XIX IMEKO World Congress
Fundamental and Applied Metrology
September 6–11, 2009, Lisbon, Portugal

CORRELATION IN UNCERTAINTY OF MEASUREMENT - A DISCUSSION OF STATE OF THE ART TECHNIQUES

Rüdiger Kessel[1], *Raghu Kacker*[1]

[1] National Institute of Standards and Technology, Gaithersburg, USA, ruediger.kessel@nist.gov

Abstract – The Guide to the expression of uncertainty has been around for 15 years and has been widely adopted by science and industry. Over time more and more complex measurements are evaluated based on these principles. As a consequence the correlation between quantities has become an important issue in the evaluation of measurement uncertainty. In this paper we will give an overview about covariance and correlation and the different state of the art techniques to handle them during the uncertainty evaluation. We will discuss the handling of observations by extending the well known calculus for the degrees of freedom for correlated cases. Next we will discuss the difficulties and limitations in handling correlations with Monte Carlo simulations together with a practical algorithm to ensure that correlation matrixes are positive semi-definite.

Keywords: measurement uncertainty, correlation

1. STANDARD GUM FRAMEWORK WITH CORRELATION

One of the biggest advantages of the GUM method is that the estimation of the measurement uncertainty is derived from the model for evaluating the result. All available knowledge about a specific measurement can be incorporated into the measurement equation. With the freedom to use an appropriate model, the GUM method can be tailored to nearly every measurement

$$Y_k = f_k(X_1,\ldots,X_n)\,. \tag{1}$$

The mainstream GUM approach is to translate the knowledge about the quantities in the model equation into a standard uncertainty and to use the law of propagation of uncertainties to propagate these standard uncertainties through a linearized model to arrive at the standard uncertainty of the result. Mutual dependencies in the knowledge about the input quantities can be expressed as a covariance or a correlation coefficient and can be used during the propagation. We recommend the usage of correlation coefficients because the normalized value of the correlation coefficients

$$r(x_i,x_j) = \frac{u(x_i,x_j)}{u(x_i)\cdot u(x_j)} \tag{2}$$

expresses the dependency between the uncertainties independent of the value of the uncertainty and can be directly used in the propagation of uncertainties:

$$\begin{aligned} u^2(y_k) &= \sum_{i=1}^{n}\left(\frac{\partial Y_k}{\partial X_i}\right)^2 u^2(x_i) \\ &+2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\frac{\partial Y_k}{\partial X_j}\frac{\partial Y_k}{\partial X_i}r(x_i,x_j)\,u(x_i)\,u(x_j) \end{aligned}\,. \tag{3}$$

Equation 1 relates result *y* with the estimates of the input quantities x_i. As a consequence the result and the input quantities get correlated and the correlation can be calculated by

$$r(y,x_i) = \sum_{j=1}^{n}\frac{\partial Y}{\partial X_j}\frac{u(x_j)}{u(y)}r(x_i,x_j)\,. \tag{4}$$

In a multiple result evaluation the evaluation model consists of a system of *m* measurement equations

$$\begin{aligned} Y_1 &= f_1(X_1,\ldots,X_n) \\ &\vdots \\ Y_m &= f_m(X_1,\ldots,X_n) \end{aligned}\,. \tag{5}$$

Since the measurement equations in Equation 5 relates input quantities to result quantities the result quantities become correlated as far as they are calculated from common input quantities. For a linearized model this correlation can be calculated based on the equation given in the GUM (H.9)

$$r(y_k,y_l) = \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial Y_k}{\partial X_i}\frac{\partial Y_l}{\partial X_j}\frac{u(x_i)u(x_j)}{u(y_k)u(y_l)}r(x_i,x_j)\,. \tag{6}$$

We can derive Equation 4 from Equation 6 as explained in [3]. For a single result the expanded uncertainty associated with the result is calculated by multiplying the standard uncertainty with a coverage factor *k*.

2. CORRELATION IN THE OBSERVED DATA

If during a measurement more than one quantity is observed and the observed values are partly mutually dependent then the correlation coefficient between any pair of quantities can be calculated by the equation given in the GUM

$$r(q_i, q_j) = \frac{1}{(n-1)\, s(q_i)\, s(q_j)} \sum_{l=1}^{n} (q_{i,l} - q_i)(q_{j,l} - q_j) \quad (7)$$

with $q_{i,l}$ and $q_{j,l}$ being values of a series ($l = 1...n$) of observations, q_i and q_j being the average value and $s(q_i)$ and $s(q_j)$ being the experimental standard deviation of the series. The degrees of freedom ν_i and ν_j are the number of observations n minus one.

One way to set the coverage factor k is by using a t-table and looking up the value based on the effective degrees of freedom. The GUM uses the Welch-Satterthwaite formula to calculate the degrees of freedom. This formula cannot be used if any input quantities with non infinite degrees of freedom are correlated. We want to propose an extended version of the Welch-Satterthwaite formula [4] which handles correlated input quantities correctly:

$$\frac{u^2(y)}{\nu_y} = \sum_{i=1}^{n} \frac{u_i^2(x_i) \cdot r^2(y, x_i)}{\nu_i} + 2 \sum_{i=1}^{n-1} \left[\frac{u_i(x_i) \cdot r(y, x_i)}{\sqrt{\nu_i}} \sum_{j=i+1}^{n} r^2(x_i, x_j) \frac{u_j(x_j) \cdot r(y, x_j)}{\sqrt{\nu_j}} \right] \quad (8)$$

In case all correlation coefficients $r(x_i,x_j)$, $i \neq j$ are zero, Equation 8 simplifies to the Welch-Satterthwaite formula.

Note that two quantities of type A which are correlated need to have the same degrees of freedom. Equation 8 has some other interesting features. If two quantities X_1 and X_2 are totally correlated ($r(x_1, x_2) = \pm 1$) then the effective degrees of freedom based on Equation 8 is equal to the degrees of freedom of X_1 or X_2 which need to be the same. This is consistent with the understanding that totally correlated quantities essential represent the same knowledge.

3. MONTE CARLO SIMULATION WITH CORRELATED INPUT QUANTITIES

The supplement 1 to the GUM [2] describes an alternative way to do the calculations to propagate the uncertainty for a given measurement model. The method can handle correlation as well as long as all quantities which are correlated are distributed normally or are totally correlated. In practice this can be an important limitation in case the distribution of the correlated quantities differs significantly from normal. This is for example the case if a t-distribution is used to bootstrap the distribution for quantities of type A. Therefore the supplement 1 uses a Bayesian evaluation of type A which is based on a normal distribution.

Another well known restriction is that the correlation matrix must be positive semi-definite (all eigenvalues larger or equal to zero) otherwise the algorithm will fail to simulate a multivariate normal distribution.

The mainstream GUM method does not have this numerical limitation. The calculation is possible as long as the uncertainty calculated from Equation 3 is greater or equal to zero although it might not make much physical sense to do so.

In practice a positive definite correlation matrix can become negative definite because of rounding errors if the correlation coefficients are rounded to 2 or 3 digits. The probability for this effect is dependent on the size of the matrix. The effect has been studied via simulations [5]. The results are shown in Figure 1. The probability that a 10×10 matrix is not positive semi-definite after rounding of the values to two significant digits is close to 1.

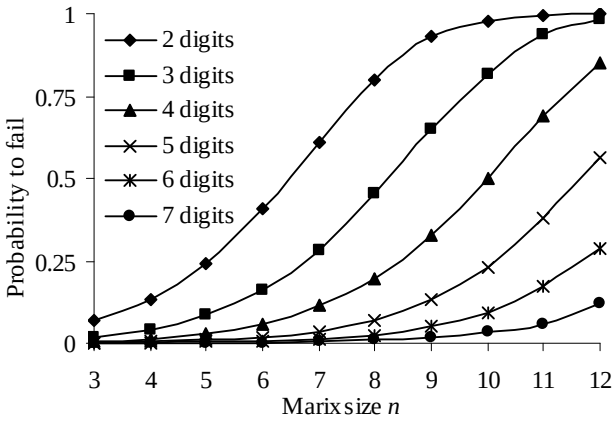

Fig. 1. Probability that a correlation matrix is not positive semi-definite after rounding based on 10000 simulations per point.

Therefore it is useful to implement a matrix-correction algorithm which ensures that all eigenvalues are positive prior to any calculation as suggested in [2]. We have studied different matrix correction methods. We think that an algorithm which shifts the negative eigenvalues (spectral decomposition) [6] is useful. However such algorithm should be combined with a check of the least maximum norm [7] of the change to ensure that the modified matrix is close enough to the original matrix. The least maximum norm can be calculated by

$$N_{\mathrm{LMN}} = \max_{i,j} \left| r_{ij} - p_{ij} \right| \quad (9)$$

with r_{ij} being the correlation coefficients of the original matrix and p_{ij} being the elements of the corrected matrix.

Simulations with 10^8 random correlations matrixes show that the corrected matrices produced by the algorithm do not differ more than ± 1 on the last significant digit of the rounded correlation matrix. This limit should be used for the least maximum norm to ensure that the correlation matrix is only negative definite because of rounding and not because of any other reasons such as incorrect data.

The matrix correction algorithm together with a limit on the least maximum norm can be incorporated into the mainstream GUM method to numerically validate a given correlation matrix. A matrix is valid if it is either positive semi-definite or if it can be corrected and the norm does not exceed ± 1 on last significant digit.

4. CONCLUSIONS

The GUM and the supplement provide a solid basis for the calculation of uncertainty in measurement including correlations and multiple results. The standard GUM

procedure can be extended by the calculation of the degrees of freedom in correlated cases and a robust matrix validation.

The Monte Carlo simulation can benefit from a robust matrix correction method. In general the number of significant digits of the values in a correlation matrix should be at least 3. Especially for correlation matrixes of larger size (greater 4×4) the number of significant digits should be increased further to avoid rounding effects.

REFERENCES

[1] BIPM *et. al.*, *Guide to the Expression of Uncertainty in Measurement*, ISBN 92-67-1011889, 1st Edition (1993)

[2] BIPM *et. al.*, *Evaluation of measurement data – Supplement 1 to the "Guide to the expression of uncertainty in measurement" – Propagation of distributions using a Monte Carlo method*, JCGM, 1st Edition (2008)

[3] Rüdiger Kessel, Michael Berglund, Philip Taylor and Roger Wellum, "How to treat correlation in the uncertainty budget when combining results from different measurements", Published on: P.Ciarlini, M.G.Cox, E.Filipe, F.PAVESE and D.Richter (eds.), *Advanced Mathematical Tools in Metrology, vol.5*, Page 232, Series on Advances in Mathematics for Applied Sciences vol. 57, World Scientific, Singapore, 2001

[4] Rüdiger Kessel and Raghu Kacker, "Effective Degrees of Freedom for Cases with Correlated Input Quantities", unpublished

[5] Rüdiger Kessel and Raghu Kacker, "Using rounded correlation matrices in Monte Carlo Simulations", unpublished

[6] Gilbert Strang and Kai Borre, *Linear algebra, geodesy, and GPS*, Wellesley-Cambridge Press, 1997, ISBN 0-9614088-8-X

[7] Sudhanshu Mishra: "Optimal solution of the nearest correlation matrix problem by minimization of the maximum norm", 2004, `http://mpra.ub.uni-muenchen.de/1783/`