Some Limitations in Uncertainty Evaluation

Speaker/Author: Dr. Henrik S. Nielsen
HN Metrology Consulting, Inc
HN Proficiency Testing, Inc.
10219 Coral Reef Way, Indianapolis, IN 46256
Phone: (317) 849 9577, Fax: (317) 849 9578
E-mail: hsnielsen@HN-Metrology.com

Abstract

Uncertainty budgets are receiving increasing attention not only from accredited laboratories, but from the metrology community at large. Often uncertainties are given with three or more significant digits, indicating that the uncertainty has been determined to within one percent or better, but how accurately can uncertainty be estimated?

The paper explores the capabilities of common statistical techniques as well as some of the tools given in the Guide to the Expression of Uncertainty in Measurement. As it turns out, these fundamental building blocks impose some very significant limitations on the uncertainty analysis.

Introduction

The International Vocabulary of Basic and General Terms in Metrology[1] (VIM) defines *measurement uncertainty* 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". It also defines *measurand* as a “particular quantity subject to measurement”.

As laboratories and metrologists become more and more interested in measurement uncertainty, and explore various ways to estimate the measurement uncertainty, the question comes to mind: “How well can we know the measurement uncertainty?”

Sometimes we see measurement uncertainty indicated with 3 - 4 - 5 or more significant digits, but can we really estimate the measurement uncertainty this well? That is the subject of this paper.

Basic Assumptions

This paper attempts to quantify the limitations of the GUM[2] method of estimating uncertainty. In order to do this, a number of assumptions have to be made.

The first assumption is that the analysis of the measuring process is generally sound, i.e. the measurement equation used is correct and the major uncertainty contributors have all been successfully identified and properly taken into account.

The second assumption is that all data used in type A estimates originate from well-designed experiments and that the significance of the variation seen in the experiments is understood.

The third assumption is that all type B estimates are based on sound engineering judgement made by knowledgeable persons understanding the measurement process in question.

If one or more of these assumptions are not met, any resemblance between the estimated uncertainty and the uncertainty of the measuring process analyzed is purely coincidental.

Uncertainty of type A evaluations

There are two main sources of uncertainty for type A evaluations. The first depends on the design of the experiment for the type A evaluation. For the evaluation to be meaningful, the experiment has to resemble the measuring process it is supposed to represent.

A short term repeatability study performed under controlled conditions in a laboratory environment by measuring parts that have been marked to ensure the participants measure in the same spot every time is not a good representation of a measuring process that takes place on the shop floor, measuring parts in random orientations as they come off a hot production machine.

But even if the experiment is well designed, there are still some limitations to how well a standard deviation can be estimated from a limited sample. Annex E.4 of GUM[2] addresses this issue and gives the following table:

Number of observations, n	$\sigma\left[s(\bar{q})\right] / \sigma(\bar{q})$ (Percent)
2	76
3	52
4	42
5	36
10	24
20	16
30	13
50	10

Table 1: *The standard deviation of the experimental standard deviation of the mean of n independent observations of a normally distributed random variable, relative to the standard deviation of that mean. From GUM*[2]

The table shows the relative uncertainty of the experimental standard deviation of the mean of a normally distributed random variable determined by a limited number of observations as a function of the number of observations.

The same relationship is true for the experimental standard deviation of the normally distributed random variable itself.

The table indicates that a standard deviation determined by 10 observations will be known within +/- 48 % at a 95% confidence level (2 standard deviations).

A standard deviation determined by 30 observations, which is a lot of observations by most peoples standards, will be known only to within +/- 26 % at a 95% confidence level (2 standard deviations).

It is surprising to most people how uncertain an experimental determination of the standard deviation of a variable is, even using relatively large data sets.

An Example

The calibration of the scale on an optical comparator using a glass reticle is an example of a measurement process that is highly dependent on operators skill. The repeatability/ reproducibility of this process are the largest contributors to the measurement uncertainty. Table 2 shows a summary of the uncertainty budget for this measurement.

Description	Evaluation	Distribution	D of F	Variation (Influence)	Variation (µm)	Multiplier	Std. Uncertainty	
							0 mm	300 mm
R & R	A	N/A	29	2.5 µm	2.5 µm	1	2.5 µm	2.5 µm
Thermal Effects	B	Rectangular	>100	7˚F	13.5 µm	0.6	N/A	8.1 µm
Resolution	B	Resolution	>100	2 µm	2 µm	0.3	0.6 µm	0.6 µm
Reticle Spec.	B	Rectangular	>100	1µm	1 µm	0.6	0.6 µm	0.6 µm
Combined Standard Uncertainty							2.6 µm	8.5 µm
Expanded Uncertainty (k=2)							5.2 µm	17 µm
Expanded Uncertainty (k=2): 5.2 µm + 39 µm/m*L, where L is the calibrated length in meters								

Table 2: *Summary of uncertainty budget for the calibration of the scale of an optical comparator 0-300 mm.*

As can be seen from table 2, the repeatability and reproducibility (R & R) is the dominating contributor for short lengths. For longer lengths, thermal effects become the dominating contributor. The repeatability and reproducibility is determined from a study where three operators each make 10 measurements. With a total of 30 measurements the degrees of freedom of this determination is 29. When such an experiment is repeated the results will vary. To illustrate this an Excel spreadsheet was set up to generate 30 sets of 30 random numbers from an underlying normal distribution with a standard deviation of 2.5 µm. A standard deviation was calculated from each of these sets of 30 numbers. The average of the standard deviations was 2.44 and the standard deviation of the standard deviations was 0.38. The highest standard deviation was 3.17 and the lowest standard deviation was 1.80. If these extremes are entered into table 2, the equation for the expanded uncertainty will vary from 4.0 µm + 42 µm/m•L to 6.6 µm + 36 µm/m•L. Usually only one set of 30 readings is generated during the evaluation of the measurement uncertainty and so it depends on the luck of the draw whether the analysis will show an uncertainty of 4.0 μm or 6.6 μm for short lengths. This represents a variation of +/-25% around the "true" uncertainty, which of course cannot be known.

It is customary in some industries to re-do repeatability and reproducibility studies e.g. on an annual basis and restate the uncertainty based on these perceived "changes in the process". As the example illustrates, as long as the changes are not more dramatic than shown above, one is better served by pooling the results of the studies and increasing the number of degrees of freedom, than discarding the original study. The example also shows that is futile to try to determine "the root cause of the change" since no change has taken place in most cases. A prudent first step would be to test the hypothesis that the two sets of data belong to different distributions with a relatively high probability, before the corrective action process is unleashed .

It is important to realize that this is not an unfortunate example of how things sometimes go wrong, but an illustration of how statistics on limited data sets works every time. The only way to reduce this effect is to increase the number of degrees of freedom by including more data in the study, but as can be seen from table 1 this effect is only reduced very slowly as the amount of data is increased.

Uncertainty of type B evaluations

For type B evaluations, there are two sources of uncertainty in the estimate:

1. Uncertainty of the estimated variation limits
2. Uncertainty of the assumed distribution.

Both depend on the amount of information available about the parameter being evaluated.

For type B evaluations it is also important to distinguish between the uncertainty due to lack of knowledge about the measurement process and/or instrumentation and the amount of error in the measured values that result from the measurement process. The errors can be significantly smaller than the uncertainty, for example in the situation where an instrument is calibrated to a relatively large tolerance, but in reality performs much closer to nominal than this tolerance implies. However, if no mechanism is available to capture this knowledge (for example from the data generated during the calibration), the uncertainty is based on the calibration tolerance.

A similar problem exists for the error distribution that can be applied to the measurement equipment. In a typical calibration process measurement equipment is evaluated against a tolerance. If all the measured deviations from nominal fall within this tolerance, the equipment passes calibration and is released for use.

The knowledge generated by this type of calibration can be modeled by rectangular distribution; it is known that all deviations are within the tolerance and none are outside the tolerance, but it is not known where within the tolerance the deviations are located.

The actual performance of the equipment may only consume a fraction of the tolerance and the deviations may follow a distribution that resembles a normal distribution, but if this is not known or tested during the calibration, the uncertainty due to the allowable tolerance is properly described by rectangular distribution with a width equal to the calibration tolerance.

Uncertainty of the estimated variation limits

The uncertainty of the estimated variation limits depends on the amount and type of information available, as well as the know-how of the person interpreting this information. Some variation limits, such as tolerance limits for measurement equipment, are very easy to identify and apply correctly. Others, such as residual temperature differences between parts, standards and measurement equipment that have been soaked together for a period of time, can only be estimated based on experience and these estimates often vary widely. Experienced and reasonable engineers may disagree by an order of magnitude for these types of effects.

Uncertainty of the assumed distribution

It must be recognized that the distributions used for type B evaluations are only approximate models, and these models are applied only for the purpose of finding a conversion factor to convert from the variation limits to the equivalent standard deviation. As most models these are not accurate representations of reality, but rather useful simplifications. Think of the assumption in Newtonian physics that the entire mass of rigid bodies is concentrated in the center of gravity. This assumption is never correct, but is nevertheless very useful.

The conversion factors for the most commonly used distributions vary from $1/\sqrt{6} \approx 0.4$ for the triangular distribution to $1/\sqrt{2} \approx 0.7$ for the U-shaped distribution. This is a difference of less than a factor of 2. If one were very pragmatic one could always use a conversion factor of 0.55 and take comfort in the knowledge that one would never be wrong by more than 27%, which in many cases would be an improvement.

In general, the differences in estimated uncertainty that are due to the differences in the assumed distributions are small compared to the differences from other sources.

The relative importance of errors in uncertainty estimates

Having reviewed and evaluated hundreds of uncertainty budgets from laboratory seeking accreditation, I have developed the following sequence of errors in terms of the magnitude of their influence on the estimated uncertainty:

1. Mathematical Errors
2. Failure to Identify Significant Uncertainty Contributors
3. Errors in the Measurement Model
4. Misinterpretation of Type A Studies
5. Misunderstanding of Statistics
6. Inappropriate Variation Limits for Type B Evaluations
7. Inappropriate Distributions for Type B Evaluations

Note that in the section “Basic Assumptions” the case is made that if any of the first 4 categories of errors are present in an uncertainty analysis, the analysis is generally of very limited, if any, value.

Mathematical errors

Mathematical errors include e.g. wrong conversions between microinches and inches. These errors can grow to many orders of magnitude. The root cause of this type of error has nothing to do with uncertainty estimation, but rather a lack of basic math and/or physics skills.

Failure to identify significant uncertainty contributors

This type of error can also grow to many orders of magnitude. The root cause of this type of error is usually a lack of understanding of the physics of the measurement in question.

Errors in the measurement model

These errors include items such as wrong models for thermal expansion effects. These errors can also grow to many orders of magnitude. The root cause of this type of error is usually a lack of understanding of the physics of the measurement in question.

Misinterpretation of type A studies

This type of error includes e.g. running a short term repeatability study and assuming it accounts for all the variation that will ever be seen in the measurement. This type of error can grow to several orders of magnitude. The root cause of this type of error is usually a lack of understanding of the purpose of the type A study or the physics of the measurement in question.

Misunderstanding of statistics

This type of error includes e.g. inappropriate use of the standard deviation of the mean rather than the standard deviation of the population. This type of error is usually not larger than an order of magnitude. The root cause is generally a lack of statistics skills.

Inappropriate variation limits for type B estimates

This type of error includes e.g. assuming that residual temperature effects are negligible. This type of error is usually not larger than an order of magnitude. The root cause is generally a lack of experience in quantifying variations and errors.

Inappropriate distributions for type B estimates

As was shown above this is typically the least significant error in an uncertainty analysis. There is a variety of root causes for this type of error.

Conclusion

This paper has discussed some of the limitations in uncertainty estimates. It has shown that the major causes of errors in uncertainty estimates are not due to lack of understanding of the uncertainty estimation process, but rather due to lack of basic skills in math and physics.

The paper has shown by example that type A estimates are much more “uncertain” than is generally recognized - purely due to statistical limitations, unless extraordinarily large data sets are compiled.

The paper has also shown that the choice of distribution for type B estimates is the least significant of all the errors that are typically committed in uncertainty estimates.

References

1. International Vocabulary of Basic and General Terms in Metrology.BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, OIML., 1993

2. Guide to the Expression of Uncertainty In Measurement. BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, OIML., 1993