified value of the standard or material. If the absolute value of this bias is less than twice the reproducibility standard deviation given in the precision statement in the test method, then the laboratory may consider that its bias is under control.

2) Interlaboratory comparison

Laboratories participating in proficiency testing schemes will have available to them data from a large number of laboratories which they can use to estimate the bias of their measurement results. Comparison of the lab mean to the grand mean or other assigned value in such programs, for example, will allow them to demonstrate that their bias is under adequate control.

2.2.2 Control of precision

To verify control of precision, the laboratory may perform a number of replicate measurements under repeatability conditions and compare its repeatability standard deviation s_l to the repeatability standard deviation s_r given in the precision statement in the test method. Comparing s_l to s_r using an F-test, for example, to test for significant difference, will allow the laboratory to demonstrate control of precision. If s_l differs significantly from s_r, then the laboratory should use s_l in its estimate of uncertainty, otherwise it can use s_r.

Once the laboratory has demonstrated control of precision and bias, it is free to use the estimates of these quantities given in the test method as the basis for its estimate of measurement uncertainty. However, it must demonstrate on an on-going basis that precision and bias continue to be under adequate control. This is accomplished, for example, by proper measurement assurance techniques including control charting. Further, as mentioned above, the reference materials or standards used in these verifications must be relevant to the levels of the test encountered by the laboratory in its routine testing. Once these conditions are fulfilled and satisfied on an on-going basis, the laboratory has all of the information it needs to estimate its measurement uncertainty.

3.0 ISO Guide to the expression of uncertainty in measurement

3.1 Specifying the measurand

Any uncertainty analysis must begin with a clear specification of the measurand. Although this step may seem trivial, it is in fact the most important and possibly the most difficult. Without a clear understanding of what the objective of the measurement is, and the factors influencing the measurement result, it is impossible to arrive at a meaningful estimate of the uncertainty.

At this point it may be worthwhile to recall the admonition found in ANSI/NCSL Z540-2-1997, the *US Guide to the Expression of Uncertainty in Measurement*:

> Although this *Guide* provides a framework for assessing uncertainty, it cannot substitute for critical thinking, intellectual honesty, and professional skill. The evaluation of uncertainty is neither a routine task nor a purely mathematical one; it depends on detailed knowledge of the nature of the measurand and of the measurement. The quality and utility of the uncertainty quoted for the result of a measurement therefore ultimately depend on the understanding, critical analysis, and integrity of those who contribute to the assignment of its value.[1]

Especially for complex tests, it is not necessarily clear what is being measured and what is influencing the measurement result. Even with industry standard test methods that set tolerances on all the various parts and features of a testing machine, places limits on the environmental conditions, and specifies the method of preparation of the samples, the nature of the material itself may be the major source of variability in test results. If the laboratory does not understand this, but sees the test or testing machine as a "black box" from which numbers are spewed by some mysterious process, then it is impossible for the laboratory to analyze the uncertainty of the measurements involved in the test.

On the other hand, the required level of detail in the definition of the measurand must be dictated by the required level of accuracy of the measurement.

The specification of a measurand may require statements about quantities such as time, temperature, and pressure.

Example. We will consider the tensile strength at break of a fiber-reinforced composite according to the method specified in ASTM D638.

Note that, implicit in this definition of the measurand, there are several factors specified- all possible contributors to the uncertainty of the tensile strength result- including:

a) specification of the accuracy and other characteristics of the testing machine (through the requirements of D638 itself and also ASTM E4 and E74 for the verification of testing machines and force indicating devices);

b) specification of the environmental conditions under which the test is to be conducted;

[1] GUM 3.4.8. Since uncertainty evaluation is neither a purely mathematical task nor a merely routine task, the details of its procedures in every particular as applied to a given test can never be codified to the point of complete unambiguousness. Reasonable and informed observers can and do disagree on the particulars of a given uncertainty analysis and it must be recognized that there is nothing that can be called the "final word" on the uncertainty analysis of a given measurement. For this reason, this guidance document, while presenting the necessary mathematical machinery needed to produce an uncertainty evaluation in accordance with the GUM, emphasizes the importance of "reasonability" (see below) in assessing the final uncertainty estimate.

c) specification of the molding conditions and dimensional characteristics of the test samples;

d) specification of the conditioning environment for the molded test samples as found, for example, in material specifications;

e) specification of the accuracy of the device used to measure the thickness and width of the molded test items.

It is important to be aware of the factors included explicitly and implicitly in the definition of a measurand. This will generally necessitate, as mentioned above, a thorough understanding of the principles, assumptions, and science underlying the measurement at hand.

3.2 Modeling the measurement

The uncertainty in the result of a measurement is a result both of our incomplete knowledge of the value of the measurand and of the factors influencing it. Even after correcting for known systematic effects, the corrected measurement result is still only an estimate of the value of the measurand because of random effects and because our knowledge of the magnitudes of the corrections is itself only an estimate. It is important to note that the result of a measurement after correction could be (unbeknownst to the analyst) very close to the value of the measurand, even though the measurement itself may have a large uncertainty. In other words, the uncertainty of a measurement must not be confused with the remaining unknown (and unknowable) "error".

There are many possible sources of uncertainty in measurement including[2]:

1) incomplete definition of the measurand;

2) imperfect realization of the definition of the measurand;

3) nonrepresentative sampling (the sample measured may not represent the defined measurand);

4) inadequate knowledge of the effects of environmental conditions on the measurement or imperfect measurement of environmental conditions;

5) personal bias in reading analog instruments, including the effects of parallax;

6) finite resolution or discrimination threshold;

7) inexact values of measurement standards and reference materials;

8) inexact values of constants and other parameters obtained from external sources and used in the data-reduction algorithm;

9) approximations and assumptions incorporated in the measurement method and procedure;

10) variations in repeated observations of the measurand under apparently identical conditions.

These sources of uncertainty are not necessarily independent and some or all of items 1-9 can contribute to the variations in repeated observations. If all of the quantities on which the result of a measurement depend could be varied, then its uncertainty could be evaluated based purely on a statistical treatment of experimental data. But this is seldom possible in practice because of the time and expense involved in such an exhaustive experimental evaluation of uncertainty. Instead, the GUM assumes that the uncertainty of a measurement result can be evaluated based on a mathematical model of the measurement. Further, it is assumed that this model can be made as accurate as needed relative to the required accuracy of the measurement.

---

[2] GUM 3.3.2.

Nevertheless, because the mathematical model is always incomplete, all relevant input quantities should be varied to the extent possible so that the uncertainty estimate can be based, as much as possible, on experimental data. Whenever possible, the use of check standards and control charts (often called *measurement assurance*) should be used to establish that a measurement system is under statistical control, and these data should be used as part of the effort to obtain a reasonable estimate of the measurement uncertainty. When the observed data shows that the mathematical model is incomplete, then the model should be revised[3].

In general, the mathematical model will be a function of several input quantities showing how the measurement result is obtained from the input quantities. If the input quantities are designated as $x_1, x_2, \ldots, x_n$, then we can write the functional relationship between the measurement result y and the input quantities x_i as

$$y = f(x_1, x_2, ..., x_n) .$$

This function is to be understood in the broadest possible context as including every possible source of variation in the measurement result including all corrections and correction factors. Consequently, the mathematical model may be a very complicated relationship between many input quantities that is never explicitly written down.

Example. Tensile strength at break, S, according to ASTM D638 is defined as the force F necessary to break a test bar divided by the cross sectional area of the bar. The cross-sectional area of the bar is defined as the product of its thickness T and its width W. Therefore, as a first approximation, we will take as the mathematical model of our tensile strength measurement:

$$S = f(F, T, W) = \frac{F}{TW} .$$

We call this a first approximation since we can immediately see that a number of potential influence quantities are not explicitly included in this function. For example, influence quantities such as temperature effects on the dimensions and strength of the bar, or inertial effects within the testing machine, or variability due to varying "necking" or "drawing" behaviors of the different test samples during the application of force, are not included. Depending on the final uncertainty estimate obtained based on this approximation, it may be necessary to revise the model function to include additional influence quantities if it is felt that the first approximation is inadequate.

3.3 Quantifying the uncertainty contributors and their associated uncertainties

3.3.1 Type A evaluation of standard uncertainty

A type A uncertainty estimate is an estimate derived from the statistical analysis of experimental data. (Note that the designation "type A" refers to the *method* by which the uncertainty estimate was obtained. "Type A" does not refer to the nature of the uncertainty contributor itself; in particular, the reader should avoid the temptation to identify type A uncertainty estimates as "random" components of uncertainty as is often done.) Because it is such a vast subject of discussion we cannot hope to cover it in all its complexity in this document. We will focus only on the standard deviation of test results as Type A uncertainty estimates. The reader is urged to consult the GUM and ISO 5725 for further details.

It will usually be the case that the best estimate of the value of a measurand will be the average of several test results. If n test results are obtained, then the average is simply the sum of the test results divided by n. If we

[3] In fact, if a laboratory has sufficient data, uncertainty analysis by the method of the GUM is not necessary and the uncertainty of measurement requirements of ISO/IEC 17025 can be satisfied via the methods of ISO 5725 or control charting as described above. The GUM will always be applicable to calibration laboratories, but it should be used by testing laboratories only in cases where the testing laboratory has insufficient data to establish a reliable estimate of its long-term reproducibility. This will be the case, for example, when the laboratory is a commercial testing facility that receives unique items to be tested. Traceability to appropriate standards will always be required regardless of the state of knowledge of reproducibility.

designate each of the test results by the symbol x_i, then we can write the following equation for the average $\bar{x}$ of n test results:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \,.$$

The experimental standard deviation s characterizes the variability, or spread, of the observed values x_i. It is given by the equation

$$s = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n-1}} \,.$$

It is best to use a calculator or spreadsheet program for these calculations. For example, the functions AVERAGE and STDEV in Excel can be used to find the average and standard deviation of test results quickly and easily.

Example. Continuing with the tensile strength example, suppose that 5 bars are tested. A "test" consists of measuring the thickness and width of each specimen and then determining the force required to break each bar. Suppose that we obtained the following data:

Bar	Thickness, T	Width, W	Force to break, F	Tensile strength, S
1	0.124 *in*	0.499 *in*	830 *lb*	13414 lb/in^2
2	0.126 *in*	0.501 *in*	900 *lb*	14257 lb/in^2
3	0.125 *in*	0.500 *in*	810 *lb*	12960 lb/in^2
4	0.126 *in*	0.500 *in*	870 *lb*	13810 lb/in^2
5	0.124 *in*	0.499 *in*	850 *lb*	13737 lb/in^2
Average, $\bar{x}$:	0.125 *in*	0.4998 *in*	852.0 *lb*	13636 lb/in^2
Stdev, s:	0.001 *in*	0.000837 *in*	34.93 *lb*	482.8 lb/in^2

From this data, we see that the best estimate[4] of the force required to break a bar is the average of the five force measurements, or 852.0 *lb*. The standard deviation of these 5 results is 34.93 *lb*. This is the type A uncertainty estimate of the force due to random variations in the testing machine, molding process, sample-curing process, and in the material itself. The standard deviation of the thickness measurements is 0.001 *in* and the standard deviation of the width measurements is 0.000837 *in* which we will take to be the type A estimates of the uncertainties of the thickness and width of an average test bar.

3.3.2 Probability distributions

A probability distribution is a mathematical function giving the probability that a random variable takes any given value or else belongs to a set of values. For example, different probability distribution functions describe the

[4] "Best", that is, relative to the extremely limited information available in this example. A laboratory with proper measurement assurance programs in place will have much more data to work with and, consequently, a much better uncertainty estimate.

probability of rolling a "1" on one roll of a six-sided die, or the probability of drawing the ace of spades from a standard 52-card deck of cards, or the probability of finding the room temperature to be 70°*F*.

There are infinitely many possible distributions but some are more useful for our purposes than others. In this section we'll derive one such distribution, simply present the others, and examine the mathematics behind them; in particular we will show how the variances of the distributions are obtained so that the reader can understand the origins of the "divisors" used in uncertainty budgets. It is assumed that the reader has a basic grasp of calculus, although this isn't necessary to understand the ideas presented.

3.3.2.1 Rectangular distribution

We will start with one of the simplest distributions, the so-called "rectangular" or "uniform" distribution. This distribution is used to model cases where the probability of obtaining any value between two limits is equal to the probability of obtaining any other value. (Discrete versions of this distribution can be used to model the dice roll and card draw examples given above, for example.)

Graphically, in cases where the distribution of possible values is continuous (as opposed to the dice roll and card draw examples), this distribution is just a horizontal line extending between two limits. (Recall that the equation of a horizontal line is $f(x) = y = c$ where c is some constant number corresponding to the point where the line intersects the y-axis.)

If the limits are $\pm a$, then we know that the probability of obtaining a number between $+a$ and $-a$, inclusive, is equal to 1 (i.e., it is certain that a number between these limits will be obtained since values outside of these limits are not included in the definition of the distribution). This is a fundamental requirement on all probability distribution functions and it is frequently stated by saying that the distribution function must be *normalized*. In order to normalize a distribution function it is necessary to integrate it with respect to the limits imposed with the condition that the integral (the area under the curve) must be equal to 1.

To see how this works, we will normalize the rectangular distribution function $f(x) = c$ with respect to the containment limits $\pm a$.

$$\int_{-a}^{+a} f(x)dx = \int_{-a}^{+a} c\,dx = cx\Big|_{-a}^{+a} = c(a-(-a)) = 2ca = 1.$$

So, from the normalization condition, we can see that the constant c must be equal to $\frac{1}{2a}$. Thus the final equation for the rectangular probability distribution function with containment limits $\pm a$ centered on the y-axis (i.e., the center of the distribution is $x = 0$) is simply $f(x) = \frac{1}{2a}$.

(We could also have obtained this same result simply by noting that the area of a rectangle is equal to the product of its height, c, and its width, $2a$, and then imposing the normalization requirement to obtain $2ac = 1$ directly. However, for more complicated probability distribution functions, such simple determinations of area aren't usually possible so it is useful to know how to normalize a function using integrals.)

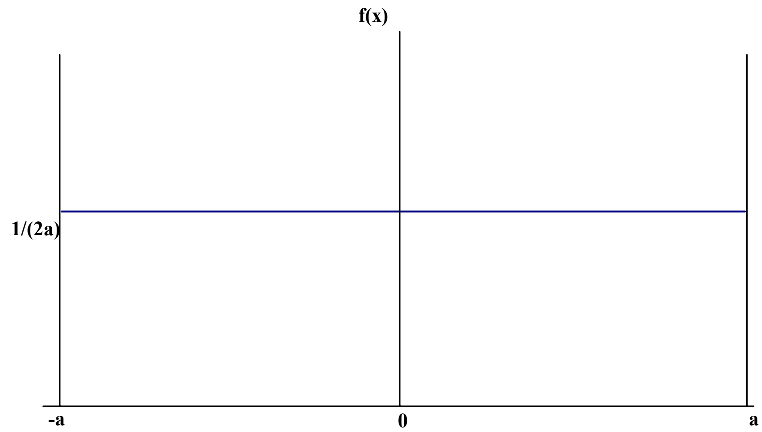

Figure 1. The rectangular or uniform distribution models cases where the probability of some

The variance, σ^2, of a probability distribution function is a measure of the dispersion, or "width", of the distribution. Formally, it is obtained from the following integral of the normalized probability distribution function:

$$\sigma^2 = \int x^2 f(x)dx \,.$$

For the rectangular probability distribution function, we obtain the variance as follows:

$$\sigma^2 = \int_{-a}^{+a} \frac{x^2}{2a} dx = \frac{1}{2a}\left(\frac{x^3}{3}\right)\Bigg|_{-a}^{+a} = \left(\frac{1}{6a}\right)\left(a^3 - (-a)^3\right) = \frac{2a^3}{6a} = \frac{a^2}{3} \,.$$

The standard deviation σ is simply the positive square root of the variance so for the rectangular distribution we obtain $\sigma = \dfrac{a}{\sqrt{3}}$.

The rectangular distribution may be used to model the statistical behavior of an uncertainty contributor if and only if all three of the following conditions are fulfilled:

1) **Uniform Probability**. By definition, the rectangular distribution assumes that the uncertainty contributor has a uniform probability of occurrence between two limits- in other words, every value between and including the limits is equally probable. Nature usually does not display this sort of behavior except in the case of discrete events (e.g., a roll of the dice, a coin toss) so this is actually a stringent requirement to fulfil.

2) **100% Containment**. It follows from the uniform probability criterion that the uncertainty contributor has zero probability of occurrence outside of the limits of the distribution. In other words, every possible value attributable to the uncertainty contributor lies within the two limits.

3) **Minimum Limits**. If the limits $\pm a$ encompass the entire distribution, then so do the limits $\pm na$ where n is any number greater than or equal to 1. It is therefore imperative that the limits we assume for a given uncertainty contributor be the minimum bounding limits. Otherwise we will derive an uncertainty estimate that is too large.

With these three criteria in mind, we can identify situations where it is appropriate to model the statistical behavior of an uncertainty contributor by the rectangular distribution.

1) **Digital Resolution**.

The uncertainty due to the finite resolution of digital indicating devices is a common uncertainty contributor. If the resolution of the device is R then we know that an indicated value x could lie anywhere between $x \pm 0.5R$. Further, unless there's some reason to believe otherwise, we can assume that the sensed value has an equal probability of lying anywhere within that interval. In this case the rectangular distribution is a good model for the uncertainty due to finite resolution and the standard uncertainty due to the finite resolution of the indicating device is

$$u_R = \frac{0.5R}{\sqrt{3}}.$$

2) **RF Phase Angle**.

RF power incident on a load will be delivered with a phase angle θ between $-\pi$ and $+\pi$ and the probability of occurrence between these limits is uniform. Consequently the standard uncertainty of the phase angle is

$$u_\theta = \frac{\pi}{\sqrt{3}} \cong 1.81.$$

3) **As an Expression of Ignorance.**

The rectangular distribution is frequently used in cases where the actual distribution is unknown. This is often the case in type B uncertainty estimates where the value and associated uncertainty of an uncertainty contributor might be taken from a reference book. For example, if we needed the coefficient of linear thermal expansion of a material, we might go to a reference book and find a value of "150 $ppm/in°C \pm 20\ ppm/in°C$". This is the only information given in the book and no information is given on how the uncertainty was derived. In cases like this, one typically treats the uncertainty as a rectangular distribution. (However, if the uncertainty is dominated by such a contributor, it is good practice to obtain more information on how it was derived and, in particular, try to determine the actual distribution that applies.)

3.3.2.2 Triangular distribution

It may be the case that we have 100% minimum containment limits but we know that there is a tendency for the values of the uncertainty contributor to be near the center of the distribution. For example, imagine two gage blocks soaking together on a soak plate. After they've reached thermal equilibrium, the most likely value for the difference in temperature between them is zero, and the distribution of possible temperatures on either side of zero tails off quickly to some limiting value (determined experimentally). The simplest probability distribution to model this behavior is the triangular distribution, which is given by the following normalized probability distribution function:

$$f(x) = \begin{cases} \dfrac{(x+a)}{a^2} & for \quad -a \le x \le 0 \\ \dfrac{(a-x)}{a^2} & for \quad 0 \le x \le a. \end{cases}$$

It can be shown that the variance of this distribution is $\sigma^2 = \dfrac{a^2}{6}$ so the standard uncertainty of an uncertainty contributor modeled by a triangular distribution with containment limits $\pm\, a$ is $u = \dfrac{a}{\sqrt{6}}$.

In the gage block example above, suppose we know from experiment that temperature differences between the blocks can be as large as 0.1°*F*. Then we would obtain as the estimate for the uncertainty of the temperature difference between the two blocks

$$u_T = \frac{0.1^\circ F}{\sqrt{6}} = 0.0408^\circ F \, .$$

Use of this distribution is subject to the same criteria given for the rectangular distribution except, obviously, the uniform probability criterion.

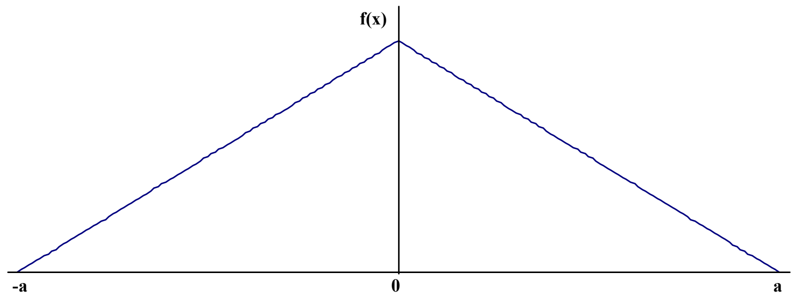

Figure 2. The triangular distribution is used to model cases where 100% containment limits are known and values are more likely to be near the mean than at the extremes.

3.3.2.3 Normal distribution

This distribution is characterized by two parameters: the mean μ, which determines the location of the center of the distribution, and the standard deviation σ, which determines the width of the distribution.

$$f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\left[\frac{(x-\mu)^2}{2\sigma^2}\right]}$$

This distribution is of fundamental importance since it represents the statistical behavior of much of what we see in nature.

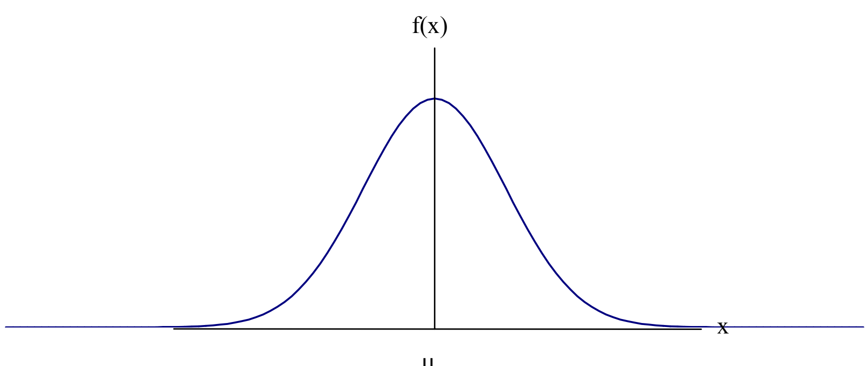

Figure 3. The normal distribution.

3.3.2.4 U distribution

This distribution models situations where the most likely value of a measurand is at or near the containment limits. For example, because of the way thermostats work, room temperature tends to be near the maximum allowed deviation from the set point, i.e., the room temperature is most likely to be too hot or too cold relative to the set point. The probability distribution function is

$$f(x) = \frac{1}{\pi\sqrt{a^2 - x^2}} \quad for \quad -a \prec x \prec a$$

where $\pm\, a$ are the containment limits. The standard deviation is given by $\sigma = \frac{a}{\sqrt{2}}$.

For example, suppose that room temperature is an uncertainty contributor and that the temperature is allowed to vary within the limits of $\pm 5°C$. Then the standard uncertainty due to room temperature variations is given by

$$u_T = \frac{5^\circ C}{\sqrt{2}} = 3.54^\circ C.$$

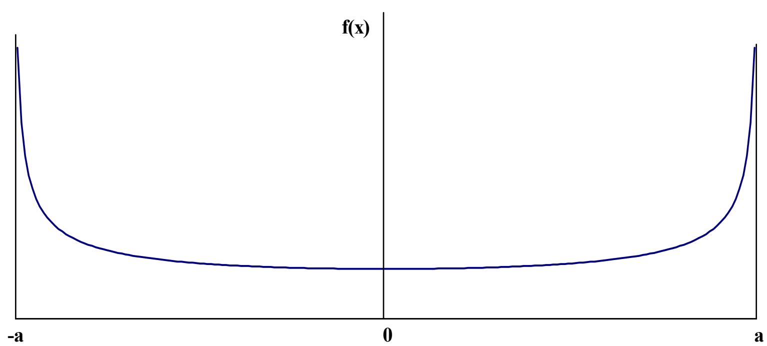

Figure 4. The U distribution models cases where the value of a measurand is likely to be near the containment limits.

3.3.2.5 Poisson distribution

The Poisson distribution is used to model the number of random occurrences of some event in a given unit of time or space. Examples of situations that can be modeled by this distribution include the number of bad parts produced by a machine in a day, the number of flaws in a bolt of fabric, or the distribution of counts detected from a radioactive sample. There are four necessary conditions under which the Poisson distribution is a good approximation:

1) The variable of interest must be a count of events (consequently, the variable of interest must be a positive integer or zero);

2) The random events must be separate and independent of each other;

3) The random events must be equally likely to occur in a given unit of time or space;

4) The random events should be rare relative to the maximum number of events possible.

If the average number of random events per unit of space or time is *m*, then the probability *P* of *x* events occurring in that unit is[5]

$$P(x;m) = \frac{m^x e^{-m}}{x!}$$

where *e* is the base of natural logarithms (*e* is approximately equal to 2.71828).

Figure 5 shows the Poisson distribution for various values of *m*. Note that as *m* increases, the distribution widens and moves to the right. In addition, as m increases the Poisson distribution approaches the normal distribution. Since the Poisson distribution is concerned with discrete counts of events, it is not actually a smooth curve and should more correctly be visualized as a frequency polygon. For our purposes, however, we have connected the dots to unclutter the graph in order to show the behavior of the distribution with increasing *m*.

Looking at the curve for *m* = 1 we see that it is extremely narrow and tails off very rapidly to the right illustrating the fact that if an event is truly rare within some specified interval of time or space, then the probability of many occurrences in the same interval is very low. Conversely, as the average number of occurrences increases, the probability of large departures from the average within the specified interval increases. In addition, since the area under the curve must be equal to 1 regardless of the value of *m*, as the distribution widens the probability of obtaining the average number of events decreases (i.e., the height of the curve decreases).

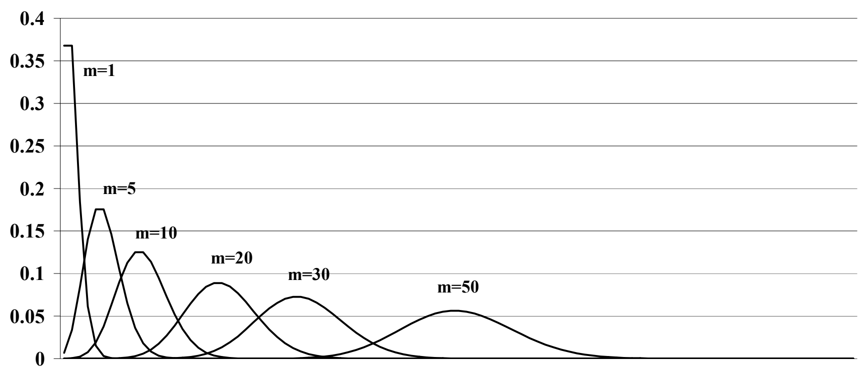

Figure 5. The Poisson distribution for various values of m.

This is one of the more difficult distributions to understand so we shall illustrate its use with several examples. For a more detailed treatment the reader should consult a text on statistics.

Example 1. A molding machine produces 1000 washers per hour and on average 10 of those washers are defective. What is the probability of this machine producing 20 defective washers in an hour?

Solution. In this case the average number of random occurrences is *m* = 10 and we would like to find the probability of obtaining *x* = 20 occurrences. From the equation for the Poisson distribution we find that the probability is

$$P = \frac{m^x e^{-m}}{x!} = \frac{10^{20} e^{-10}}{20!} \cong 0.019 .$$

[5] The symbol "!" is used to denote "factorial". The factorial of a number n is equal to the product of the positive integers lesser than or equal to n: n! = (1)(2)(3)…(n-1)(n)= $\prod_{i=1}^{n} i$. By definition, 0! = 1.

This is a pretty low probability, so if 20 defective washers were ever produced in an hour, we can be quite confident that it was not due to chance: we would look to the molding machine, the material, or the operator for the cause.

Example 2. The average number of counts detected by a radiation counter is 1 per second. What is the probability of obtaining 5 counts in one second?

Solution. This is the same problem as Example 1 but with $m = 1$ and $x = 5$. So the probability of obtaining 5 counts in one second is

$$P = \frac{m^x e^{-m}}{x!} = \frac{1^5 e^{-1}}{5!} \cong 0.0031 .$$

The Poisson distribution has the interesting property that its standard deviation is equal to the square root of the average number of occurrences:

$$\sigma = \sqrt{m} .$$

With this property in mind we can answer such questions as in Example 3.

Example 3. 0.1 *ml* of a 10^5 dilution of a bacterial culture is spread on a nutrient plate. The following day, 57 colonies are observed. How many bacteria were in the original culture?

Solution. We would like to express our answer in terms of *cfu*/*ml* ± *s*, where *s* is the standard deviation of the probability distribution function governing the distribution of bacterial colonies on the surface of a plate. This distribution is the Poisson distribution so the standard deviation is $s = \sqrt{57} \cong 7.55$. The number of "colony forming units" in the original culture must have been

$$\frac{(57cfu)(10^5)}{0.1ml} = 5.70x10^7\, cfu/ml$$

with a standard deviation of

$$\frac{(7.55cfu)(10^5)}{0.1ml} \cong 0.76x10^7\, cfu/ml$$

so our result is $(5.7 \pm 0.8) \times 10^7$ *cfu*/*ml*. We round the final estimate to one decimal place so as not to convey the impression of great accuracy in our estimate.

3.3.2.6 Summary

For containment limits ±*a*, the standard uncertainty estimates associated with the various probability distributions are as follows:

Rectangular: $\frac{a}{\sqrt{3}} \cong 0.5774a$

Triangular: $\frac{a}{\sqrt{6}} \cong 0.4082a$

U: $\frac{a}{\sqrt{2}} \cong 0.7071a$.

For the Poisson distribution, the standard uncertainty estimate is $\sqrt{m}$ where m is the average number of random occurrences in a given time or space interval.

3.3.3 Type B evaluation of standard uncertainty

Some uncertainty contributors cannot be evaluated statistically, or else a statistical evaluation would be impractical, or a statistical evaluation may simply be unnecessary. In these cases, the magnitude and associated uncertainty of an influence quantity has to be estimated based on past experience, taken from a handbook, extracted from a calibration report, etc. Estimates obtained in this way are called type B estimates. (As with type A uncertainty estimates, the reader should note that the designation "type B" refers to the *method* by which the uncertainty estimate was obtained. "Type B" does not refer to the nature of the uncertainty contributor itself; in particular, the reader should avoid the temptation to identify type B uncertainty estimates as "systematic" components of uncertainty.)

Example. The type B estimates of uncertainties in the tensile strength case we are examining come from the uncertainty due to the finite resolution of the micrometer and the calibration uncertainty of the testing machine.

We record the dimensions of the test bar only to the nearest 0.001 *in*, but suppose that the micrometer can read to 100 μin. We have then a rectangular distribution with containment limits ±50 μin and a standard uncertainty of

$$u_R = \frac{50\mu in}{\sqrt{3}} = 28.9\mu in \ .$$

The standard deviation of the thickness measurements was 1000 μin and the standard deviation of the width measurements was 837 μin. The uncertainty due to the finite resolution of the micrometer is so much smaller than the standard deviation of the thickness and width measurements that we shall ignore it and take as the uncertainty of the thickness and width measurements the standard deviation of these measurements. Clearly, our micrometer is much better than our molding and curing process.

The calibration certificate of the testing machine says only that it meets D638 requirements. D638 requires that the indication of the testing machine has to be correct to within ±1%. The average of the force indications in this example was 852 *lb*, so we know that the average of the force indication is good to within approximately ±8.5 *lb*. This is much smaller than the standard deviation of the 5 test results, so we will neglect the calibration uncertainty of the testing machine and take as the uncertainty in the force needed to break the bars the standard deviation of the 5 test results, or 34.93 *lb*. Clearly, variability in the material itself overwhelms any uncertainty contribution from the testing machine.

In this case it turns out that the variability in the material itself, as well as variability in the molding and curing processes, leads to type A estimates of uncertainties that are much larger than possible type B estimates of uncertainties. This is often the case with materials testing, particularly plastics and rubber, but it is not necessarily the case and in general the relative sizes of type A and type B estimates of uncertainties should be at least roughly estimated for different materials before neglecting one or the other.

3.4 Sensitivity coefficients

Sensitivity coefficients are essentially conversion factors that allow one to convert the units of an input quantity into the units of the measurand. For example, if our measurand is "resistance" (measured in ohms, Ω) and if temperature (measured in degrees Celsius, $°C$) is an input quantity, then we can see immediately that to convert a temperature into a resistance we will have to multiply the temperature by some constant c with units of $\Omega/°C$.

Sensitivity coefficients are also, and more importantly, measures of how much change is produced in the measurand by changes in an input quantity. Mathematically, sensitivity coefficients are obtained from partial derivatives of the model function f with respect to the input quantities. In particular, the sensitivity coefficient c_i of the input quantity x_i is given by

$$c_i = \frac{\partial f}{\partial x_i}$$

which expresses mathematically how much f changes given an infinitesimal change in x_i.

Example. The model function we're using for the tensile strength determination is

$$S = \frac{F}{TW}$$

where S is the tensile strength, F is the force needed to break a test bar, and T and W are the thickness and width respectively of the test bar. We obtain the sensitivity coefficients as follows:

$$c_F = \frac{\partial S}{\partial F} = \frac{1}{TW} = \frac{S}{F}; \;\; c_T = \frac{\partial S}{\partial T} = \frac{-F}{T^2 W} = \frac{-S}{T}; \;\; c_W = \frac{\partial S}{\partial W} = \frac{-F}{TW^2} = \frac{-S}{W}\,.$$

The average thickness of the test bars is 0.125 *in*, the average width is 0.4998 *in*, the average force needed to break the bars is 852.0 *lb* and the average tensile strength is 13673 lb/in^2. With these values we can determine the values of each of the sensitivity coefficients:

$$c_F = \frac{S}{F} = \frac{13637\, {}^{lb}\!/_{in^2}}{852 lb} = +16.01 in^{-2};$$

$$c_T = \frac{-S}{T} = \frac{-13637\, {}^{lb}\!/_{in^2}}{0.125 in} = -109100\frac{lb}{in^3}\,;$$

$$c_W = \frac{-S}{W} = \frac{-13637\, {}^{lb}\!/_{in^2}}{0.4998 in} = -27285\frac{lb}{in^3}\,.$$

Sensitivity coefficients can also be evaluated experimentally. In cases where the model function is not known, obviously it is not possible to determine the sensitivity coefficients mathematically so it would be necessary to devise suitable experiments to determine the sensitivity coefficients. In this tensile strength measurement, for example, we could evaluate experimentally the effect of varying thickness on the tensile strength result on a given material while holding the width and force constant. But experiment has its limits. In this example, holding the thickness and width constant while varying the force would be extremely difficult, if not impossible, since, as we have seen, variability in the material itself is a major factor in the uncertainty of the tensile strength.

In principle, however, the experimental evaluation of sensitivity coefficients is simple and works like this:

If the variable of interest is x, we begin by selecting a small range centered on the expected or typical value of x. For example, we might vary the thickness of the tensile strength sample over the range 0.124 *in* to 0.126 *in*. Having selected the range and prepared suitable test specimens, perform the complete test method on each of the test samples, and plot the results with the value of x on the x-axis of the graph and the test results on the y-axis. Fit a least-squares line through the data points. The slope of this line will be the desired sensitivity coefficient.

This method will work if the change in test result is approximately linear over the range of varying x selected. If the behavior is highly non-linear over the selected range, this method will not work. However, if it is possible to narrow the range over which x is varied, then this narrower range will be more nearly linear. (In other words, a suitably small segment of a curve will be approximately linear.)

Alternatively, sensitivity coefficients can be estimated numerically if the model function is known. Since the partial derivatives given above simply express how much the measurand changes given an infinitesimal change in one of the input quantities, while holding the other input quantities constant, we can numerically estimate the sensitivity coefficients as follow.

1) Estimating the force sensitivity coefficient.

We are interested in estimating how much the tensile strength changes given small changes in the force, while holding the thickness and width constant. To do this, we can use our model function to determine the tensile strength at the average thickness and width at three different forces: one force will be the average force determined experimentally as above; the other two forces will be slightly lesser and slightly greater than the average force.

The average thickness of the test bars is 0.125 *in*, the average width is 0.4998 *in*, and the average force needed to break the bars is 852 *lb*. With these values, we can determine numerically the value of the force sensitivity coefficient:

Force	Thickness	Width	Tensile Strength	Difference
851 *lb*	0.125 *in*	0.4998 *in*	13621 lb/in^2	16 lb/in^2
852 *lb*	0.125 *in*	0.4998 *in*	13637 lb/in^2	
853 *lb*	0.125 *in*	0.4998 *in*	13653 lb/in^2	16 lb/in^2

From these calculations, we can see that a 1 *lb* change in the force causes a change of 16 lb/in^2 in the tensile strength. In other words, the sensitivity coefficient is 16 $(lb/in^2)/lb$, or 16 in^{-2}. This happens to be almost exactly the same sensitivity coefficient as determined above using the partial derivatives directly.

2) Estimating the thickness sensitivity coefficient.

To determine the thickness sensitivity coefficient, we proceed exactly as we did with the force: we hold the force and width constant and calculate the tensile strength for small changes in the thickness.

Force	Thickness	Width	Tensile Strength	Difference
852 *lb*	0.1249 *in*	0.4998 *in*	13648 lb/in^2	-11 lb/in^2
852 *lb*	0.1250 *in*	0.4998 *in*	13637 lb/in^2	
852 *lb*	0.1251 *in*	0.4998 *in*	13627 lb/in^2	-10 lb/in^2

If we average these two differences, we see that a 0.0001 *in* change in the thickness causes a -10.5 lb/in^2 change in the tensile strength. In other words, the thickness sensitivity coefficient is:

$$c_T = \frac{-10.5 \, lb/in^2}{0.0001 in} = -105000 \, lb/in^3 .$$

This estimate differs from the estimate obtained above using partial derivatives by less than 4%. This is pretty good, but we could improve this estimate by considering smaller changes in the thickness and retaining more significant figures in the intermediate calculations as follows:

Force	Thickness	Width	Tensile Strength	Difference
852 *lb*	0.12499 *in*	0.4998 *in*	13638.55 lb/in^2	-1.10 lb/in^2
852 *lb*	0.12500 *in*	0.4998 *in*	13637.45 lb/in^2	
852 *lb*	0.12501 *in*	0.4998 *in*	13636.36 lb/in^2	-1.09 lb/in^2

Averaging these two differences, we see that a 0.00001 *in* change in the thickness produces a change in the tensile strength of -1.095 lb/in^2. So this estimate of the thickness sensitivity coefficient is -109500 lb/in^3. The estimate obtained using partial derivatives was -109100 lb/in^3, so we can see that just by considering smaller changes in the input quantity and retaining more significant figures in the intermediate calculations we can greatly improve the numerical estimate of a sensitivity coefficient. We could go even further considering ever-smaller changes in the thickness and retaining more significant figures in the intermediate calculations:

Force	Thickness	Width	Tensile Strength	Difference
852 *lb*	0.124999 *in*	0.4998 *in*	13637.564 lb/in^2	-0.109 lb/in^2
852 *lb*	0.125000 *in*	0.4998 *in*	13637.455 lb/in^2	
852 *lb*	0.125001 *in*	0.4998 *in*	13637.346 lb/in^2	-0.109 lb/in^2

This yields an estimate of the thickness sensitivity coefficient of -109000 lb/in^3.

3) Estimating the width sensitivity coefficient.

To determine the width sensitivity coefficient, we proceed exactly as we did with the force and thickness: we hold the force and thickness constant and calculate the tensile strength for small changes in the width:

Force	Thickness	Width	Tensile Strength	Difference
852 *lb*	0.125 *in*	0.49979 *in*	13637.73 lb/in^2	-0.28 lb/in^2
852 *lb*	0.125 *in*	0.49980 *in*	13637.45 lb/in^2	
852 *lb*	0.125 *in*	0.49981 *in*	13637.18 lb/in^2	-0.27 lb/in^2

(In this example, we have retained more significant figures than in the previous two examples because the tensile strength is relatively insensitive to small changes in width, as can be seen in the table.)

Averaging these two differences, we see that a 0.00001 *in* change in the width causes a -0.275 lb/in^2 change in the tensile strength. Therefore, the width sensitivity coefficient is

$$c_W = \frac{-0.275 \, {lb}/{in^2}}{0.00001 in} = -27500 \, {lb}/{in^3} .$$

This differs from the estimate obtained using partial derivatives by less than 1%.

The important things to remember when evaluating sensitivity coefficients numerically are:

1) Evaluate the sensitivity coefficients using *small* changes in the parameter of interest. It would not do, for example, to evaluate the tensile strength at 0.025, 0.125, and 0.225 inches when estimating the thickness sensitivity coefficient: these changes are far too gross to approximate the infinitesimal changes presupposed in the determination by partial derivatives. For example, using these numbers, we might try to estimate the thickness sensitivity coefficient as follows:

Force	Thickness	Width	Tensile Strength	Difference
852 *lb*	0.025 *in*	0.4998 *in*	68187 *lb*/*in*^{2}	-54550 *lb*/*in*^{2}
852 *lb*	0.125 *in*	0.4998 *in*	13637 *lb*/*in*^{2}	
852 *lb*	0.225 *in*	0.4998 *in*	7576 *lb*/*in*^{2}	-6061 *lb*/*in*^{2}

Averaging these differences shows that a 0.1 *in* change in the thickness produces a change in the tensile strength of -30306 *lb*/*in*^{2} for a sensitivity coefficient of -303060 *lb*/*in*^{3}, which is almost three times larger than the estimate obtained using partial derivatives.

On the other hand, this stricture is important only when the measurand is sensitive to changes in an input quantity. For example, tensile strength is not very sensitive to changes in the force. If we had estimated the force sensitivity coefficient at 752 *lb*, 852 *lb*, and 952 *lb* we would still have obtained a sensitivity coefficient of about 16 *in*$^{-2}$. Still, to be safe, it is best to consider small changes in the input quantity.

2) Retain enough significant figures in the intermediate calculations so that small changes in the measurand can be detected, as in the width and thickness examples above. Using width as an example, suppose that we used the following data obtained simply by rounding the tensile strength estimate to the nearest pound per square inch:

Force	Thickness	Width	Tensile Strength	Difference
852 *lb*	0.125 *in*	0.49979 *in*	13638 *lb*/*in*^{2}	-1 *lb*/*in*^{2}
852 *lb*	0.125 *in*	0.49980 *in*	13637 *lb*/*in*^{2}	
852 *lb*	0.125 *in*	0.49981 *in*	13637 *lb*/*in*^{2}	0 *lb*/*in*^{2}

Then we would obtain as our estimate of the width sensitivity coefficient –50000 *lb*/*in*^{3}, which is clearly an inadequate estimate given what we already know about its best estimated value. The reason this estimate is so poor is that we didn't retain enough significant figures in the intermediate calculations of tensile strength to see accurately the small changes that changes in width produce in the tensile strength.

If the model function is known, it is probably simpler just to obtain the sensitivity coefficients directly from the partial derivatives. However, if the model function is complex then numerical approximation may be simpler. For example, the function defining Brinell hardness B in terms of applied force F, diameter of the indenter D, and diameter of the indentation d

$$B = \frac{0.204F}{\pi D\left(D - \sqrt{D^2 - d^2}\right)}$$

is differentiable in D and d, but a good deal of work is involved and the probability of making a mistake when taking the partial derivatives of B with respect to D and d is not negligible. (See Appendix 3, however, for an example of an uncertainty estimate for Brinell hardness. It would be a good exercise for the reader to estimate numerically the sensitivity coefficients in that example.)

3.5 Combining the contributors

3.5.1 Non-correlated input quantities

Once all of the values of the uncertainty contributors u_i have been estimated and reduced to one standard deviation, and the sensitivity coefficients c_i have been determined, it is usually necessary only to "root-sum-square" their products, i.e., take the square root of the sum of the squares of the uncertainty estimates multiplied by the squares of their corresponding sensitivity coefficients, in order to determine the combined standard uncertainty u_c:

$$u_c = \sqrt{\sum_i c_i^2 u_i^2} \ .$$

Alternatively, we can determine the combined variance, which is simply the square of the combined standard uncertainty[6]:

$$u_c^2 = \sum_i c_i^2 u_i^2 \ .$$

In what follows, the equations will be simpler if we look at the combined variance.

3.5.2 Correlated input quantities

An important complication arises when input quantities are correlated. Correlation occurs when the values of input quantities are not independent. For example, in the tensile strength measurement we've been examining, the measurements of the thickness and width of the test bar are correlated because both quantities are measured with the same micrometer. Or suppose that a load cell is calibrated with a set of 10-pound deadweights all calibrated at the same calibration laboratory. In this case, amongst the input quantities are the uncertainties of the various combinations of deadweights and these uncertainties are correlated- the errors from the calibration lab are passed on to the calibration uncertainty of each of the weights which in turn impact the uncertainty of the load cell calibration. Correlated input quantities are common in testing so, although the subject is complicated, we have no choice but to examine how to handle them. For a more detailed treatment of the subject, the reader is urged to consult the GUM.

In the case of correlated input quantities, the combined variance is given by

$$u_c^2 = \sum_{i=1}^{n} c_i^2 u_i^2 + 2\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} c_i c_j u_i u_j r(x_i, x_j) \ .$$

[6] A simplified derivation of this expression can be found in the GUM section E.3.

The correlation coefficient $r(x_i,x_j)$ characterizes the degree of correlation between the input quantities x_i and x_j. For noncorrelated (independent) input quantities, r will be equal to zero. For perfectly correlated input quantities r will equal ±1. For varying degrees of correlation, r will vary between +1 and –1.

To estimate the correlation coefficient, we first must estimate the "covariance" of the two correlated input quantities. The covariance of the means of two correlated input quantities is estimated by

$$s(\overline{q},\overline{r}) = \frac{1}{n(n-1)}\sum_{k=1}^{n}(q_k-\overline{q})(r_k-\overline{r}) .$$

The correlation coefficient is then given by

$$r(\overline{x}_i,\overline{x}_j) = \frac{s(\overline{x}_i,\overline{x}_j)}{s(\overline{x}_i)s(\overline{x}_j)}$$

where $s(\overline{x}_i,\overline{x}_j)$ is as given above, and $s(\overline{x}_i)$ and $s(\overline{x}_j)$ are the experimental standard deviations of the input quantities x_i and x_j.

Example. The measurements of the thickness and width of the test bars are correlated, as discussed above, since T and W are measured independently and nearly simultaneously by the same micrometer. Since we are dealing with correlated input quantities we must use the full expression for the combined variance:

$$u_c^2 = \sum_{i=1}^{N}c_i^2u_i^2 + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}c_ic_ju_iu_jr(x_i,x_j) .$$

Consequently, we have to determine the estimated correlation coefficients from the means and standard deviations of the thickness, width, and force:

$$\overline{T} = 0.125in;\ \ \overline{W} = 0.4998in;\ \ s_T = 0.001in;\ \ s_W = 8.37x10^{-4}in$$

$$s(\overline{T},\overline{W}) = \frac{1}{n(n-1)}\sum_{i=1}^{n}(T_i-\overline{T})(W_i-\overline{W}) = 1.5x10^{-7}in^2$$

$$r(\overline{T},\overline{W}) = \frac{s(\overline{T},\overline{W})}{s(\overline{T})s(\overline{W})} = \frac{1.5x10^{-7}in^2}{(0.001in)(8.37x10^{-4}in)} = 0.179$$

(Although these quantities are indeed correlated, an examination of the correlation coefficient $r(\overline{T},\overline{W})$ shows that the degree of correlation is fairly small. We could probably ignore correlation effects without greatly changing the final uncertainty estimate, but we will proceed to consider correlation effects. Later, we'll see how the final estimate compares to an estimate that ignores correlation effects.)

Using the information we've gathered so far, we find that the combined variance is

$$
\begin{aligned}
u_c^2 &= c_F^2 u_F^2 + c_T^2 u_T^2 + c_W^2 u_W^2 + 2 c_T c_W u_T u_W r(\overline{T},\overline{W}) \\
&= \left(312570 + 11902.73 + 521.1613 + 893.0621\right)\frac{lb^2}{in^4} \\
&= 325887\frac{lb^2}{in^4}
\end{aligned}
$$

which yields a combined standard uncertainty of

$$u_c = 571\frac{lb}{in^2}.$$

As noted above, although we are dealing with correlated input quantities, the degree of correlation is small and it may be safe to ignore the correlation. If we ignore correlation, and simply root-sum-square the contributors and their associated sensitivity coefficients, as in section 3.5.1 above, then we obtain for the combined standard uncertainty

$$
\begin{aligned}
u_c &= \sqrt{c_F^2 u_F^2 + c_T^2 u_T^2 + c_W^2 u_W^2} \\
&= \sqrt{(16.01)^2 (34.93)^2 + (-109100)^2 (0.001)^2 + (-27286)^2 (0.000837)^2} \\
&= 570\frac{lb}{in^2}.
\end{aligned}
$$

This differs from the more complete estimate by less than 0.2%. This is not meant to imply that it is always safe to ignore correlation effects. It just happens that in this example it is a safe approximation.

3.5.3 Relative combined variance

If the model function f is of the form

$$f = c x_1^{p_1} x_2^{p_2} \ldots x_n^{p_n}$$

and the exponents p_i are known positive or negative numbers, and if the uncertainty contributors are not correlated, then the *relative combined variance* can be determined from

$$\left[\frac{u_c(y)}{y}\right]^2 = \sum_{i=1}^{n}\left[\frac{p_i u(x_i)}{x_i}\right]^2.$$

Example. The model function for our tensile strength measurement is

$$S = \frac{F}{TW}$$

which can also be written as

$$S = F T^{-1} W^{-1}.$$

Since we know that the degree of correlation between the thickness and width measurements is small, we can ignore correlation effects and find the relative combined variance and relative combined standard uncertainty as follows:

$$\left[\frac{u_c(S)}{S}\right]^2 = \left[\frac{(1)(34.93)}{852}\right]^2 + \left[\frac{(-1)(0.001)}{0.125}\right]^2 + \left[\frac{(-1)(0.000837)}{0.4998}\right]^2 \cong 0.001748;$$

$$\frac{u_c(S)}{S} = \sqrt{0.001748} \cong 0.042 = 4.2\%.$$

The estimate of the tensile strength result was 13636 lb/in^2 and 4.2% of this is equal to about 570 lb/in^2 as expected.

3.6 Calculating the expanded uncertainty

The GUM method of uncertainty estimation relies on identifying and quantifying the uncertainties of the input quantities and expressing those uncertainties as one standard deviation. The combined standard uncertainty is consequently a standard deviation and for a normal distribution one standard deviation encompasses approximately 68% of possible values of the measurand.

Although the combined standard uncertainty can be used to express the uncertainty of a measurement result, in some commercial, industrial, or regulatory applications, or when health and safety are concerned, it is often necessary to give a measure of uncertainty that defines an interval about the measurement result that may be expected to encompass a larger fraction of the values that could reasonably be attributed to the measurand than does a single standard deviation.

The additional measure of uncertainty that encompasses a large fraction of expected values of the measurand is called expanded uncertainty and is denoted by U. The expanded uncertainty U is obtained by multiplying the combined standard uncertainty by a coverage factor k:

$$U = ku_c(y).$$

The procedure for determining the coverage factor will simply be presented here. The reader is urged to consult the GUM for more information and the rationale behind the procedure.

3.6.1 Estimating the coverage factor *k*.

To obtain the value of the coverage factor k, it is necessary to take into account the uncertainty of the estimate of the combined standard uncertainty $u_c(y)$. This uncertainty is characterized by the "effective degrees of freedom" ν_{eff} of $u_c(y)$ which is a measure of how much information was available to make the estimate of $u_c(y)$. A large number of degrees of freedom implies that more information was available for the estimate of $u_c(y)$.

To encompass approximately 95% of the possible values of the measurand (95% is just a conventional level of confidence), it is usually the case that the coverage factor k will be a number in the range of 2 to 3. For large values of ν_{eff}, k will be close to 2. This corresponds to the fact that two standard deviations encompass approximately 95% of a normal distribution. If limited information was available in making the uncertainty estimate, so that the uncertainty of the estimate is large, this will be reflected by a small number of degrees of freedom and a large value of k.

The four-step procedure for calculating k is:

1) Obtain the estimate of the measurand y and the estimate of the combined standard uncertainty $u_c(y)$.

2) Estimate the effective degrees of freedom ν_{eff} using the Welch-Satterthwaite formula:

$$\nu_{eff} = \frac{u_c^4(y)}{\sum_{i=1}^{n} \frac{c_i^4 u^4(x_i)}{\nu_i}}$$

where ν_i is the degrees of freedom of the estimate of the magnitude of the uncertainty contributor x_i. The degrees of freedom of a Type A evaluation based on n repeated measurements is simply $\nu = n - 1$. If n independent observations are used to determine both the slope and intercept of a straight line by the least squares method, then the degrees of freedom of their respective standard uncertainties is $\nu = n - 2$. In general, for a least-squares fit of m parameters to n data points the degrees of freedom of the standard uncertainty of each parameter is $\nu = n - m$.

Estimating the degrees of freedom of a type B estimate of uncertainty can be quite difficult but the GUM gives an expression that can be used to estimate it[7]:

$$\nu = \frac{1}{2}\left[\frac{\Delta u(x_i)}{u(x_i)}\right]^{-2}.$$

The quantity in large brackets is just the relative uncertainty of the estimate of $u(x_i)$. For a Type B evaluation of standard uncertainty, it is a subjective quantity that may be estimated based on professional judgement.

This is an important problem since, for example, the uncertainty of virtually every test will be affected by the calibration uncertainty of a testing machine or measuring device. This calibration uncertainty is obtained as a Type B estimate since its value is obtained from the calibration report. However, few calibration labs will report the number of degrees of freedom associated with their estimate of the calibration uncertainty, and it should not simply be assumed that it is "infinite" (as is often done) since estimates of calibration uncertainties might be good only to within 25% - maybe better, maybe worse - depending on the calibration lab. (This doesn't necessarily mean that the calibration lab is not doing a good job, it does mean that uncertainty estimates are themselves subject to uncertainty.)

Although this is the procedure recommended by the GUM, there are serious concerns about the idea of trying to quantify the uncertainty of an uncertainty estimate[8]. While such an idea is certainly intellectually tractable, as a practical matter it should be recognized that it can be difficult enough to evaluate the uncertainty of a measurement result in a meaningful way, let alone the uncertainty of that uncertainty. The authors of the CODATA paper base their concerns on three points:

i) "… although carrying out Type B evaluations of uncertainty is rarely easy, … such evaluations are usually done reliably for known effects. The difficulty with [an uncertainty estimate] most often arises from an unrecognized effect, that is, an effect for which no component of uncertainty has been evaluated because its existence was not realized. Trying to assign an 'uncertainty to an uncertainty' based only on known components of uncertainty is not necessarily reliable.

ii) "… if there are doubts about the reliability of an initially assigned uncertainty, then one should use the information on which the doubts are based to reevaluate it (which in most cases means increasing the uncertainty) so that the doubts are removed. In short, all available information should be used in the evaluation of components of uncertainty.

iii) "The third and final observation concerns the possibility of including a margin of safety in the [uncertainty estimates] as is sometimes suggested. In particular, should the [uncertainty estimates] include an extra component so that they are 'certain' to be correct? … [such extra components of

[7] GUM G.4.2.

[8] P. J. Mohr and B. N. Taylor, *CODATA recommended values of the fundamental physical constants: 1998*, Rev. Mod. Phys., Vol. 72, No. 2, April 2000, p. 358.

uncertainty should not be included] but rather, the best values based on all the available information [should be used], which in some cases means relying on the validity of the result of a single experiment or calculation. This approach … provides a faithful representation of our current state of knowledge with the unavoidable element of risk that that knowledge may include an error or oversight."

So, while it is best to expend some effort in determining the best value of a type B estimate of uncertainty and associated degrees of freedom to the degree commensurate with the importance and intended use of the uncertainty estimate, we shall illustrate the GUM method with an example.

Example. Suppose that one's knowledge of how input quantity estimate x_i was determined and how its standard uncertainty was evaluated leads one to judge that the value of $u(x_i)$ is good to within about 10%. This may be taken to mean that the relative uncertainty is 0.10 and so we obtain as the estimate of the degrees of freedom of this uncertainty estimate

$$\nu = \frac{(0.10)^{-2}}{2} = 50.$$

If the available information leads one to conclude that the estimate was good to within only about 25%, then one would have obtained

$$\nu = \frac{(0.25)^{-2}}{2} = 8.$$

In the case of type B estimates of the uncertainties of input quantities modeled by probability distribution functions with finite containment limits $\pm a$, the equation for estimating the degrees of freedom of the uncertainty estimate implies that $\nu_i \to \infty$ since the containment limits $-a$ and $+a$ are chosen in such a way that the probability of the quantity in question lying outside the limits is extremely small or zero. Allowing ν_i to approach infinity poses no difficulty in applying the Welch-Satterthwaite formula since we have the following well-known limit:

$$\lim_{\nu \to \infty} \frac{1}{\nu} = 0.$$

If all of the uncertainty contributors are obtained via Type B estimates with infinite degrees of freedom, the estimate of the combined standard uncertainty will also have infinite degrees of freedom.

3) Obtain the *t*-factor $t_p(\nu_{eff})$ from the Student's *t*-table (see Appendix 2) corresponding to the degrees of freedom ν_{eff} and the required level of confidence p. If ν_{eff} is not an integer, as will usually be the case, one should round or truncate ν_{eff} down to the nearest integer.

4) Take $k = t_p(\nu_{eff})$ and find the expanded uncertainty $U = ku_c(y)$.

Example. Continuing with the tensile strength example, we calculate the number of effective degrees of freedom ν_{eff} as follows:

Recall that the combined standard uncertainty u_c was found to be 571 lb/in^2. The uncertainties of the force, thickness and width measurements were 34.93 *lb*, 0.001 *in*, and 0.000837 *in* respectively. Each of these three uncertainties were estimated on the basis of 5 repeat measurements. Consequently, we have Type A estimates of uncertainties with $n = 5$ and $\nu_i = n - 1 = 4$ degrees of freedom. Finally, the sensitivity coefficients of these input quantities were found to be $c_F = 16.01\ in^{-2}$, $c_T = -109100\ lb/in^3$, and $c_W = -27286\ lb/in^3$.

Using the Welch-Satterthwaite equation we obtain the following estimate of the effective degrees of freedom of the estimate of the combined standard uncertainty:

$$\nu_{eff} = \frac{u_c^4}{\frac{c_F^4 u_F^4}{\nu_F} + \frac{c_T^4 u_T^4}{\nu_T} + \frac{c_W^4 u_W^4}{\nu_W}}$$

$$= \frac{(571)^4}{\frac{(16.01)^4 (34.93)^4}{4} + \frac{(-109100)^4 (0.001)^4}{4} + \frac{(-27286)^4 (0.000837)^4}{4}} \approx 4 .$$

Consulting the Student's *t*-table (see Appendix 2), we find that the value of *t* corresponding to 4 degrees of freedom at the 95% level of confidence is $t = 2.78$. So the expanded uncertainty of the tensile strength result is

$$U = ku_c = (2.78)(571\frac{lb}{in^2}) = 1587\frac{lb}{in^2} \approx 1600\frac{lb}{in^2} .$$

This is about 12% of our test result of 13600 psi. We round the final uncertainty estimate to no more than two significant figures so as not to convey the impression of greater accuracy than is warranted. Standard rounding practice such as the one found in section 6.4 of ASTM E29 should be followed, although it is common practice always to round uncertainty estimates *up* to one or two significant figures so as to obtain a "conservative" estimate. Guidelines for retaining significant figures in test results can be found in section 7 of E29. It is best to do these uncertainty calculations on a spreadsheet so that intermediate-rounding errors can be avoided. In this case, rounding errors were negligible.

3.7 Reasonability

Although this uncertainty estimate involved a good deal of work, we can be reasonably confident that we have the best estimate of the final tensile strength result given the available information. For a material destined for a critical automotive component, for example, we may be justified in investing such effort in our uncertainty estimation. (Of course, in such a case it would be even more important to collect far more data than we did in this example.) However, it is often the case that, even for critical components, engineering tolerances are established such that the finished product can withstand fairly large variations in material component properties. Which is to say that we are unlikely ever to need to devote such effort to an uncertainty analysis as we have done here.

The above is not meant to trivialize the need for rigorous estimation of measurement uncertainty, but let's take a "cruder" approach to the estimation of the uncertainty of our tensile strength result and see how it compares to the more rigorous approach:

Suppose that we know, from long experience with the mechanical properties of plastics, that the calibration uncertainties of the testing machine and micrometer, as well as the finite resolution of these devices, are negligible compared to the variability of the material itself. Then, as a quick first approximation to the uncertainty of a tensile strength measurement, we might look to the standard deviation of the 5 test results. In this case, we have the standard deviation as 482.8 lb/in^2. This standard deviation was derived based on 5 measurements, so we know that we have 4 degrees of freedom. Consulting the Student's *t*-table for 4 degrees of freedom we find that $k = 2.78$ is the coverage factor corresponding to the 95% level of confidence. Multiplying the standard deviation of the tensile strength results by 2.78 we obtain a quick uncertainty estimate of approximately 1300 lb/in^2.

This estimate is 19% smaller than the more rigorous estimate of 1600 lb/in^2- the price we pay for estimating uncertainty based on limited information[9]. However, depending on the application and the needs of the customer, it

[9] Although this is often the case, it isn't a general rule. It could be the case that uncertainty estimates based on limited information could be too large. In this case, it happens that the estimate is too small.

could be entirely adequate[10]. The amount of effort invested in an uncertainty estimate must be dictated by the needs of the user of the measurement result and associated uncertainty estimate. If the user is satisfied with a quick estimate then he also must accept the fact that he is assuming the risks associated with making decisions based on an uncertainty estimate that is itself quite uncertain. The important thing is that the uncertainty must be reported correctly, without exaggerating its reliability, and with as much information as is needed to tell the user of the uncertainty how it was estimated. This way, the user can form his own judgements concerning the nature of the estimate and its reliability.

In the end, every uncertainty estimate should be subjected to a "reasonability check". The analyst should ask questions such as "Is this estimate reasonable?" "Is this estimate in line with what I know about the nature of the measurement and of the material?" "Can this estimate be supported with proficiency testing data, or data accumulated as part of a measurement assurance program?" Uncertainty estimates that look strange- either too big or too small- should be re-evaluated, looking first for mathematical blunders, second for uncertainty contributors whose magnitudes may have been poorly estimated or completely neglected. Finally, it may be necessary to revise the mathematical model.

In this tensile strength example, the reasonability test is based on a comparison of the rigorous estimate with the quick estimate and on our own experience with tensile strength determinations of plastic compounds. Based on these criteria, we conclude that either estimate obtained in this example is equally reasonable. Other situations may involve different criteria. But in all cases, reasonability is finally based on "gut feel" and experience[11].

3.8 Reporting uncertainty

When reporting the result of a measurement, at a minimum one should

1) Give a full description of how the measurand Y is defined;

2) State the result of the measurement as $Y = y \pm U$ and give the units of y and U;

3) Give the value of the coverage factor k used to obtain U;

4) Give the approximate level of confidence associated with the interval $y \pm U$ and state how it was determined.

The numerical values of the estimates of the measurand and expanded uncertainty should not be given with an excessive number of significant digits. It usually suffices to quote uncertainty estimates to no more than two significant figures.

When stating measurement results and uncertainty estimates it is always advisable to err on the side of providing too much information rather than too little and this information must be stated as clearly as possible.

Example. The statement of our tensile strength result might take the following form:

---

[10] In fact, we could have dispensed with the rigorous analysis entirely and looked only to the quick estimate. If the quick estimate is not overly optimistic, and if it satisfies the customer's tolerance requirements, we need go no further. If the estimate is felt to be too optimistic based on our experience with the measurement in question, or if it is unsuitable for customer requirements, then we must delve into a more rigorous estimate.

[11] The reader should not allow this fact to lead him to question the "mathematical objectivity" of uncertainty estimates. Although there is undoubtedly a large element of subjectivity involved, a rigorous uncertainty estimate is based on objective mathematical machinery (albeit presented in essentially "recipe" form in this document). But mathematics is just as subject to the "garbage in, garbage out" syndrome as is, for example, computer programming. Human judgement based on sound technical experience and professional integrity is of paramount importance in evaluating the merits of any uncertainty estimate.

"The tensile strength at break of this material as defined by ASTM D638 was found to be (13600 ± 1600) pounds per square inch. The number after the ± symbol is the numerical value of an expanded uncertainty $U = ku_c$, with U determined from a combined standard uncertainty u_c = 571 pounds per square inch using a coverage factor of k = 2.78 and defines an interval estimated to have a level of confidence of 95 percent with 4 effective degrees of freedom."

3.9 Summary of the method

1. Specify the measurand

   Clearly specify the quantity to be determined and the method by which it is to be determined. This specification must include all of the factors that could significantly affect the measurement result including but not limited to environmental conditions, sampling procedures, sample preparation procedures, as well as the test or calibration method to be used.

2. Derive the mathematical model

   Express mathematically the relationship between the measurand and all of the input quantities upon which the measurand depends. This mathematical model should contain every quantity, including all corrections, that can contribute significantly to the uncertainty of the result of the measurement.

3. Quantify the influence quantities

   Estimate the value of each input quantity either by the statistical analysis of repeated observations or by other means such as taking the uncertainty of a reference standard from a calibration certificate, estimating temperature effects on test or calibration results based on theoretical predictions, estimating the uncertainty of physical constants based on data in reference books and so on.

4. Evaluate the standard uncertainty of each influence quantity

   Evaluate the standard uncertainty of the estimates of the values for the input quantities. For the uncertainty of an input estimate based on the statistical analysis of repeated observations, the estimate is made as described in section 2.3.1 (Type A). For estimates obtained by other means, the standard uncertainty is evaluated as described in section 2.3.3 (Type B).

5. Evaluate sensitivity coefficients and covariances

   For noncorrelated input quantities, evaluate their respective sensitivity coefficients either directly by differentiation of the function used to model the measurement or indirectly by experiment. For correlated input quantities, evaluate the associated covariances.

6. Calculate the measurement result

   Calculate the result of the measurement based on the mathematical model adopted and the estimates of the input quantities.

7. Determine the combined standard uncertainty

   Determine the combined standard uncertainty of the measurement result from the standard uncertainties and covariances of the input estimates.

8. Determine the expanded uncertainty

   If it is necessary to provide an expanded uncertainty, first estimate the effective degrees of freedom of the estimate of the combined standard uncertainty. Consulting a t-table, multiply the combined standard uncertainty by the coverage factor associated with the estimated degrees of freedom and desired level of confidence.

9. Report the measurement result and associated uncertainty estimate

   Report the measurement result and either the combined standard uncertainty or the expanded uncertainty as described in section 2.8.

3.10 Uncertainty "budgets"

It should be apparent by now that uncertainty analysis is not a trivial task. Even for the simplest measurements it is often the case that a great deal of thought must go into the estimate of the measurement uncertainty. The person who undertakes an uncertainty analysis has to be thoroughly familiar with the principles of the measurement in question- familiarity with the requirements of the test procedure or standard alone is not sufficient- otherwise significant uncertainty contributors are likely to be poorly estimated or else missed entirely.

Every uncertainty analysis will include some assumptions and it is important that these assumptions be documented and justified. Uncertainty contributors taken to be negligible must still be documented and justified.

It is, unfortunately, common practice to regard uncertainty analysis as the pursuit of an "uncertainty budget", the ubiquitous tables appearing in guidance documents such as UKAS M3003 and even in the GUM itself.

The tables found in these documents are the least important part of the analysis. They merely summarize the data in a convenient format. What is most important is that a clear, well-documented narrative be available for each uncertainty analysis. Such a detailed exposition of an uncertainty analysis is not needed every time an analysis is undertaken, however. If the conditions and assumptions used to estimate an uncertainty are the same in one case as they were in a past case, then the narrative developed for a past case is applicable to the present case and need not be duplicated. At most what may be necessary is to update the values of specific uncertainty contributors if new information becomes available.

Specifically, what is not acceptable as documentation of an uncertainty estimate is a table, often generated by commercially available uncertainty analysis software or freeware, with no indication of any of the assumptions embodied in the table, and/or no indication of where the data in the table came from. In addition, since it is often the case that software developers will use differing terminologies, a narrative accompanying the table is even more vital since part of the narrative will have to be devoted to explaining the terminology used by the software's author.

At a minimum, a well-documented uncertainty evaluation will contain the following elements:

1) The identity and value of each input estimate and its standard uncertainty together with a description of how they were obtained;

2) The estimated covariances or estimated correlation coefficients (preferably both) associated with all input estimates that are correlated, and the methods used to obtain them, or a statement to the effect that they were assumed or found to be negligible and were omitted;

3) The degrees of freedom for the standard uncertainty of each input estimate and how they were obtained and, where appropriate, a calculation of the effective degrees of freedom of the calculated combined standard uncertainty;

4) The functional relationship between the measurand and the input quantities and, if deemed useful, the partial derivatives or sensitivity coefficients, or a statement that they were assumed or found to be negligible. Any experimentally determined coefficients should always be given together with a description of how they were obtained.

Appendix 1. Definitions

accuracy (of measurement) (VIM 3.5): closeness of the agreement between the result of a measurement and a true value of the measurand

NOTES:

1. "Accuracy" is a qualitative concept
2. The term precision should not be used for "accuracy".

{Author's note: "an accepted reference value" may be used in place of "a true value" in this definition.}

bias (ISO 3534-1): the difference between the expectation of the test results from a particular laboratory and an accepted reference value

NOTE: Bias is the total systematic error as contrasted to random error. There may be one or more systematic error components contributing to the bias. A larger systematic difference from the accepted reference value is reflected by a larger bias value.

combined standard uncertainty (GUM 2.3.4): standard uncertainty of the result of a measurement when that result is obtained from the values of a number of other quantities, equal to the positive square root of a sum of terms, the terms being the variances or covariances of these other quantities weighted according to how the measurement result varies with changes in these quantities

correlation (ISO 3534-1): the relationship between two or several random variables within a distribution of two or more random variables

NOTE: Most statistical measures of correlation measure only the degree of linear relationship.

coverage factor (GUM 2.3.6): numerical factor used as a multiplier of the combined standard uncertainty in order to obtain an expanded uncertainty

NOTE: A coverage factor, k, is typically in the range of 2 to 3.

error (of measurement) (VIM 3.10): result of a measurement minus a true value of the measurand

NOTES:

1. Since a true value cannot be determined, in practice a conventional true value is used.
2. When it is necessary to distinguish "error" from "relative error", the former is sometimes called "absolute error of measurement". This should not be confused with "absolute value of error", which is the modulus of the error.

{Author's note: "an accepted reference value" may be used in place of "a true value" in this definition.}

expanded uncertainty (GUM 2.3.5): quantity defining an interval about the result of a measurement that may be expected to encompass a large fraction of the distribution of values that could reasonably be attributed to the measurand

NOTES:

1. The fraction may be viewed as the coverage probability or level of confidence of the interval.
2. To associate a specific level of confidence with the interval defined by the expanded uncertainty requires explicit or implicit assumptions regarding the probability distribution characterized by the measurement

result and its combined standard uncertainty. The level of confidence that may be attributed to this interval can be known only to the extent to which such assumptions may be justified.

influence quantity (VIM 2.7): quantity that is not the measurand but that affects the result of the measurement

Examples:

a) temperature of a micrometer used to measure length;
b) frequency in the measurement of the amplitude of an alternating electric potential difference;
c) bilirubin concentration in the measurement of hemoglobin concentration in a sample of human blood plasma.

level of confidence (GUM C.2.29): The value of the probability associated with a confidence interval or a statistical coverage interval

NOTE: The value is often expressed as a percentage.

measurand (VIM 2.6): particular quantity subject to measurement

Example: Vapor pressure of a given sample of water at 20°*C*.

NOTE: The specification of a measurand may require statements about quantities such as time, temperature, and pressure.

measurement (VIM 2.1): set of operations having the object of determining a value of a quantity

precision (ISO3534-1): the closeness of agreement between independent test results obtained under stipulated conditions

NOTES:

1. Precision depends only on the distribution of random errors and does not relate to the true value or the specified value.
2. The measure of precision is usually expressed in terms of imprecision and computed as a standard deviation of the test results. Less precision is reflected by a larger standard deviation.
3. "Independent test results" means results obtained in a manner not influenced by any previous result on the same or similar test object. Quantitative measures of precision depend critically on the stipulated conditions. Repeatability and reproducibility conditions are particular sets of extreme conditions.

repeatability (VIM 3.6): closeness of the agreement between the results of successive measurements of the same measurand carried out under the same conditions of measurement

NOTES:

1. The conditions are called repeatability conditions.
2. Repeatability conditions include: the same measurement procedure; the same observer; the same measuring instrument used under the same conditions; the same location; repetition over a short period of time.
3. Repeatability may be expressed quantitatively in terms of the dispersion characteristics of the results.

reproducibility (VIM 3.7): closeness of the agreement between the results of measurements of the same measurand carried out under changed conditions of measurement

NOTES:

1. A valid statement of reproducibility requires specification of the conditions changed.

2. The changed conditions may include but are not limited to: principle of measurement; method of measurement; observer; measuring instrument; reference standard; location; conditions of use; time.
3. Reproducibility may be expressed quantitatively in terms of the dispersion characteristics of the results.
4. Results are here usually understood to be corrected results.

standard uncertainty (GUM 2.3.1): uncertainty of the result of a measurement expressed as a standard deviation

trueness (ISO 3534-1): the closeness of agreement between the average value obtained from a large series of test results and an accepted reference value

NOTES:

1. The measure of trueness is usually expressed in terms of bias.
2. Trueness has been referred to as “accuracy of the mean”. This usage is not recommended.

type A evaluation of uncertainty (GUM 2.3.2): method of evaluation of uncertainty by the statistical analysis of observations

type B evaluation of uncertainty (GUM 2.3.3): method of evaluation of uncertainty by means other than the statistical analysis of a series of observations

uncertainty of measurement (VIM 3.9): parameter, associated with the result of a measurement, that characterizes the dispersion of the values that could reasonably be attributed to the measurand

NOTES:

1. The parameter may be, for example, a standard deviation (or a given multiple of it), or the half-width of an interval having a stated level of confidence.
2. Uncertainty of measurement comprises, in general, many components. Some of these components may be evaluated from the statistical distribution of the results of series of measurements and can be characterized by experimental standard deviations. The other components, which can also be characterized by standard deviations, are evaluated from assumed probability distributions based on experience or other information.
3. It is understood that the result of the measurement is the best estimate of the value of the measurand, and that all components of uncertainty, including those arising from systematic effects, such as components associated with corrections and reference standards, contribute to the dispersion.

This definition is that of the “Guide to the expression of uncertainty in measurement” in which its rationale is detailed (see in particular 2.2.4 and Annex D).

Appendix 2. Student's *t*-distribution

Value of $t_p(\nu)$ from the *t*-distribution for degrees of freedom ν that defines an interval $-t_p(\nu)$ to $+t_p(\nu)$ that encompasses the fraction p of the distribution.

Degrees of freedom ν	Fraction p in percent					
	68.27[(a)]	90.00	95.00	95.45[(a)]	99.00	99.73[(a)]
1	1.84	6.31	12.71	13.97	63.66	235.8
2	1.32	2.92	4.30	4.53	9.92	19.21
3	1.20	2.35	3.18	3.31	5.84	9.22
4	1.14	2.13	2.78	2.87	4.60	6.62
5	1.11	2.02	2.57	2.65	4.03	5.51
6	1.09	1.94	2.45	2.52	3.71	4.90
7	1.08	1.89	2.36	2.43	3.50	4.53
8	1.07	1.86	2.31	2.37	3.36	4.28
9	1.06	1.83	2.26	2.32	3.25	4.09
10	1.05	1.81	2.23	2.28	3.17	3.96
11	1.05	1.80	2.20	2.25	3.11	3.85
12	1.04	1.78	2.18	2.23	3.05	3.76
13	1.04	1.77	2.16	2.21	3.01	3.69
14	1.04	1.76	2.14	2.20	2.98	3.64
15	1.03	1.75	2.13	2.18	2.95	3.59
16	1.03	1.75	2.12	2.17	2.92	3.54
17	1.03	1.74	2.11	2.16	2.90	3.51
18	1.03	1.73	2.10	2.15	2.88	3.48
19	1.03	1.73	2.09	2.14	2.86	3.45
20	1.03	1.72	2.09	2.13	2.85	3.42
25	1.02	1.71	2.06	2.11	2.79	3.33
30	1.02	1.70	2.04	2.09	2.75	3.27
35	1.01	1.70	2.03	2.07	2.72	3.23
40	1.01	1.68	2.02	2.06	2.70	3.20
45	1.01	1.68	2.01	2.06	2.69	3.18
50	1.01	1.68	2.01	2.05	2.68	3.16
100	1.005	1.660	1.984	2.025	2.626	3.077
∞	1.000	1.645	1.960	2.000	2.576	3.000

[(a)]For a quantity z described by a normal distribution with expectation μ_z and standard deviation σ, the interval $\mu_z \pm k\sigma$ encompasses p = 68.27, 95.45, and 99.73 percent of the distribution for k = 1, 2, and 3 respectively.

Appendix 3. Uncertainty Estimate for a Brinell Hardness Test

ASTM E10 equation 1 defines Brinell hardness (designated as B for the purposes of this example) as a function of the test force (F, measured in newtons), the diameter of the indenter (D, measured in mm), and the mean diameter of the indentation (d, measured in mm):

$$B = \frac{0.204F}{\pi D\left(D - \sqrt{D^2 - d^2}\right)}$$

The unit for Brinell hardness is thus N/mm^2. The combined variance is

$$u^2(B) = \left(\frac{\partial B}{\partial F}\right)^2 u^2(F) + \left(\frac{\partial B}{\partial D}\right)^2 u^2(D) + \left(\frac{\partial B}{\partial d}\right)^2 u^2(d)$$

and the sensitivity coefficients c_i are given by:

$$c_F = \frac{\partial B}{\partial F} = \frac{0.204}{\pi D\left(D - \sqrt{D^2 - d^2}\right)}$$, with unit of mm^{-2},

$$c_D = \frac{\partial B}{\partial D} = \frac{0.204F\left(\dfrac{D^2}{\sqrt{D^2 - d^2}} + \sqrt{D^2 - d^2} - 2D\right)}{\pi D^2\left(D - \sqrt{D^2 - d^2}\right)^2}$$, with unit of N/mm^3, and

$$c_d = \frac{\partial B}{\partial d} = \frac{-0.204Fd}{\pi D\sqrt{D^2 - d^2}\left(D - \sqrt{D^2 - d^2}\right)^2}$$, also with units of N/mm^3.

For the purpose of this model budget we'll make the following assumptions:

1) Assume that, according to the calibration certificate, the force indication of the testing machine "meets ASTM E10 tolerances". Section 15.1.1 of E10 specifies that "a Brinell hardness testing machine is acceptable for use when the test force error does not exceed ±1%." With a 3000 kgf (29400 N) test force, the uncertainty in the indication is then 294 N.

2) Let's assume that the test is performed with a nominally 10 mm diameter steel ball. Section 5.2.1 of E10 specifies that the maximum permitted deviation from this nominal value in any diameter is 0.005 mm and the calibration certificate states that the ball "meets ASTM E10 tolerances". We'll take 0.005 mm as the uncertainty in the diameter of the ball.

3) In the Brinell hardness test, two perpendicular diameters of the indentation are measured and the mean value of those two diameters is used as the value of d in equation (1). Since one reading is not necessarily representative of the test piece as a whole, let's assume that, in this case, 5 indentations were made on the test piece and that the mean diameters obtained for each indentation were 3.00, 3.10, 2.90, 3.05, and 2.95 mm. The mean of these five results is 3.00 mm and the standard deviation is 0.079 mm. We'll use the standard deviation of 0.079 mm as the uncertainty of the mean of the diameter of any given indentation.

Given these assumptions, we see that, for a 10 mm indenter with a 3000 kgf test force and a 3.00 mm diameter mean indentation, the Brinell hardness number comes out to 415 HBS. We can also evaluate the sensitivity coefficients: $c_F = \frac{\partial B}{\partial F} = 0.0141 mm^{-2}$; $c_D = \frac{\partial B}{\partial D} = 2.001 N \cdot mm^{-3}$; and $c_d = \frac{\partial B}{\partial d} = -283.0 N \cdot mm^{-3}$. The combined variance is then $u^2(B) = 505.6 N^2 \cdot mm^{-4}$ which yields a combined standard uncertainty of $u(B) = 22.5 N \cdot mm^{-2}$.

To determine the appropriate coverage factor to use to find the expanded uncertainty we have to determine the effective degrees of freedom ν_{eff} using the Welch-Satterthwaite equation. For the test force and ball diameter we'll assume infinite degrees of freedom, but for the indentation diameter we have four degrees of freedom so that

$$\nu_{eff} = \frac{u^4(B)}{\frac{c_F^4 u^4(F)}{\infty} + \frac{c_D^4 u^4(D)}{\infty} + \frac{c_d^4 u^4(d)}{4}} = 4 * \frac{22.5^4}{(-283.0)^4 (0.079)^4} \cong 4 \qquad (6).$$

Referring to the t-table, we find that the coverage factor k corresponding to four degrees of freedom at the 95% confidence level is 2.78. We then obtain for the expanded uncertainty U

$$U = k \cdot u(B) = (2.78)(22.5 N \cdot mm^{-2}) \cong 63 N \cdot mm^{-2} \qquad (7).$$

We can now state our measurement result as follows:

"The Brinell hardness of this material as defined in ASTM E10 was found to be 415 HBS ± 63 HBS. The stated uncertainty of 63 HBS is an expanded uncertainty expressed at approximately the 95% confidence level using a coverage factor of k=2.78 corresponding to 4 effective degrees of freedom."

Table 1. Summary Uncertainty Budget.

Uncertainty Source	Estimated Value	Distribution/ Divisor	Sensitivity Coefficient	Degrees of Freedom	Standard Uncertainty
Force Indication	294 N	Rectangular/ √3	0.0141 mm^{-2}	∞	2.393 N/mm^2
Ball Diameter	0.005 mm	Rectangular/ √3	2.001 N/mm^3	∞	0.006 N/mm^2
Indentation Diameter	0.079 mm	Normal/1	-283 N/mm^3	4	22.36 N/mm^2
Effective Degrees of Freedom / Coverage Factor (for 95% level of confidence)			4 / 2.78		
Combined Standard Uncertainty			22.5 N/mm^2		
Expanded Uncertainty			63 N/mm^2 ≅ 15%		

Looking at this table, we can see that the greatest uncertainty contributor by far is the diameter of the indentation (about 10 times larger than the next largest contributor which is the uncertainty due to the force indication), reflecting the fact that, in this example at least, variability in the material itself is the greatest contributor to the uncertainty. Therefore, we can obtain an uncertainty estimate based on just the standard deviation of the 5 hardness determinations.

In this example, the diameters obtained were 3.00, 3.10, 2.90, 3.05, and 2.95 mm corresponding to Brinell hardness values of 414, 388, 444, 401, and 429 HBS. The standard deviation of these 5 hardness values is 22 HBS. Assuming that these 5 results were drawn from a normal distribution with mean 415 HBS and standard deviation 22 HBS, and applying the coverage factor k = 2.78 from the Student's t-table for 4 degrees of freedom at the 95% level of confidence yields an expanded uncertainty of 62 HBS, which is only 1.6% smaller than the initial estimate of 63 HBS.

The lesson to be learned here is that when material variability is the largest uncertainty contributor, it is usually unnecessary to exert all of the effort needed to determine sensitivity coefficients, etc., as we have done in this example. The statistics of the test results will usually provide a sufficient basis for forming a reasonable uncertainty estimate. In this case, the dominant influence quantity can reasonably be described by a normal distribution.

References and Useful Web Sites

References:

ANSI/NCSL, Z540-2-1997, “U.S. Guide to the Expression of Uncertainty in Measurement”, 1st ed., October 1997.

Eurachem/CITAC, “Quantifying Uncertainty in Analytical Measurement”, 2nd edition, 2000.

European cooperation for Accreditation, EA-3/02, “The Expression of Uncertainty in Quantitative Testing”.

European cooperation for Accreditation, EA-4/02, “Expression of the Uncertainty of Measurement in Calibration”, December 1999. (Replaces EAL-R2).

ILAC-G17:2002 Introducing the Concept of Uncertainty of Measurement in Testing in Association with the Application of the Standard ISO/IEC 17025

ISO, “International Vocabulary of Basic and General Terms in Metrology” (VIM), 2nd ed., 1993.

ISO 3534-1, “Statistics - Vocabulary and symbols – Part 1: Probability and general statistical terms”.

ISO 3534-2, “Statistics – Vocabulary and symbols – Part 2: Statistical quality control”.

ISO 3534-3, “Statistics – Vocabulary and symbols – Part 3: Design of experiments

ISO 5725-1, “Accuracy (trueness and precision) of measurement methods and results – Part 1: General principles and definitions”.

ISO 5725-2, “Accuracy (trueness and precision) of measurement methods and results – Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method”.

ISO 5725-3, “Accuracy (trueness and precision) of measurement methods and results – Part 3: Intermediate measures of the precision of a standard measurement method”.

ISO 5725-4, “Accuracy (trueness and precision) of measurement methods and results – Part 4: Basic methods for the determination of the trueness of a standard measurement method”.

ISO 5725-5, “Accuracy (trueness and precision) of measurement methods and results – Part 5: Alternative methods for the determination of the precision of a standard measurement method”.

ISO 5725-6, “Accuracy (trueness and precision) of measurement methods and results – Part 6 Use in practice of accuracy values”.

ISO/DTS 21748, “Guide to the use of repeatability, reproducibility and trueness estimates in measurement uncertainty estimation”.

ISO/IEC Guide 33:1989, “Uses of certified reference materials”.

ISO/IEC 17025:1999, “General requirements for the competence of calibration and testing laboratories”.

United Kingdom Accreditation Service, “The Expression of Uncertainty in Testing”, UKAS publication ref: LAB 12, 1st ed., September 2000.

United Kingdom Accreditation Service, M3003, “The Expression of Uncertainty and Confidence in Measurement”, 1st ed., December 1997.

Useful Web Sites:

American National Standards Institute (ANSI): www.ansi.org

American Society for Testing and Materials (ASTM): www.astm.org

American Society of Mechanical Engineers (ASME): www.asme.org

Co-Operation on International Traceability in Analytical Chemistry (CITAC): www.citac.ws

Eurachem: www.eurachem.bam.de

European cooperation for Laboratory Accreditation (EA): www.european-accreditation.org

International Laboratory Accreditation Cooperation (ILAC): www.ilac.org/

International Organization for Standardization (ISO): www.iso.ch

International vocabulary of basic and general terms in metrology (VIM): www.cornnet.nl/~mlbroens/vim.htm

National Conference of Standards Laboratories International (NCSLI, formerly known as NCSL, the National Conference of Standards Laboratories): www.ncslinternational.org

National Institute of Standards and Technology (NIST): www.nist.gov

NIST-SEMATECH Engineering Statistics Internet Handbook: www.itl.nist.gov/div898/handbook/index.htm

Uncertainty Analysis: www.itl.nist.gov/div898/handbook/mpc/section5/mpc5.htm

United Kingdom Accreditation Service (UKAS): www.ukas.com