

## SHORT COMMUNICATION

# About the probability distribution of a quantity with given mean and variance

Stefano Olivares<sup>1,2,3</sup> and Matteo G A Paris<sup>2,3,4</sup>

<sup>1</sup> Dipartimento di Fisica, Università degli Studi di Trieste, I-34151 Trieste, Italy

<sup>2</sup> Dipartimento di Fisica, Università degli Studi di Milano, I-20133 Milano, Italy

<sup>3</sup> CNISM, UdR Milano Statale, I-20133 Milano, Italy

<sup>4</sup> INRIM, Strada delle Cacce 91, 10135 Torino, Italy

E-mail: [stefano.olivares@fisica.unima.it](mailto:stefano.olivares@fisica.unima.it) and [matteo.paris@fisica.unimi.it](mailto:matteo.paris@fisica.unimi.it)

Received 7 October 2011, in final form 13 March 2012

Published 4 April 2012

Online at [stacks.iop.org/Met/49/L14](http://stacks.iop.org/Met/49/L14)

## Abstract

Supplement 1 to the GUM (GUM-S1) recommends the use of the maximum entropy principle (MaxEnt) in determining the probability distribution of a quantity having specified properties, e.g., specified central moments. When we only know the mean value and the variance of a variable, GUM-S1 prescribes a Gaussian probability distribution for that variable. When further information is available, in the form of a finite interval in which the variable is known to lie, we indicate how the distribution for the variable in this case can be obtained. A Gaussian distribution should only be used in this case when the standard deviation is small compared with the range of variation (the length of the interval). In general, when the interval is finite, the parameters of the distribution should be evaluated numerically, as suggested by Lira (2009 *Metrologia* 46 L27). Here we note that the knowledge of the range of variation is equivalent to a bias of the distribution towards a flat distribution in that range, and the principle of minimum Kullback entropy (mKE) should be used in the derivation of the probability distribution rather than the MaxEnt, thus leading to an exponential distribution with non-Gaussian features. Furthermore, up to evaluating the distribution negentropy, we quantify the deviation of mKE distributions from MaxEnt ones and, thus, we rigorously justify the use of the GUM-S1 recommendation also if we have further information on the range of variation of a quantity, namely, provided that its standard uncertainty is sufficiently small compared with the range.

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

Supplement 1 to the GUM (GUM-S1) [1] provides assignments of probability density functions for some common circumstances. In particular, it is stated that if we know only the mean value  $\bar{x}$  and the variance  $\sigma_x^2$  of a certain quantity  $X$ , we should assign a Gaussian probability distribution to that quantity, according to the principle of maximum entropy (MaxEnt) [2, 3]. The derivation is quite simple, as one has to look for the distribution  $p(x)$  maximizing the Shannon entropy:

$$S[p] = - \int_{\mathbb{R}} dx p(x) \log p(x), \quad (1)$$

which is given by

$$p(x) = \exp\{-\lambda_0 - \lambda_1 x - \lambda_2 x^2\}, \quad (2)$$

where the values of the coefficients  $\lambda_k$  should be determined to satisfy the constraints:

$$\int_{\mathbb{R}} dx p(x) x^k = M_k, \quad (3)$$

with

$$M_0 = 1, \quad M_1 = \bar{x}, \quad M_2 = \sigma_x^2 + \bar{x}^2. \quad (4)$$

However, sometimes we also know the range of the possible values of the quantity  $X$ . Two relevant examples are given by the phase-shift in interferometry, which is topologically confined in a  $2\pi$  window, and by the displacement amplitude of a harmonic oscillator, whose range of variation is dictated

by energy constraints. In this case, it has been noticed by Lira in [4] that a Gaussian probability distribution with support on the real axis can be rigorously justified only if the standard uncertainty is sufficiently small with respect to the range of variation of the quantity. In more detail, if we have any information about the range of variation, then this information should be employed in deriving the distribution maximizing the entropy as well as in evaluating the values of the coefficients  $\{\lambda_0, \lambda_1, \lambda_2\}$  of the distribution.

We denote by  $\mathbb{B} \subset \mathbb{R}$  the range of the quantity  $X$ , i.e. the subset of the real line where the values of  $X$  have non-zero probability to occur. The functional form of the distribution is still given by the exponential function in equation (2), but with non-zero support only in  $\mathbb{B}$ , whereas the coefficients are to be determined by formulae like those in equation (3), again with  $\mathbb{R}$  replaced by  $\mathbb{B}$ . It then follows, e.g., that for a variable which is known *a priori* to lie in a given interval, the MaxEnt distribution is not Gaussian, and the Gaussian approximation may be employed only if the standard deviation is small compared with the range of possible values of the quantity.

Here we point out that having information about the range of variation may be expressed as a bias of the distribution towards a flat distribution in that range and the reasoning presented in [4] may be subsumed by the minimum Kullback entropy principle (mKE) [5, 6, 7]. The Kullback entropy, or relative entropy, or Kullback–Leibler divergence, of two distributions  $p(x)$  and  $q(x)$  reads

$$K[p|q] = \int_{\mathbb{R}} dx p(x) \log[p(x)/q(x)]. \quad (5)$$

According to the mKE, in order to find the distribution  $p(x)$  given a bias towards  $q(x)$ , we should minimize the function

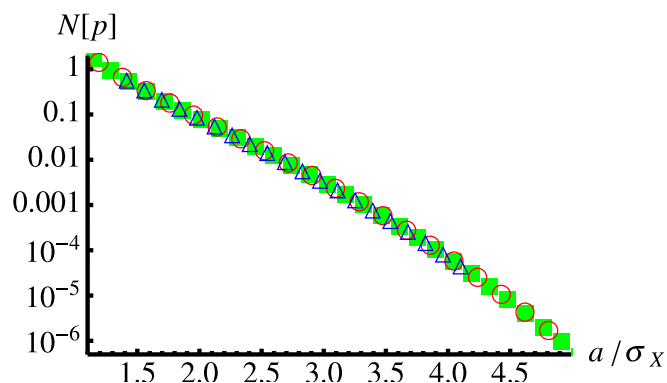
$$K[p] = K[p|q] + \sum_{k=0}^2 \lambda_k \left[ \int_{\mathbb{R}} dx p(x) x^k - M_k \right], \quad (6)$$

with respect to the function  $p(x)$ , obtaining

$$p(x) = q(x) \exp\{-\lambda_0 - \lambda_1 x - \lambda_2 x^2\}, \quad (7)$$

where the parameters  $\lambda_k$  can be still (numerically) computed using equation (3). Equation (7) represents the probability distribution satisfying the given constraints, but with a bias towards the distribution  $q(x)$ , which, for instance, may contain the information about the range of the variable  $x$ . This information, which in the case of the MaxEnt is not explicitly taken into account, is now naturally considered from the beginning. Remarkably, this is a different scenario from that covered in GUM-S1, i.e. when further information on the quantity is available, namely, the interval of values within which the quantity is known to lie is finite.

Indeed, as mentioned above, if the standard uncertainty is sufficiently small with respect to the range of variation of the quantity, we can adopt a Gaussian probability distribution over the whole real axis and, thus, use the GUM-S1 recommendation. In order to rigorously justify this statement, which has been qualitatively addressed in [4], we assess



**Figure 1.** Scaling of the negentropy of mKE distribution for zero mean value and a variable known to lie in a symmetric interval  $[-a, a]$ . We report the negentropy of the distribution as a function of the ratio  $a/\sigma_X$  for different values of the variance:  $\sigma_X^2 = 0.5$  (green squares), 1 (red circles), 2 (blue triangles).

quantitatively how the knowledge of the range of variation influences the assignment of a probability distribution by considering the deviation of the mKE distribution from a Gaussian distribution, which would represent the MaxEnt solution in the absence of any information about the range of variation. The deviation from normality of the mKE distribution (7) may be quantified by its negentropy [8]:

$$N[p] = \frac{1}{2} [1 + \log(2\pi\sigma_X^2)] - S[p], \quad (8)$$

where  $S[p]$  is the Shannon entropy (1) of the distribution (7). As for example, for a variable known to lie in a given interval  $[a, b] \subset \mathbb{R}$ ,  $a < b$ , that corresponds to a bias of  $p(x)$  towards the flat distribution:

$$q(x) = \begin{cases} (b-a)^{-1} & \text{if } x \in [a, b], \\ 0 & \text{otherwise,} \end{cases} \quad (9)$$

the negentropy (8) reads

$$N[p] = \frac{1}{2} [1 + \log(2\pi\sigma_X^2)] - \log(b-a) - \lambda_0 - \lambda_1 \bar{x} - \lambda_2 (\sigma_X^2 + \bar{x}^2). \quad (10)$$

In the simplest case, namely when  $\bar{x} = 0$  and  $x \in [-a, a]$ , the dependence of the coefficients  $\lambda_0$  and  $\lambda_2$  is such that we have a scaling law for negentropy, which depends only on the ratio  $a/\sigma_X$ . This is illustrated in figure 1, where we report the negentropy as a function of  $a/\sigma_X$  for different values of  $\sigma_X$ .

In conclusion, we have shown that the determination of the probability distribution of a variable for which we know the first two moments and its range of variation may be effectively pursued using the mKE. Furthermore, the negentropy of the distribution may be used to quantify how much the mKE solution differs from the MaxEnt one, i.e. to assess how the knowledge of the range of variation influences the assignment of a probability distribution. Our analysis quantitatively supports the conclusions of [4] and rigorously justifies the use of the GUM-S1 recommendation also in the presence of further information on the range of variation of a quantity, namely, provided that its standard uncertainty is sufficiently small compared with the range.

## Acknowledgments

The authors thank M Genovese and I P Degiovanni for useful discussions. This work has been supported by MIUR (FIRB 'LiCHIS' - RBFR10YQ3H), MAE (INQUEST) and the University of Trieste (FRA 2009).

## References

- [1] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP and OIML 2008 *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 [http://www.bipm.org/utis/common/documents/jcgm/JCGM\\_101\\_2008\\_E.pdf](http://www.bipm.org/utis/common/documents/jcgm/JCGM_101_2008_E.pdf)
- [2] Jaynes E T 1957 *Phys. Rev.* **106** 620  
Jaynes E T 1957 *Phys. Rev.* **108** 171
- [3] Wöger W 1987 *IEEE Trans. Instrum. Meas.* **IM-36** 655
- [4] Lira I 2009 *Metrologia* **46** L27
- [5] Kullback S 1959 *Information Theory and Statistics* (New York: Wiley)
- [6] Jaynes E T 1968 *IEEE Trans. Syst. Sci. Cybernet.* **SSC-4** 227
- [7] Olivares S and Paris M G A 2007 *Phys. Rev. A* **76** 042120
- [8] Hyvärinen A 1998 *Advances in Neural Information Processing Systems* vol 10, ed M I Jordan, M J Kearns and S A Solla (Cambridge, MA: MIT Press) p 273  
Hyvärinen A and Oja E 2000 *Neural Networks* **13** 411