POLISH ACADEMY OF SCIENCES - MATERIALS SCIENCE COMMITTEE
SILESIAN UNIVERSITY OF TECHNOLOGY OF GLIWICE
INSTITUTE OF ENGINEERING MATERIALS AND BIOMATERIALS
ASSOCIATION OF ALUMNI OF SILESIAN UNIVERSITY OF TECHNOLOGY

Conference Proceedings

AMME'2002

11th INTERNATIONAL SCIENTIFIC CONFERENCE
ACHIEVEMENTS IN MECHANICAL & MATERIALS ENGINEERING

Information in metric space

M. Pappalardo

DIMEC Department of Mechanical Engineering, University of Salerno
Via P. don Melillo, 84084 Fisciano (SA). Italy

The idea of information, in the theories of Fisher and Wiener-Shannon [1] [2] , is a measure only on probabilistic and repetitiveness events. The idea of information is larger than the probability. It is possible the formulation of an Extend Theory of Information for probabilistic and non-probabilistic events. In this paper we extend the Wiener–Shannon's axioms to the non-probabilistic and repetitiveness events. On the basis of so called Laplace's Principle of insufficient reason, the *MaxInf* principle is defined for to choose solutions in absence of knowledge. In this paper, as example of application, the value of information, as a measure of equality of data among a set of values, is applied in numeric analysis as method for approximation of data.

1. INTRODUCTION

The design of a complex system involves analysis, organisation and calculation of various elements under external constraints. The definition of time and space in terms of independent variables is:

$$t \in [0,T] \subseteq \mathrm{R}^+ \quad \mathbf{x}=\{x_1, x_2, x_3\} \in D \subseteq \ ^3 \tag{1}$$

A physical system can be observed in an interval time $[0,T] \subseteq \mathrm{R}^+$ and in a volume $D \subseteq \ ^3$.
The vector $\mathbf{u}$ of the state of a system is:

$$\mathbf{u}=(t,\mathbf{x}): [0,T] \times D \mapsto \ ^{\mathrm{n}} \tag{2}$$

If $\Omega \subseteq \ ^2$ is the field of application of the system, and ω a remarked event at time *t*, then the measure of ω will be always incorrect [7]. The knowledge of ω is not given by its coordinates in Ω, but it is possible to assert only that ω is limited in a subset A_i of Ω. In a field of application, one event is represented by a subset $A_i \subseteq \Omega$ with

$$\bigcap_i A_i = \phi \text{ and } \bigcup_i A_i = \Omega \tag{3}$$

If the density of probability function $\varphi(\omega)$ is subjected to the constraints $\varphi(\omega) \geq 0$ then the probability is given by

$$p(A_i) = \int_{A_i} \varphi(\omega) d\varpi \qquad \int_\Omega \varphi(\omega) d\omega = 1 \tag{4}$$

The Shannon's expression Entropy, on the partition is given by

$$\pi : \Omega = \bigcup_i A_i \qquad S(A) = -c\sum_{i=1}^{n} p(A_i) \log p(A_i) \tag{5}$$

Entropy is a generic value with very large meaning not connected to probabilities. If we are dealing in a space of probability distributions the *distance* between two probability distribution **p** and **q** is given by the distance $D(\mathbf{p}:\mathbf{q})$. If **q** *is a priori distribution* we need to select the distribution **p** *closeness to* **q.** For satisfying the constraints of probability distribution we can use the measure of *cross-entropy* developed by Kullbach and Leibler [6]

$$D(\mathrm{p}:\mathrm{q}) = \sum_{i=1}^{n} p_i \, ln \frac{p_i}{q_i} \tag{6}$$

D(**p:q**) is the *distance* of the a priori distribution **q** from the distribution **p**. In order to optimisation a distribution we can use the minimum cross entropy principle: *From all probability distribution satisfying given constraint we must choose the distribution* ***p*** *that minimise the measure* D(**p:q**). This is the well know *MinEnt* principle. If the a priori distribution is of maximum uncertainty, $\mathbf{q} = \underbrace{1/n, ..., 1/n}_{n}$ then

$$D(\mathbf{p,q}) = \sum_{i=1}^{n} p_i \ln \frac{p_i}{1/n} = \ln(n) - \left(-\sum_{i=1}^{n} p_i \ln p_i \right) \tag{7}$$

where $S(\mathrm{p}) = -\sum_{i=1}^{n} p_i \ln p_i$ *is the Shannon's entropy*. Therefore, when the a priori distribution **q** has the maximum uncertainty, for minimising D(**p:q**) we should choose values of **p** *maximising entropy S(***p***)*. This is the *Jaynes' maximum principle (MaxEnt)* [8]. Jaynes, the principal proponent of *MaxEnt Principle* in axiomatic way, suggests that in all probability distribution, when we have *only* the constraint that $\{p_i \geq 0\}$ and $\sum_i p_i = 1$, we should choose that one has maximum entropy. The use of probability distribution with less then maximum entropy implies the use of additional information [6]. In order to optimisation a distribution can use the Minimum Cross Entropy Principle: *From all probability distribution satisfying given constraint we must choose the distribution p that minimise the measure D(p:q).* This is the well know *MinEnt* principle. If the a priori distribution is of maximum uncertainty then $q = \underbrace{1/n, ..., 1/n}_{n}$. Therefore, when the a priori distribution q has the maximum uncertainty, for minimising *D(p:q)* we should choose values of *p maximising entropy S(p)*. This is the *Jaynes' maximum principle (MaxEnt).* Jaynes, the principal proponent of *MaxEnt Principle* in axiomatic way, suggests that in all probability distribution, when we have *only* the constraint that $\{p_i \geq 0\}$ and $\sum_i p_i = 1$, we should choose that one has maximum entropy. The use of probability distribution with less then maximum entropy implies the use of additional information.

2. AXIOMS OF NEW THEORY OF INFORMATION (EXTENDED THEORY)

The idea of information, in the theories of Fisher and Wiener-Shannon [1][2], is a measure only on probabilistic and repetitiveness events. The idea of information is larger than the probability and the axioms of Wiener–Shannon can be extended to the non-probabilistic and repetitiveness events. Let Ω to be the field of all events ω, probabilistic or non-probabilistic, and $\Im$ a class of parts of Ω, $\Im \subset \wp_{arts}(\Omega)$. With $A \subset \Im$ we can assume the next two axioms:

AXIOM I: The value of information *J(A)* is always non negative*:*

$J(A) : \Im \rightarrow {}^{+}$ (8)

AXIOM II: The value of information *J(A)* is monotonous in regard to inclusion:

$\forall A, B \in \Im \quad , \quad B \subset A \quad , \quad J(B) \geq J(A)$ (9)

Now it is possible the construction of new algorithms in terms of information, founded only on the first and second axioms[3]. For independent events it is opportune to assume a third axiom:

AXIOM III: If the events $A, B \in \Im$ are independent for all the values of information we have:

$\forall A, B \in \Im \quad \mathrm{J}(B \cap A) = J(B) + J(A)$ (10)

The third axiom shows that when we are in presence of independent events it is possible to add up information. If Ω is a certain event and ϕ the impossible event than, for an universal validity of $J(A)$ and $J(\phi)$, for all $\Omega, \Im$ and J must be:

$J(\Omega) = 0\,, J(\phi) = +\infty$ (11)

The expression $J(\Omega) = 0$ means that Ω is a certain event without needs of information. The expression $J(\phi) = +\infty$ means that if ϕ is an impossible event with the needs of infinite information. In a metric space Ω, if ω is an event in $\Im \subset \wp_{arts}(\Omega)$, its measure will be always incorrect. The knowledge of ω is not given by its coordinates in Ω, but it is possible only to assert that ω is limited in a subset $A_i \in \Im$. If $d(A_i)$ is the diameter of set A_i, than, more is the precision of measures, less is the measure of diameter of event A_i. If we assume that is a set $\{P_{x,y}\}$ of ideal data in a continuous closed bounded subset $\Omega \in [D]$, given any $\varepsilon > 0$, there is a set $\{M_{x,y}\}$ of values of measures with sufficiently high precision such that

$\left|P_{x,y} - M_{x,y}\right| < \varepsilon \;$ for $(x, y) \in \Omega$ (12)

But the probability p of an exact measure is in inverse proportion to the precision, so the ideal measure of point's coordinates of has null probability to be obtained: it is an impossible event. The impossible event ϕ and the certain event Ω are always independent from J and *A*: they are universal values. All three axioms have correspondent axioms in Wiener-Shannon theory.

2.1. New Theory of Information in Metric Space

With the axioms I, II, and III it is possible to construct models for information very useful in applications. For every event $A \in \Im$ we can have a measure of information using the mathematical expression :

$$J(A) = \frac{1}{d(A)} \tag{13}$$

This definition of information has a natural application in metric space [3]. Let be $\Omega =$ and $\Im \mapsto$, than we have that better must be the result of a measure than smaller is the diameter of A_i, and larger will be *J(A)*. [9] If we assume that all the measures are made with equal care, and for any value of ω the data have a normal distribution, the probability that the error $d(A_i)$ will fall in a small interval δ_y is given for ω_i

$$P(\omega_i) = \frac{1}{\sigma\sqrt{2\pi}}\left[\exp-\left(d(A_i)\right)^2/2\sigma^2\right]\delta_y \tag{14}$$

Similar expression can be written for all ω_i in Ω. The standard deviation σ is a measure of precision of the measurements and it is a constant for all the data. As the separate measurements are independent for all events, the probability for all is the product

$$P=\prod_{k=0}^{N} P(\omega_i)=\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^{N+1}\left[exp-\frac{1}{2\sigma^2}\sum_{i=0}^{N}\left(d(A_i\ )\right)^2\right]\delta_y^{\ N+1} \tag{15}$$

Maximum of P is the sign of the goodness of the measures of the diameters of subsets A_i. This will occur when

$$\sum_{i=0}^{N}\left(d(A_i\ )\right)^2 = \text{minimum} \tag{16}$$

In general, the criterion that the sum of diameters $\sum_{i=0}^{N}\left|d(A_i\ )\right|$ of sets A_i be as small as possible would have given the better result. We have that the same information can be valued by the probability and by the non-probabilistic measures of diameters. So we can have the measure of information from non-probabilistic data. Now, for to ascribe some information to the realised event A_i, we can assume as measure of information a non-probabilistic function

$$J(A)\overset{def}{=}\Psi\left[\sum\left|d(A_i)\right|\right] \quad J(\ A_i\ )\propto\frac{1}{\left|d(\ A_i\ )\right|} \tag{17}$$

Any vector $\overline{x}=\left(x_1,x_2,\cdots,x_n\right)$ representing proportions of some whole is subject to the unit sum constraints $\sum_i x_i=1$. One of most usual dissimilarities and distance $d\left(x_i,x_j\right)$ to measure the difference between two compositions are Minkowski's distances. In general, it is possible to have the measures of information for probabilistic and non-probabilistic events using empirical or non-empirical functions non-attached to the probability and to the repetitiveness.

2.2. The Extended Principle of Max-Information (MaxInf)

It is possible to define a new principle on basis of the New Theory of Information. On the analogy of *MaxEnt principle*, the name is *Max Information Principle* (*MaxInf*) [6]. In the New Theory of Information, instead of probability, it is possible to utilise of a finite number of appropriate proportion subject to a set of constraints that add up to one. In observance of the Axioms, let $d_1,d_2,...,d_n$ be n non-negative real numbers, let [6] [7]

$$\sum_{i=1}^{n} d_i \neq 0 \qquad \rho_i=\frac{d_i}{d_1+d_2+...+d_n} \qquad \sum_{i=1}^{n}\rho_i=1 \ \ (\ \rho_i\geq 0\ \ \forall i\ ) \tag{18}$$

We can use the measure of information the relation

$$\mathrm{J}(\rho)=J\left(\rho_1,\quad \rho_2,....,\quad \rho_n\right)=\ -\sum_{i=1}^{n}\rho_i\ln\rho_\mathrm{i} \tag{19}$$

So that: $J(\rho)$ is maximum when $\rho_1=\ \ \rho_2=....=\ \ \rho_n$

$J(\rho)$ is minimum when: $\forall i$ only one number is $\neq$ zero

In metric space, using Euclidean's distances the information maybe

$$J\ =-\sum_{ij}\left|d\left(x_i,x_j\right)\right|\Big/\sum_{ij}\left|d\left(x_i,x_j\right)\right|\log\left(\left|d\left(x_i,x_j\right)\right|\Big/\sum_{ij}\left|d\left(x_i,x_j\right)\right|\right) \tag{20}$$

The value of information *J(ρ)* is a measure of equality of numbers among themselves. Appling the same formalism of MaxEnt Principle it is easy to define the MaxInf Principle on the basis of so called *Laplace's Principle* of insufficient reason.

MaxInf Principle: *Out of all knowledge, choose the solution closet to the uniform distribution of information.*

In the situations in which we have no reasons for to prefer a solution, it is better choose the solution with uniform distribution, or the closet to the uniform distribution of information.

3. APPLICATION

One application of the *MaxInf* principle is in problems of approximation as criteria to find polynomials for to represent a given set $E = \{(x_i, y_i),..\}$ of empirical points. Ideally, this process should take in account the reliability of the observations, so the more reliable points will have grater weight on approximating function. In absence of knowledge, on basis of *MaxInf* principle, we must use a polynomials, which in representing points, the deviation from them choose the solution closet the uniform distribution of information. In metric space, let be $y = f(x_i)$ the approximating function from which we obtain, from the points (x_i, y_i), the *n* deviation $d_i = (f(x_i) - y_i)$. The estimator vector is

$$\bar{d} = (d_1, d_2,.., d_n)^T \tag{21}$$

As function for to measure the information we can use the function

$$J = \sum_i \frac{1}{|(f(x_i) - y_i)|} \tag{22}$$

From MaxInf we have the max value for *J* when

$$J_1 = J_2 = ... = J_n \qquad \frac{1}{|(f(x_1) - y_1)|} = \frac{1}{|(f(x_2) - y_2)|} = ... = \frac{1}{|(f(x_n) - y_n)|} \qquad \forall i \tag{23}$$

The max of information is obtained when the approximating function $y = f(x_i)$ *has the same error* from all the *n* points (x_i, y_i).

$$|(f(x_1) - y_1)| = |(f(x_2) - y_2)| = ... = |(f(x_n) - y_n)| = h \tag{24}$$

The estimator vector has the distributions

$$\bar{d} = \left(\underbrace{h,\ h,..,h}_{n}\right)^T \tag{25}$$

If the approximation function is a polynomial

$$f(x) = a + bx + cx^2 + ... + dx^n + ... \tag{26}$$

The deviations from the points $\{(x_i, y_i),..\}$ of the function $f(x_i)$ evaluated at certain abscissa and the given ordinate corresponding to the same abscissa:

$$\begin{aligned} f(x_1) - y_1 &= a + bx_1 + cx_1^{\ 2} + ... + dx_1^{\ k} + ... - y_1 = (-1)^1 h \\ &\vdots \qquad\qquad \vdots \qquad\qquad\qquad \vdots \\ f(x_n) - y_n &= a + bx_n + cx_n^{\ 2} + ... + dx_n^{\ k} + ... - y_n = (-1)^n h \end{aligned} \tag{27}$$

From the solution of the linear system $A\overline{x} = \overline{b}$ with $\overline{b} = (y_1, y_2, ..., y_n)^T$ and $\overline{x} = (a, b, c, ..., h)^T$ we have the solution of polynomials and the value of h from which can be valuated the approximation with the max information.
Example- Let us illustrate the method by a very simple numerical problem. Using as polynomial the straight line $(y = a + bx)$, approximate the points, (0,1), (1,4.5) and (2,6.8).
Solving for $\overline{x} = (a, b, h)^T$ would be the task:

$$A\overline{x} = \overline{b} \quad \overline{x} = A^{-1}\overline{b} \quad \overline{x} = |1.3, 2.9, 0.3|^T \tag{28}$$

The straight line approximation is $y = 1.3 + 2.9x$. The value of *h,* obtained from *MaxInf* principle is *h*=0.30. And the estimator vector is $\overline{d} = (0.3, 0.3, 0.3)^T$ with spherically distributions.

4. CONCLUSION

Considering that the idea of information is larger than the probability, we have extended the axioms of Wiener–Shannon to the non-probabilistic and repetitiveness events. On the basis of so called Laplace's Principle of insufficient reason, it is defined the *MaxInf* Principle for to choose solutions when we are out of all knowledge. The MaxInf principle, applied in numeric analysis for approximation of data has shown the possibility of analysing data on the basis of the extended theory of information.

REFERENCES

1. N. Wiener, *Cybernetics,* Paris Herman (Act. Sc. 1053) 1948.
2. C. E. Shannon, *The Bell System Technical Journal* –Vol. 27, July, October, 1948.
3. J. Kampe de Feriet, *Note di Teoria dell'Informazione,* Anno Accademico 1971-72-Annali dell'Istituto di Matematica Applicata dell'Università di Roma. 1972.
4. N. Siddal, *Probabilistic Engineer Design,* M.Dekker JHG N.York 1983.
5. Nam P. Suh: *The Principles of Design.* Oxford University Press New York 1990.
6. N. Kapur, K. Kesavan, *Entropy Optimization Principles*, Acad. Press Inc. Boston 1992.
7. M. Pappalardo, *Information in Design Process,* 2nd International Conference on Planned Maintenance, Reliability and Quality - Oxford , England 2nd -3rd April 1998.
8. Jaynes E.T. *Probability Theory: The logic of Science*. http://omega.albany.edu: 8008/ Jaynes Book.html
9. A. Donnarumma, M. Pappalardo, A. Pellegrino, *Measure of independence in soft design*- Jurnal of Material Processing Technology – 124 ELSEVIER (2002).