Measurement 57 (2014) 102–107

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

Maximum entropy distribution under moments and quantiles constraints

Bartosz Barzdajn [1]

Główny Urząd Miar, Laboratorium Długości Zakładu Długości i Kąta, ul. Elektoralna 2, 00-139 Warszawa, Poland

ARTICLE INFO

Article history:
Received 10 February 2014
Received in revised form 24 June 2014
Accepted 23 July 2014
Available online 1 August 2014

Keywords:
Relative entropy
Maximum entropy principle
Moments
Quantiles
Lagrange multipliers
Uncertainty
Measurement

ABSTRACT

When the results of a measurement are transferred from one stage in the chain of traceability to the next, the information gathered about the measurement is summarised. The summary involves, for example, details about applied measurement methods, environmental conditions, and measurement results including measurement uncertainty. The information about uncertainty usually takes the form of summary statistics such as an estimate, a standard deviation and a coverage interval specified by two quantiles. The information is used to construct a probability distribution for a given property or characteristic of an artefact, which is needed when the artefact is used as a reference in a subsequent stage. But in order to ensure impartiality in the process to establish the probability distribution, a general rule should be applied, for example, the principle of maximum entropy. In this paper, the application of this principle to establish a probability distribution when the mentioned summary statistics are available will be discussed, and its extension to moment constraints to satisfy the requirements of metrology will be introduced.

1. Introduction

Establishing traceability in metrology requires that measurement results can be linked to references through a documented unbroken chain ([1]). Consider the following simple example. At the first stage of a traceability chain, an artefact has been calibrated, for example, a steel rule. As a result of the calibration, the expectation of a parameter characterising the artefact, for example, the deviation from nominal value, is given with the associated standard deviation and two quantiles, but without any information about the probability distribution for the parameter. The artefact is then used in a subsequent stage as a reference standard. At the formulation stage of uncertainty evaluation of the subsequent stage, the information about the parameter is used to assign a probability distribution for the parameter [2]. To ensure impartiality in the process to assign the probability distribution, it is recommended in Section 6 of the GUMS1 [3] to use the principle of maximum entropy. Since the GUM [4] and its supplements are key documents giving harmonised procedures for uncertainty evaluation in metrology, it is important to provide explicit tools for the assignment of a probability distribution for a quantity when the available information is that most commonly encountered in metrology, i.e., an estimate, a standard deviation and quantiles, regardless of whether they ware calculated according to GUM, GUMS1 or to any other guidelines. It is worth emphasising, that the method presented in this paper will be interesting only in cases when quantiles and higher moments do not introduce redundant information. An example might be a situation when we are interpreting the results of Monte Carlo simulations applied to the model for which the conditions of applicability of central limit theorem have

E-mail address: b.barzdajn@imperial.ac.uk
URL: http://www.gum.gov.pl

[1] Currently at Imperial College London, Department of Materials, South Kensington Campus, London SW7 2AZ, UK.

http://dx.doi.org/10.1016/j.measurement.2014.07.012
0263-2241/

not been met. This may be due to non-linearity of the measurement model or the impact of non-Gaussian influence factors, but what is most important here is that often, the distribution cannot be expressed explicitly in a closed form.

The paper is organised as follows. First, the concept of entropy and maximum entropy principle will be introduced including several potential applications. Next, the general framework for evaluating maximum entropy distribution using Lagrange multipliers will be discussed. Further, it will be elaborated on how this technique can be applied to cases when moments of a distribution are known. This topic is very well known and widely described in the literature. However, it will provide background for further considerations. In the following section it will be explained how the mentioned technique can by applied to evaluate maximum entropy distribution specified by moments and quantiles. Those considerations will be backed by two numerical experiments constructed in such a way that they will allow verification of presented results.

2. Entropy and the maximum entropy principle

In this paper, the concept of entropy taken from information theory, the so-called Shannon entropy [5], will be considered. Since the concept of entropy originated in physics, and was then strongly developed for use in information theory, it is not an easy task to provide an intuitive interpretation of the concept appropriate to the topic of uncertainty evaluation. Entropy is often regarded as an expectation of information content. Information content itself is a measure of "informativeness" of a given possible outcome. Therefore, by selecting the maximum entropy distribution we select a least informative distribution. Principle of maximum entropy was introduced by Jaynes in [6]. Since then, this principle found many applications for example [7–10].

For a continuous one-dimensional random variable X described by a probability density function $p(x)$ that has infinite support, the entropy h is defined by the functional

$$h[p(x)] = -K\int_{-\infty}^{\infty} p(x)\log p(x)\mathrm{d}x, \tag{2.1}$$

where K is a constant value. Here, squared brackets are used to embrace the argument of a functional. We will restrict ourselves in this paper to the case described above, although the presented framework can be applied to other cases as well, e.g. for a random variable that has support bounded from both sides. Furthermore it will be assumed that $K = 1$.

Using the principle of maximum entropy, summary statistics may be used to determine the distribution of the random variable. For example, if only available information is that it has finite support, then according to the principle of maximum entropy a rectangular distribution should be assigned to the variable. Many other commonly encountered cases are also well known. For example, GUMS1 [3] treats situations when we have different knowledge about the random variable and its distribution. Furthermore, maximum entropy distributions under moment constraints have been widely studied. Nevertheless, there is still a lack of study about cases when other summary statistics are known, such as the median, quartile range, quantiles of various orders together with classical moments. Such statistics can be the outcome of a calculation of uncertainty when an expectation and expanded uncertainty are insufficient to summarise the results. The development of a general framework will not only provide the means to reconstruct the distribution from such summary information, but also could popularise the application of more robust summary statistics based on ranks, and overall give rise to new important results. However, one have to keep in mind that when applying principle of maximum entropy, the random variable, for which the distribution was evaluated on the basis of limited information, by no means, is identical to the state-of-knowledge distribution. The difference between these two distributions will reflect the loss of information due to an imperfect way of summarising information about uncertainty. This fact can be used to asses applicability of various frameworks.

3. Maximum entropy framework

Recall that entropy, as most summary statistics, is a functional. Maximisation of a functional is an optimisation problem and we may use the method of Lagrange multipliers to obtain its solution. The general form for the Lagrangian, in the case when we wish to determine the argument $p(x)$ that minimises the function $f[p(x)]$ under N constraints $g_i[p(x)] = c_i$, is

$$L[p(x),\boldsymbol{\lambda}] = f[p(x)] - \sum_{i=1}^{N}\lambda_i(g_i[p(x)] - c_i),$$

where λ_i are Lagrange multipliers. To find a maximum of $f[p(x)]$, we seek to minimise the objective function $-f[p(x)]$ by solving the system of equations

$$\frac{\partial L[p(x),\boldsymbol{\lambda}]}{\partial p(x)} = 0, \tag{3.1}$$

$$\frac{\partial L[p(x),\boldsymbol{\lambda}]}{\partial \lambda_i} = 0, \tag{3.2}$$

where Eq. (3.1) defines a stationary point and Eq. (3.2) defines arguments for which the constraints are satisfied. By solving the above equations with $f[p(x)]$ replaced by the entropy h and the constraints expressing the given information about the distribution, we can determine the maximum entropy distribution. Furthermore, by exploring the properties of those equations we may also specify the family of distributions to which the maximum entropy distribution belongs.

4. Maximum entropy distribution determined by moments

It is assumed that if the moments of the distribution are known, it is possible to specify the family of distributions to which the corresponding maximum entropy distribution