



## ISTITUTO NAZIONALE DI RICERCA METROLOGICA Repository Istituzionale

The least informative distribution and correlation coefficient of measurement results

*Original*

The least informative distribution and correlation coefficient of measurement results / Mana, G; Pizzocaro, M. - In: METROLOGIA. - ISSN 0026-1394. - 58:1(2021), p. 015012. [10.1088/1681-7575/abcbe9]

*Availability:*

This version is available at: 11696/71910 since: 2021-11-11T17:07:39Z

*Publisher:*

IOP publishing

*Published*

DOI:10.1088/1681-7575/abcbe9

*Terms of use:*

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

*Publisher copyright*

(Article begins on next page)



---

PAPER • OPEN ACCESS

# The least informative distribution and correlation coefficient of measurement results

To cite this article: G Mana and M Pizzocaro 2021 *Metrologia* **58** 015012

View the [article online](#) for updates and enhancements.

# The least informative distribution and correlation coefficient of measurement results

G Mana\*  and M Pizzocaro 

INRIM—Istituto Nazionale di Ricerca Metrologica, Str. delle cacce 91, 10135 Torino, Italy

E-mail: [g.mana@inrim.it](mailto:g.mana@inrim.it)

Received 7 May 2020, revised 30 October 2020

Accepted for publication 18 November 2020

Published 11 January 2021



## Abstract

Correlations play a significant role in data analysis and the evaluation and expression of the uncertainty, yet estimating them is often difficult. This paper provides examples of how to infer the measurand value, given only the uncertainties and correlation ranges of the measurement results. The least informative data-distribution is not Gaussian, but the marginal distributions are. Explicit results are given in the case of a data pair, where the inferred correlation coefficient is the midpoint of the given range.

Keywords: probability theory, inference methods, measurement and error theory, data analysis, correlations

(Some figures may appear in colour only in the online journal)

## 1. Introduction

Correlations play an important role in data analysis and evaluating and expressing the uncertainty [1–7]. For instance, when averaging correlated data with unequal uncertainties, contrary to what might be expected, both large positive and negative correlations reduce the uncertainty [8]. Moreover, when the correlation coefficient tends to one, the weighted mean lies outside the data interval and the associated uncertainty tends to zero.

Estimating the correlation coefficients of literature data can be difficult or impossible [4, 5]. The authors do not always give them, nor the information needed for an estimate. When data are subject to accounting identities, which express a datum as the sum of the other, or they have specified marginals correlations considered in [9, 10].

Here, we consider the evaluation of the least-informative distribution and correlation coefficient by considering a data

pair, where no more information is available than the standard uncertainties and the range of the correlation coefficient. We determine the sought distribution by ensuring that, subject to any contextual information, it is minimally informative. Only in this way can we be sure that the distribution and correlation coefficient take all the information available into account, but no uncontrolled assumptions have been introduced.

In the discrete case, the maximum entropy principle, which minimizes the Shannon information encoded in a distribution, solves the problem. For continuous distributions the Shannon information is meaningless. In this case, we minimize the Kullback–Liebler divergence, which measures the difference between the sought distribution and a distribution which is assumed to encode the absence of testable information [11, 12]. A concise introduction to classical (and quantum) information theory is in [13].

In section 2, we set the problem for two variables and derive the distribution and correlation coefficient consistent with the contextual information. The following section gives the measurand posterior-distribution. Next, in sections 4 and 5, we present the posterior inference of the measurand and its standard uncertainty and discuss from where constraints on the correlation arise. The general case of more than two variables

\* Author to whom any correspondence should be addressed.

is considered in section 6. Section 7 provides some application examples.

## 2. Data likelihood

### 2.1. Problem statement

Let  $x_1$  and  $x_2$  be two measurement results having variance  $\sigma_1^2$  and  $\sigma_2^2$ , but unknown correlation coefficient  $\rho$ . The interval of the possible  $\rho$  values might be known to be smaller than the default  $[-1, 1]$  one, as will be illustrated in section 5 and examples. Since shifted and scaled variables have the same  $\rho$  value, we can assume a zero mean and unit variance without loss of generality. The challenge is to find the least informative sampling distribution and  $\rho$  value consistent with  $\rho \in [\rho_1, \rho_2]$ .

### 2.2. Solution

Since, by marginalization, the sought sampling distribution of the data is

$$P(x_1, x_2) = \int_{\rho_1}^{\rho_2} p(x_1, x_2, \rho) d\rho, \quad (1)$$

the problem is to find the joint distribution of the data and correlation coefficient,  $p(x_1, x_2, \rho)$ , given  $\langle x_1 \rangle = \langle x_2 \rangle = 0$ ,  $\sigma_1 = \sigma_2 = 1$ , and  $\rho \in [\rho_1, \rho_2]$ . Also,  $p(x_1, x_2, \rho)$  can be written in terms of conditional distributions as

$$p(x_1, x_2, \rho) = L(x_1, x_2 | \rho) \pi(\rho), \quad (2)$$

where  $L(x_1, x_2 | \rho)$  is the data likelihood (given the mean, variance, and correlation coefficient) and  $\pi(\rho)$  is the coefficient distribution, given  $\rho \in [\rho_1, \rho_2]$ .

The  $(x_1, x_2)$  distribution having the minimum Kullback–Liebler divergence from the uniform one and zero mean, unit variance, and specified  $\rho$  value,

$$L(x_1, x_2 | \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{x_1^2+x_2^2-2\rho x_1 x_2}{2(1-\rho^2)}}. \quad (3)$$

is binormal. The distribution of  $\rho \in [\rho_1, \rho_2]$ ,

$$\pi(\rho) = \frac{\text{if}(\rho_1 < \rho < \rho_2)}{\rho_2 - \rho_1}, \quad (4)$$

where  $\text{if}(\cdot)$  is one if its argument is true and zero otherwise, follows trivially from minimizing the divergence from a uniform distribution.

Eventually, by using (2)–(4) in (1), the sought distribution is

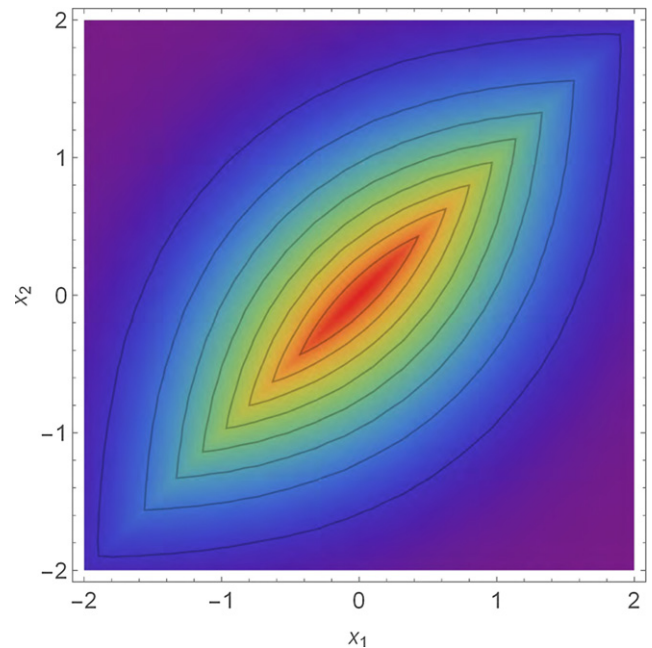
$$P(x_1, x_2) = \frac{1}{\rho_2 - \rho_1} \int_{\rho_1}^{\rho_2} \frac{e^{-\frac{x_1^2+x_2^2-2\rho x_1 x_2}{2(1-\rho^2)}}}{2\pi\sqrt{1-\rho^2}} d\rho. \quad (5)$$

Figure 1 shows its contour plot when  $\rho_1 = 0$  and  $\rho_2 = 1$ .

The least-informative correlation coefficient of the data,

$$\rho_{12} = (\rho_1 + \rho_2)/2, \quad (6)$$

is obtained from (2) by carrying out the relevant integrations. Also, the marginal distributions of the data are normal, with zero mean and unit variance.



**Figure 1.** Contour plot of the distribution of two correlated data having zero mean, unit variance, and (unknown)  $\rho \in [0, 1]$ .

The equation (6) can be generalized to the case where more detailed measurable information on  $\rho$  is available and, consequently,  $\pi(\rho)$  is any distribution. By proceeding as before, we still have  $\langle x_1 \rangle = \langle x_2 \rangle = 0$ ,  $\sigma_1^2 = \sigma_2^2 = 1$ . Furthermore,  $\rho_{12} = \langle \rho \rangle_\pi$ , where the average  $\langle \cdot \rangle_\pi$  is taken with respect to  $\pi(\rho)$ .

## 3. Posterior distribution of the measurand

Let us consider a data pair having the same mean, the measurand  $\mu$ , and variance-covariance matrix

$$\Sigma = \begin{pmatrix} a^2 & a\rho \\ a\rho & 1 \end{pmatrix} \sigma_2^2, \quad (7)$$

where  $0 \leq a^2 \leq 1$  is the variances' ratio. Without loss of generality, we can set  $x_1 = 0$  and  $x_2 = 1$ , which can be done by shifting and rescaling the data. Therefore,  $\sigma_1 = a\sigma_2$  and  $\sigma_2$  are the fractional standard uncertainties of the most and least accurate datum, respectively, relative to their difference. Also, if  $\sigma_2 \ll 1$ , the data are inconsistent.

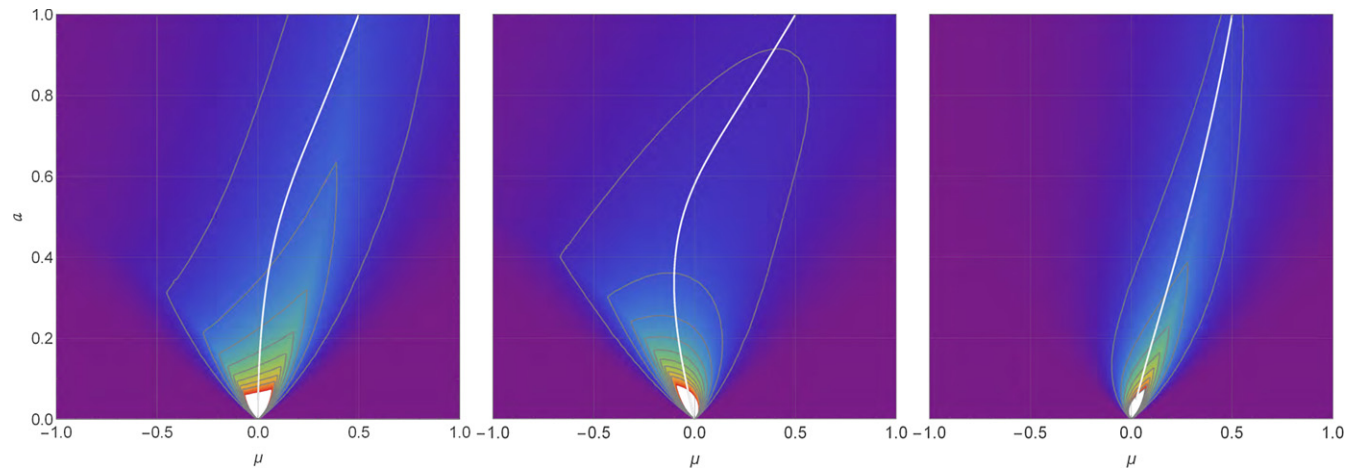
### 3.1. Fixed correlation

If  $\rho$  is known, the least-informative distribution and likelihood of the reduced data  $x_1 = 0$  and  $x_2 = 1$  is

$$L(\mathbf{x} | \mu) = \frac{e^{-\mathbf{x}^T \Sigma^{-1} \mathbf{x} / 2}}{2\pi\sqrt{|\Sigma|}}, \quad (8)$$

where  $\mathbf{x} = (-\mu, 1 - \mu)^T$  and  $|\cdot|$  indicates the determinant.

To take informed decisions, we need the posterior distribution of the measurand values. By mapping the prior information on  $\mu$  into a uniform distribution, as its support tends to the reals, the sought posterior converges to a normal one. Hence,



**Figure 2.** Representation of the posterior probability densities (9) of the measurand values—horizontal sections of the density plots—as it varies with increasing uncertainty ratio  $a$ , when  $\sigma_2 = 1$ ,  $\rho \in [-1, 1]$  (left),  $\rho \in [0, 1]$  (center), and  $\rho \in [-1, 0]$  (right). The data have been normalized so that  $x_1 = 0 \pm a$  and  $x_2 = 1 \pm 1$ . The gray contour-lines are the loci of constant density-values. The white lines are the posterior mean of  $\mu$ . The white area is where the density extends beyond the plot range; in fact, it goes to the infinite.

$\mu \sim N(\mu|\mu_0, \sigma_\mu^2)$ , where the posterior mean and variance,  $\mu_0$  and  $\sigma_\mu^2$ , are given in section 4.1.

3.2. Unknown correlation

If  $\rho \in [\rho_1, \rho_2]$ , by the same reasoning that delivered us to (3), the least-informative distribution and likelihood of the reduced data  $x_1 = 0$  and  $x_2 = 1$  is

$$L(\mathbf{x}|\mu) = \frac{1}{\rho_2 - \rho_1} \int_{\rho_1}^{\rho_2} \frac{e^{-\mathbf{x}^T \Sigma^{-1} \mathbf{x} / 2}}{2\pi \sqrt{|\Sigma|}} d\rho. \tag{9}$$

If we give to the  $\mu$  values a sequence uniform prior distribution having increasingly large supports, the posterior one converges to  $\mu \sim L(\mathbf{x}|\mu)/Z$ , where the normalization constant  $Z$  is given in the appendix.

Figure 2 shows the posterior distribution of the measurand for increasing uncertainty ratio, when  $\sigma_2 = 1$ ,  $\rho \in [-1, 1]$  (left),  $\rho \in [0, 1]$  (center), and  $\rho \in [-1, 0]$  (right). If the data have different uncertainties, both positive and negative correlations shrink the mean to the most accurate datum, while reducing the uncertainty. Also, unless  $\rho < 0$ , the posterior mean approaches the most accurate datum from the outside of the data range.

4. Posterior mean and variance

4.1. Known correlation

If  $\rho$  is known in advance, the posterior mean,

$$\mu_0 = \frac{a(a - \rho)}{1 - 2a\rho + a^2}, \tag{10}$$

and variance,

$$\sigma_\mu^2 = \frac{(1 - \rho^2)\sigma_1^2}{1 - 2a\rho + a^2}, \tag{11}$$

of the measurand equal the maximum likelihood estimates [8].

4.2. Unknown correlation

When  $\rho$  is unknown, the analytic expressions of the posterior mean and variance of the measurand (given in the appendix) are too complex to give useful insights. They differ from (10) and (11) and from the maximum likelihood estimates. This difference is not surprising. The maximum likelihood estimates are the optimal compressions of the measurement results; the posterior mean and variance are the updated expectation and uncertainty of the measurand.

We give some asymptotic behavior that might help to check the correctness of (A.2) and (A.3). If the uncertainty ratio tends to one, the posterior mean is the data midpoint. If it tends to zero, the posterior mean,

$$\lim_{a \rightarrow 0} \mu_0 = -(\rho_1 + \rho_2)a/2 \tag{12}$$

approaches the most accurate datum from above or below according to the midpoint of the interval of the possible  $\rho$  values. The posterior variance approaches

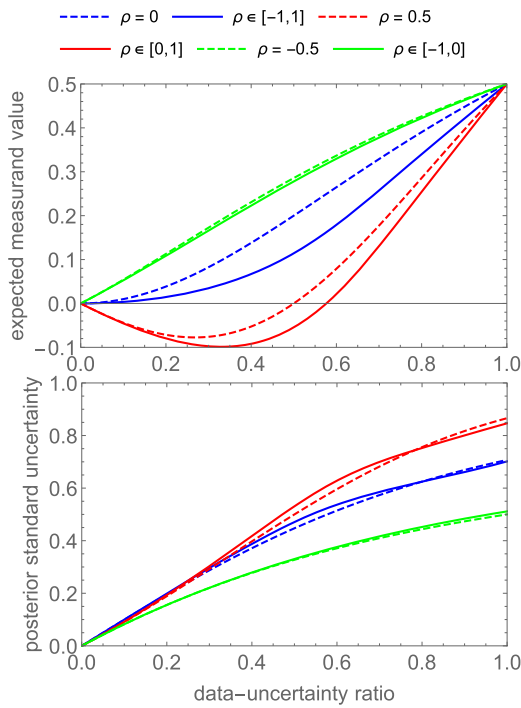
$$\lim_{a \rightarrow 0} \sigma_\mu^2 = (1 - \rho_1\rho_2)\sigma_1^2. \tag{13}$$

It is worth noting that when  $\rho_1 = \rho_2 = \rho$ , (12) and (13) are the same limits as per (10) and (11). Eventually, if  $\sigma_2 \gg 1$  the posterior mean is

$$\lim_{\sigma_2 \rightarrow \infty} \mu_0 = \begin{cases} -\frac{1 + a - \sqrt{1 + a^2}}{2\sqrt{1 + a^2}} & \text{if } \rho \in [0, 1] \\ 0 & \text{if } \rho \in [-1, 1] \\ -\frac{1 - a - \sqrt{1 + a^2}}{2\sqrt{1 + a^2}} & \text{if } \rho \in [-1, 0] \end{cases}$$

which, when  $a \rightarrow 0$ , repeats (12).

Figure 3 compares the posterior mean and standard deviation (A.2) and (A.3) to the (10) and (11) ones, when  $\sigma_2 = 1$ . Positive correlations bias the mean toward the most accurate datum more than the negative ones. Unless  $\rho \in [-1, 0]$ , (A.2) shifts more quickly than (10). This occurs because—as shown



**Figure 3.** Posterior mean (top) and standard uncertainty (bottom) of the measurand value. Solid lines: unknown correlations, the ranges are given in the legend. Dashed lines: fixed correlations, the least-informative values are given in the legend. The reduced data are  $x_1 = 0 \pm a$  and  $x_2 = 1 \pm 1$ .

by the blue and red lines—both positive and negative correlations bias the mean in the same way. In the cases considered, the measurand variance (A.3) equals (11) for all practical purposes. If the data have the same uncertainty, the variance is  $1/2$  when  $\rho = 0$ ,  $3/4$  when  $\rho = 0.5$ , and  $1/4$  when  $\rho = -0.5$ .

### 5. Correlation range

Correlations arise because, for example, the same standards or input data are used. Other sources are noise transients in time-series data and constraints, for instance, to match an aggregate datum.

To see how contextual information constrains the  $\rho$  range, without loss of generality, let us suppose that the data model is

$$\begin{aligned} x_1 &= \mu + \eta_1 + \eta_0 \\ x_2 &= \mu + \eta_2 + h\eta_0 \end{aligned} \tag{14}$$

where  $\eta_{1,2,0}$  are zero-mean uncorrelated errors having variances  $u_1^2$ ,  $u_2^2$  and  $u_0^2$ , respectively and  $\eta_0$  and  $h$  take the uncertainty of the common references and input data (or the noise memory) and the (possibly) different sensitivities into account.

If  $h = 1$ , (14) describe data corrected for the same quantity, where the quantity magnitude is uncertain. If  $h \neq 1$ , the data are corrected for different values of the same quantity, where their responsiveness is uncertain. For example, different thermal expansions, where the expansion coefficient is uncertain. In this case,  $u_0^2$  is the variance of the expansion coefficient and  $h$  the ratio of the differences between the measurement temperatures and a reference one.

The variance-covariance matrix of the data is given by (7), where

$$\sigma_2 = \sqrt{u_2^2 + h^2 u_0^2}, \tag{15a}$$

$$a = \sigma_1 / \sigma_2 = \sqrt{u_1^2 + u_0^2} / \sigma_2, \tag{15b}$$

$$\rho = \frac{h u_0^2}{a \sigma_2^2}. \tag{15c}$$

The inequalities  $0 < u_0 / \sigma_2 < a < 1$ ,  $0 < |\rho| < 1$ , and  $|h| < \sigma_2 / u_0$ , can be solved to see that [14]

$$|\rho| < \min [ |h| a, (|h| a)^{-1} ]. \tag{16}$$

This inequality defines the range of the possible  $\rho$  values when there is no other information than the total standard deviations  $\sigma_{1,2}$ . The data can be maximally correlated only if  $|h| a = 1$ . If  $h = 1$ , the maximum  $\rho$  value is equal to the uncertainty ratio. If  $a = 1$ , the magnitude of the sensitivity ratio or its inverse bound the correlation.

### 6. Extension to multiple data

To extend our results to more than two data, it is necessary to take the positivity of the covariance matrix into account. A criterion, named after James Joseph Sylvester [15], to test the matrix positivity is that all its principal minors are positive. For instance, in the case of three data, the first two principal minors are always positive and the Sylvester’s criterion reduces to  $1 - \rho_{12}^2 - \rho_{23}^3 - \rho_{13}^2 + 2\rho_{12}\rho_{23}\rho_{13} > 0$ . Figure 4 shows the three-dimensional region in which this inequality is true.

By assuming that the results were corrected for the same systematic effects,

$$x_i = \mu + \eta_i + h_i \eta_0 \tag{17}$$

where  $\eta_{i,0}$  are zero-mean uncorrelated errors having variances  $u_{i,0}^2$ . In (17),  $\eta_0$  and  $h_i$  take the common contributions to the error budgets and the different  $x_i$ ’s sensitivities to them into account. For the sake of simplicity, we consider only the case  $\forall i, h_i > 0$  and assume that every common contribution to the error budgets is not greater than the total standard uncertainty of the most accurate datum, say  $\sigma_1$ . Hence, when there is no other information than the total standard deviations  $\sigma_i$ , the ranges of the  $\rho_{ij}$  values are constrained by

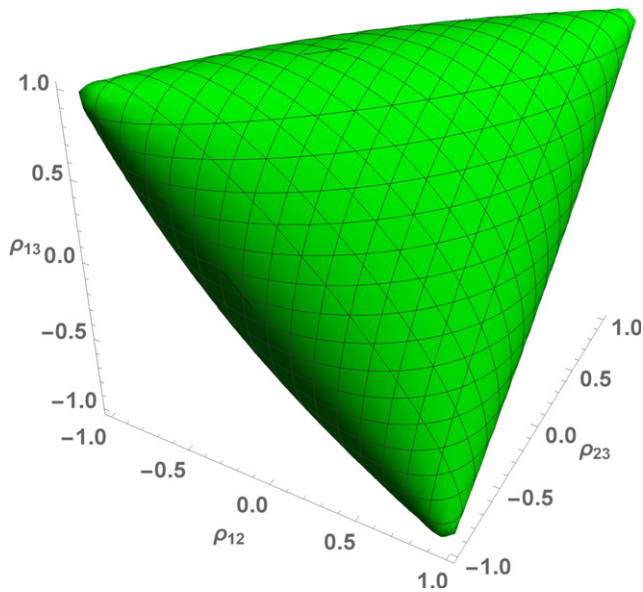
$$0 < \rho_{ij} = \frac{h_i h_j u_0^2}{\sigma_i \sigma_j} < \frac{\sigma_1^2}{\sigma_i \sigma_j}. \tag{18}$$

By proceeding as in section 3.2, the posterior distribution of the  $\mu$  value is

$$\mu \sim \frac{1}{Z} \int \dots \int \frac{e^{-\mathbf{x}^T \Sigma^{-1} \mathbf{x} / 2}}{\sqrt{|\Sigma|}} d\rho_{1,2} \dots d\rho_{n-1,n}, \tag{19}$$

$0 < \rho_{ij} < \sigma_1^2 / (\sigma_i \sigma_j)$   
 $\Sigma > 0$

where  $\mathbf{x} = (x_1 - \mu, x_2 - \mu, \dots, x_n - \mu)^T$  and  $\Sigma > 0$  limits the integration region to where  $\Sigma$  is positive definite (for an



**Figure 4.** A 3 × 3 covariance matrix is positive definite if and only if the correlation coefficients are inside an inflated tetrahedron.

example, in the case of three data, see figure 4). The normalization constant  $Z$ , posterior mean, and variance are given in ways similar to (A.1)–(A.3), but no analytical integration is possible.

### 7. Application examples

#### 7.1. Example 1

The two repeated measurements of the Avogadro constant,  $N_A = 6.022\,140\,99(18) \times 10^{23} \text{ mol}^{-1}$  and  $N_A = 6.022\,140\,76(12) \times 10^{23} \text{ mol}^{-1}$ , given in [16] are correlated by a number of (nearly) equal corrections, made for the effect of the same influence quantities. To make these  $N_A$  values usable for the estimation of a self-consistent set of values of the constants of physics by the Committee on Data for Science and Technology (CODATA), one of the authors estimated their correlation coefficient as equal to 0.17 [4].

Taking this value into account, the posterior inference and maximum likelihood estimate of the Avogadro constant are both equal to [4]

$$N_A = 6.022\,140\,82(11) \times 10^{23} \text{ mol}^{-1}. \quad (20)$$

If the analyst does not know the  $\rho$  value—by assuming that the data model is (14), where  $h = 1$ —is still possible to infer the  $N_A$  values as follows. Since the uncertainty ratio is  $a = 2/3$ , according to (16), the  $\rho$  value must be in the  $[0, 2/3]$  interval and its least informative value is  $\rho = 1/3$ , not so far from the estimated 0.17 value. The posterior mean and standard uncertainty (A.2) and (A.3) are

$$N_A = 6.022\,140\,81(11) \times 10^{23} \text{ mol}^{-1}. \quad (21)$$

This value is in excellent agreement with (20), which required a detailed analysis of each contribution to the total uncertainties of the measured values.

**Table 1.** Reduced results  $x_i$  (first row and column), uncertainty ratios  $a_i = \sigma_i/\sigma_4$  (diagonal), and correlation coefficients (upper triangle, blue). The lower triangle (red) gives the correlation upper bounds estimated via the model (18). The largest reduced standard-uncertainty is  $\sigma_4 = 0.424$ . Adapted from [6].

$x_i$	0.000	0.410	0.599	1.000
0.000	0.389	0.303	0.205	0.188
0.419	0.583	0.667	0.276	0.245
0.599	0.467	0.272	0.833	0.134
1.000	0.389	0.227	0.181	1.000

#### 7.2. Example 2

Since 2011, the international Avogadro coordination determined  $N_A$  by counting the atoms in the same  $^{28}\text{Si}$ -enriched monocrystals.

Reference [6] provides guidance on how these results must be used to infer an updated value. Table 1 shows the reduced data,

$$0 \leq x_i = \frac{N_i - N_1}{N_4 - N_1} \leq 1, \quad (22)$$

sorted according to ascending uncertainties, the uncertainty ratios  $a_i = \sigma_i/\sigma_4 \leq 1$ , and the correlation coefficients (typed blue).  $N_4$  and  $N_1$  are the least and most accurate results, respectively, and

$$\sigma_i = \frac{\sigma_{N_i}}{N_4 - N_1}. \quad (23)$$

Taking these data into account, the posterior inference and maximum likelihood estimate are both equal to  $\mu_0 = x_{\text{ML}} = 0.15(15)$ , or  $N_A = 6.022\,140\,588(65) \times 10^{23} \text{ mol}^{-1}$  [6]. If the correlations were set to zero, because unknown,  $\mu_0 = x_{\text{ML}} = 0.25(13)$ , or  $N_A = 6.022\,140\,631(54) \times 10^{23} \text{ mol}^{-1}$ , are significantly different.

If the correlations were unknown, by making the simplest assumption that the results were equally corrected for the same systematic effects, the data model is the same as (17). The upper bounds to the correlation coefficients given by (18) are shown in table 1 (typed red). Two of them are slightly inconsistent. This inconsistency might originate from having neglected the differences between the systematic-correction variances.

The posterior mean and standard uncertainty have been numerically evaluated via (19) and the relevant integrations. They are  $\mu_0 = 0.17(15)$ , or  $N_A = 6.022\,140\,598(64) \times 10^{23} \text{ mol}^{-1}$ . Not so dissimilar from the inference using the maximally informative data.

#### 7.3. Example 3

As an example from the interlaboratory comparisons, which produce a consensus value for the common measurand that averages the measurement results of the participants, we considered a bilateral comparison of stainless steel mass standards [17], where the participants supplied an estimate of the correlation coefficients.

Reference [17] and table 2 provide the differences between the true and nominal masses measured by the pilot (before and after the circulation) and participating laboratory, as well

**Table 2.** Weighing results for the 500 g mass standard as differences (expressed in mg) from the nominal mass value (first row and column), associated standard uncertainties (diagonal), and correlation coefficients (upper triangle, blue). The lower triangle (red) gives the correlation upper bounds estimated via the model (18). Adapted from [17].

$x_i$	−0.237	−0.222	−0.244
−0.237	0.086	0.60	0.60
−0.222	0.86	0.10	0.60
−0.244	0.78	0.67	0.11

as the associated standard uncertainties,  $\sigma_i$ , and correlation coefficients (typed blue). The upper bounds to the correlation coefficients, calculated by application of (18), are typed red.

The posterior mean and standard uncertainty, −0.234(37) mg, have been numerically evaluated via (19) and the relevant integrations. It is worth noting that, in this case, the domain  $0 < \rho_{ij} < \sigma_1^2/(\sigma_i\sigma_j)$  is not fully included in  $\Sigma \succ 0$ . The maximum likelihood estimate and the associated standard uncertainty given in [17] are −0.234(41) mg.

#### 7.4. Example 4

As a last example, we consider the results of the frequency measurements of the unperturbed transition  $^1S_0-^3P_0$  in the  $^{171}\text{Yb}$  atom [18, 19],

$$\nu_{\text{Yb}} = 518\,295\,836\,590\,863.71(11) \text{ Hz} \quad (24a)$$

$$\nu_{\text{Yb}} = 518\,295\,836\,590\,863.61(13) \text{ Hz}, \quad (24b)$$

where the traceability to the international system was provided by a link to the international atomic time (TAI) [20]. The authors of [18, 19] calculated an uncertainty contribution from the primary frequency standards that contributed to TAI during the measurements of 0.07 Hz and 0.06 Hz, respectively.

The systematic contributions to the uncertainty of each TAI's standard are correlated, though different averages might have been used in the two measurements. The data published by the International Bureau of Weights and Measures allowed the estimate of the  $\rho$  value. We calculated  $\rho = 0.27$ , after a careful evaluation of the contributions of the relevant primary and secondary standards. Taking this value into account, the posterior inference and maximum likelihood estimate of the  $^1S_0-^3P_0$  transition-frequency are  $\nu_{\text{Yb}} = 518295836590863.671(94)$  Hz.

By using the data model (14), where  $\sigma_1 = 0.11$  Hz,  $\sigma_2 = 0.13$  Hz,  $u_0 \leq 0.07$  Hz, and  $hu_0 \leq 0.06$  Hz, the maximum data-correlation is  $\rho = 0.29$ . According to (A.2) and (A.3), the posterior inference of the transition frequency is  $\nu_{\text{Yb}} = 518295836590863.670(90)$  Hz, which is again in excellent agreement with the maximum likelihood estimate.

## 8. Conclusions

Correlation coefficients give the appropriate weights in analyzing the measurement results and expressing the

measurand uncertainty. However, the information reported in the literature is often insufficient to estimate them, and additional inputs would be necessary. Therefore, we provided examples of the inference of the measurand value, when only the total uncertainties and correlation limits are associated with the measurement results.

To do this, we determined the least informative sampling distribution by minimizing the Kullback–Liebler divergence relative to a uniform distribution. In the case of a data pair, the inferred correlation coefficient is the midpoint of the possible values. In particular, the correlation coefficient inferred from the absence of constraints is zero. However, the data distribution is not Gaussian, though its marginals are.

Our results apply when common contribution(s) are included in the error budgets, but their amounts are unknown. Obviously, our results do not apply to correlation of dark uncertainties [21].

## Acknowledgments

Support was received from the project EMPIR 18SIB05 ROCIT and Ministero dell'Istruzione, dell'Università e della Ricerca. This project received funding from the EMPIR programme co-financed by the Participating States and the European Union's Horizon 2020 research and innovation programme.

## Appendix A. Posterior mean and variance

The normalization constant of the posterior distribution (9) is [14]

$$Z = \frac{\int_{-\infty}^{+\infty} \int_{\rho_1}^{\rho_2} e^{-\frac{(1-2a\rho+a^2)\mu^2-2a(a-\rho)\mu+a^2}{2a^2(1-\rho^2)\sigma_2^2}} \frac{d\rho d\mu}{2\pi\sqrt{1-\rho^2}a\sigma_2^2}}{\rho_2 - \rho_1} = \frac{\tau_1 E_{3/2}(\tau_1^{-2}) - \tau_2 E_{3/2}(\tau_2^{-2})}{2\sqrt{\pi}a(\rho_2 - \rho_1)\sigma_2^2}, \quad (A.1)$$

where  $\tau_{1,2}^2 = 2\sigma_2^2(a^2 - 2a\rho_{1,2} + 1)$  and  $E_n(\cdot)$  is the exponential integral function [22, 23].

The posterior mean is [14]

$$\mu_0 = \frac{\int_{-\infty}^{+\infty} \int_{\rho_1}^{\rho_2} \mu e^{-\frac{(1-2a\rho+a^2)\mu^2-2a(a-\rho)\mu+a^2}{2a^2(1-\rho^2)\sigma_2^2}} \frac{d\rho d\mu}{2\pi\sqrt{1-\rho^2}a\sigma_2^2}}{Z(\rho_2 - \rho_1)} = \frac{\left\{ \frac{\tau_1 \exp(-\tau_1^{-2}) - \tau_2 \exp(-\tau_2^{-2})}{\sqrt{\pi} [(a^2 - 1)\sigma_2^2 - 1] [\text{erf}(\tau_1^{-1}) - \text{erf}(\tau_2^{-1})]} \right\}}{2 [\tau_1 E_{3/2}(\tau_1^{-2}) - \tau_2 E_{3/2}(\tau_2^{-2})]}, \quad (A.2)$$

where  $\text{erf}(\cdot)$  is the error function [22, 23].



The posterior variance is [14]

$$\sigma_{\mu}^2 = \frac{\int_{-\infty}^{+\infty} \int_{\rho_1}^{\rho_2} \mu^2 e^{-\frac{(1-2a\rho+a^2)\mu^2-2a(a-\rho)\mu+a^2}{2a^2(1-\rho^2)\sigma_2^2}} d\rho d\mu}{Z(\rho_2 - \rho_1)} - \mu_0^2$$

$$= \frac{\left\{ \sqrt{\pi} (3\sigma_2^2 + 1) \tau_1 \tau_2 [\operatorname{erf}(\tau_1^{-1}) - \operatorname{erf}(\tau_2^{-1})] + 2 [v_1 \tau_2 \exp(-\tau_1^{-2}) - v_2 \tau_1 \exp(-\tau_2^{-2})] \sigma_2^2 \right\}}{3\tau_1 \tau_2 [\tau_1 E_{3/2}(\tau_1^{-2}) - \tau_2 E_{3/2}(\tau_2^{-2})]} - \mu_0^2, \quad (\text{A.3})$$

where

$$v_{1,2} = 1 + a^2 - 2a\rho_{1,2} + [2 + a^2 + 2a^4 - a\rho_{1,2}(2 + 2a^2 + a\rho_{1,2})] \sigma_2^2. \quad (\text{A.4})$$

## ORCID iDs

G Mana  <https://orcid.org/0000-0002-4109-7254>

M Pizzocaro  <https://orcid.org/0000-0003-2353-362X>

## References

- [1] Zorn C J W 2001 *Am. J. Polit. Sci.* **45** 470–90
- [2] Hanley J A, Negassa A, deBurgh Edwardes M D and Forrester J E 2003 *Am. J. Epidemiol.* **157** 364–75
- [3] Bergoglio M, Malengo A and Mari D 2011 *Measurement* **44** 1461–7
- [4] Mana G *et al* 2015 *J. Phys. Chem. Ref. Data* **44** 031209
- [5] Margolis H S and Gill P 2015 *Metrologia* **52** 628–34
- [6] Fujii K, Massa E, Bettin H, Kuramoto N and Mana G 2017 *Metrologia* **55** L1–4
- [7] Mohr P J, Newell D B, Taylor B N and Tiesinga E 2018 *Metrologia* **55** 12–146
- [8] Cox M G, Eiø C, Mana G and Pennecci F 2006 *Metrologia* **43** S268–75
- [9] Larralde H 2012 *Phys. Rev. E* **86** 061117
- [10] Rodrigues J D F 2016 *J. Bus. Econ. Stat.* **34** 357–67
- [11] Jaynes E and Bretthorst G 2003 *Probability Theory: The Logic of Science* (Cambridge: Cambridge University Press)
- [12] Sivia D and Skilling J 2006 *Data Analysis: A Bayesian Tutorial Oxford Science Publications* (Oxford: Oxford University Press)
- [13] Witten E 2020 *Riv. Nuovo Cimento* **43** 187–227
- [14] Wolfram Research, Inc 2017 *Mathematica, Version 12* (Champaign, IL: Wolfram Programming Lab) [www.wolfram.com](http://www.wolfram.com)
- [15] Gilbert G T 1991 *Am. Math. Mon.* **98** 40–6
- [16] Azuma Y *et al* 2015 *Metrologia* **52** 360–75
- [17] Mitsas C and Mangutova-Stoilkovska B 2012 *Metrologia* **49** 07013
- [18] McGrew W F *et al* 2019 *Optica* **6** 448–54
- [19] Pizzocaro M, Bregolin F, Barbieri P, Rauf B, Levi F and Calonico D 2019 *Metrologia* **57** 035007
- [20] Panfilo G and Arias F 2019 *Metrologia* **56** 042001
- [21] Thompson M and Ellison S L R 2011 *Accred Qual. Assur.* **16** 483–7
- [22] Olver F W J, Olde Daalhuis A B, Lozier D W, Schneider B I, Boisvert R F, Clark C W, Miller B R and Saunders B V (ed) 2019 NIST digital library of mathematical functions <https://dlmf.nist.gov/8.19.E3,%http://dlmf.nist.gov/7.2.E1>
- [23] Marichev O and Trott M 2019 Mathematical functions site <http://functions.wolfram.com/06.34.02.0001.01>