

## Evaluation of Measurement Uncertainty Using Bayesian Inference

Gaetano Iuculano<sup>1</sup>, Gabriella Pellegrini<sup>2</sup>, Andrea Zanobini<sup>1</sup>

<sup>1</sup> Department of Electronics and Telecommunication, University of Florence, Via di S. Marta, 3 – Florence – I 50139 phone: +39 055 4796276, fax +39 055 494569,  
e-mail: [iuculano@ingfi1.ing.unifi.it](mailto:iuculano@ingfi1.ing.unifi.it), [willis@ingfi1.ing.unifi.it](mailto:willis@ingfi1.ing.unifi.it)

<sup>2</sup> Department of Applied Mathematics, Engineering Faculty, University of Florence, Via di S. Marta, 3–Florence–I 50139 phone: +39 055 4796248, fax: +39 055 471787,  
e-mail: [gualtieri@dma.unifi.it](mailto:gualtieri@dma.unifi.it)

**Abstract**-The Bayesian inference provide a natural and consistent way to make best use of all relevant historical information which characterizes the measurand in a measurement process. In this paper the use of Bayesian method is described to include prior information in the evaluation of the measurement uncertainty. The proposed approach gives an implementation on the evaluation of the confidence region with respect the usual techniques described in the ISO *Guide to the Expression of Uncertainty in Measurement* (GUM) [1]. Experimental model is studied and numerical results are reported to assess the validity of the proposed formulae.

### I. Introduction

In the orthodox Neymann-Pearson approach we are given a set of observed data from which we are to decide whether some hypothesis about the real world (in particular about parameters of primary interest) is true (or, put more cautiously, whether to act as if it were true). The first thing orthodox statistics does is to imbed the observed data set in a "sample space" which is an imaginary collection containing other data sets that one thinks might have been observed but were not. Then one introduces the "sampling distribution" defined by the probability  $p(D_i/H)$  that the generic data set  $D_i$  would be observed if the hypothesis H were true; this probability is interpreted as the theoretic abstraction of the frequency with which the data set  $D_i$  would be observed in the long run if the measurement were made repeatedly with H constantly true. When one asserts the long run results of an arbitrarily long sequence of measurements that have not been performed it would appear that he is on a rather large hidden fund of prior knowledge about the analysed phenomenon.

If we are not told what that knowledge is and how it was obtained, we might be excused for doubting its existence. Further, in the long run, how often would it lead us to a correct conclusion, or how large would the average error of estimation be?

Finally when making inferences about a set of parameters of interest we must also take into account of nuisance or incidental parameters. Except when suitable sufficient statistics exist for all the parameters difficulties arise in dealing with nuisance parameters by orthodox approach.

Sufficient statistics play a vital role in sampling theory. For, if inferences about fixed parameters are to be made using the distributional properties of statistics which are functions of the data, then, to avoid inefficiency due to the leakage of information, it is essential that a small minimally sufficient set of statistics be available containing all the information about the parameters. By happy mathematical accident such set of sufficient statistics do exist for a number of important distributions and, in particular, for the Normal distribution. However, serious difficulties can accompany the exploration of less restricted models which may be motivated by scientific interest, but for which no convenient set of sufficient statistics happens to be available. Because Bayesian analysis is concerned with the distribution of parameters, given known (fixed) data, it does not suffer from this artificial constraint. It does not matter whether or not the distribution of interest happens to have the special form which yields sufficient statistics.

Furthermore, even when sufficient statistics are available, examples can occur in sampling theory where there is difficulty in eliminating nuisance parameters.

Using sampling theory it is difficult to take account of constraints which occur in the specification of the inference-parameter space.

By contrast, in Bayesian analysis, inferences are based on probabilities associated with different values of parameters which could have given rise to the fixed set of data which has actually occurred. In calculating such probabilities we must make assumptions about prior distributions, but we are not dependent upon the existence of sufficient statistics, and no difficulty occurs in taking account of parameters constraints. In other terms the Bayesian approach sets us free from the joke of sufficiency. In the pure minimum cross-entropy (noiseless Bayes) methods, our reasoning format is almost the opposite of sampling theory. Instead of considering the class of all data sets  $\{D_1, \dots, D_N\}$  consistent with a hypothesis  $H$ , we consider the class of all hypotheses  $\{H_1, \dots, H_N\}$  consistent with the one data set  $D_{obs}$  that was actually observed. In addition we use prior information  $I$  that represents our knowledge (from physical law) of the possible ways in which Nature could have generated the various  $H_i$ . Out of the class  $C$  of hypotheses consistent with our data, we pick the one favoured by prior information, which means, usually, having the minimum cross-entropy.

Each successive piece of data that one obtains is a new constraint that restricts the possibilities permitted by our previous information.

At any stage an honest description of what we know must take into account (that is, assign non zero probability to every possibility that is not ruled out by our prior information and data.

It is not possible to extract all that detail from the data alone. The minimum cross-entropy gives us more information only because we have put more information into it. Often prior information is available about nuisance parameters, but orthodox ideology does not recognize it because it does not consist of frequencies in any random experiment. In the problems where pure minimum cross-entropy is appropriate we are concerned, not with frequencies in any random experiment, but with rational thinking in a situation where our information is incomplete. In orthodox thinking a frequency is considered "objective" and therefore respectable, while a mere state of knowledge is subjective and unscientific.

But in the real-world problems of measurement faced by every experimenter it is evidently his state of knowledge that determines the quality of the decisions he is able to make in situations that will never be repeated; and it is the frequencies that are pigments of the imaginations.

Interestingly, the sampling distribution that orthodox theory does allow us to use is nothing more than a way of describing our prior knowledge about the "noise". Thus, orthodox thinking is in the curious position of holding it decent to use prior information about noise, but indecent to use prior information about the signal of interest.

If we have both noise and prior information neither of sampling theory and MINCENT principle is adequate. But both are only limiting cases of a more general method that applies in all cases. Adding prior information capabilities to orthodox methods, or noise capabilities to MINCENT, we arrive in either case at the Bayes method, which is actually simpler conceptually and older historically than either of these special cases.

The proposed evaluation of measurement uncertainty is based completely on Bayesian analysis and on the principle of minimum cross-entropy. The theory is universally applicable to most measurement tasks including complex non linear adjustment and, in particular, in case where the well-established least-squares or maximum likelihood techniques fail.

## II. General remarks

Assume that in a measurement process all the sources of uncertainty are characterized by probability distribution functions, the form of which is assumed to either be known from experience or unknown and so conjectured.

In Bayesian approach prior to obtaining the measurement results the experimenter considers his degrees of belief and individual experience for the possible models and represents them in the form of initial probabilities (called "prior"). Once the results are obtained Bayes' theorem enables the experimenter to calculate a new set of probabilities which represents revised degrees of belief in the possible models, taking into account the new information provided by the measurements results. The results are the linkage between "posterior" and "prior". In brief, results do not create beliefs, rather they modify existing ones. The revised probabilities are referred as "posterior".

For the Bayesian experimenter it is natural to regard subjective beliefs as primary and "objective consensus" as a special case arising when the amount of the data available and the explicit statement of the data structure is dominant enough to overwhelm every one's prior beliefs forcing them into the same posterior shape and location. The Bayesian inference is appealing for its sequential nature alternation between conjecture and experiment, since it allows us to continually update information

about the same measurand as more measurements are undertaken in different occasions. In fact today's posterior for the first measurement occasion can play the role of tomorrow's prior for the second one. Let us to come back to the measurement process whose involved quantities are summarized in two principal sets represented by row vectors: the input quantities  $\underline{X}$  in number of  $n$ , and the output quantities  $\underline{Y}$ , in number of  $m$ . Let  $(\underline{x}, \underline{y})$  the actual realizations of  $(\underline{X}, \underline{Y})$  in a particular occasion, they represent a state of the measurement process in that occasion. The process has a set  $D$  of possible states  $(\underline{x}, \underline{y}) \in D$  which identify the joint domain of the random variables  $(\underline{X}, \underline{Y})$ .

The link between the input and output quantities is expressible through a functional relationships of type:  $Y_i = g_i(X_1, \dots, X_n)$ ;  $i=1, \dots, m$  with  $m \leq n$ .

The mutual behaviour between the input quantities  $\underline{X}$  and  $\underline{Y}$  is statistically drawn by the joint probability  $f(\underline{x}, \underline{y})$  which can be written as:

$$f(\underline{x}, \underline{y}) = f(\underline{x})f(\underline{y}|\underline{x}) \quad (1)$$

where  $f(\underline{x})$  is the marginal joint density of the input quantities  $\underline{X}$  and  $f(\underline{y}|\underline{x})$  is the conditional joint density of the output quantities  $\underline{Y}$ , given  $\underline{X} = \underline{x}$ .

In the Bayesian approach we have another important linkage between  $f(\underline{x})$  and  $f(\underline{x}, \underline{y})$ , that is:

$$f(\underline{x}|\underline{y}) = c f(\underline{x})f(\underline{y}|\underline{x}) \quad (2)$$

where  $f(\underline{x}|\underline{y})$  is the posterior joint density of the input quantities  $\underline{X}$ , whose  $f(\underline{x})$  is interpreted as the prior joint density and  $f(\underline{y}|\underline{x}) = \ell(\underline{x}, \underline{y})$ , regarded as a function of  $\underline{x}$  for prefixed output values  $\underline{y}$ , represents the well-known likelihood;  $c$  is a "normalizing" constant necessary to ensure that the posterior joint density integrates, with respect to  $\underline{x}$ , to one.

In practical situations the measurement process represents a controlled learning process in which various aspects on uncertainty analysis are illuminated as the study proceeds in an up-to-date alternation between conjecture and experiment carried out via experimental design and data analysis. A measurement process is performed if information supplied by it is likely to be considerably more accurate, stable and reliable than the pool of information already available. The substantial amount of information, got with respect to the conditions prior to the result after the measurement process is performed, can be connected to the "Kullback's principle of minimum cross-entropy". This, as it is known, is a correct method of inductive inference when no sufficient knowledge about the statistical distributions of the involved random variables is available before the measurement process is carried out except for the permitted ranges, the essential model relationships and some constraints, gained in past experience, valuable usually in terms of expectations of given functions or bounds on them.

### III The Entropy optimisation principles and the Bayesian Inference

The evaluation of the measurement uncertainty is analysed by applying the entropy optimisation principles to the joint density function  $f(\underline{x}, \underline{y})$  expressed by (1), taking into account the Bayesian inference, given by (2), which discriminates between the prior densities of the input quantities  $\underline{X}$  and the conditional densities of the output quantities given the input ones.

In order to produce the best estimation for  $f(\underline{x}, \underline{y})$ , subjected to given constraints, we introduce the joint cross entropy functional:

$$S = \int_D f(\underline{x}, \underline{y}) \ln \frac{f(\underline{x}, \underline{y})}{f_0(\underline{x}, \underline{y})} d\underline{x} d\underline{y} = E \left\{ \ln \frac{f(\underline{X}, \underline{Y})}{f_0(\underline{X}, \underline{Y})} \right\} \quad (3)$$

where we have passed to the compact notations:  $d\underline{x} = dx_1 \dots dx_n$ ,  $d\underline{y} = dy_1 \dots dy_m$  and where  $f_0(\underline{x}, \underline{y})$ , which a priori must be known, is defined by Jaynes [1] an "invariant measure" function.

According to the principle of minimum cross entropy introduced by Kulback under the name of the principle of minimum discrimination information, we research the joint probability function that minimize the functional  $S$  using the Euler-Lagrange method by the calculus of variation, taking into account the given constraints.

If  $f_0(\underline{x}, \underline{y})$  is assumed to be constant, it is equivalent to apply the principle of maximum entropy to the functional:

$$H = - \int_D f(x, y) \ln f(x, y) dx dy$$

### A priori distribution: Exponential model

Before to apply the Bayesian approach, we have to find the probability distribution function (the prior density),  $f(x)$ , of a quantity X whose previous data are taken from manufacturer's specifications or handbooks. We suppose that the value  $\mu_0$  (the best available estimate) is given and it is stated that "the smallest possible value is  $a_- = \mu_0 - b_-$  and the largest possible value is  $a_+ = \mu_0 + b_+$ ", with  $b_-, b_+ > 0$ . The known upper and lower bounds,  $a_-$  and  $a_+$  for the quantity X are not symmetric with respect to its best known estimate  $\mu_0$ , assumed to be the expectation of X, that is:

$$E\{X\} = \int_{a_-}^{a_+} x f(x) dx = \mu_0 \quad (4)$$

We consider the cross-entropy of the probability distribution function  $f(x)$  with respect to a given prior one  $f_0(x)$ :

$$S = \int_{a_-}^{a_+} f(x) \ln \frac{f(x)}{f_0(x)} dx \quad (5)$$

and according to the Kulback's cross-entropy principle, the best estimate of  $f(x)$  is the distribution minimizing S.

We suppose  $f_0(x)$  constant; it is really uniform if the basic available information concerns only the bounds without involving the most plausible value  $\mu_0$ . Consequently, instead of (5), we must maximize the well-known Jaynes entropy:

$$H = - \int_{a_-}^{a_+} f(x) \ln f(x) dx \quad (6)$$

subject to the constraint (4) and to the normalizing condition  $\int_{a_-}^{a_+} f(x) dx = 1$ , using Euler-Lagrange method by the calculus of variation.

After simple manipulation we obtain:

$$f(x) = c e^{-\lambda(x-\mu_0)} \quad a_- \leq x \leq a_+ \quad (7)$$

that is the a priori distribution of X is truncated exponential with expectation  $\mu_0$ .

The constant c is such that:

$$c \int_{a_-}^{a_+} e^{-\lambda x} dx = 1 \quad \text{i.e.} \quad c = \frac{\lambda}{e^{\lambda b_-} - e^{-\lambda b_+}} \quad (8)$$

The Lagrange multiplier  $\lambda$  is determined through the constraints (4). In fact we have:

$$c \int_{a_-}^{a_+} (x - \mu_0) e^{-\lambda(x-\mu_0)} dx = c \int_{a_-}^{a_+} x e^{-\lambda x} dx = 0 \quad (9)$$

For  $b_- \neq b_+$  using the integral's property and taking into account (8) we deduce:

$$\lambda = \frac{e^{\lambda b_-} - e^{-\lambda b_+}}{b_- e^{\lambda b_-} + b_+ e^{-\lambda b_+}} \quad (10)$$

Substituting (10) into (8) we obtain:

$$c = [b_- e^{\lambda b_-} + b_+ e^{-\lambda b_+}]^{-1} \quad (11)$$

The variance of X may be evaluated by:

$$\text{Var}\{X\} = c \int_{a_-}^{a_+} (x - \mu_0)^2 e^{-\lambda(x-\mu_0)} dx = c \int_{a_-}^{a_+} x^2 e^{-\lambda x} dx = \frac{b_-^2 e^{\lambda b_-} - b_+^2 e^{-\lambda b_+}}{e^{\lambda b_-} - e^{-\lambda b_+}} \quad (12)$$

Finally, by (10) and (12) we obtain:

$$\text{Var}\{X\} = b_+ b_- - \frac{b_+ - b_-}{\lambda} \quad (13)$$

according to the GUM [1] note 2 in paragraph 4.3.8.

We will work through the example 1 in paragraph 4.3.7 taken from the GUM. The value of the

coefficient of linear thermal expansion of pure copper at 20<sup>o</sup>C  $\alpha_{20}(CU)$ , as  $16.52 \times 10^{-6} \text{ } ^\circ\text{C}^{-1}$ , assuming that “the error in this value should not exceed  $0.40 \times 10^{-6} \text{ } ^\circ\text{C}^{-1}$ , it is reasonable to assume that the value of  $\alpha_{20}(CU)$  lies with equal probability in the interval  $16.12 \times 10^{-6} \text{ } ^\circ\text{C}^{-1}$  to  $16.92 \times 10^{-6} \text{ } ^\circ\text{C}^{-1}$ .

So that if we assume  $a_- = 16.4 \times 10^{-6} \text{ } ^\circ\text{C}^{-1}$  and  $a_+ = 16.92 \times 10^{-6} \text{ } ^\circ\text{C}^{-1}$ , it is interesting to study the variability of the standard uncertainty and its relative value with respect all the possible values of  $\alpha_{20}(CU)$  inside the given interval  $[a_-, a_+]$ . Computational results are reported in table 1:

In column (1) is given the value  $\mu_0$  attributed to  $\alpha_{20}(CU)$  inside the interval  $[a_-, a_+]$ , in column (2) the evaluation of the scale parameter  $\lambda$  satisfying eq. (10), in column (3) the standard uncertainty  $\sigma = \sqrt{\text{Var}\{X\}}$  given by eq. (12) and finally in column (4) the coefficient  $c = \sigma/\mu_0$  representing the relative variability of the standard uncertainty with respect the  $\mu_0$  value:

$\mu_0$	$\lambda$	$\sigma = \sqrt{\text{Var}\{X\}}$	$c = \sigma/\mu_0$
$16.50 \times 10^{-6}$	$9.6683 \times 10^6$	$9.4351 \times 10^{-8}$	$5.7182 \times 10^{-3}$
$16.52 \times 10^{-6}$	$7.7177 \times 10^6$	$1.0826 \times 10^{-7}$	$6.5533 \times 10^{-3}$
$16.62 \times 10^{-6}$	$1.8009 \times 10^6$	$1.4689 \times 10^{-7}$	$8.8381 \times 10^{-3}$
$16.82 \times 10^{-6}$	$-9.6683 \times 10^6$	$9.4351 \times 10^{-8}$	$5.6095 \times 10^{-3}$

**Table 1.**

### A posterior distribution by Bayesian approach

According to Bayes' theorem for continuous variables, the posterior distribution function may be written as:

$$f_x(x|y_0) = ce^{-x}l(x, y_0) \quad y_0 \geq 0 \quad (14)$$

where  $y_0$  is the value of the output  $Y \in [0, +\infty)$  obtained in a particular occasion and  $l(x, y) = f_y(y|x)$  for a generic  $y$  is the well known likelihood.

By applying the maximum entropy principle to the likelihood  $l(x, y)$  subjected to the constraints:

$$\int_0^{+\infty} l(x, y)dy = 1 \quad \text{and} \quad \int_0^{+\infty} yl(x, y)dy = E\{Y|x\} = x$$

we obtain

$$l(x, y) = \frac{1}{x} e^{-\frac{y}{x}} \quad (15)$$

substituting (15) into (14) we have:

$$f_x(x|y_0) = c \frac{1}{x} e^{-\left(\lambda x + \frac{y_0}{x}\right)} \quad (16)$$

now the constant  $c$  may be evaluated by imposing the constraint  $\int_0^{+\infty} f_x(x/y_0)dy = 1$ .

So that we can write:

$$c \int_0^{+\infty} \frac{1}{x} e^{-\left(\lambda x + \frac{y_0}{x}\right)} dx = 1$$

and consequently

$$c = \frac{1}{2K_0\left(2\sqrt{\lambda y_0}\right)} \quad (17)$$

being  $K_\nu(z)$  the modified Bessel functions.

Expectation and variance for the posterior distribution are evaluated respectively by:

$$E\{X|y_0\} = c \int_0^{+\infty} x e^{-\left(\lambda x + \frac{y_0}{x}\right)} dx = \sqrt{\frac{y_0}{\lambda}} \frac{K_1\left(2\sqrt{\lambda y_0}\right)}{K_0\left(2\sqrt{\lambda y_0}\right)} \quad (18)$$

and

$$\text{Var}\{X\|y_0\} = E\{X^2\|y_0\} - E^2\{X\|y_0\} = \frac{1}{\lambda} \left( y_0 + \sqrt{\frac{y_0}{\lambda} \frac{K_1(2\sqrt{\lambda y_0})}{K_0(2\sqrt{\lambda y_0})}} \right) - \frac{y_0}{\lambda} \left( \frac{K_1(2\sqrt{\lambda y_0})}{K_0(2\sqrt{\lambda y_0})} \right)^2 \quad (19)$$

By referring to the interval  $a_- = 0, a_+ = +\infty$ , if the prior expectation value is assumed to be  $\mu_0 = 16.52 \times 10^{-6}$ , we have the prior distribution scale parameter  $\lambda = \frac{1}{\mu_0} = 60533$ .

In Table 2 are reported the posterior expectation value  $\bar{\mu}_0$ , the posterior standard uncertainty  $\bar{\sigma} = \sqrt{\text{Var}\{X\|y_0\}}$  and its relative value with respect  $\bar{\mu}_0$  coming from equations (18) and (19) for different values of  $y_0$ , representing the output  $Y \in [0, +\infty)$  obtained in a particular occasion:

$y_0$	$\bar{\mu}_0 = E\{X\ y_0\}$	$\bar{\sigma} = \sqrt{\text{Var}\{X\ y_0\}}$	$\bar{\sigma}/\bar{\mu}_0$
$16.38 \times 10^{-6}$	$20.216 \times 10^{-6}$	$1.3996 \times 10^{-5}$	0.69232
$16.58 \times 10^{-6}$	$20.318 \times 10^{-6}$	$1.4026 \times 10^{-5}$	0.69036
$16.62 \times 10^{-6}$	$20.338 \times 10^{-6}$	$1.4033 \times 10^{-5}$	0.68997
$16.72 \times 10^{-6}$	$20.388 \times 10^{-6}$	$1.4048 \times 10^{-5}$	0.68901
$16.82 \times 10^{-6}$	$20.439 \times 10^{-6}$	$1.4063 \times 10^{-5}$	0.68805
$16.92 \times 10^{-6}$	$20.489 \times 10^{-6}$	$1.4078 \times 10^{-5}$	0.68710

Table 2

### Conclusion

Rigorous formulae for the evaluation of measurement uncertainty have been developed. The formulae were established using entropy optimisation principles and Bayesian inference. Examples for exponential model were presented.

### References

- [1] *Guide to the Expression of Uncertainty in Measurement, first edition (1993)*, corrected and reprinted (1995), International Organization for Standardization (Geneva, Switzerland).
- [2] *International Standard ISO 3534-1 Statistics-Vocabulary and Symbols-Part I: Probability and General Statistical Terms, first edition (1993)*, International Organization for Standardization (Geneva, Switzerland).
- [3] Kapur J.N., *Maximum-Entropy Models in Science and Engineering*, John Wiley & Sons
- [4] Weise K., Wöger W., "A Bayesian theory of measurement uncertainty", *Meas. Science Tech.*, 1993, **4**, 1-11
- [5] Lira I., Kyriazis G., "Bayesian inference from measurement information", *Metrologia*, 1999, **36**, 163-169.