XVIII IMEKO WORLD CONGRESS
Metrology for a Sustainable Development
September, 17 – 22, 2006, Rio de Janeiro, Brazil

ON THE USE OF MINIMUM CROSS ENTROPY PRINCIPLE AND BAYES' THEOREM FOR THE UNCERTAINTY EVALUATION IN A MEASUREMENT PROCESS

G. Iuculano [1] *,G. Pellegrini* [2]*, A. Zanobini* [1]

[1] Department of Electronics and Telecommunications, University of Florence,
Via di S. Marta, 3 – Florence – I 50139 E-mail: iuculano@ingfi1.ing.unifi.it
[2] Department of Applied Mathematics, Engineering Faculty, University of Florence,
Via di S. Marta, 3 – Florence – I 50139 E-mail: gabriella.pellegrini@unifi.it

Abstract: In this paper the evaluation of measurement uncertainty in a multivariate model is carried out by applying the principle of minimum cross entropy (MINCENT) and Bayes' theorem.
In particular the MINCENT optimization procedure is used to translate the information contained in the known form of likelihood into a prior distribution for Bayesian inference.
The methodology is adapted and tested on a recalibration model. Some basic ideas and general remarks on the Bayesian probability theory and entropy optimization principles are reported too.

Keywords: Measurement Uncertainty, Bayesian Inference, Minimum cross Entropy.

1. INTRODUCTION

The problem of assigning probability distributions which reflect the prior information about all the quantities involved in the measurement process, before the measurement results are available is a critical task in the use of inference methods of data analysis based on Bayes' theorem.
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.
So the principle of minimum cross entropy may be used to translate the information contained in the known form of likelihood into a prior distribution for Bayesian inference.
In this paper a procedure to evaluate the measurement uncertainty in a multivariate model is focused and carried out by applying the indicated principles. To asses the validity of the proposed approach, the methodology is applied to a recalibration model.

2. GENERAL REMARKS

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 theoric 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 mathematical accident such site 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 $\{D_1,\cdots,D_N\}$ consistent with a hypothesis H, we consider the class of all hypotheses $\{H_1,\cdots,H_N\}$ 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 favored by prior information I-, which means, usually, having the minimum cross-entropy (abbreviated "MINCENT").
Each successive piece of data that one obtains is a new constraint that, if cogent, 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. MINCENT 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 MINCENT 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.

3. THE MEASUREMENT MODEL

The measurement model concern multiple measurands and yields simultaneously multiple results. All the relevant quantities involved in the process are regarded as random variables so that their standard and mutual uncertainties are characterized through probability distributions with general multivariate joint density functions. We represent the input and output quantities through row vectors: $\mathbf{X} = (X_1,..., X_n)$, and $\mathbf{Y} = (Y_1,..., Y_m)$, belonging to the domain $\mathbf{D}_X, \mathbf{D}_Y$, respectively. Let $(\mathbf{x}, \mathbf{y})$, with $\mathbf{x} = (x_1,..., x_n)$ and $\mathbf{y} = (y_1,..., y_n)$, represent either the generic possible values or the actual realizations of $(\mathbf{X}, \mathbf{Y})$ in a particular occasion, characterizing the state of the measurement process in that occasion.

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

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

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

The Bayesian Approach

As it is known [1] in Bayesian approach all the involved quantities possess a prior distribution reflecting the accumulate state of knowledge, experience and also non-statistical information before the measurement process is carried out. What distinguishes the Bayesian paradigm from other statistical approach is that prior to obtaining the measurement results the experimenter considers his degrees of rational 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 (called "posterior"), which represents revised degrees of belief in the possible models, taking into account the new information provided by the measurements results. In the Bayesian approach we have an important linkage with the probability density function defined by (1):

$$f(\mathbf{x}|\mathbf{y}) = c\, f(\mathbf{x}) f(\mathbf{y}|\mathbf{x}) \qquad (2)$$

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

The Minimum Cross-Entropy Principle

In this work the aspects of the measurement uncertainty are investigated through the Bayesian approach given by eq. (2), and using the minimum cross-entropy principle (the Kullback's principle) [2], to determine the joint density function expressed by eq.(1) and under specified constraints, that is when given new information, in terms of certain expectations and/or domains of the involved quantities, by taking also into account the normalization properties of the probability density functions.

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

$$S = \int_{\mathbf{D}} f(\mathbf{x},\mathbf{y}) \ln \frac{f(\mathbf{x},\mathbf{y})}{f_0(\mathbf{x},\mathbf{y})} \,\mathrm{d}\,\mathbf{x}\,\mathrm{d}\,\mathbf{y} = \mathrm{E}\left\{ \ln \frac{f(\mathbf{X},\mathbf{Y})}{f_0(\mathbf{X},\mathbf{Y})} \right\} \qquad (3)$$

where we have passed to the compact notations: $d\,\mathbf{x} = dx_1 \ldots dx_n$, $d\,\mathbf{y} = dy_1 \ldots dy_n$ and where $f_0(\mathbf{x},\mathbf{y})$, in conformity with eq.(1), can be written as follows:

$$f_0(\mathbf{x},\mathbf{y}) = f_0(\mathbf{x}) f_0(\mathbf{y}|\mathbf{x}) \qquad (4)$$

The joint density $f_0(\mathbf{x},\mathbf{y})$, which a priori must be known, is defined by Jaynes [2] an "invariant measure" function.

In fact, since $f(\mathbf{x},\mathbf{y})$ and $f_0(\mathbf{x},\mathbf{y})$ transform in the same way under a change of variables, S remains invariant to any coordinate transform. It can be shown that $S \geq 0$; the equality sign will hold if $f(\mathbf{x},\mathbf{y}) = f_0(\mathbf{x},\mathbf{y})$ 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 $f_0(\mathbf{x},\mathbf{y})$, into a more circumstantial posterior joint distribution described by $f(\mathbf{x},\mathbf{y})$.

It also appears that, in some sense, the larger the divergence between $f(\mathbf{x},\mathbf{y})$ and $f_0(\mathbf{x},\mathbf{y})$, 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 $f(\mathbf{x},\mathbf{y})$, which minimizes the cross entropy given by eq.(3) is favoured 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 [3].

Let us note the crucial meaning of the function $f_0(\mathbf{x},\mathbf{y})$. Suppose only the domain D of all the involved quantities $(\mathbf{x},\mathbf{y})$ is known but we have no other information before the measurement is carried out.

In this case the alone constraint is the normalization condition on densities, that is:

$$\int\!\!\int_D f(\mathbf{x},\mathbf{y})\,\mathrm{d}\,\mathbf{x}\,\mathrm{d}\,\mathbf{y} = 1 \qquad (5)$$

where $D = D_{\mathbf{x}} \times D_{\mathbf{Y}}$, the solution which minimizes the entropy S, given by eq.(3) is:

$$f(\mathbf{x},\mathbf{y}) \propto \left(\int\!\!\int_D f(\mathbf{x},\mathbf{y})\,\mathrm{d}\,\mathbf{x}\,\mathrm{d}\,\mathbf{y} \right)^{-1} f_0(\mathbf{x},\mathbf{y}) \qquad (6)$$

Except for a constant factor, the function $f_0(\mathbf{x},\mathbf{y})$ is also the joint density describing "the complete ignorance" with exception the permitted domain $D = D_{\mathbf{x}} \times D_{\mathbf{Y}}$, before the measurement is carried out. If nothing it can be said about the shape of $f_0(\mathbf{x},\mathbf{y})$, it is likely to assume that it is constant over the domain $D = D_{\mathbf{x}} \times D_{\mathbf{Y}}$, tailing to zero outside that domain.

When the "invariant measure" function is constant, the Kulback's principle of minimum cross entropy is equivalent to the Jaynes's principle of maximum entropy [4].

By taking into account eq.s (1), (4), (5), eq. (3) can be written as:

$$S = \int_{\mathbf{D}_{\mathbf{x}}} f(\mathbf{x}) \ln \frac{f(\mathbf{x})}{f_0(\mathbf{x})} \,\mathrm{d}\,\mathbf{x} + \int_{\mathbf{D}_{\mathbf{x}}} f(\mathbf{x}) S_1(\mathbf{x}) \mathrm{d}\,\mathbf{x}$$

or, in a more compact and succinct form:

$$S = S_o + E\{S_1(\mathbf{X})\} \qquad (7)$$

with

$$S_0 = E\left\{\ln\frac{f(\mathbf{X})}{f_0(\mathbf{X})}\right\} = \int_{\mathbf{D}_x} f(\mathbf{x})\ln\frac{f(\mathbf{x})}{f_0(\mathbf{x})}\mathrm{d}\,\mathbf{x} \quad (8)$$

and

$$E\{S_1(\mathbf{X})\} = \int_{D_x} f(\mathbf{x})d\,\mathbf{x}\int_{\mathbf{D}_y} f(\mathbf{y}|\mathbf{X})\ln\frac{f(\mathbf{y}|\mathbf{X})}{f_0(\mathbf{y}|\mathbf{X})}\mathrm{d}\,\mathbf{y} \quad (9)$$

being:

$$S_1(\mathbf{X}) = \int_{\mathbf{D}_y} f(\mathbf{y}|\mathbf{X})\ln\frac{f(\mathbf{y}|\mathbf{X})}{f_0(\mathbf{y}|\mathbf{X})}\mathrm{d}\,\mathbf{y} \quad (10)$$

At this point eq.(7) must be minimized, as we said, under specified constraints.
If the constraints on the input quantities do not interfere with those ones of the output quantities, as usually it is in practice, in order to minimizing the general joint cross-entropy we can minimize before with respect to $f(\mathbf{y}|\mathbf{x})$ and then with respect to $f(\mathbf{x})$.
If we refer to Bayes's theorem in eq.(2), we minimize the cross-entropy before with respect to the likelihood and then with respect to the prior density of the input quantities.
In a first step we minimize the functional $S_1(\mathbf{X})$ given by eq.(10), with respect to $f(\mathbf{y}|\mathbf{x})$, subject either to the normalizing condition:

$$\int_{D_y} f(\mathbf{y}|\mathbf{x})\mathrm{d}\,\mathbf{y} = 1 \quad (11)$$

and to additional t constraints of type:

$$\int_{D_y} g_i(\mathbf{y}|\mathbf{X})f(\mathbf{y}|\mathbf{X})\mathrm{d}\,\mathbf{y} = g_i(\mathbf{X}); \quad \mathrm{i} = 1,...\mathrm{t} \quad (12)$$

where the functions $g_i(\mathbf{y}|\mathbf{X})$ and the values $g_i(\mathbf{X})$ are a priori known and in general they can depend on $\mathbf{X}$.
It is important to emphasize that the minimization of $S_1(\mathbf{X})$ with respect to $f(\mathbf{y}|\mathbf{x})$, implies the minimization of its expectation $E\{S_1(\mathbf{X})\}$ with respect the same one.
The minimum of cross-entropy $S_1(\mathbf{X})$ subject to constraints (11) and (12) is found through variational methods, by solving the well-known associated Euler's equation.
The conditional joint density function that minimize the functional $S_1(\mathbf{X})$ is given by:

$$f(\mathbf{y}|\mathbf{X}) = c\, f_0(\mathbf{y}|\mathbf{X})\exp\left(-\sum_{i=1}^{t}\lambda_i g_i(\mathbf{y}|\mathbf{X})\right) \quad (13)$$

being $c > 0$ a normalization constant and $\lambda_k, k = 1,...t$ the well known Lagrange's multipliers.
The Lagrange multipliers have a deep meaning: the generic λ_i is the "potential" of the datum $g_i(\mathbf{X})$ that measures how important a constraint it represents. Redundant data are at zero potential and are therefore "invisible" in the distribution involved in the principle of minimum cross-entropy and in the prediction that comes from it. A highly relevant datum $f_i(\mathbf{X})$ is one without which our predictions would be very different; then its λ_i is large and its presence greatly influences the entropy given by eq. (10).
By solving the well-known Euler's equation associated to the functional in eq.(10), we deduce, that the minimum of cross-entropy $S_1(\mathbf{X})$ satisfying the constraints eq.s (11) and (12) is:

$$S_{1\min}(\mathbf{X}) = \ln c - \sum_{i=1}^{t}\lambda_i g_i(\mathbf{y}|\mathbf{X}) \quad (14)$$

From eq. (9), taking into account the normalization condition:

$$\int_{D_x} f(\mathbf{x})\mathrm{d}\,\mathbf{x} = 1 \quad (15)$$

we get:

$$E_{\min}\{S_1(\mathbf{X})\} = E\{S_{1\min}(\mathbf{X})\} = \ln c - \sum_{i=1}^{t}\lambda_i g_i(\mathbf{y}|\mathbf{X}) \quad (16)$$

Now we must minimize with respect to $f(\mathbf{x})$ the cross entropy:

$$S = \int_{\mathbf{D}_x} f(\mathbf{x})\left[\ln\frac{f(\mathbf{x})}{f_0(\mathbf{x})} + S_{1\min}(\mathbf{x})\right]\mathrm{d}\,\mathbf{x} \quad (17)$$

by considering, besides the normalization condition eq. (15) and additionals constraints of type:

$$\int_{\mathbf{D}_x} w_i(\mathbf{x})f(\mathbf{x})\mathrm{d}\,\mathbf{x} = \mathrm{w}_i, \mathrm{i} = 1,...\mathrm{s} \quad (18)$$

where the functions $w_i(\mathbf{x})$ and the values w_i are known a priori.
The minimum of S with respect to $f(\mathbf{X})$ subject to constraints eq.s (15) and (18) is given, as previously, through the calculus of variations, by introducing new Lagrange's multipliers $\mu_i, i = 1,...s$.
After simple manipulation we obtain the solution of the variational problem:

$$f(\mathbf{x}) = c' f_0(\mathbf{x})\mathbf{exp}\left(-\sum_{\mathbf{i}=1}^{\mathbf{t}}\lambda_i g_i(\mathbf{x}) - \sum_{\mathbf{i}=1}^{s}\mu_i w_i(\mathbf{x})\right) \quad (19)$$

4. EXAMPLE

In this section we show the use of the proposed approach in a recalibration model.
The quantities involved in the measurement process are supposed to be random variables and classified into three different sets:

1) The input quantities: $\mathbf{X} = \{\mathrm{M}, \mathrm{V}\}$, with generic realizations $\mathbf{x} = \{\mathrm{m}, \mathrm{v}\}$

where:

- M is the measurand $-\infty < \mathrm{M} < \infty$
- V is the nuisance parameter due to the ignorance about the variability of output quantities $0 \le \mathrm{V} < \infty$

2) The output quantities: $\mathbf{Y} = \mathbf{I}$, with the generic realizations $\mathbf{y} = \mathbf{i}$ being $\mathbf{I} = \{\mathrm{I}_1, \mathrm{I}_2, ..., \mathrm{I}_n\}$ the indications and where the generic $\mathrm{I}_r \in (-\infty, \infty), \forall \mathrm{r} = 1,...\mathrm{n}$

3) The data $\mathbf{y} = \mathbf{i}$

and the indicated values: $\mathbf{i_0} = \{i_{01},...i_{0r},..., i_{0n}\}$

We assume a priori non informative densities, that is:

$$f_0(\mathrm{M,V,I}) = f_0(\mathrm{M,V}) f_0(\mathbf{I}|\mathrm{M,V}) = \text{constant}$$

and in a contextual connection

$$f_0(\mathrm{M,V}) \quad \text{and} \quad f_0(\mathbf{I}|\mathrm{M,V})$$

are supposed to be constant too.

TheConstraints:

- Actual constraints for the output quantities:

$$\mathrm{E}\{\mathrm{I_r}|\mathrm{M,V}\} = \mathrm{M}, \forall \mathrm{r} = 1,...\mathrm{n}$$

$$\mathrm{Var}\{\mathrm{I_r}|\mathrm{M,V}\} = \mathrm{V}, \forall \mathrm{r} = 1,...\mathrm{n} \tag{20}$$

$$\mathrm{Cov}\{\mathrm{I_r}, \mathrm{I_s}|\mathrm{M,V}\} = 0, \forall \mathrm{r} \neq \mathrm{s} = 1,...\mathrm{n}$$

- A priori constraints for the output quantities:

$$E\{\mathrm{I_r}\} = \mu, \forall \mathrm{r} = 1,...\mathrm{n} \qquad Var\{\mathrm{I_r}\} = \sigma^2, \forall \mathrm{r} = 1,...\mathrm{n} \tag{21}$$

The constraints (20) and (21) imply those ones for the input quantities, that is:

$$E\{\mathrm{M}\} = \mu \; ; \;\; E\{M^2\} + E\{V\} = \sigma^2 + \mu^2 \tag{22}$$

In a first step, minimizing $S_1(\mathrm{M,V})$, we get:

$$S_{1\,\mathbf{min}}(\mathrm{M,V}) = -\ln\left(c\mathrm{V}^{n/2}\right) \tag{23}$$

In a second step, by imposing the constraints in eq.s (22) with Lagrange multipliers λ_1 and λ_2 respectively and the normalization property:

$$\int_{-\infty}^{\infty}\int_{0}^{\infty} f(m,v)\,dm\,dv = 1$$

with Lagrange multiplier λ_0, the Eulero's equation can be writes as:

$$ln\, f(m,v) + ln\, c_1 - ln\, V^{\frac{n}{2}} + \lambda_1 m + \lambda_2\left(i^2 + v\right) = 0$$

The optimal solution of the variational problem is:

$$f(m,v) = c_1\, exp\left(-\lambda_1 m - \lambda_2 m^2\right) v^{\frac{n}{2}}\, exp(-\lambda_2 v) \tag{24}$$

where λ_1, λ_2 are the Lagrange multipliers.

The equation (24) confirms that M and V are mutually independent and , by imposing $c_1 = c_2 c_3$, we can write:

$$f(m,v) = f(m) f(v) \tag{25}$$

where:

$$f(m) = c_2\, exp\left(-\lambda_1 m - \lambda_2 m^2\right) \tag{26}$$

with:

$$c_2 = \frac{1}{\sqrt{2\pi}\sqrt{var\{M\}}}\, exp\left(\frac{-\left(E\{M\}\right)^2}{2Var\{M\}}\right)$$

$$\lambda_1 = -\frac{E\{M\}}{Var\{M\}}, \; \lambda_2 = -\frac{1}{2Var\{M\}} \tag{27}$$

so that eq.(26) will be written as:

$$f(m) = \frac{1}{\sqrt{2\pi}\sqrt{var\{M\}}}\, exp\left(\frac{-\left(m - E\{M\}\right)^2}{2Var\{M\}}\right) \tag{28}$$

that is a normal density with expectation $E\{M\} = \mu$ and variance $Var\{M\}$ which can be deduced from eq. (22) (see after) and where:

$$f(v) = c_3 v^{\frac{n}{2}}\, exp(-\lambda_2 v) = \frac{v^{\alpha-1}\, exp(-v/\beta)}{\beta^\alpha \Gamma(\alpha)} \tag{29}$$

is a gamma density with, taking into account the third one eq.s (27):

$$c_3 = \frac{1}{\beta^\alpha \Gamma(\alpha)}, \qquad \alpha = \frac{n}{2} + 1, \qquad \beta = \frac{1}{\lambda_2} = 2\,Var\{M\} \tag{30}$$

It is important to emphasize that the Gamma distribution is characterized by n and $Var\{M\}$.

In order to evaluate this last, we can consider the eq.s (22) which are equivalent to:

$$Var\{M\} + E\{V\} = \sigma^2 \tag{31}$$

On the other end $E\{V\}$ is the expectation of the Gamma distribution defined by (11) and taking into account eq.s (30) we deduce:

$$E\{V\} = \alpha\beta = (n+2)Var\{M\} \tag{32}$$

By substituting eq. (32) into eq. (31) we have:

$$Var\{M\} = \frac{\sigma^2}{n+3} \tag{33}$$

Bayesian approach:

The posterior density of M given $\mathbf{I} = \mathbf{i}_0$ is:

$$f(m|\mathbf{i}_0) = kf(m) f(\mathbf{i}_0|m) \tag{34}$$

where $f(m)$ is given by (28) and $f(\mathbf{i}_0|m)$ is deduced by recalling that:

$$f(\mathbf{i}|M) = \int_0^{+\infty} f(\mathbf{i}|M,v) f(v|m)\,dv = \int_0^{+\infty} f(\mathbf{i}|M,v) f(v)\,dv \tag{35}$$

since M and V are mutually independent and where $f(v)$ is given by (29) and $f(\mathbf{i}|\,\mathrm{M,v})$, as well known, is the multivariate normal distribution given by:

$$f(\mathbf{i}|M, v) = \frac{1}{\sqrt{(2\pi)^n}} \frac{1}{\sqrt{v^n}} exp\left(\frac{1}{2v} \sum_{r=1}^{n} (i_r - M)^2 \right)$$

Finally we obtain:

$$f(\mathbf{i}_0|m) = k \int_0^\infty \exp(-a_0/v) \exp(-v/\beta) dv = \sqrt{4a_0\beta} K_1 \left(\sqrt{\frac{4a_0}{\beta}} \right) \qquad (36)$$

where

$$a_0 = \frac{1}{2} \sum_{r=1}^{n} (i_{r0} - m)^r > 0$$

and

$$\beta = \frac{2}{m+3} \sigma^2$$

From eq. (36) with n=1, we deduce:

$$f(\mathbf{i}_0|m) = k_2 \frac{1}{\sqrt{2\pi}\sqrt{\sigma^2/4}} \exp(-2(m-\mu)/\sigma^2) \left[\frac{\sigma^2}{2} |i_{10} - m| K_1(|i_{10} - m|\sigma) \right]$$

5. CONCLUSION

In this work the aspects of the measurement uncertainty in a multivariate model are investigated through the Bayesian inference approach and using the minimum cross-entropy principle to determine the joint prior density function under some specified constraints.

Acknowledgment

The authors are gratefull to Prof. Wolfgang Woeger, Senior Scientist at Physikalisch-Technische Bundesanstalt (PTB) Braunschweig, for his important support and suggestions to this research.

REFERENCES

[1] K. Weise, W. Woger : A Bayesian theory of measurement uncertainty, Meas. Sci. Technol., (1993),4, 1-11.

[2] E.T.Jaynes: Information theory and statistical mechanics I, Phys.Rev., (1957),108, 171-190.

[3] John E. Shore, Rodney W. Johnson: Axiomatic Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross-Entropy, IEEE Trans on Information Theory, (1980), IT-26, n.1, 26-37.

[4] J.N. Kapur: Maximum-Entropy Models in Science and Engineering, John Wiley-Sons.

.