

# UNCERT 2003

International Conference on the Uncertainty of Measurement

St Catherine's  
College,  
Oxford, UK

9-10 April 2003

## UNCERT Proceedings

Technical  
Conference



**NPL**   
National Physical Laboratory

**Order Statistics and Uncertainty  
Evaluation**

**F Pennecchi  
CNR-IMGC**

# Order statistics and uncertainty evaluation

Francesca Pennechi<sup>1</sup>, Maurice Cox<sup>2</sup>, Walter Bich<sup>1</sup>

<sup>1</sup>Istituto di Metrologia "G Colonnetti", Torino, Italy, <sup>2</sup>National Physical Laboratory, Teddington, UK

## 1 Abstract

Order statistics are considered as a basis for key comparison (KC) data evaluation, in particular for evaluating KC reference values (KCRVs) and degrees of equivalence (DoEs), including the associated uncertainties. The results obtained for actual KC data are compared with those given by the application of recent guidelines.

## 2 Introduction

Implementations of the CIPM *Mutual Recognition Arrangement* (MRA) [1] are required for KC data evaluation, especially for the determination of the KCRV and the DoEs of the participating national measurement institutes (NMIs). Various approaches have been suggested for KC data evaluation, but consensus has not been reached on a specific technique. KC data evaluation is a significant task in metrology, because of its relevance to global trade and because such comparisons are intended to test the principal measurement techniques in the field.

This paper reports on a study of order statistics (Section 3.1) as a basis for KC data evaluation. When sorted, a set of mutually independent KC measurements can be regarded as a sample of realisations of order statistics, taken from a parent distribution suitably chosen to model the sample itself. Both the KCRV, taken as the median, and the consequent DoEs (Section 3.2) can be expressed as functions of order statistics.

A family of parent distributions is chosen to mimic actual KC data (Section 3.3) and the estimation of the parameters of these distributions (Section 3.4) considered.

The results obtained for some mass KC data are compared with those obtained by an application of guidelines provided by *BIPM Director's Advisory Group on Uncertainties* [2] (Section 4).

## 3 Theory

### 3.1 Order statistics

Consider a random sample  $X_1, \dots, X_N$ , drawn from a population with (continuous) cumulative distribution function (cdf)  $F_X$ . The arrangement of the sample  $X_{(1)} \leq \dots \leq X_{(N)}$  constitutes the *order statistics* of the original sample [3].  $X_{(r)}$  is called the *r*th *order statistic*. Although the  $X_i$  are independent, the corresponding order statistics  $X_{(r)}$  are necessarily mutually dependent as a consequence of the ordering.

Using the method of Jacobians for transformations [3, p. 14] or simple probabilistic arguments, formulae can be obtained for probability density functions (pdfs), cdfs, moments and joint moments for order statistics and also for functions of order statistics. Such formulae consist of integrals involving the sample size, the rank of the order statistic(s) and the cdf for the parent population.

For the uniform as the parent distribution, the formulae can easily be evaluated; for the standard normal, the results have been tabulated [4]. In general, methods of numerical approximation are necessary.

Common applications of order statistics are in the estimation of location and scale parameters, in *distribution-free* and *nonparametric* inference, and in the study of censored samples and extreme observations (outliers).

### 3.2 Model for key comparison data

Let  $x_i$  be the measurement provided by the *i*th NMI ( $i = 1, \dots, N$ ) and  $x_{ref}$  the KCRV. The DoE for the measurement  $x_i$  is formed by the deviation  $d_i = x_i - x_{ref}$  and the associated uncertainty [1].

Widely used estimators for  $x_{ref}$  are some form of average of  $x$ , the set of values  $x_i$ ,  $i = 1, \dots, N$  (e.g., a weighted mean or the median). While the methods of calculating such KCRVs are well known, in some cases the manner in which the associated uncertainties and the DoEs are evaluated is less well understood. This is so for the median.

In general,  $x_{\text{ref}}$  is correlated with the  $x_i$ , because it is determined from those values. This correlation must be taken into account in forming the DoE. Let  $u(\alpha)$  denote the standard uncertainty associated with an estimate  $\alpha$  and  $r(\alpha, \beta)$  the correlation coefficient associated with estimates  $\alpha$  and  $\beta$ . Then [5, formula (13)],

$$u^2(d_i) = u^2(x_i) + u^2(x_{\text{ref}}) - 2u(x_i)u(x_{\text{ref}})r(x_i, x_{\text{ref}}). \quad (1)$$

$u(x_i)$  would be provided by the  $i$ th NMI.  $u(x_{\text{ref}})$  and  $r(x_i, x_{\text{ref}})$  would depend on the choice of  $x_{\text{ref}}$ .

Consider the use of the median of the  $x_i$  as  $x_{\text{ref}}$ . In determining the median, the set of measurements is ordered and can then be regarded as a realisation of a sample of order statistics. As the sample median is a realisation of the central order statistic or of the semi-sum of the two central order statistics, according to whether  $N$  is odd or even, the whole theory of order statistics can be applied when the median is chosen as  $x_{\text{ref}}$ .

Applying formula (1),  $u(x_{\text{ref}})$  can be approximated by using a scale estimator such as the MAD [6] (the robust counterpart of the median), and the correlation coefficient  $r(x_i, x_{\text{ref}})$  determined by evaluating the corresponding formula for the correlation between order statistics. The latter is a main consideration of this paper.

### 3.3 Contaminated normal distribution

To determine such correlations requires knowledge of the parent distribution from which the data is perceived to be drawn. In many contexts, the underlying distribution is taken as Gaussian, but, for a KC, practical experience often indicates discrepancies from this distribution, especially in the presence of outliers. A contaminated normal distribution [7] is considered here in the belief that it may be a more appropriate distribution to model KC data in the case that the data seems normally distributed except for a small number of extreme values.

This distribution consists of the weighted mean of two Gaussian distributions: a standard one  $F(x)$  and a longer-tailed one  $F(x/k)$ , with variance  $k^2$  ( $k > 1$ ). The resulting cdf is  $H(x) = (1 - \varepsilon)F(x) + \varepsilon F(x/k)$ , where the parameter  $\varepsilon$  is the probability of obtaining an element from the longer-tailed distribution.

The contaminated normal distribution has mean zero and variance  $\text{Var}_H = 1 - \varepsilon + \varepsilon k^2$ . Its pdf has heavier tails than a Gaussian pdf. Other features, such as the *Kolmogorov distance*, can be used to study the departure of  $H$  from a Gaussian cdf [8]. Such features depend only on  $\varepsilon$  and  $k$ .

### 3.4 Choice of the parameters of a contaminated normal distribution

In the case of actual KC data, if two normal distributions are chosen, one with mean  $\mu$  and standard deviation  $\sigma$  to model the "good" data and the other with the same mean but standard deviation  $k\sigma$  to model the outliers, the sum of the two distributions produces an appropriate contaminated normal distribution. To work with this distribution it is necessary to estimate its parameters  $\varepsilon$  (indicating, within this model, the proportion of outliers in the sample) and  $k$ .

In order to estimate  $\varepsilon$ , various methods for detecting outliers within a sample can be used, such as those that label suspected outliers (e.g., the modified z-score method and the boxplot method [9]) and those that identify outliers using hypothesis testing (e.g., the Rosner, Dixon, Grubbs and discordance tests [10]).

The values identified as outliers decide which data is modelled by the normal distribution with standard deviation  $\sigma$ , and which by the longer-tailed distribution. A value for  $k$  is then obtained as the quotient of the estimates of the standard deviations of the two normal distributions. It is recommended that each estimate, as well as the estimate of  $\mu$ , be obtained robustly because of the presence of outliers within the sample.

## 4 Computational results

For a CCM KC of 1 kg mass standards, carried out by  $N = 15$  NMIs, the sample median  $m$  was taken as the KCRV, and the sample MAD used to evaluate the associated uncertainty (both estimators have valuable robustness properties) [11]. The DoEs, however, were incorrectly determined: no account was taken of the correlation between each NMI's measurement and the KCRV.

The theory of Section 3 was applied to this KC in order to account for these correlations. Table 1 shows the ordered  $x_i$  and their standard uncertainties. Outliers were identified by those z-scores  $0.6745|x_i - m|/\text{MAD}(x)$  that exceeded 3.5. Since three values, 1.764, 1.8469 and 1.8581, were so identified,  $\varepsilon$  was taken as  $3/15 = 0.2$ .

The 12 "good" data values were modelled by a normal distribution with mean  $\mu$  and standard deviation  $\sigma$ . The three outliers were modelled by a normal distribution with the same mean  $\mu$  and a larger standard deviation  $k\sigma$ .  $\mu$  was estimated by  $m$ .  $\sigma$  and  $k\sigma$  were estimated by the MADs of the two sub-samples, but calculated with respect to  $m$  rather than the medians of the two sub-samples themselves. The quotient of these two MAD estimators gave 18.81 as an estimate of  $k$ .

For these values of  $\varepsilon$  and  $k$ , numerical approximations to all the principal moments of the distributions for the 15 order statistics drawn from a contaminated normal distribution having these parameters were formed, viz., expectation values  $E_i$  and variances  $\text{Var}_i$ , and also covariances  $\text{Cov}_{i,8}$  and correlation coefficients  $\text{Corr}_{i,8}$  between each order statistic and the central one, the median. The results are given in Table 2. Due to the symmetric behaviour of the order statistics with respect to the median, only the relationships between the first seven order statistics and the eighth (the median) are shown.

For the purpose of this work, the most useful results are the seven correlation coefficients used in estimating the variances of the deviations between each laboratory value and the median using formula (1). The standard uncertainties, formed as the square roots of these variances, are shown in Table 3 and compared with those provided by the guidelines of the *BIPM Director's Advisory Group on Uncertainties* [2, Procedure B].

| $i$ | $(x_i - 1 \text{ kg})/\text{mg}$ | $u(x_i)/\text{mg}$ |
|-----|----------------------------------|--------------------|
| 1   | 1.7640                           | 0.0165             |
| 2   | 1.7775                           | 0.014              |
| 3   | 1.7800                           | 0.0195             |
| 4   | 1.7825                           | 0.0185             |
| 5   | 1.7961                           | 0.023              |
| 6   | 1.7963                           | 0.014              |
| 7   | 1.7966                           | 0.0125             |
| 8   | 1.7980                           | 0.0135             |
| 9   | 1.7999                           | 0.0165             |
| 10  | 1.8000                           | 0.0175             |
| 11  | 1.8012                           | 0.012              |
| 12  | 1.8025                           | 0.015              |
| 13  | 1.8041                           | 0.0125             |
| 14  | 1.8469                           | 0.024              |
| 15  | 1.8581                           | 0.022              |

Table 1. NMI measurements and standard uncertainties.

## 5 Conclusions

The use of the median as a KCRV and hence to determine DoEs has been presented. The approach is intrinsically different from the use of the median given in recent recommendations [2]. The approach presented here regards the NMIs' measurements as realisations of order statistics drawn from a contaminated normal cdf. The other approach assigns a cdf (here a Gaussian) to each measurand realised by an NMI's measurement, and then uses Monte Carlo simulation to approximate the distributions for the median and the DoEs. The standard uncertainties associated with the deviations (NMI value - KCRV) for the two approaches, however, are quite similar for the mass KC studied (also see Figure 1).

| $i$ | $E_i$   | $\text{Var}_i$ | $\text{Cov}_{i,8}$ | $\text{Corr}_{i,8}$ |
|-----|---------|----------------|--------------------|---------------------|
| 1   | -15.949 | 168.048        | 1.070              | 0.203               |
| 2   | -5.934  | 53.563         | 0.786              | 0.263               |
| 3   | -2.428  | 13.479         | 0.427              | 0.285               |
| 4   | -1.211  | 2.807          | 0.227              | 0.332               |
| 5   | -0.724  | 0.620          | 0.155              | 0.484               |
| 6   | -0.440  | 0.241          | 0.143              | 0.717               |
| 7   | -0.212  | 0.176          | 0.149              | 0.872               |
| 8   | 0       | 0.166          | 0.166              | 1                   |

Table 2. Expectation values, variances, covariances and correlations for the order statistics.

Both procedures require intensive computation, while being very different in character. The presented method requires the numerical evaluation of single and double integrals [8] and Procedure B a large number of Monte Carlo trials [2]. The results for the first method were obtained using either the adaptive quadrature routine provided by Matlab or the trapezoidal rule with a small interval size, and in some cases both approaches were used for validation purposes.

Although the method uses recognized concepts of order statistics, its reliability depends on the adequacy of a contaminated normal distribution to model KC data, and on the estimation of its parameters. On the other hand, it was observed that the method does not seem very sensitive to small changes in these parameters.

The theory of order statistics offers an approach to be considered alongside other candidates for KC data evaluation. It also has potential in other metrology disciplines where estimates of location and scale are required. Further developments will concern the determination of the DoEs at a 95% level of confidence, as required by the MRA [1].

Sarantis Kamvissis (NPL) made valuable review comments on a draft of this paper.

| $u(x_i - x_{ref})/\text{mg}$ |                  |             |
|------------------------------|------------------|-------------|
| $i$                          | Order statistics | Procedure B |
| 1                            | 0.016            | 0.017       |
| 2                            | 0.014            | 0.015       |
| 3                            | 0.019            | 0.019       |
| 4                            | 0.018            | 0.018       |
| 5                            | 0.022            | 0.022       |
| 6                            | 0.013            | 0.014       |
| 7                            | 0.011            | 0.012       |
| 8                            | 0.011            | 0.013       |
| 9                            | 0.015            | 0.016       |
| 10                           | 0.016            | 0.017       |
| 11                           | 0.011            | 0.012       |
| 12                           | 0.014            | 0.015       |
| 13                           | 0.012            | 0.013       |
| 14                           | 0.024            | 0.024       |
| 15                           | 0.022            | 0.023       |

Table 3. Standard uncertainties associated with the deviations (NMI value - KCRV), obtained using order statistics (A) and recent guidelines (B).

## 6 References

1. CIPM. Mutual recognition of national measurement standards and of calibration and measurement certificates issued by national metrology institutes, Bureau International des Poids et Mesures, Sèvres, 1999.
2. Cox M. G., *Metrologia*, 39, p. 589-595, 2002.
3. Gibbons J. D. and Chakraborti S., *Nonparametric statistical inference*, Marcel Dekker, New York, 1992.
4. Institute of Mathematical Statistics. *Selected Tables in Mathematical Statistics*, American Mathematical Society, Providence, 1977.
5. BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, and OIML. *Guide to the Expression of Uncertainty in Measurement*, 1995. ISBN 92-67-10188-9, Second Edition.
6. Müller J. W., *Rapport BIPM-95/2*, Possible advantages of a robust evaluation of comparisons, BIPM, Sèvres, 1995.
7. Wilcox R.R., *Introduction to robust estimation and hypothesis testing*, Academic Press, San Diego, 1997.
8. Pennechi F., *Rapporto Tecnico IMGC No. 85*, Statistiche ordinate estratte da una distribuzione contaminata, Torino, 2002.
9. Iglewicz B. and Hoaglin D.C., *How to detect and handle outliers*, American Society for Quality Control, Milwaukee, 1993.
10. David H.A., *Order statistics*, John Wiley & Sons, New York, 1981.
11. BIPM Key Comparison Database ([www.bipm.org/kcdb](http://www.bipm.org/kcdb)) - CCM.M-K1.

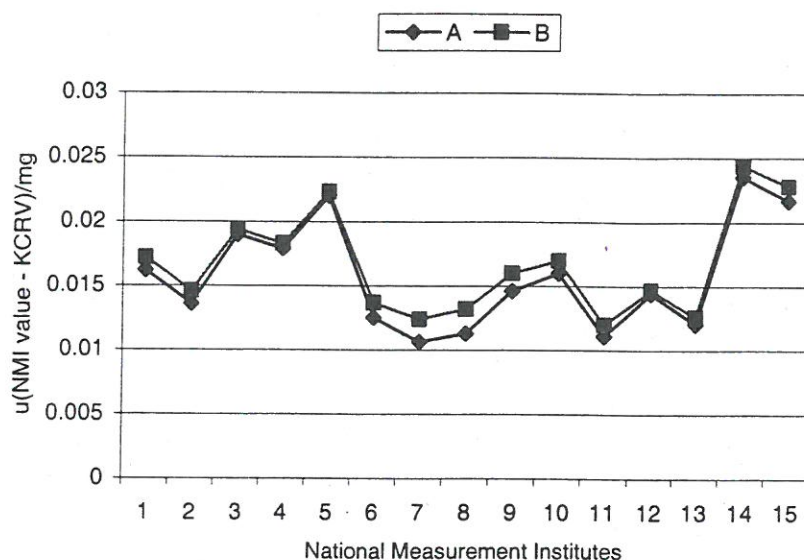


Figure 1. Pictorial representation of the data in Table 3. The points are connected by straight lines for visualization purposes only.