XVII IMEKO World Congress
Metrology in the 3rd Millennium
June 22-27,2003,Dubrovnik, Croatia

THE EVALUATION OF MEASUREMENT UNCERTAINTY AND THE PRINCIPLE OF MINIMUM JOINT CROSS-ENTROPY

G. Iuculano[(1)], A. Lazzari[(2)], G. Pellegrini[(3)], and A. Zanobini[(1)]

[(1)] Engineering Faculty, Department of Electronics and Telecommunication, University of Florence-Italy
[(2)] Engineering Faculty, Department of Electrical Engineering, University of Bologna, Bologna - Italy
[(3)] Engineering Faculty, Department of Applied Mathematics, University of Florence, Florence-Italy

Abstract – A measurement process represents a controlled learning process in which various aspects on uncertainty analysis are investigated.

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.

In this paper the authors pointed out the connection between the evaluation of the uncertainty in a repeated measurements process and the "Kullback's principle of minimum cross-entropy".

Keywords: Uncertainty, Minimum Entropy Principle.

1. THE PRINCIPLE OF MINIMUM JOINT CROSS-ENTROPY

We consider a general measurement process which treats multiple measurands and yields simultaneously multiple results or estimates.

A measurement process has imperfections that give rise to uncertainty in each measurement result. The assessment of uncertainties associated to the results is given by statistical tools only if all the relevant quantities involved in the process are interpreted or regarded as random variables either they are random in nature or purposely randomized.

In other terms all the sources of uncertainty are characterized by probability distribution functions, the form of which is assumed to either be known from measurements or unknown and so conjectured.

We classify all the involved quantities into two principal sets represented by row vectors:

1) output quantities $\underline{Y}$, in number of m

2) input quantities $\underline{X}$ which comprise the rest of quantities, in number of n.

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 $\lfloor(\underline{x},\underline{y})\in D\rfloor$ which identify the joint domain of the random variables $(\underline{X},\underline{Y})$.

Further, often, it is possible in a measurement process to individuate mathematical and/or empirical models which link input and output quantities through functional relationships of type:

$$Y_i = g_i(X_1,\cdots,X_n);\ \text{i=1, ..., m} \qquad \text{with } m \le n \tag{1}$$

In practical situations the transformation defined by (1) is differentiable and invertible.

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:

$$f(\underline{x},\underline{y}) = f(\underline{x})f(\underline{y}\mid\underline{x}) \tag{2}$$

where $f(\underline{x})$ is the marginal joint density of the input quantities $\underline{X}$ and $f(\underline{y}\mid\underline{x})$ 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:

$$f_{\underline{x}}(\underline{x},\underline{y}) = c\, f(\underline{x})f(\underline{y}\mid\underline{x}) \tag{3}$$

where $f_{\underline{x}}(\underline{x},\underline{y})$ is the posterior joint density of the input quantities $\underline{X}$, whose $f(\underline{x})$ is interpreted as the prior joint density and $f(\underline{y}\mid\underline{x}) = \ell(\underline{x},\underline{y})$, 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.

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.

1.1 The evaluation of measurement uncertainty based on the principle of minimum joint cross-entropy (MINCENT)

The authors propose to apply the Kullback's principle to the joint density function $f\left(\underline{x},\underline{y}\right)$ expressed by (2) but taking into account the Bayesian relation (3) which discriminates between the prior densities of the input quantities $\underline{X}$ and the conditional densities of the output quantities given the input ones.

To this end we introduce the joint cross entropy in the following manner:

$$S=\int_D f\left(\underline{x},\underline{y}\right)\ln\frac{f\left(\underline{x},\underline{y}\right)}{f_0\left(\underline{x},\underline{y}\right)}\underline{dx}\underline{dy}=E\left\{\ln\frac{f\left(\underline{X},\underline{Y}\right)}{f_0\left(\underline{X},\underline{Y}\right)}\right\} \tag{4}$$

where $\underline{dx}=dx_1\cdots dx_n$, $\underline{dy}=dy_1\cdots dy_m$ and where $f_0\left(\underline{x},\underline{y}\right)$, which a priori must be known, is defined as an "invariant measure" function.

In fact, since $f\left(\underline{x},\underline{y}\right)$ and $f_0\left(\underline{x},\underline{y}\right)$ transform in the same way under a change of variables, S remains invariant to any coordinate transformation.

It can be shown that $S\geq 0$; the equality sign will hold if $f\left(\underline{x},\underline{y}\right) = f_0\left(\underline{x},\underline{y}\right)$ almost everywhere (except possibly on a set of measure zero).

The joint cross entropy is an adequate information measure since, in the space of probability distributions, measures some kind of information amount necessary to change a prior poor knowledge on the measurement process, represented by $f_0\left(\underline{x},\underline{y}\right)$, into a more circumstanciated posterior joint distribution described by $f\left(\underline{x},\underline{y}\right)$.

It also appears that, in some sense, the larger the divergence between $f(\cdot)$ and $f_0(\cdot)$, the larger will be the value of S; this justifies our calling S also measure of directed divergence.

Further one can prove, up to a constant factor, that the joint density $f\left(\underline{x},\underline{y}\right)$ which minimizes the cross entropy given by (4) is favored over other possible densities since minimizing that entropy, subject to arbitrary constraints, leads to satisfy axioms that are accepted as requirements for an efficient information measure.

These axioms are all based on one fundamental principle: if a problem can be solved in more than one way, the decisions should be consistent.

2. THE REPEATED MEASUREMENTS

We repeat n times the measurements process on the same measurand according to the "repeatability conditions" estabilished by the "Guide to the expression of uncertainty in measurement" (ISO 1993).

We may conveniently regard any set of measurement results $\underline{y}=\left(y_1,\cdots,y_n\right)$ as the n-dimensional realization of an induced random vector $\underline{Y}=\left(Y_1,\cdots,Y_n\right)$ which we can call output vector.

Let us now assign the random variable X to the measurand and express its occasional realization as x which we suppose constant during all the replications. It is correct to introduce the conditioned random vector $\underline{Y}|X=x$ constituted by n conditioned random variables $\left\{Y_i|X=x;i=1,\cdots,n\right\}$ which are conditionally independent and identically distributed, assuming X=x.

We suppose that the conditional statistic parameters are given by:

$$\begin{cases} E\left\{Y_i|X=x\right\}=x \\ E\left\{\left(Y_i-x\right)^2|X=x\right\}=\sigma^2 \\ E\left\{\left(Y_i-x\right)\left(Y_j-x\right)|X=x\right\}=0 \end{cases} \tag{5}$$

for i≠j and i,j=1, ...,m.

Obviously, from (5) we can verify that:

$$\begin{cases} E\left\{Y_i\right\}=E\left\{X\right\} \\ E\left\{\left(Y_i-x\right)^2\right\}=\sigma^2 \\ E\left\{\left(Y_i-x\right)\left(Y_j-x\right)\right\}=0 \end{cases} \tag{6}$$

We consider the unknown joint probability density of the measurand X and the random vector $\underline{Y}$, that is:

$$f\left(x,\underline{y}\right)= f(x)f\left(\underline{y}|x\right) \tag{7}$$

We suppose that all the previous knowledge and experience is memorized into a prior joint density as follows:

$$f_0\left(x,\underline{y}\right)= f_0(x)f_0\left(\underline{y}|x\right) \tag{8}$$

We introduce the cross-entropy or the directed divergence (or the discrimination function) between $f_0\left(x,\underline{y}\right)$ and $f\left(x,\underline{y}\right)$ defined by:

$$S = \int\int_{-\infty}^{+\infty} f(x) f(\underline{y}|x) \ln \frac{f(x) f(\underline{y}|x)}{f_0(x) f_0(\underline{y}|x)} dx d\underline{y} = S_X + E\{S_{\underline{Y}}(X)\} \quad (9)$$

with the compact notation $\underline{dy} = dy_1 dy_2 \cdots dy_n$ and where:

$$S_X = \int_{-\infty}^{+\infty} f(x) \ln \frac{f(x)}{f_0(x)} dx \quad (10)$$

is the divergence corresponding to the measurand and $E\{S_{\underline{Y}}(X)\}$ is the expectation of conditional the output divergence, that is:

$$E\{S_{\underline{Y}}(X)\} = \int_{-\infty}^{+\infty} S_{\underline{Y}}(x) f(x) dx \quad (11)$$

and

$$S_{\underline{Y}}(x) = \int_{-\infty}^{+\infty} f(\underline{y}|x) \ln \frac{f(\underline{y}|x)}{f_0(\underline{y}|x)} d\underline{y} \quad (12)$$

In deriving the first term at the member of (9) we have taken into account the normalizing condition:

$$\int_{-\infty}^{+\infty} f(\underline{y}|x) d\underline{y} = 1 \quad (13)$$

Due to the conditional indipendence of the output quantities $Y_1, \cdots, Y_n$ we can write:

$$\begin{cases} f(\underline{y}|x) = \prod_{i=1}^{n} f(y_i|x) \\ f_0(\underline{y}|x) = \prod_{i=1}^{n} f_0(y_i|x) \end{cases} \quad (14)$$

$f(y_i|x)$ and $f_0(y_i|x)$ being the common marginal conditional densities at the generic argument y_i, assuming X=x.

Obviously the first one of (14) makes inessential the third constraint (5).

By substituting (14) into (12) we deduce:

$$S_{\underline{Y}}(x) = n \int_{-\infty}^{+\infty} f(y|x) \ln \frac{f(y|x)}{f_0(y|x)} dy \quad (15)$$

where we have again used the normalizing condition but for the marginal density, that is:

$$\int_{-\infty}^{+\infty} f(y|x) dy = 1 \quad (16)$$

Now, by referring to (15), we minimize the marginal directed divergence:

$$S(x) = \int_{-\infty}^{+\infty} f(y|x) \ln \frac{f(y|x)}{f_0(y|x)} dy \quad (17)$$

subject to the first of constraints (5) and to (16).

It is known that:

$$S(x) \geq 0 \quad (18)$$

and the equality sign will hold if $f(y|x) = f_0(y|x)$ almost everywhere, that is, except on a set of measure zero.

The divergence S(x) given by (17) can be considered as a functional of $f(y|x)$ in the distribution space and for brevity we can write: S(x)≡I(f).

It can be shown the convexity of the directed divergence functional.

In fact by considering in the distribution space two distinct densities $f_1(y|x)$ and $f_2(y|x)$ we get:

$$I[\tau f_1 + (1-\tau) f_2] \leq \tau I(f_1) + (1-\tau) I(f_2) with 0 \leq \tau \leq 1 \quad (19)$$

Consequently the functional has an unique minimum with respect to $f(y|x)$.

Using Lagrange's method of undetermined multipliers the density which minimizes S(x) is given by:

$$f(y|x) = f_0(y|x) e^{-\left[\lambda_0 + \lambda_1 y + \lambda_2 (y-x)^2\right]} \quad (20)$$

where the Lagrange multipliers $\lambda_0, \lambda_1, \lambda_2$, are determined by using the given constraints.

In order to develope a concrete function for $f(y|x)$ it is necessary to define the prior density $f_0(y|x)$.

For simplicity we impose:

$$f_0(y|x) \equiv \text{constant} = k \quad (21)$$

Taking (21) for an infinite interval implies assuming an "improper" prior distribution. In this case minimization of directed divergence (17) is equivalent to maximization on Jaynes entropy defined by:

$$H = -\int_{-\infty}^{+\infty} f(y|x) \ln f(y|x) dy \quad (22)$$

The minimizing distribution is however still proper since it has to satisfy the given constraints.
If we adopt (21) by referring to (20) and to the given constraints, after simple manipulation we obtain:

$$f(y|x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-x)^2}{2\sigma^2}} \quad (23)$$

By substituting (23) into the first of (14) we obtain:

$$f(\underline{y}|x) = \frac{1}{2\pi^{\frac{n}{2}} \sigma^n} e^{-\frac{\chi^2}{2}} \quad (24)$$

where χ^2 is the realization of:

$$\chi^2 = \frac{\sum_{i=1}^{n} (Y_i - x)^2}{\sigma^2} \quad (25)$$

which is a chi-squared random variable with n degrees of freedom.

Now we consider the Bayesian inference and introduce the posterior density of the measurand that is:

$$f_X\left(x|\underline{y}_0\right)=c\,f(x)\,f\left(\underline{y}_0|x\right) \tag{26}$$

where $\underline{y}_0=\left(y_{10},\cdots,y_{n0}\right)$ is the set of effective results at the particular occasion when X=x and c is a convenient normalizing constant.

By introducing (24) we obtain:

$$f_X\left(x|\underline{y}_0\right)=c'\,f(x)\,e^{-\frac{\sum_{i=1}^{n}(y_{i0}-x)^2}{2\sigma^2}} \tag{27}$$

If also $f(x)$ is supposed constant we have:

$$f_X\left(x|\underline{y}_0\right)=c''e^{-\frac{\sum_{i=1}^{n}(y_{i0}-x)^2}{2\sigma^2}}=c'''e^{-\frac{nx^2-2\sum_{i=1}^{n}y_{i0}\,x}{2\sigma^2}} \tag{28}$$

where c''' is the up-to-date final normalizing constant.

Due to the normalizing condition, it is easy to verify that:

$$\int_{-\infty}^{+\infty} f_X\left(x|\underline{y}_0\right)dx=c'''\int_{-\infty}^{+\infty} e^{-\frac{nx^2-2\sum_{i=1}^{n}y_{i0}\,x}{2\sigma^2}}\,dx=1 \tag{29}$$

The posterior expectation of the measurand is given by:

$$E\left\{X|\underline{y}_0\right\}=c'''\int_{-\infty}^{+\infty} x e^{-\frac{nx^2-2\sum_{i=1}^{n}y_{i0}\,x}{2\sigma_x^2}}=\frac{\sum_{i=1}^{n}y_{i0}}{n}=\overline{y}_0 \tag{30}$$

and the posterior variance is given by:

$$Var\left\{X|\underline{y}_0\right\}=\frac{\sigma^2}{n} \tag{31}$$

as would be expected.

If σ^2 is unknown an additional assumption is needed. We can use the so-called conformity property, that is:

$$E\left\{\frac{\sum_{i=1}^{n}\left(X-y_{i0}\right)^2}{\sigma^2}|\underline{y}_0\right\}=n \tag{32}$$

or equivalently, by imposing

$$\left(X-y_{i0}\right)=\left(X-\overline{y}_0\right)-\left(y_{i0}-\overline{y}_0\right)$$

and using (30),we have:

$$\frac{nVar\left\{X|\underline{y}_0\right\}}{\sigma^2}+\sum_{i=1}^{n}\frac{\left(y_{i0}-\overline{y}_0\right)^2}{\sigma^2}=n \tag{33}$$

which, recalling (31), finally yields the well known formula:

$$\sigma^2=\sum_{i=1}^{n}\frac{\left(y_{i0}-\overline{y}_0\right)^2}{n-1} \tag{34}$$

3. CONCLUSIONS

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.

REFERENCES

[1] *Guide to the Expression of Uncertainty in Measurement,* first edition, 1993, corrected and reprinted 1995, International Organisation for Standardisation (Geneva, Switzerland).

[2] International Standard ISO 3534-1 *Statistics-Vocabulary and Symbols-Part I: Probability and General Statistical Terms,* first edition, 1993, International Organisation for Standardisation (Geneva, Switzerland).

[3] *Expression of the Performance of Electrical and Electronic Measuring Instruments*, International Standards-revision of IEC 359 drafts July 1996.

AUTHORS:
G. Iuculano, 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
A. Lazzari, Department of Electrical Engineering, University of Bologna, Viale Risorgimento, 2 – Bologna – I 40136
Phone: +39 051 2093488, Fax: +39 051 2093470, E-mail: annarita.lazzari@mail.ing.unibo.it
G. Pellegrini Gualtieri, Department of Applied Mathematics, Engineering Faculty, University of Florence, Via di S. Marta, 3– Florence–I 50139 Phone: +39 055 4796248, Fax: +39 055 494569, E-mail: Gualtieri@dma.unifi.it
A. Zanobini, Department of Electronics and Telecommunication, University of Florence, Via di S. Marta, 3 – Florence – I 50139 Phone: +39 055 4796392, Fax: +39 055 494569, E-mail: willis@ingfi1.ing.unifi.it