THE EVALUATION OF THE UNCERTAINTY ASSOCIATED WITH THE CONSTRUCTION AND USE OF A CALIBRATION FUNCTION

Ignacio Lira

Pontificia Universidad Católica de Chile, Santiago, Chile, ilira@ing.puc.cl

Abstract: A Monte Carlo procedure is presented for computing the joint state-of-knowledge probability distribution to be assigned to the coefficients of a calibration function. The procedure is fully in line with the approach in Supplement 1 to the Guide to the Expression of Uncertainty in Measurement. It consists of propagating the joint probability distribution of the calibration quantities through the mathematical model of the measurement by which the coefficients are defined. Usually this model is derived from a least-squares adjustment procedure, but other fitting criteria are also possible. The probability distribution for a given stimulus can also be obtained by Monte Carlo from the distributions for the coefficients of the curve and for the response that corresponds to the stimulus. An example illustrates the application of the procedure.

Key words: Calibration functions, Monte Carlo method, curve fitting, probability distributions.

1. INTRODUCTION

The evaluation of the uncertainty associated with the construction and use of a calibration function is of interest in most areas of metrology. Theoretical considerations about the calibrated instrument, or the way the calibration data look in a scatter plot, determine the form of the calibration function. Its parameters are then obtained by some fitting procedure. When the instruments is in use, a particular value of the response is measured. By inverting the calibration function it is possible to find the value of the stimulus that produced the measured response.

However, the measured value is just an estimate of the response. An appropriate statistical model for the generation of this value should be used for obtaining a state-of-knowledge distribution associated with the actual response, which is not perfectly known.

Furthermore, the coefficients of the calibration function are also not perfectly known. A joint state-of-knowledge distribution should be associated with them. This is done by using the uncertainties associated with the calibration data.

It follows that the value of the stimulus determined by inverting the calibration function is uncertain. In this paper, a Monte Carlo procedure for establishing the probability distribution for the stimulus is proposed, through which its uncertainty may be derived.

2. NOTATION

The notation in the sequel is as follows. Upper case Roman letters denote quantities to which state-of-knowledge probability distributions are assigned. Any such distribution is denoted with the letter f followed by a subscript indicating the quantity or quantities to which the distribution refers. Lower case Greek letters denote the argument or arguments of the distribution, which represent the possible values of the corresponding quantities. Lower case Roman letters denote particular values of quantities. Bold face letters represent vectors. The letters p and u are reserved for sampling distributions and standard uncertainties, respectively.

Let $\boldsymbol{X}=(X_1,\ldots,X_n)^{\mathrm{T}}$ be the vector of quantities representing the various stimula during calibration. The corresponding responses are $\boldsymbol{Y}=(Y_1,\ldots,Y_n)^{\mathrm{T}}$. These two vectors will be referred to as the calibration quantities. The calibration curve is of the form

$$Y = F(X, \boldsymbol{A}), \tag{1}$$

where X represents a generic stimulus, Y represents the corresponding response and the vector $\boldsymbol{A}$ consists of the m coefficients of the function. For the ranges of X and Y for which this function is assumed to be valid, it has to be one-to-one in order for it to be useful in practice.

During calibration the values $\boldsymbol{x}=(x_1,\ldots,x_n)^{\mathrm{T}}$ are submitted to the instrument and the values $\boldsymbol{y}=(y_1,\ldots,y_n)^{\mathrm{T}}$ are recorded. The function (1) may be chosen from the way the data pairs (x_i,y_i) appear in a scatter plot or from the assumed behavior of the instrument. In the former case a linear polynomial in the coefficients $\boldsymbol{A}$ is usually chosen, but this is not necessary; in fact, the function (1) may be of any form.

When the instrument is in use, a response Y_0 is measured and the corresponding the stimulus X_0 is determined by using the inverse of the calibration function. The goal of the analysis is to determine the probability distribution associated with X_0, whose mean is taken as the best estimate of this quantity and whose standard deviation is taken as the standard uncertainty associated with the estimate.

3. ANALYSIS

The first stage of the analysis is to obtain the joint distribution $f_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{\xi},\boldsymbol{\eta})$ that encodes the degree of belief about the calibration quantities $\boldsymbol{X}$ and $\boldsymbol{Y}$. Next, the distribution

$f_{\boldsymbol{A}}(\boldsymbol{\alpha})$ for the coefficients $\boldsymbol{A}$ needs to be determined. Finally, the distribution $f_{X0}(\xi_0)$ should be obtained. Some details of this sequence are described next.

3.1. First stage: finding the distribution $f_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{\xi},\boldsymbol{\eta})$

Obtaining the distribution $f_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{\xi},\boldsymbol{\eta})$ may be an easy task or a rather complicated undertaking. It cannot be formulated in general, because it depends strongly on how the quantities $\boldsymbol{X}$ and $\boldsymbol{Y}$ are defined, on the way their estimates are obtained and on whether they are independent or not. The situation where all these quantities depend on one another may become quite difficult to analyze, but fortunately it arises infrequently. Most often the information of one vector can be gathered independently from that of the other. In that case, $f_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{\xi},\boldsymbol{\eta})$ factorizes into the product $f_{\boldsymbol{X}}(\boldsymbol{\xi})\, f_{\boldsymbol{Y}}(\boldsymbol{\eta})$.

Independence *within* these vectors cannot in general be taken for granted. For simplicity, however, it is often assumed that the effects of the possible correlations are negligible. This simplification allows taking any one quantity as independent from the rest, so $f_{\boldsymbol{X}}(\boldsymbol{\xi})$ and $f_{\boldsymbol{Y}}(\boldsymbol{\eta})$ become equal to the product of the distributions for the individual calibration quantities

If these quantities are all independent and there are no data of X_i, say, its distribution must be either given as a result of a previous analysis, or sufficient information about this quantity must be available so that its distribution can be constructed accordingly, e.g. using the principle of maximum entropy. For example, it might only be known that the true value of X_i is within a given interval, in which case the distribution of this quantity is taken as uniform over that interval.

If there are data, they are usually assumed to have been drawn from sampling distributions that depend in general on various parameters. A common situation is that where measured values $\boldsymbol{x}_i=(x_{i1},\ldots,x_{ik})^{\mathrm{T}}$ pertaining to the quantity X_i are available, each of which is assumed to be independently generated by some random process. For example, if X_i is the rate at which some event occurs and the data are a series of counts of occurrence of the event during non-overlapping time intervals, the sampling distribution, derived from a Poisson process, is

$$p\left(x_i\,|\xi_i\right)=\left(\prod\nolimits_j x_{ij}!\right)^{-1}\left(\xi_i t\right)^{x_i}\exp\left(-\xi_i t\right), \qquad (2)$$

where $\xi_i>0$, $x_i=\sum_j x_{ij}$ and t is the total counting time.

"Prior" functions for all imperfectly known parameters in the sampling distributions are required. The priors may be state-of-knowledge probability distributions that are given or that can be constructed from existing information. However, situations arise when *a priori* there is no information whatsoever about some or all of the parameters. In such situations non-informative priors must be used. Choosing such priors is a controversial subject; at present it is an accepted practice to derive them by formal rules based on the structure of the sampling distributions [1]. For example, if the sampling distribution is given by equation (2) and there is no prior information about X_i, one would use the prior $f_{X_i}\left(\xi_i\right)\propto\xi_i^{-1/2}$ [2]. As this example illustrates, very often non-informative priors are improper functions, i.e., they cannot be normalized.

All priors, proper or improper, are updated through their multiplication with the corresponding sampling distributions followed by normalization. This operation, based on Bayes' theorem, results in the "posterior" joint distributions of the unknown parameters. For example, multiplication of the sampling distribution (2) with the prior $\xi_i^{-1/2}$ yields for X_i a gamma posterior distribution with shape parameter $x_i+0.5$ and rate parameter t, whose mean is $(x_i+0.5)/t$. These results are widely available in the literature, see e.g. [3].

3.2. Second stage: finding the distribution $f_{\boldsymbol{A}}(\boldsymbol{\alpha})$

The problem of finding the distribution $f_{\boldsymbol{A}}(\boldsymbol{\alpha})$ in the second stage does not depend on how the distribution $f_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{\xi},\boldsymbol{\eta})$ was obtained, so it can be described in very general terms. This problem is conditioned by the criterion that is used for fitting the function to the calibration data. The most common criterion is that of least-squares, which consists of solving the generally implicit system

$$H(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{A})=0\,, \qquad (3)$$

that results from the minimization

$$\sum_{i=1}^{n} w_i R_i^2\rightarrow\min\,, \qquad (4)$$

where the quantities $R_i=Y_i-F(X_i,\boldsymbol{A})$ are called residuals and the weights w_i (which may all be equal to one) are chosen according to an appropriate prescription.

The system (3) is an evaluation model, in the sense that for given values $\boldsymbol{\xi}$ of $\boldsymbol{X}$ and $\boldsymbol{\eta}$ of $\boldsymbol{Y}$, the values $\boldsymbol{\alpha}$ of the vector $\boldsymbol{A}$ are fixed. Usually there will be neither data nor prior knowledge about the coefficients, in which case their joint distribution follows analytically by transformation of the distribution $f_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{\xi},\boldsymbol{\eta})$ and subsequent marginalization, or numerically by propagating the latter through the model (3). In some situations there may be relevant data or prior knowledge about the coefficients $\boldsymbol{A}$, but this case falls outside the scope of this paper.

Note that in the present approach the calibration data $\boldsymbol{x}=(x_1,\ldots,x_n)^{\mathrm{T}}$ and $\boldsymbol{y}=(y_1,\ldots,y_n)^{\mathrm{T}}$ can be given from the outset or can be computed afterwards as the means of the marginal distributions for X_i and Y_i, respectively. However, these data are not strictly needed, all that is required is that information for obtaining the joint distribution $f_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{\xi},\boldsymbol{\eta})$ be available. Similarly, the estimates $\boldsymbol{a}=(a_1,..,a_m)^{\mathrm{T}}$ of the coefficients $\boldsymbol{A}$ can be computed as the means of the marginal distributions for the individual coefficients A_i, but again this computation may not always be necessary.

3.3. Consistency

After obtaining the distribution $f_{\boldsymbol{A}}(\boldsymbol{\alpha})$ it should be checked whether or not the residuals R_i are consistent with the uncertainties embodied by the distributions $f_{\boldsymbol{X,Y}}(\boldsymbol{\xi},\boldsymbol{\eta})$ and $f_{\boldsymbol{A}}(\boldsymbol{\alpha})$. In the extreme case, if the uncertainties associated with the estimates of the calibration quantities are negligible there would be no distributions $f_{\boldsymbol{X,Y}}(\boldsymbol{\xi},\boldsymbol{\eta})$ and $f_{\boldsymbol{A}}(\boldsymbol{\alpha})$, or rather, they would become multivariate delta functions [4]. The values of the coefficients would then be obtained by minimizing the sum $\sum(y_i-F(x_i,\boldsymbol{a}))^2$, which would yield a deterministic system of equations $H(\boldsymbol{x},\boldsymbol{y},\boldsymbol{a})=0$. From the solution of this system the residuals $r_i= y_i-F(x_i,\boldsymbol{a})$ would be computed; there would be no uncertainty associated with them. However, if all input coordinates are perfectly known the curve should not be "fitted": it should pass through all points, or equivalently, all residuals should vanish. If this does not happen the curve does not adequately represent the physics of the problem, there is a mistake in the process of acquisition of the calibration data or the assumption of the uncertainties being negligible is inconsistent with the chosen form of the curve.

Naturally, this extreme case is of no interest. More generally the residuals $R_i= Y_i-F(X_i,\boldsymbol{A})$ will be quantities having a joint state-of-knowledge distribution $f_{\boldsymbol{R}}(\boldsymbol{\rho})$, which may be determined by expressing the coefficients in terms of the calibration quantities through the model $H(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{A})=0$ and subsequently propagating the distributions $f_{\boldsymbol{X,Y}}(\boldsymbol{\xi},\boldsymbol{\eta})$ and $f_{\boldsymbol{A}}(\boldsymbol{\alpha})$. Consistency can then be established by calculating the area of the marginal distributions $f_{Ri}(\rho_i)$ on the side of the origin opposite to that where the respective mean r_i is situated. The data would be consistent with the function $Y=F(X,\boldsymbol{A})$ if all probabilities computed in this way are greater than some arbitrary small value, say 1%. Since the distributions $f_{Ri}(\rho_i)$ will often be Gaussians, or nearly so, this computation can be avoided by just checking if the ratios $|r_i|/u(r_i)$ are all smaller than some arbitrary number, say 3, where $u(r_i)$ is the standard deviation (or standard uncertainty) of $f_{Ri}(\rho_i)$.

3.4. Third stage: obtaining the distribution $f_{X0}(\xi_0)$

The third stage applies when the instrument is in use. It indicates a value y_0, which is taken as the best estimate of the response Y_0 whose distribution $f_{Y0}(\eta_0)$ must be established in the manner indicated in subsection 3.1, that is, by drawing a statistical model for the generation of the datum y_0 and subsequently applying Bayes' theorem, with a non-informative prior if necessary.

The distribution $f_{X0}(\xi_0)$ for the stimulus X_0 is then obtained by simply propagating the distributions $f_{Y0}(\eta_0)$ and $f_{\boldsymbol{A}}(\boldsymbol{\alpha})$ through the model

$$X_0 = G(Y_0, \boldsymbol{A}), \tag{5}$$

where G is the inverse of the function F.

3.5. Numerical computation

From this general outline it may be seen that the formulation of the problem of evaluating the uncertainty associated with the construction and use of a calibration curve is relatively simple. However, the numerical implementation of the method may present difficulties. Indeed, it will rarely be possible to obtain the distribution $f_{\boldsymbol{A}}(\boldsymbol{\alpha})$ in the second stage by using e.g. the procedure in [4] which combines analytical formulas with numerical integration routines. Moreover, the distribution $f_{\boldsymbol{A}}(\boldsymbol{\alpha})$ is seldom the desired outcome. Rather, this distribution is needed in the third stage for computing the distribution $f_{X0}(\xi_0)$.

Fortunately, at least in principle the distributions $f_{\boldsymbol{A}}(\boldsymbol{\alpha})$ and $f_{X0}(\xi_0)$ can be numerically approximated by using a Monte Carlo method [5] which can summarily be explained as follows: (i) draw random samples $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ from the distribution $f_{\boldsymbol{X,Y}}(\boldsymbol{\xi},\boldsymbol{\eta})$; (ii) compute the values $\boldsymbol{\alpha}$ that follow from the system $H(\boldsymbol{\xi},\boldsymbol{\eta},\boldsymbol{\alpha})=0$; (iii) draw a random sample η_0 from the distribution $f_{Y0}(\eta_0)$determined in the third stage; (iv) calculate the value ξ_0 that results from the equation $\xi_0=G(\eta_0,\boldsymbol{\alpha})$; (v) repeat a large number of times. According to [5], about one million trials can often be expected to deliver statistically stable results. However, since there is no guarantee that this or any specific pre-assigned number will suffice, [5] describes a procedure for selecting the number of trials adaptively.

4. EXAMPLE

Suppose a particle detector is calibrated by independently counting the number of particles coming from seven samples of radioactive reference materials having very long half-lives. The masses of all samples are equal and their certified activities x_i are in the range 1 to 50 becquerels. These values are shown in table 1, together with their associated standard uncertainties $u(x_i)$. The number of particles N_i counted by the detector during perfectly known times t_i are also shown in table 1. The best estimate of the efficiency E of the setup is e=0.300 with an associated standard uncertainty $u(e)$=0.005. From equation (3) in [6] it may be seen that this factor enters into the pertinent equations only as a factor multiplying the counting time. The background count level is also measured during calibration and is found to be negligible.

A scatter plot of the values x_i and $y_i=N_i/(e\ t_i)$ shows that the detector is not quite linear (see figure 1). This behavior can always be accommodated by choosing a polynomial as the calibration function; in the present case perhaps of second or third order. Just for demonstration purposes we shall choose instead the curve

$$Y = A_1 \exp\left(\frac{A_2}{A_3 + X}\right), \tag{6}$$

where X is the activity of the sample and Y is the corresponding measured activity.

Table 1. Calibration data for a particle detector. Samples with certified activities x and associated standard uncertainties $u(x)$ were used. The number of counts N registered by the detector during times t are also shown.

Sample	x / Bq	$u(x)$ / Bq	N / counts	t / s
1	1.178	0.015	105	240
2	3.542	0.040	234	197
3	7.213	0.079	340	160
4	14.01	0.17	550	140
5	23.66	0.28	720	110
6	33.91	0.33	690	73
7	48.85	0.51	480	40

The Monte Carlo method was applied by drawing values ξ_i from Gaussian distributions with means x_i and standard uncertainties $u(x_i)$. In accordance with the last paragraph of subsection 3.1, values η_i were drawn from gamma distributions with shape parameters N_i+0.5 and rate parameters $\varepsilon_i\, t_i$. Since the ratio $e/u(e)$ is greater than three, the values ε_i of the efficiency were drawn from a Gaussian distribution with mean e and standard deviation $u(e)$ [7]. Triplets $\boldsymbol{\alpha}$ were then computed as

$$(\alpha_1, \alpha_2, \alpha_3) = \text{Arg Min} \sum_{i=1}^{7} \left[\eta_i - \alpha_1 \exp\left(\frac{\alpha_2}{\alpha_3 + \xi_i} \right) \right]^2 . \quad (7)$$

By repeating this process 10^6 times the distribution $f_{\boldsymbol{A}}(\boldsymbol{\alpha})$ was obtained. The means and standard deviations of the coefficients are shown in table 2, and the correlation coefficients are shown in table 3. Figure 1 shows the fitted curve together with the data points. All residual ratios $r_i/u(r_i)$ are less than 3, so the chosen calibration function is consistent with the data. The fact that it does not pass through the origin, however, suggests that this function was not a very good choice. At any rate, it should not be used to predict activities below 1 Bq.

Table 2. Means and standard deviations of the coefficients $\boldsymbol{\alpha}$ in equation (6).

Coefficient	Mean / Bq	St. Dev / Bq
A_1	94.9	17.7
A_2	–51.8	11.0
A_3	12.3	2.7

Table 3. Correlation coefficients .

	A_1	A_2	A_3
A_1	1	–0.984	0.956
A_2	–0.984	1	–0.991
A_3	0.956	–0.991	1

Suppose now that a sample of unknown activity X_0 is measured and that N_0=328 gross counts are measured by the detector during t_0=170 s at a site where background activity cannot be ignored. A blank sample is then also measured, yielding N_b=27 counts during t_b=350 s. Both samples are identical in mass and shape to those used in the calibration. To obtain the activity of the sample, 10^6 values η_0 and η_b were drawn from appropriate gamma distributions and triplets $\boldsymbol{\alpha}$ were taken from the stored values of $f_{\boldsymbol{A}}(\boldsymbol{\alpha})$. The corresponding values ξ_0 were then found by solving

$$\eta_0 - \eta_b = \alpha_1 \exp\left(\frac{\alpha_2}{\alpha_3 + \xi_0} \right) . \quad (8)$$

It turned out that none of the differences $\eta_0 - \eta_b$ was negative. Figure 2 shows the numerical approximation to the distribution $f_{X0}(\xi_0)$ of the activity of the sample. The mean is x_0=6.63 Bq with standard uncertainty $u(x_0)$=0.48 Bq. It is seen from this figure that $f_{X0}(\xi_0)$ is well approximated by a Gaussian distribution, which facilitates the analytic computation of a coverage interval, should this be needed.

It should be noted that the facts that the chosen calibration function is non-linear in the coefficients and that the statistical models of the quantities $\boldsymbol{Y}$ are non-Gaussian offer absolutely no difficulties to the proposed method, other than the necessity of numerically minimizing equation (7).

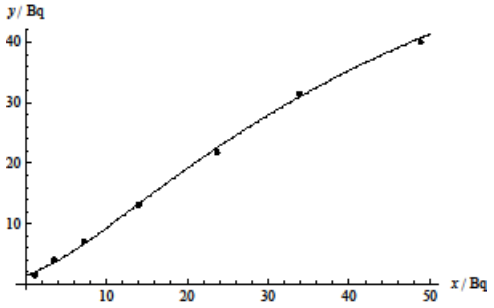

Fig. 1. Calibration curve of the particle detector using the means a in table 2 as estimates of the coecients A. The markers are located at the data points x_i and $y_i = (e\, t_i)^{-1}\, N_i$.

Fig. 2. Normalized histogram obtained by the Monte Carlo method to approximate the distribution $f_{X0}(\xi_0)$ of the activity of a measured sample. The data are N_0 = 328 counts during t_0 = 170 s. A blank sample is also measured, yielding N_b = 27 counts during t_b = 350 s. The continuous curve is a Gaussian distribution centered at ξ = 6.63 Bq and standard uncertainty 0.48 Bq.

5. CONCLUSION

A procedure for obtaining the joint state-of-knowledge probability distribution associated with the coefficients $\boldsymbol{A}$ of a curve fitted to two vectors of calibration quantities $\boldsymbol{X}$ and $\boldsymbol{Y}$ has been presented. The procedure starts by finding the joint distribution associated with these vectors in such a way

that all available information about the quantities they represent is taken into account. Next, a mathematical model of the measurement that relates the coefficients to the calibration quantities is drawn up. The model is obtained by minimizing the weighted or unweighted sum of the squared differences between the components of one vector of quantities and the members of the set formed by the curve evaluated from the other vector. The distribution for the calibration quantities, propagated through the model, determines the distribution for the coefficients (about which neither data nor prior knowledge are assumed to exist). This can be conveniently achieved by the Monte Carlo method described in Supplement 1 to the GUM [5]. Finally, the probability distribution for a stimulus X_0 can also be obtained by Monte Carlo from the distributions for the coefficients $\boldsymbol{A}$ and for the response Y_0.

The ensuing probability distributions $f_{\boldsymbol{A}}(\boldsymbol{\alpha})$ and $f_{X0}(\xi_0)$ are obtained in a somewhat subjective manner, as they depend on several assumptions and choices that have to be made throughout. One of these choices may be the form of the calibration function itself. Many different curves may be deemed to be reasonable. In any case, a test should be conducted to check whether or not the selected curve is consistent with the data and the *a priori* information. Least-squares is the most common adjustment criterion, but the method allows other possibilities to be contemplated. Finally, the sampling models used to obtain the distribution $f_{X,Y}(\boldsymbol{\xi},\boldsymbol{\eta})$ have to be chosen from the assumed characteristics of the data generation process. Therefore, the state of knowledge represented by the distributions $f_{\boldsymbol{A}}(\boldsymbol{\alpha})$ and $f_{X0}(\xi_0)$ must be interpreted as being conditional on all these choices and assumptions.

AKNOWLEDMENT

The support of Fondecyt (Chile) research grant 1095160 is gratefully acknowledged.

REFERENCES

[1] J. O. Berger, J. M. Bernardo, D.Sun, The formal definition of reference priors. *Ann. Statist.* **37**, 905-938, 2009.

[2] R. Yang, J. O. Berger. A catalog of noninformative priors. ISDS Discussion paper 97-42, Parexel International and Purdue University, 1998.

[3] G. E. P. Box, G. C. Tiao. *Bayesian Inference in Statistical Analysis*. Addison Wesley, Reading, Mass, 1973 (Reprinted 1992 Wiley Classics Library Edition).

[4] I. Lira, D. Grientschnig. Bayesian assessment of uncertainty in metrology: a tutorial. *Metrologia*, **47**, R1-R14, 2010.

[5] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. *Evaluation of measurement data / Supplement 1 to the "Guide to the expression of uncertainty in measurement" / Propagation of distributions using a Monte Carlo method*. Joint Committee for Guides in Metrology, JCGM 101, 2008.

[6] J. P. Laedermann, J. F. Valley, F. C. Bochud. Measurement of radioactive samples: application of the Bayesian statistical decision theory. *Metrologia*, **42**, 442-448, 2005.

[7] I. Lira. The probability distribution of a quantity with given mean and variance. *Metrologia*, **46**, L27-L28, 2009