Advance Physics Letter

Generalized Entropy and its Application in Information Technology

[1]Sarat K. Parhi, [2]L. Das

[1]Vijayanjali Institute of Technology, Balasore, Odisha, India- 756043.
[2]Dept. of Applied Mathematics Technological University, New Delhi- 110042.

Abstract: In this paper an algorithm that can generate target information from the set of given information is discussed. Subsequently, the procedure for obtaining entropy for an information theoretic function with more than two variables is discussed. Finally, it has been shown that this theory can be applied in the information transmission.

I. INTRODUCTION

The study of information science involves with the technique of storage, retrieval, and manipulation of information. In this context, the concept of information can be defined as the recorded or the communicated data or the materials that has some specific meaning associated with the symbolic representation, digital representation, measurement of uncertainty and the derivation of any particular information from the set of available or unavailable information.

From the above mentioned research topics, measurement of uncertainty are to be studied here. Suppose a sequence of assumptions is given as atomic information and from this sequence, one has either to derive the consequences of some specific information (say target information). For this purpose, it is essential to store information in digital code and then these are to be processed under certain operations.

The algorithm that can generate all possible information or some specific target information is to be discussed in this chapter. Unavailable information in a document source usually named as entropy is also discussed in this chapter. The entropy that is to be discussed in this chapter is of more than two variables, that means the generalizations of usual entropy for one variable. Finally a brief discussion is made on the application of entropy in the information transmission.

II. NOTATIONS AND TERMINOLOGY

Information means certain impressions or assumptions. The data are the components of binder to any information. But data are nothing but the signals in the form of electromagnetic, heart, sound, lights or ethers. To codify those data various notations are used various researchers. Here we specify a few notations those are used in this chapter.

We consider the sequence of assumptions as the points in $\Gamma \subset R^n$ and these points may be called as the population set. In our purpose we will assume the components of any point vector are either zero or one. (The algorithm that is to generate all or the specific targeted information is basically, take a point from the population set and modify it to the reach the target. The details procedure is stated in section 8.3). Though the points in the assumptions set are vectors with binary components, these can be further represented as the binary strings.

A monomial binary string is a vector whose components are single binary string (of length n). A polynomial binary string is a vector whose components are more than binary strings. Let M_n denotes the set of all binary string, defined on $\{0,1\}^n$. $M_{n,k}$ denote the space of functions represented by the monomials which have at most k- intervals. $D_{n,k}$ is assumed as the space of functions (Boolean functions of n-variables) and represented by a disjunction of monomials in $M_{n,k}$. The phrases “Learning Algorithm”, “Training Algorithm”, “Learning Samples” etc. are used in the subsequent texts. Those are defined as follows,

Definition 2.1

The unavailable of random data in an information is called entropy. For example it may happen in Chanel error during transmission of information.

Definition 2.2

A learning algorithm is a procedure that to accept the training samples and their functions.

Definition 2.3

A training samples “s” is a member of the set $S\Xi(Xx\{0,1\}^m)$ or

$S=\{(x_1,b_1),(x_1,b_2)\ldots(x_m,b_m)$ for $b_i \in \{0,1\}^m$ for $i=1,2\ldots m\}$. In the other words, the input strings of a learning sample are called the training sample. Occasionally, $\{0,1\}^m$ is represented by $\sum^m$ which is not different to the m-times Cartesian product of the set of alphabets (symbolically represented as $\sum=\{0,1\}$).

Definition 2.4

The output of a learning algorithm is called learning samples.

The concept of positive formula and negative formula are usually defined as per the context. In this chapter we assume the meanings of positive and negative formula in the following manner.

Definition 2.5

The formula that gives the assurance of correct information is called the positive formula.

Definition 2.6

The formula that assurance of wrong information is called the negative formula.

The truth-values usually associated with the positive and negative formula are one and zero respectively.

The binary strings of length m is associated with a vertex of the m-cube and it is also called a formula. The formulas are usually labeled by the assignment of truth values namely zero and one and the procedures for such assignment is due to the proximity with the target information. The decision for the proximity is usually calculated by the hypothesis or by the judicious guessing.

No matter what is the ways of selection of labels, it is usually a mapping (say) C:X→{0,1}. Thus if x ϵ X is a formula C(x)=0 implies x is negative formula else it is positive. In practice, there exist some formulas whose labels are neither zero nor one. Those are not discussed here.

Definition 2.7

A sequence labeled formulas is called "training sample". And it is usually belong to the product space (Xx$\{0,1\}^m$).

Sometimes it is required to enfold the given string to obtain the specific information, for which we used an operator F and it operates on a string of length n. In fact F is a linear transformation such that F:B^n→B($\sqrt{n}$,$\sqrt{n}$) and can be represented in the form of a square matrix. The generation of new strings can be explained in the following algorithm (of the transformation).

III. ALGORITHM FOR FOLDING STRING

1. If the string is of length n, then it is to be folded at k$\sqrt{n}$ positions, where k ϵ {1,2,…|$\sqrt{n}$|} the symbol |$\sqrt{n}$|indicates the lowest integer among all possible integers greater than $\sqrt{n}$.

2. The output substrings are of length |$\sqrt{n}$| each.

3. A string of length greater than one and less than or equal to 4 is expressed as 2 x 2 square matrices.

4. j-th column of the folding matrix is the transpose of j-th substring after j-th partition.

5. In general, if n x n is not a perfect square then |$\sqrt{n}$ x$\sqrt{n}$ |can be consider as |$\sqrt{n}$| x |$\sqrt{n}$| and this is the order of the corresponding folding matrix. As some of the proposition of the folding matrix are blank. These blank position can be filled by insertion of zeros.

6. Particularly, the string of length 3 say x_1 x_2 x_3 can be expressed as$\begin{bmatrix} x_1 & x_3 \\ x_2 & 0 \end{bmatrix}$, here the position 2 x 2 is inserted to zero as there are shortage of one element to become a perfect square of order 2 x 2. Similarly the string of length 2 can be expressed as the matrix$\begin{bmatrix} x_1 & 0 \\ x_2 & 0 \end{bmatrix}$.

Sometimes the folding matrix can be derived from the reaction function. Whether the reaction function Θ is defined as follows.

Let F(x_1,x_2,…x_n)

$$= \begin{pmatrix} x_1 & x_{1+\lceil\sqrt{n}\rceil} \cdot & x_{n-\lceil\sqrt{n}\rceil} \\ x_2 & x_{2+\lceil\sqrt{n}\rceil} \cdot & x_{n-1-\lceil\sqrt{n}\rceil} \\ x_{\lceil\sqrt{n}\rceil} & x_{2\lceil\sqrt{n}\rceil} . & x_{\lceil n\rceil} \\ & & \end{pmatrix}$$

be the folding string and it can be defined by the reaction function ⊕:$\sum^{\lceil\sqrt{n}\rceil\cdot\lceil\sqrt{n}\rceil}$x $\sum^{\lceil\sqrt{n}\rceil} \rightarrow \sum^{\lceil\sqrt{n}\rceil}$, such that ⊕ : (Θ ,x')= x", Θ =F(x)∊ $\sum^{\lceil\sqrt{n}\rceil\cdot\lceil\sqrt{n}\rceil}$, x'∊$\sum^n$ and x"=(x_1', x_2', ... x_n') is the new string in $\sum^n$.

Corollary 3.1

If all the training sample contain all positive formula of the target information, Then the algorithm will output the accurate target information.

Corollary 3.2

New information from the output algorithm (1) can be obtained by deleting at least one literal in each iteration, provided a positive formula in the training sample is selected suitable.

IV. INFORMATION GENERATION BY ESTIMATION OF ERROR

New information are generated by manipulating the stored information codes or strings. During Such manipulation ,the resultant string size may be exceed to the space allocated for the restoration, which causes an overflow of string bits. In most of the situations, the length of the strings are kept *Affixed* in order to economize the storage capacity as well as to smoothening the transactions .In this way, the truncation error are committed and is such errors are propagated exponentially many new other (even absurd) information can be obtained. But no doubt this helps to generate the information pool very quickly .To generate the new information accepting a specific information vector as its input.

Let B={0,1} be the set of alphabets and $B^N=\{0,1\}^N$ be the set of binary string of length N obtained from B. Now x ϵ B^N implies $x=(x_1,x_2,\ldots x_N)$ Where x_i ϵ B for all i ϵ{1,2,…N}. Thus the binary string of size N is associated with a vertex of N-cube for N=3, $B^3= B^2 \times B=\{$ (0,0),(0,1),(1,0),(1,1)}×{(0,1)}={(0,0,0),(0,1,0),(1,0,0),(1,1,0),(0,0,1),(0,1,),(1,0,1),(1,1,1)}={000,010,100,110,001,011,101,111}

Let $F(x)=F(x, c_{1,}c_{2,\ldots.}c_k)$ be a polynomial of degree k-1. It can be written as $F(x)=c_{1+}c_2x+c_3x^{2+\cdots+}c_kx^{k-1}$ where $c_{1,}c_{2,\ldots.}c_k$ are real constants whose numerical values are associated with the binary strings can be determined. The reoccurring expression

$F^*(x)= c_1^*\varphi_1(x) + c_2^*\varphi_{2(x)} + c_3^*\varphi_{3(x)} + \ldots + c_k^*\varphi_k$ for $\varphi_{i(x)}=x^{i-1}$ is a function of the form $F^*(x)= F(\varphi_i(x), c_1^*, c_{2,}^*\cdots c_k^*)$ The calculation $c_1^*, c_{2,}^*\cdots c_k^*$ can be made from the relationship $\sum f_i = N c_1^* + c_2^*\sum x_i$, where N is the number of partitions in X-interval. That means $c_1^*=\frac{\sum f_i}{N}$. The calculation of c_2^* is obtained from the quadratic expressions $\sum x_if_i = c_1^*\sum x_i + c_2^*\sum x_i^2$ and this implies $c_2^*=\frac{\sum x_if_i}{\sum x_i^2}$. Similarly the calculation of c_3^* can be obtained from the cubic and quadratic expressions $\sum x_if_i = c_1^*\sum x_i + c_2^*\sum x_i^2 + c_3^*\sum x_i^3$ and $\sum x_i^2f_i = c_1^*\sum x_i^2 + c_2^*\sum x_i^3 + c_3^*\sum x_i^4$ implies $c_3^* =[N\sum x_i^2f_i - \sum x_i^2 \sum f_i]/$ $N\sum x_i^4-(\sum x_i^2)^2$

The other coefficient c_4^* can be obtained from the reoccurring expression $f(c_4^*) =\frac{(c_4^*-c_2^*)(c_4^*-c_3^*)}{(c_1^*-c_2^*)(c_1^*-c_3^*)} f(c_1^*)+\frac{(c_4^*-c_1^*)(c_4^*-c_3^*)}{(c_2^*-c_1^*)(c_2^*-c_3^*)} f(c_2^*) + \frac{(c_4^*-c_1^*)(c_4^*-c_2^*)}{(c_3^*-c_1^*)(c_3^*-c_2^*)} f(c_3^*)$.

The values $c_1^*, c_{2,}^*\cdots c_k^*$ of are the approximation of the of the original coefficients $c_{1,}c_{2,\ldots.}c_k$. The derivations satisfy the following reoccurring formula

$d_k= f_k - F(x, c_{1,}c_{2,\ldots.}c_k)$ and $d_k^1 = f_k - F(\phi i(x), c_1^*, c_{2,}^*\cdots c_k^*)$. For k=3, the deviations can be obtained as the vertices of the cube and the respective function is a convex function. Using the following norms can approximate the average deviations, $\| d_k^1 \| = \sum_{k=1}^{N} d_k$. This process is also useful for useful for the calculation of the elements of the matrices.

The above mentioned approximation algorithm is summarized in the following manner.

1. S is the set of vectors, each having m components, and this set is also named as population set. Define a mapping f:S → T, for T⊆ R, Where R is the set of real numbers.

2. Calculate the values of $f(x)$ at x_i such that for the increment h of x_k , $f(x_k+h)= f_k+\epsilon_k$, for k= 1,2,3,…, N and ϵ_k are known errors.

3. The data of f_k contain the slowly varying components. To overcome this difficulty, we modify it by introducing the functions F(x) and F*(x) . The domain of F(x) and F*(x) are $B^N \subseteq$S.

4. The reoccurring expression of F(x) is a polynomial $F(x)=F(x, c_{1,}c_{2,\ldots.}c_k)$.

5. The linear form of the function polynomial $F(x)= F(x,c_{1,}c_{2,\ldots.}c_k)$ is defined as $F^*(x)= c_1^*\varphi_1(x) + c_2^*\varphi_2(x) + c_3^*\varphi_3(x) + \ldots + c_k^*\varphi_k(x)$ where ϕi :S→ R, and $\{\phi_i\}$ is a priori selected set of functions such that $\varphi_{i(x)}=x^{i-1}$ for i= 1,2,…,k.

6. If K>N, the said approximation allows the spread of errors, otherwise , no error propagates.

7. For obtaining the linear form of F(x) the computation of the values $c^*_1, c^*_2, c^*_3, \ldots, c^*_k$ is made as per the formula $c1^* = \frac{\sum f_i}{N}$, $c_2^* = \frac{\sum x_if_i}{\sum x_i^2}$ and $c^*_3 = [N\sum x_i^2f_i - \sum x_i^2\sum f_i]/ N\sum x_i^4-(\sum x_i^2)^2$

8. The deviations satisfy the following reoccurring formula $d_k= f_k - F(x, c_{1,}c_{2,\ldots.}c_k)$ and $d_k^1 = f_k - F(\phi i(x), c_1^*, c_{2,}^*\cdots c_k^*)$.

9. The average deviation can be calculated by the formula $\| d_k^1 \| = \sum_{k=1}^{N} d_k$

10. The elements of the matrix are to be obtained in the appropriate manner.

V. SQUARE MATRIX METHOD FOR THE INFORMATION REPRESENTATION

In this section we define another algorithm for the determination of square matrix that can represent the information. In fact this method has some similarity with the algorithm that has already discussed in Section 4. But the notations and some other concepts of the new algorithm are different that to the previous algorithm.

Consider the word $W= l_1(w)\, l_2(w) \ldots l_N(w)$ of finite sequence of length N, Where li∊{0,1}, for i= 1,2…N. Suppose $\Omega =\{0,1\}^N =\{0,1\}x\{0,1\}x\ldots\ldots\ldots.x\{0,1\}$.

N-3

We consider the shift map d:→ Ω such that $d(w)=l_2(w)l_3(w) \ldots\ldots l_N(w)$.The full shift of

N-3

N-symbols can be represented as (Ω,d).

In this way the restriction on the length can be interpreted.

Algorithm 5.1

This algorithm determines the creation of generating function that can be obtained from any known binary string as an input. The binary string is the digital representation of certain wave signal. The major scheme of this algorithm is interpreted in the following manner;

1. Obtain the decimal number from the input binary string.

2. Assume this number as the determinant of some square matrix.

3. The elements of this square matrix are the function of the coefficients of the polynomial function.

4. Express this polynomial function as a reoccurrence relation.

5. Use this reoccurrence relation as the sources of generation of other binary strings, which can be used for the resources of the pool of .

Let A =[A_{lk}], for l,k∈{1,2,… |$\sqrt{N}$|} be a square matrix of order

|$\sqrt{N}$|x|$\sqrt{N}$|. Let the elements of this matrix A be either zero or one. The grammars for creation of such matrix are defined as follows.

A_{lk}=0, if the words of length is written as l_n(w)l_{n+1}(w) =*lk, else* A_{lk} =1. Considered the set Ω_A ={w: $A(l_n(w), l_{n+1}(w))$ =1for l_n(w),l_{n+1}(w) ∈ {0,1}, for n∈{1,2,… |$\sqrt{N}$|} }. The system ((Ω_A, d /Ω_A) is called a sift of finite type. To obtain a full sift operator we assume the matrix A is a primitive matrix and there exists no such n, for which $A_{lk}^{n} \geq 0$.

VI. RELATION BETWEEN ENTROPY AND THE ERRORS IN INFORMATION QUANTIFICATIONS

The unavailable information in a set of documents is called entropy and the unavailability of information is usually occurred due to error or the lack of sufficient data. Thus estimated error and the entropy have a close relationship. One may be complementary of other. In the subsequent articles we describe the entropy and its application in the information processing.

VII. ENTROPY AND JOINT ENTROPY OF MUTUAL INFORMATION

In this section we are to discuss the entropy of an information generating function with three variables. Usually, the probability distribution of the entropy and the relevant mutual information are calculated from the finite amount of data. The process of such calculation are available in the research publication of Roulston [1997,1999], Crossberger [1998], Fraser[1998], Pompe [1993]. Most of the litrature discussed the entropy of N-points of a discrete one- dimensional variables.

$H_{obs}=H_{\infty}-\frac{c^*-1}{2N}$ ……………….. (7.1)

Where c^* is the number of states for which p_i the probability of i^{th} state is nonzero. Using the entropy formula expressed in equation 7.1, we are to analyze the errors of an observed mutual information in three dimensional space. The notation for this analysis are given below.

I_{obs}: observed mutual information.

P_{ijk}: probability of i^{th} , j^{th} and k^{th} states in x, y, and z axes

X_{lmn}: A variable point in three dimensional space.

The variance of the observed entropy H_{obs} is denoted as V [H_{obs}] . The standard error formula is considered as V[H_{obs}] = $\sum_{k=1}^{c}(\frac{\partial H_{obs}}{\partial n_k})$V[$n_k$], where $\frac{\partial H_{obs}}{\partial n_k}$ is the partial differentiation of the observed entropy with respect to the variable n_k . The joint entropy of three discrete variables x, y, and z is defined as follows.

H(x,y,z) = - $\sum_{i=1}^{c_x}\sum_{j=1}^{c_y}\sum_{k=1}^{c_z} p_{ijk} \log p_{ijk}$.

The sum over the c_x, c_y and c_z states number of variables along the three co-ordinate axes. The concerned joint entropy of the mutual information of three discrete variables x ,y and z can be defined as

I(x,y,z) = H(x)+H(y)+H(z)-H(x,y)-H(y,z)-H(z,x)+H(x,y,z) …………….(7.2)

If the hypothesis concerning the entropy and the joint entropy of the variables are to be tested, then the determination of uncertainty relating to the values of these quantities is to be calculated.

VIII. ESTIMATION OF ERRORS DURING THE QUANTIFICATION OF INFORMATION IN THE FUNCTION OF THREE VARIABLES

The systematic and random errors of an entropy function are calculated in the following manner. Suppose the value x_{lmn} is assigned to one of the variables states. The binomial distribution of expected values and the variance can be written in the following formula.

E[x_{ijk}] = Np_{ijk} and V[x_{ijk}]= Np_{ijk} (1-p_{ijk})……… (7.3)

The observed mutual information is

I_{obs}= H_{obs}(x) + H_{obs}(y) + H_{obs}(z)-H_{obs}(x,y) - H_{obs}(y,z) - H_{obs}(z,x) + H_{obs}(x,y,z)....(7.4)

If there are N-triplets (x,y,z), then the expected values of I_{obs} i9s given by I_{obs}= H_∞(x)-$\frac{c_x^*-1}{2N}$ + H_∞ (y)- $\frac{c_y^*-1}{2N}$ +H_∞ (z)- $\frac{c_z^*-1}{2N}$ -H_∞ (x,y)+ $\frac{c_x^*c_y^*-1}{2N^2}$ -H_∞ (y,z)+ $\frac{c_y^*c_z^*-1}{2N^2}$ -H_∞ (z,x)+ $\frac{c_z^*c_x^*-1}{2N^2}$ +H_∞ (x,y,z) -$\frac{c_x^*c_y^*c_z^*-1}{2N^3}$ ………….. (7.4)

I_{obs}= I_∞ -$\frac{c_x^*-1}{2N}$ -$\frac{c_y^*-1}{2N}$ - $\frac{c_z^*-1}{2N}$ -+ $\frac{c_x^*c_y^*-1}{2N^2}$ +$\frac{c_y^*c_z^*-1}{2N^2}$ + $\frac{c_z^*c_x^*-1}{2N^2}$ - $\frac{c_x^*c_y^*c_z^*-1}{2N^3}$

Where, I_∞ is the three mutual information measured when n → ∞, if c_{xy}^*, c_{yz}^* *and* c_{zx}^* are the number of states in the xy, yz, zx planes. Where the joint probabilities are non-zero. Similarly, if c_{xyz}^* is the number of states in the three dimensional space where the joint probability $p_{ijk} \neq 0$ then the joint probability can be expressed as $q_{ijk}=\frac{x_{ijk}}{N}$ …………..(7.5)

More over equation 7.5 can be expressed as

$$q_{ijk}=\frac{x_{ijk}}{\sum_{i=1}^{c_x}\sum_{j=1}^{c_y}\sum_{k=1}^{c_z}x_{ijk}} \quad \ldots\ldots\ldots\ldots\ldots\ldots.. \quad (7.6)$$

Differentiating partially with respect to x_{lmn} we have,

$$\frac{\partial q_{ijk}}{\partial x_{lmn}}=\frac{N\frac{\partial x_{ijk}}{\partial x_{lmn}}-x_{ijk}}{N^2}=\frac{\delta_{il}\delta_{kn}\delta_{jn}}{N}-\frac{q_{ijk}}{N}$$

The variations of Probability measures along X,Y and Z axes can be estimated by using the following partial differential equations

$$\frac{\partial}{\partial x_{lmn}}\left(\sum_{i=1}^{c_x}q_{ijk}\right)=\frac{\delta_{jm}\delta_{kn}}{N}-\sum_{i=1}^{c_x}\frac{q_{ijk}}{N} \ldots\ldots\ldots\ldots(7.7)$$

$$\frac{\partial}{\partial x_{lmn}}\left(\sum_{i=1}^{c_y}q_{ijk}\right)=\frac{\delta_{ij}\delta_{kn}}{N}-\sum_{j=1}^{c_y}\frac{q_{ijk}}{N} \ldots\ldots\ldots\ldots(7.8)$$

$$\frac{\partial}{\partial x_{lmn}}\left(\sum_{k=1}^{c_z}q_{ijk}\right)=\frac{\delta_{il}\delta_{jm}}{N}-\sum_{k=1}^{c_z}\frac{q_{ijk}}{N} \ldots\ldots\ldots\ldots(7.9)$$

The variations of the probability measures of the joint random variables along the XY, YZ and XZ planes are started as follows,

$$\frac{\partial}{\partial x_{lmn}}\left(\sum_{i=1}^{c_x}\sum_{j=1}^{c_y}q_{ijk}\right)=\frac{\delta_{kn}}{N}-\sum_{i=1}^{c_x}\sum_{j=1}^{c_y}\frac{q_{ijk}}{N}\ldots..(7.10)$$

$$\frac{\partial}{\partial x_{lmn}}\left(\sum_{j=1}^{c_y}\sum_{k=1}^{c_z}q_{ijk}\right)=\frac{\delta_{ij}}{N}-\sum_{j=1}^{c_y}\sum_{k=1}^{c_z}\frac{q_{ijk}}{N}\ldots..(7.11)$$

$$\frac{\partial}{\partial x_{lmn}}\left(\sum_{i=1}^{c_x}\sum_{k=1}^{c_z}q_{ijk}\right)=\frac{\delta_{jm}}{N}-\sum_{i=1}^{c_x}\sum_{k=1}^{c_z}\frac{q_{ijk}}{N}\ldots..(7.12)$$

The corresponding changes in the observed mutual information can be written as

$$I_{obs}=H(x)+H(y)+H(z)-H(x,y)-H(y,z)+H(z,x)+H(x,y,z)$$

$$-\sum_{i=1}^{c_x}\sum_{k=1}^{c_z}\left(\sum_{k=1}^{c_z}q_{ijk}\right)\ln\left(\sum_{k=1}^{c_z}q_{ijk}\right) - \sum_{i=1}^{c_x}\sum_{k=1}^{c_z}\left(\sum_{j=1}^{c_y}q_{ijk}\right)\ln\left(\sum_{j=1}^{c_y}q_{ijk}\right) - \sum_{i=1}^{c_x}\sum_{k=1}^{c_z}\left(\sum_{k=1}^{c_z}q_{ijk}\right)\ln\left(\sum_{k=1}^{c_z}q_{ijk}\right) + \sum_{i=1}^{c_x}\left(\sum_{j=1}^{c_y}\sum_{k=1}^{c_z}q_{ijk}\right)\ln\left(\sum_{j=1}^{c_y}\sum_{k=1}^{c_z}q_{ijk}\right)+ \sum_{j=1}^{c_y}\left(\sum_{i=1}^{c_x}\sum_{k=1}^{c_z}q_{ijk}\right)\ln\left(\sum_{i=1}^{c_x}\sum_{k=1}^{c_z}q_{ijk}\right)+ \sum_{k=1}^{c_z}\left(\sum_{i=1}^{c_x}\sum_{j=1}^{c_y}q_{ijk}\right)\ln\left(\sum_{i=1}^{c_x}\sum_{j=1}^{c_y}q_{ijk}\right) - \left(\sum_{i=1}^{c_x}\sum_{j=1}^{c_y}\sum_{k=1}^{c_z}q_{ijk}\right)\ln\left(\sum_{i=1}^{c_x}\sum_{j=1}^{c_y}\sum_{k=1}^{c_z}q_{ijk}\right) \ldots (7.13)$$

In order to estimate the errors in the observed information, again Equation 7.3 is to be differentiated and it yields the variations of the observed information function. This implies

$$\frac{\partial}{\partial x_{lmn}}(I_{obs})=-\sum_{j=1}^{c_y}\sum_{k=1}^{c_z}\left(1+\ln\left(\sum_{i=1}^{c_x}q_{ijk}\right)\right)\frac{\partial}{\partial x_{lmn}}\left(\sum_{i=1}^{c_x}q_{ijk}\right)-\sum_{i=1}^{c_x}\sum_{k=1}^{c_z}\left(1+\ln\left(\sum_{j=1}^{c_y}q_{ijk}\right)\right)\frac{\partial}{\partial x_{lmn}}\left(\sum_{j=1}^{c_y}q_{ijk}\right)-\sum_{i=1}^{c_x}\sum_{j=1}^{c_y}\left(1+\ln\left(\sum_{k=1}^{c_z}q_{ijk}\right)\right)\frac{\partial}{\partial x_{lmn}}\left(\sum_{k=1}^{c_z}q_{ijk}\right)+\sum_{i=1}^{c_x}\left(1+\ln\left(\sum_{j=1}^{c_y}\sum_{k=1}^{c_z}q_{ijk}\right)\right)\frac{\partial}{\partial x_{lmn}}\left(\sum_{j=1}^{c_y}\sum_{k=1}^{c_z}q_{ijk}\right)+$$

$$\sum_{j=1}^{c_y}\left(1+\ln\left(\sum_{i=1}^{c_x}\sum_{k=1}^{c_z}q_{ijk}\right)\right)\frac{\partial}{\partial x_{lmn}}\left(\sum_{i=1}^{c_x}\sum_{k=1}^{c_z}q_{ijk}\right)+$$

$$\sum_{k=1}^{c_z}\left(1+\ln\left(\sum_{i=1}^{c_x}\sum_{j=1}^{c_y}q_{ijk}\right)\right)\frac{\partial}{\partial x_{lmn}}\left(\sum_{i=1}^{c_x}\sum_{j=1}^{c_j}q_{ijk}\right)-\sum_{j=1}^{c_y}\sum_{j=1}^{c_y}\sum_{k=1}^{c_z}\left(1+\ln q_{ijk}\right)\frac{\partial}{\partial x_{mn}}\left(\sum_{i=1}^{c_x}\sum_{j=1}^{c_y}\sum_{k=1}^{c_z}q_{ijk}\right)\ldots\ldots\ldots\ldots..(7.14)$$

Relating all above mentioned equations we obtain

$$-\frac{\partial}{\partial x_{lmn}}(I_{obs})=\sum_{j=1}^{c_y}\sum_{k=1}^{c_z}\left(1+\ln\left(\sum_{i=1}^{c_x}q_{ijk}\right)\right)\left(\frac{\delta_{jm}\delta_{kn}}{N}-\sum_{i=1}^{c_x}\frac{q_{ijk}}{N}\right)+\sum_{i=1}^{c_x}\sum_{k=1}^{c_z}\left(1+\ln\left(\sum_{j=1}^{c_y}q_{ijk}\right)\right)\left(\frac{\delta_{il}\delta_{kn}}{N}-\sum_{j=1}^{c_y}\frac{q_{ijk}}{N}\right)+\sum_{i=1}^{c_x}\sum_{j=1}^{c_y}\left(1+\ln\left(\sum_{k=1}^{c_z}q_{ijk}\right)\right)\left(\frac{\delta_{il}\delta_{jm}}{N}-\sum_{k=1}^{c_z}\frac{q_{ijk}}{N}\right)-\sum_{i=1}^{c_x}\left(1+\ln\left(\sum_{j=1}^{c_y}\sum_{k=1}^{c_z}q_{ijk}\right)\right)\left(\frac{\delta_{il}}{N}-\sum_{j=1}^{c_y}\sum_{k=1}^{c_z}\frac{q_{ijk}}{N}\right)-\sum_{j=1}^{c_y}\left(1+\ln\left(\sum_{i=1}^{c_x}\sum_{k=1}^{c_z}q_{ijk}\right)\right)\left(\frac{\delta_{jm}}{N}-\sum_{i=1}^{c_x}\sum_{k=1}^{c_z}\frac{q_{ijk}}{N}\right)-\sum_{k=1}^{c_z}\left(1+\ln\left(\sum_{j=1}^{c_y}\sum_{k=1}^{c_z}q_{ijk}\right)\right)\left(\frac{\delta_{in}}{N}-\sum_{i=1}^{c_x}\sum_{j=1}^{c_y}\frac{q_{ijk}}{N}\right)+\sum_{i=1}^{c_x}\sum_{j=1}^{c_y}\sum_{k=1}^{c_z}\left(1+\ln q_{ijk}\right)\left(\frac{\delta_{ij}\delta_{jm}\delta_{kn}}{N}-\sum_{i=1}^{c_x}\sum_{j=1}^{c_y}\sum_{k=1}^{c_z}\frac{q_{ijk}}{N}\right)\ldots\ldots\ldots\ldots.(7.15)$$

Or

$$-\frac{\partial}{\partial x_{lmn}}(I_{obs})=\sum_{j=1}^{c_y}\sum_{k=1}^{c_z}\left(\frac{\delta_{jm}\delta_{kn}}{N}-\sum_{i=1}^{c_x}\frac{q_{ijk}}{N}\right)+\sum_{j=1}^{c_y}\sum_{k=1}^{c_z}ln\left(\sum_{i=1}^{c_x}q_{ijk}\right)\frac{\delta_{jm}\delta_{kn}}{N}-\sum_{j=1}^{c_y}\sum_{k=1}^{c_z}ln\left(\sum_{i=1}^{c_x}q_{ijk}\right)\sum_{j=1}^{c_y}\frac{q_{ijk}}{N}+\sum_{i=1}^{c_x}\sum_{k=1}^{c_z}\left(\frac{\delta_{il}\delta_{kn}}{N}-\sum_{j=1}^{c_y}\frac{q_{ijk}}{N}\right)+\sum_{j=1}^{c_y}\sum_{k=1}^{c_z}ln\left(\sum_{i=1}^{c_x}q_{ijk}\right)\frac{\delta_{il}\partial_{kn}}{N}-\sum_{i=1}^{c_x}\sum_{k=1}^{c_z}ln\left(\sum_{j=1}^{c_y}q_{ijk}\right)\sum_{j=1}^{c_y}\frac{q_{ijk}}{N}+\sum_{i=1}^{c_x}\sum_{j=1}^{c_y}\left(\frac{\delta_{il}\delta_{jm}}{N}-\sum_{k=1}^{c_z}\frac{q_{ijk}}{N}\right)+\sum_{i=1}^{c_x}\sum_{j=1}^{c_y}ln\left(\sum_{k=1}^{c_z}q_{ijk}\right)\frac{\delta_{il}\delta_{jm}}{N}-\sum_{i=1}^{c_x}\sum_{j=1}^{c_y}ln\left(\sum_{j=1}^{c_y}q_{ijk}\right)-\sum_{k=1}^{c_z}\frac{q_{ijk}}{N}-\sum_{i=1}^{c_x}\frac{\delta_{il}}{N}+\sum_{i=1}^{c_y}\sum_{j=1}^{c_j}\sum_{k=1}^{c_k}\frac{q_{ijk}}{N}-\sum_{i=1}^{c_x}ln\left(\sum_{j=1}^{c_y}\sum_{k=1}^{c_z}q_{ijk}\right)\frac{\delta_{il}}{N}+\sum_{i=1}^{c_x}ln\left(\sum_{j=1}^{c_y}\sum_{k=1}^{c_z}q_{ijk}\right)\left(\sum_{j=1}^{c_y}\sum_{k=1}^{c_z}\frac{q_{ijk}}{N}\right)-\sum_{k=1}^{c_z}\frac{\delta_{kn}}{N}+\sum_{i=1}^{c_x}\sum_{j=1}^{c_y}\sum_{k=1}^{c_z}\frac{q_{ijk}}{N}-\sum_{k=1}^{c_z}ln\left(\sum_{i=1}^{c_x}\sum_{j=1}^{c_y}q_{ijk}\right)\frac{\delta_{kn}}{N}+\sum_{k=1}^{c_z}ln\left(\sum_{j=1}^{c_y}\sum_{k=1}^{c_z}q_{ijk}\right)\left(\sum_{i=1}^{c_x}\sum_{j=1}^{c_y}\frac{q_{ijk}}{N}\right)-\sum_{t=1}^{c_z}\frac{\delta_{ln}}{N}+\sum_{i=1}^{c_x}\sum_{j=1}^{c_y}\sum_{k=1}^{c_z}\frac{q_{ijk}}{N}-\sum_{k=1}^{c_z}ln\left(\sum_{i=1}^{c_x}\sum_{j=1}^{c_y}q_{ijk}\right)\frac{\delta_{ln}}{N}+\sum_{k=1}^{c_z}ln\left(\sum_{j=1}^{c_y}\sum_{k=1}^{c_z}q_{ijk}\right)\left(\sum_{i=1}^{c_x}\sum_{k=1}^{c_y}\frac{q_{ijk}}{N}\right)+\sum_{i=1}^{c_x}\sum_{j=1}^{c_y}\sum_{k=1}^{c_z}\frac{\delta_{il}\delta_{jm}\delta_{kn}}{N}-\sum_{i=1}^{c_x}\sum_{j=1}^{c_y}\sum_{k=1}^{c_z}\frac{q_{ijk}}{N}+\left(\sum_{i=1}^{c_x}\sum_{j=1}^{c_y}\sum_{k=1}^{c_z}\ln q_{ijk}\right)\frac{\delta_{il}\delta_{jm}\delta_{kn}}{N}-\left(\sum_{i=1}^{c_x}\sum_{j=1}^{c_y}\sum_{k=1}^{c_z}\ln q_{ijk}\right)\frac{q_{ijk}}{N}$$

On simplification of the above equation the following equation is an approximation to the obtained equation

$$-\frac{\partial}{\partial x_{lmn}}(I_{obs})=\frac{1}{N}\left(\ln\left(\sum_{k=1}^{c_z}q_{ijk}\right)+\ln\left(\sum_{j=1}^{c_y}q_{ijk}\right)+\ln\left(\sum_{i=1}^{c_x}q_{ijk}\right)-\ln\left(\sum_{j=1}^{c_y}\sum_{k=1}^{c_z}q_{ijk}\right)-\ln\left(\sum_{i=1}^{c_x}\sum_{k=1}^{c_z}q_{ijk}\right)-\ln\left(\sum_{j=1}^{c_y}\sum_{i=1}^{c_x}q_{ijk}\right)+\ln q_{lmn}\right.$$

The standard errors caused in this approximation can be formulated in the following way

$$V[I_{obs}]=\sum_{i=1}^{c_x}\sum_{j=1}^{c_y}\sum_{k=1}^{c_z}\left(\frac{\partial I_{obs}}{\partial x_{lmn}}\right)^2 V[x_{lmn}]$$

It can be simplified in to the following expressions

$V[l_{obs}] = \sum_{i=1}^{c_x}\sum_{j=1}^{c_y}\sum_{k=1}^{c_z}(\frac{1}{N})^2($ $(\ln(\sum_{k=1}^{c_z} q_{ijk})+$ $\ln(\sum_{j=1}^{c_y} q_{ijk})+\ln(\sum_{i=1}^{c_x} q_{ijk})-\ln(\sum_{j=1}^{c_y}\sum_{k=1}^{c_z} q_{ijk})-\ln$ $(\sum_{i=1}^{c_x}\sum_{k=1}^{c_z} q_{ijk}) - \ln(\sum_{j=1}^{c_y}\sum_{i=1}^{c_x} q_{ijk})+$ $] = \sum_{i=1}^{c_x}\sum_{j=1}^{c_y}\sum_{k=1}^{c_z} q_{ijk} + I_{obs})^2 * V[x_{lmn}]$

The variation of the arbitrary point x_{ijk} can be estimated from the observed distribution, which is formulated in the following manner

$V[x_{lmn}]=Nq_{lmn}(1-q_{lmn}) +\Theta(\epsilon_{lmn})$

The estimated errors during quantification and the representation of information in the distribution function consisting of three random variable form are discussed in this action.

IX. APPLICATION

Let the set of alphabet be $\Lambda=\{a_1,a_2\}$. The binary basis of the set is $\Lambda=\{0,1\}$. The probability measure for the use of the symbols a_1 and a_2 are $p(a_1)$ and $p(a_2)$ respectively. The binary entropy function is $H(x)=-p(a_1)\log p(a_2)$. If $p(a_1)=\frac{1}{2}$, the probable errors occurred during the information transmission over a noisy binary information channels can be calculated in the following manner. If the channel is symmetric that means the outcome of the low and high voltages are equally likely, then this channel is called BSC or binary symmetric channel and the respective matrix is Q= $\begin{bmatrix} 1-p_1 & p_1 \\ p_1 & 1-p_1 \end{bmatrix}$

The probability of receiving output matrix

$B=\{b_1,b_2\} = \{0,1\}$

X. CONCLUSION

This paper describes the entropy for more than three variables is discussed and this theory can applied in the information transmission. The algorithm can express requisite information square matrix in the appropriate manner. This application gives the better clarification of this paper.

REFERENCES

[1] W. P. Dittrich Banzhaf, and B. Eller, self organisation in a binary string with spatial interactions, Physica- D, 125, 1999, P 85-104.

[2] M.S. Roulstone, Estimating error on measured entropy and mutual information Physica-D, 125, 1999, P 285-294.

[3] J. R. Chozottes, E.Floriani and R lima, Relative entropy and indification of Gibbes measures in dynamical systems, journal of Stat. Physics, 90, 1998 P. 697-725.

[4] P. Grassberger, Finite sample correlations to entropy and dimension estimates. Physics letters A, 128, 1998, P369-373.

[5] M.S. Roulstone, Significance testing of information theoric functional, Physica- D, 110, 1997, P, 62-66.

[6] Y.C. Tian, and F. Gao, Extraction at delay information from chaotic time series based on information entropy, Physica-D, 108, 1997, P113-118.

[7] E. Kreyszig, Advanced Engineering Mathematics, 8th Edition, Wiely India Edition, 1999.

[8] Shunsuke Ihare, Information Theory for Continuous System, World Scientific Publishing Co Pvt. Ltd., Singapore, 1993.

[9] Leon Brillouin, Science and Information Theory, Member of the National Academic of Science, Academic Press London, Second Edition.

◈◈◈