

Some Coding Theorems on Fuzzy Entropy Function Depending Upon Parameter R and V

M.A.K. Baig and Mohd Javid Dar

P.G. Department of Statistics University of Kashmir, Hazratbal, Srinagar-190006 (INDIA)

Abstract: In the literature of information theory several types of coding theorems involving fuzzy entropy functions exists. In this paper, some new fuzzy coding theorems have been obtained involving utilities. The fuzzy coding theorems obtained here are not only new but also generalizes some well known results available in the literature.

Key Words: Fuzzy Set, Fuzzy Entropy Function, Fuzzy Useful Entropy Function, Fuzzy code word length and Fuzzy average useful codeword length.

AMS Subject Classification: 60E15, 62N05, 90B25, 94A17, 94A24.

I. Introduction

The notion of fuzzy sets was proposed by Zadeh [1965] with a view to tackling problems in which indefiniteness arising from a sort of intrinsic ambiguity plays a significant role. Fuzziness, a feature of uncertainty, results from the lack of sharp distinction of the boundary of a set, i.e., an individual is neither definitely a member of the set nor definitely not a member of it. The first to qualify the fuzziness was made by Zadeh [1968], who based on probabilistic framework introduced the entropy combining probability and membership function of a fuzzy event as weighted Shannon entropy.

Definition 1: Let $X = (x_1, x_2, \dots, x_n)$ be a discrete universe of discourse. A fuzzy set 'A' on X is defined by a characteristic function $\mu_A(x_i) = \{x_1, x_2, x_3, \dots, x_n\} \rightarrow [0,1]$. The value of $\mu_A(x)$ of A at $x \in X$ stands for the degree of membership of x in A. If every element of the set 'A' is '0' or '1', there is no uncertainty about it and a set is said to be crisp set.

Definition 2: A fuzzy set A^* is called the sharpened version of fuzzy set A, if the following conditions are satisfied.

$$\mu_{A^*}(x) \leq \mu_A \quad \text{if } \mu_A(x) \leq 0.5; \quad \forall i$$

and

$$\mu_{A^*}(x) \geq \mu_A(x), \quad \text{if } \mu_A(x) \geq 0.5; \quad \forall i$$

De Luca and Termini [1972] formulated a set of four properties and these properties are widely accepted as criterion for defining any fuzzy entropy. In fuzzy set theory, the entropy is a measure of fuzziness which expresses the amount of average ambiguity in making a decision whether an element belongs to a set or not. So, a measure of average fuzziness in fuzzy set $H(A)$ should have the following properties to be valid fuzzy entropy.

P₁(Sharpness): $H(A)$ is minimum if and only if A is a crisp set i.e., $\mu_A(x) = 0$ or $1; \quad \forall x$.

P₂(Maximality): $H(A)$ is maximum if and only if A is most fuzzy set i.e., $\mu_A(x) = 0.5; \quad \forall x$.

P₃(Resolution): $H(A) \geq H(A^*)$, where A^* is sharpened version of A.

P₄(Symmetry): $H(A) = H(\bar{A})$, where \bar{A} is the complement of A i.e., $\mu_{\bar{A}}(x_i) = 1 - \mu_A(x_i)$.

II. Basic Concepts

Let X be discrete random variable taking on a finite number of possible values $X = (x_1, x_2, \dots, x_n)$ with respective membership function $A = \{\mu_A(x_1), \mu_A(x_2), \dots, \mu_A(x_n)\} \rightarrow [0,1]$, $\mu_A(x_i)$ give of the element s the degree of belongingness x_i to the set A. The function $\mu_A(x_i)$ associates with each $x_i \in R^n$ a grade of membership to the set A and is known as membership function.

Denote

$$X = \begin{bmatrix} x_1 & x_2 & \dots & x_n \\ \mu_A(x_1) & \mu_A(x_2) & \dots & \mu_A(x_n) \end{bmatrix} \quad (2.1)$$

We call the scheme (2.1) as a finite fuzzy information scheme. Every finite scheme describes a state of uncertainty. De Luca and Termini [1972] introduced a quantity which, in a reasonable way to measures the amount of uncertainty (fuzzy entropy) associated with a given finite scheme. This measure is given by

$$H(A) = - \sum_i^n [\mu_A(x_i) \log \mu_A(x_i) + (1 - \mu_A(x_i)) \log(1 - \mu_A(x_i))] \quad (2.2)$$

The measure (2.2) serve as a very suitable measure of fuzzy entropy of the finite information scheme(2.1).

Let a finite source of n source symbols $X = (x_1, x_2, \dots, x_n)$ be encoded using alphabet of D symbols, then it has been shown by Feinstein [1958] that there is a uniquely decipherable/ instantaneous code with lengths l_1, l_2, \dots, l_n iff the following Kraft [1949] inequality is satisfied

$$\sum_i^n D^{-l_i} \leq 1 \tag{2.3}$$

Belis and Guiasu [1968] observed that a source is not completely specified by the probability distribution P over the source alphabet X in the absence of qualitative character. So it can be assumed (Belis and Guiasu [1968]) that the source alphabet letters are assigned weights according to their importance or utilities in view of the experimenter.

Let $U = (u_1, u_2, \dots, u_n)$ be the set of positive real numbers, u_i is the utility or importance of x_i . The utility, in general, is independent of probability of encoding of source symbol x_i , i.e. p_i . The information source is thus given by

$$X = \begin{bmatrix} X_1 & X_2 & \dots & X_n \\ p_1 & p_2 & \dots & p_n \\ u_1 & u_2 & \dots & u_n \end{bmatrix}, u_i > 0, p_i \geq 0, \sum_i^n p_i = 1 \tag{2.4}$$

Belis and Guiasu [1968] introduced the following quantitative- qualitative measure of information

$$H(P, U) = - \sum_i^n u_i p_i \log p_i \tag{2.5}$$

Which is a measure for the average of quantity of ‘variable’ or ‘useful’ information provided by the information source(2.4).

Guiasu and Picard [1971] considered the problem of encoding the letter output by the source (2.4) by means of a single letter prefix code whose codeword’s c_1, c_2, \dots, c_n are of lengths l_1, l_2, \dots, l_n respectively and satisfy the Kraft’s inequality(2.3), they included the following ‘useful’ mean length of the code

$$L(U) = \frac{\sum_i^n u_i p_i l_i}{\sum_i^n u_i p_i} \tag{2.6}$$

Further they derived a lower bound for (2.6).

Now, corresponding to (2.5) and(2.6), we have the following fuzzy measures

$$H(A, U) = - \sum_{i=1}^n u_i \{ \mu_A(x_i) + (1 - \mu_A(x_i)) \} \log \{ \mu_A(x_i) + (1 - \mu_A(x_i)) \} \tag{2.7}$$

and

$$L(U) = \frac{\sum_{i=1}^n u_i \{ \mu_A(x_i) + (1 - \mu_A(x_i)) \} l_i}{\sum_{i=1}^n u_i \{ \mu_A(x_i) + (1 - \mu_A(x_i)) \}} \tag{2.8}$$

respectively.

In the next section, fuzzy coding theorem have been obtained by considering a new parametric fuzzy entropy function involving utilities and generalized useful fuzzy code word mean length. The result obtained here are not only new, but also generalized some well known results available in the literature of information theory. .

III. Fuzzy Coding Theorem on New Fuzzy Parametric Entropy Function

Consider a function

$$H_R(\mu_A^v(x)_i + (1 - \mu_A(x_i))^v, U) = \frac{R}{R-1} \left[1 - \left(\frac{\sum_{i=1}^n u_i (\mu_A^{R+v-1}(x_i) + (1-\mu_A(x_i))^{R+v-1})}{\sum_{i=1}^n u_i (\mu_A^v(x_i) + (1-\mu_A(x_i))^v)} \right)^{\frac{1}{R}} \right] \tag{3.1}$$

Where $R > 0 (\neq 1), v > 0, \sum_{i=1}^n (\mu_A(x_i) + (1 - \mu_A(x_i))) = 1, (\mu_A(x_i) + (1 - \mu_A(x_i))) \geq 0$.

Remark (3.1).

- (1) When $v = 1$, (3.1) reduces to the ‘useful’ R-norm fuzzy information measure corresponding to Singh, Kumar and Tuteja [2003].
- (2) When $v = 1, u_i = 1 \forall i = 1, 2, \dots, n$, (3.1) reduces to the R-norm fuzzy informatin measure corresponding to Boekee and Lubbee [1980].
- (3) When $R \rightarrow 1, v = 1$ and $u_i = 1 \forall i = 1, 2, \dots, n$, (3.1) reduces to the De Luca and Termini [1972] measure of fuzzy entropy corresponding to the Shannon [1948] measure of entropy.

Further, consider a generalized ‘useful’ codeword mean length

$$L_R(\mu_A^v(x)_i + (1 - \mu_A(x_i))^v, U) = \left[1 - \frac{\sum_{i=1}^n u_i (\mu_A^v(x_i) + (1-\mu_A(x_i))^v) D^{-l_i (\frac{R-1}{R})}}{\sum_{i=1}^n u_i (\mu_A^v(x_i) + (1-\mu_A(x_i))^v)} \right] \tag{3.2}$$

Where $R > 0 (\neq 1), v > 0, \sum_{i=1}^n (\mu_A(x_i) + (1 - \mu_A(x_i))) = 1, (\mu_A(x_i) + (1 - \mu_A(x_i))) \geq 0$. D is the size of the code alphabet.

Remark (3.2).

- (1) When $v = 1$, (3.2) reduces to the fuzzy ‘useful’ codeword mean length corresponding to the Sing, Kumar and Tuteja [2003].
- (2) When $v = 1$, $u_i = 1 \forall i = 1, 2, \dots, n$, (3.2) reduces to the fuzzy codeword mean length corresponding to the Boekee et al [1980].
- (3) when $R \rightarrow 1$, $v = 1$ and $u_i = 1 \forall i = 1, 2, \dots, n$, (3.2) reduces to the fuzzy optimal codeword mean length corresponding to the Shannon [1948]

We now establish a result that in a sense, gives a characterization of

$$H_R(\mu_A^v(x)_i + (1 - \mu_A(x_i))^v, U) \text{ Under the condition}$$

$$\sum_{i=1}^n u_i (\mu_A^{v-1}(x_i) + (1 - \mu_A(x_i))^{v-1}) D^{-l_i} \leq \sum_{i=1}^n u_i (\mu_A^v(x_i) + (1 - \mu_A(x_i))^v) \quad (3.3)$$

Remark (3.3).

When $v = 1$, $u_i = 1 \forall i = 1, 2, \dots, n$, and $\sum_{i=1}^n (\mu_A(x_i) + (1 - \mu_A(x_i))) = 1$, (3.3) is a generalization of (2.3) which is Kraft’s [1949] inequality.

Theorem (3.1). For every code whose lengths l_1, l_2, \dots, l_n satisfies (3.3), then the average fuzzy codeword length satisfies

$$L_R(\mu_A^v(x)_i + (1 - \mu_A(x_i))^v, U) \geq H_R(\mu_A^v(x)_i + (1 - \mu_A(x_i))^v, U) \quad (3.4)$$

Equality holds iff

$$l_i = -\log(\mu_A^R(x_i) + (1 - \mu_A(x_i))^R) + \log \frac{\sum_{i=1}^n u_i (\mu_A^{R+v-1}(x_i) + (1 - \mu_A(x_i))^{R+v-1})}{\sum_{i=1}^n u_i (\mu_A^v(x_i) + (1 - \mu_A(x_i))^v)} \quad (3.5)$$

Proof: By holder’s inequality [1967]

$$\sum_{i=1}^n x_i y_i \geq \left(\sum_{i=1}^n x_i^p \right)^{\frac{1}{p}} \left(\sum_{i=1}^n y_i^q \right)^{\frac{1}{q}} \quad (3.6)$$

$\forall x_i, y_i > 0, i = 1, 2, \dots, n$ and $\frac{1}{p} + \frac{1}{q} = 1, p < 1 (\neq 0), q < 0$ or $q < 1 (\neq 0), p < 0$. We see the equality holds iff there exists a positive constant c such that

$$x_i^p = c y_i^q \quad (3.7)$$

Setting

$$x_i = u_i^{\frac{R}{R-1}} \left(\mu_A^{\frac{vR}{R-1}}(x_i) + (1 - \mu_A(x_i))^{\frac{vR}{R-1}} D^{-l_i} \right), \quad y_i = u_i^{\frac{1}{1-R}} \left(\mu_A^{\frac{R+v-1}{1-R}}(x_i) + (1 - \mu_A(x_i))^{\frac{R+v-1}{1-R}} D^{-l_i} \right),$$

$$p = \frac{R-1}{R} \text{ and } q = 1 - R$$

In (3.6) and using (3.3), we get

$$\left[\sum_{i=1}^n u_i (\mu_A^v(x)_i + (1 - \mu_A(x_i))^v) D^{-l_i \left(\frac{R-1}{R} \right)} \right]^{\frac{R}{1-R}} \geq \frac{\left[\sum_{i=1}^n u_i (\mu_A^{R+v-1}(x)_i + (1 - \mu_A(x_i))^{R+v-1}) \right]^{\frac{1}{1-R}}}{\sum_{i=1}^n u_i (\mu_A^v(x)_i + (1 - \mu_A(x_i))^v)} \quad (7.8)$$

Dividing both sides of (3.8) by $(\sum_{i=1}^n u_i (\mu_A^v(x)_i + (1 - \mu_A(x_i))^v))^{\frac{R}{1-R}}$, we get

$$\left[\frac{\sum_{i=1}^n u_i (\mu_A^v(x)_i + (1 - \mu_A(x_i))^v) D^{-l_i \left(\frac{R-1}{R} \right)}}{\sum_{i=1}^n u_i (\mu_A^v(x)_i + (1 - \mu_A(x_i))^v)} \right]^{\frac{R}{1-R}} \geq \left[\frac{\sum_{i=1}^n u_i (\mu_A^{R+v-1}(x)_i + (1 - \mu_A(x_i))^{R+v-1})}{\sum_{i=1}^n u_i (\mu_A^v(x)_i + (1 - \mu_A(x_i))^v)} \right]^{\frac{1}{1-R}}$$

Taking $0 < R < 1$, raising both sides to the power $\frac{1-R}{R}, R \neq 1$, also $\frac{R}{R-1} < 0$ for $0 < R < 1$ and after suitable operations, we obtain the result (3.4). For $R > 1$, the inequality (2.3.4) can be obtained in a similar fashion.

Theorem (3.2). For every code with lengths l_1, l_2, \dots, l_n satisfies (3.3). Then

$L_R(\mu_A^v(x)_i + (1 - \mu_A(x_i))^v, U)$ can be made to satisfy the inequality

$$L_R(\mu_A^v(x)_i + (1 - \mu_A(x_i))^v, U) < H_R(\mu_A^v(x)_i + (1 - \mu_A(x_i))^v, U) D^{\left(\frac{R-1}{R} \right)} + \frac{R}{R-1} \left(1 - D^{\left(\frac{R-1}{R} \right)} \right) \quad (3.9)$$

Proof: Let l_i be the positive integer satisfying the inequality

$$\begin{aligned}
 & -\log \left(\mu_A^R(x_i) + (1 - \mu_A(x_i))^R \right) + \log \frac{\sum_{i=1}^n u_i \left(\mu_A^{R+v-1}(x_i) + (1 - \mu_A(x_i))^{R+v-1} \right)}{\sum_{i=1}^n u_i \left(\mu_A^v(x_i) + (1 - \mu_A(x_i))^v \right)} \leq l_i < \\
 & \quad -\log \left(\mu_A^R(x_i) + (1 - \mu_A(x_i))^R \right) + \log \frac{\sum_{i=1}^n u_i \left(\mu_A^{R+v-1}(x_i) + (1 - \mu_A(x_i))^{R+v-1} \right)}{\sum_{i=1}^n u_i \left(\mu_A^v(x_i) + (1 - \mu_A(x_i))^v \right)} + \\
 & \quad 1 \tag{3.10}
 \end{aligned}$$

Consider the interval

$$\delta_i = \left[\begin{aligned} & -\log \left(\mu_A^R(x_i) + (1 - \mu_A(x_i))^R \right) + \log \frac{\sum_{i=1}^n u_i \left(\mu_A^{R+v-1}(x_i) + (1 - \mu_A(x_i))^{R+v-1} \right)}{\sum_{i=1}^n u_i \left(\mu_A^v(x_i) + (1 - \mu_A(x_i))^v \right)}, \\ & -\log \left(\mu_A^R(x_i) + (1 - \mu_A(x_i))^R \right) + \log \frac{\sum_{i=1}^n u_i \left(\mu_A^{R+v-1}(x_i) + (1 - \mu_A(x_i))^{R+v-1} \right)}{\sum_{i=1}^n u_i \left(\mu_A^v(x_i) + (1 - \mu_A(x_i))^v \right)} + 1 \end{aligned} \right] \tag{3.11}$$

of length 1. In every δ_i , there lies exactly one positive integer l_i such that

$$\begin{aligned}
 & 0 < -\log \left(\mu_A^R(x_i) + (1 - \mu_A(x_i))^R \right) + \log \frac{\sum_{i=1}^n u_i \left(\mu_A^{R+v-1}(x_i) + (1 - \mu_A(x_i))^{R+v-1} \right)}{\sum_{i=1}^n u_i \left(\mu_A^v(x_i) + (1 - \mu_A(x_i))^v \right)} \leq l_i < \\
 & -\log \left(\mu_A^R(x_i) + (1 - \mu_A(x_i))^R \right) + \log \frac{\sum_{i=1}^n u_i \left(\mu_A^{R+v-1}(x_i) + (1 - \mu_A(x_i))^{R+v-1} \right)}{\sum_{i=1}^n u_i \left(\mu_A^v(x_i) + (1 - \mu_A(x_i))^v \right)} + 1 \tag{3.12}
 \end{aligned}$$

We will first show that the sequence $\{l_1, l_2, \dots, l_n\}$, thus defined satisfies (3.3).

From (3.12), we have

$$0 < -\log \left(\mu_A^R(x_i) + (1 - \mu_A(x_i))^R \right) + \log \frac{\sum_{i=1}^n u_i \left(\mu_A^{R+v-1}(x_i) + (1 - \mu_A(x_i))^{R+v-1} \right)}{\sum_{i=1}^n u_i \left(\mu_A^v(x_i) + (1 - \mu_A(x_i))^v \right)} \leq l_i$$

or

$$\begin{aligned}
 & -\log \frac{\mu_A^R(x_i) + (1 - \mu_A(x_i))^R}{\left(\frac{\sum_{i=1}^n u_i \left(\mu_A^{R+v-1}(x_i) + (1 - \mu_A(x_i))^{R+v-1} \right)}{\sum_{i=1}^n u_i \left(\mu_A^v(x_i) + (1 - \mu_A(x_i))^v \right)} \right)} \leq -\log_D D^{-l_i} \\
 & \frac{\mu_A^R(x_i) + (1 - \mu_A(x_i))^R}{\left(\frac{\sum_{i=1}^n u_i \left(\mu_A^{R+v-1}(x_i) + (1 - \mu_A(x_i))^{R+v-1} \right)}{\sum_{i=1}^n u_i \left(\mu_A^v(x_i) + (1 - \mu_A(x_i))^v \right)} \right)} \geq D^{-l_i} \tag{3.13}
 \end{aligned}$$

Multiplying both sides $\sum_{i=1}^n u_i \left(\mu_A^{v-1}(x_i) + (1 - \mu_A(x_i))^{v-1} \right)$ and summing over $i = 1, 2, \dots, n$, we get (3.3). The last inequality in (3.12) gives

$$\begin{aligned}
 & l_i < -\log \left(\mu_A^R(x_i) + (1 - \mu_A(x_i))^R \right) + \log \frac{\sum_{i=1}^n u_i \left(\mu_A^{R+v-1}(x_i) + (1 - \mu_A(x_i))^{R+v-1} \right)}{\sum_{i=1}^n u_i \left(\mu_A^v(x_i) + (1 - \mu_A(x_i))^v \right)} + 1 \\
 & l_i < -\log \left(\mu_A^R(x_i) + (1 - \mu_A(x_i))^R \right) + \log \frac{\sum_{i=1}^n u_i \left(\mu_A^{R+v-1}(x_i) + (1 - \mu_A(x_i))^{R+v-1} \right)}{\sum_{i=1}^n u_i \left(\mu_A^v(x_i) + (1 - \mu_A(x_i))^v \right)} + \log_D D
 \end{aligned}$$

or

$$\begin{aligned}
 & l_i < -\log \frac{\left(\mu_A^R(x_i) + (1 - \mu_A(x_i))^R \right)}{\frac{\sum_{i=1}^n u_i \left(\mu_A^{R+v-1}(x_i) + (1 - \mu_A(x_i))^{R+v-1} \right)}{\sum_{i=1}^n u_i \left(\mu_A^v(x_i) + (1 - \mu_A(x_i))^v \right)}} + \log_D D \\
 & D^{-l_i} > \frac{\left(\mu_A^R(x_i) + (1 - \mu_A(x_i))^R \right)}{\frac{\sum_{i=1}^n u_i \left(\mu_A^{R+v-1}(x_i) + (1 - \mu_A(x_i))^{R+v-1} \right)}{\sum_{i=1}^n u_i \left(\mu_A^v(x_i) + (1 - \mu_A(x_i))^v \right)}} D^{-1}
 \end{aligned}$$

Taking $0 < R < 1$ and raising both sides to the power $\frac{R-1}{R}$, we get

$$D^{-l_i \left(\frac{R-1}{R} \right)} < \left(\frac{\left(\mu_A^R(x_i) + (1 - \mu_A(x_i))^R \right)}{\frac{\sum_{i=1}^n u_i \left(\mu_A^{R+v-1}(x_i) + (1 - \mu_A(x_i))^{R+v-1} \right)}{\sum_{i=1}^n u_i \left(\mu_A^v(x_i) + (1 - \mu_A(x_i))^v \right)}} \right)^{\frac{R-1}{R}} D^{\frac{R-1}{R}}$$

Multiplying both sides by $\frac{u_i \left(\mu_A^v(x_i) + (1 - \mu_A(x_i))^v \right)}{\sum_{i=1}^n u_i \left(\mu_A^v(x_i) + (1 - \mu_A(x_i))^v \right)}$ and summing over $i = 1, 2, \dots, n$ and after simplifying, gives (3.9).

For $R > 1$, the proof follows along the similar lines.

Reference:

- [1]. **Belis, M. and Guiasu, S. [1968]:** ‘A quantitative and qualitative measure of information in cybernetic system’, IEEE Transaction on information theory, Vol.IT-14, pp. 593-594.
- [2]. **Boekee, E. and Van Der Lubbe, J.C.A. [1980]:** “The R- Norm Information measure”, Information and Control, Vol. 45, pp. 136-155.
- [3]. **De Luca and S. Termini [1972]:** A Definition of Non-probabilistic Entropy in the Setting of fuzzy sets theory, Information and Control, Vol.20, pp. 301-312.
- [4]. **Guiasu, S. and Picard,C.F. [1971]:** Borne inferieure de la longer de certain codes, C.R. Academic Sciences, Paris,Vol. 273, pp. 248-251
- [5]. **Hooda, D. S. and Ram, A. [1998]:** ‘Characterization of Non-additive ‘useful’ information measure’, Recent Advances in information theory, Statistics and Computer applications, CCC Haryana, Agriculture University, Hisar, pp. 64-77
- [6]. **Kraft, L.J. [1949]:** “A device for quantizing grouping and coding amplitude modulates pulses”, M. S. Thesis, Department of Electrical Engineering, MIT, Cambridge.
- [7]. **Longo,G. [1976]:** “A noiseless coding theorem for source having utilities”, SIAM Journal of Applied Mathematics, Vol.30, pp.739-748
- [8]. **Shannon, C.E. [1948]:** The Mathematical theory of Communication, Bell Syst. Tech. Journal, Vol.27, pp.423-467.
- [9]. **Shisha, O. [1967]:** “Inequalities”, Academic press, New York.
- [10]. **Sing, R.P., Kumar, R. and Tuteja, R.K. [2003]:** “Application of Holder’s inequality in information theory” ,Information Sciences, Vol. 152, pp. 145-154
- [11]. **Zadeh, L. A. [1965]:** Fuzzy sets. Information and Control, Vol.8, pp. 338-353.
- [12]. **Zadeh, L. A. [1968]:** Probability measures of fuzzy events. Journal of Mathematical Analysis and Applications, Vol.23, pp.421-427.