International Journal of Pure and Applied Mathematics
Volume 2 No. 2 2002, 213-226

UNCERTAINTY MEASURES OF TYPE β UNDER SIMILARITY RELATIONS

Gurdial Arora[1], Frederick Petry[2 §], Theresa Beaubouef[3]

[1]Dept. of Mathematics
Xavier University of Louisiana
New Orleans, LA 70125, USA
e-mail: gdial@xula.edu

[2]Center for Intelligence and Knowledge - Based Systems
Electrical Engineering and Computer Science
Tulane University
204 Stanley Thomas Hall
New Orleans, Louisiana 70118-5674, USA
e-mail: petry@eecs.tulane.edu

[3]Dept. of Computer Science
Southeastern Louisiana University, Hammond, LA 70402, USA
e-mail: tbeaubouef@selu.edu

Abstract: In this paper, we will consider the generalization of the information measure of type β under similarity relation. In other words the generalization will be considered under the situations in which the underlying objects of concern have a similarity relationship and a probability distribution is also defined on the objects. The results will generalize the results in the case of Shannon's entropy proved by Yager.

AMS Subject Classification: 94A17
Key Words: entropy, similarity relation, fuzzy subsets, measures of fuzziness

Received: June 5, 2002

[§]Correspondence author

1. Introduction

The concept of entropy or uncertainty plays an important role in the field of information theory. Information theory and the concept of entropy or uncertainty have found applications in diverse fields such as: electrical engineering, psychology, biology, economics, social sciences, ecology, statistics, computer science and fuzzy set theory among other fields. Shannon's entropy [10] is formulated in terms of a probability distribution. In fact Shannon proved [1] that the measure of the amount of information, Shannon's entropy, can be uniquely determined by some rather natural properties or postulates. Also we can say that entropy measures, essentially, the degree of uncertainty associated with a probability distribution.

Assume that there are n outcomes of a random experiment and the i-th outcome can occur with probability p_i. In other words, let $P = \{p_1, p_2, ..., p_n\}$ be the probability distribution. Then Shannon's measure of uncertainty or entropy, denoted by $H(P)$, is defined as

$$H(P) = -\sum_{i=1}^{n} p_i \ln p_i.$$

Renyi [9] considered the problem of finding other measures of uncertainty that may prove to be suitable in other areas. He introduced a measure of uncertainty involving one parameter. Since 1961, various researchers have generalized the concept of entropy involving parameters that are based on the probability theory. Some of the measures have found applications in different disciplines. For a complete discussion of generalized measures of entropy or uncertainty refer [1] and for applications refer [6] .

As pointed out in the previous paragraph, there exist other generalized measures of entropy in the literature of information theory. One reason to consider alternative measures of uncertainty or entropy is to have the flexibility which may be necessary in a variety of applications. Also different measures of entropy lead to a unique model for each situation. Another reason to consider generalized measures of entropy is that some probability distributions can be obtained by maximizing Shannon entropy but with complicated and artificial constraints. Thus, if we have at our disposal a variety of measures, then we can obtain a variety of models and the model that is closest to observations will emerge as the satisfactory one. In the literature, some of the generalized measures have been applied successfully to different fields such as marketing and accounting [6]. In this paper we will consider the measure involving one parameter proposed in [5]. The uncertainty measure proposed by Havrda and Charvat, called uncertainty (or entropy) of type β, is defined as

$$H_\beta(P) = A_\beta \left(\sum p_i (p_i^{\beta-1} - 1 \right), \qquad \beta > 0,\ \beta \neq 1\,,$$

where $A_\beta = \left(e^{1-\beta} - 1\right)^{-1}$.

One of the reasons for considering this measure is that it has been used by various researchers in different fields.

Yager [11] considered the following problem as a motivation to generalize the Shannon's measure of uncertainty. Assume that during an election for a president we have four candidates and let p_1, p_2, p_3, and p_4 be their probabilities of winning the election. The amount of uncertainty as to who shall win the election can be computed both by Shannon's entropy and entropy of type β. Let us assume that three candidates have similar ideas on domestic agenda. We are interested in the candidate's position on domestic agenda rather than the actual identity of the person. Assume that the first three candidates have the same position on domestic agenda but the fourth candidate has a different position from other candidates. We are interested in measuring the uncertainty with regards to the domestic agenda. It is clear that the uncertainty in this situation would be less than the uncertainty concerning the original situation with regard to who shall be the president. The simple reason being that the three candidates' positions are the same.

In order to handle this problem, Yager [11] used the concept of a measure of similarity introduced by Zadeh [12]. The concept of similarity relation is an extension of the concept of equivalence relation. Let us assume that X is a set containing n elements $x_1, x_2, \ldots, x_n$. Then a fuzzy subset A of X is characterized by a membership function $\mu_A(\cdot)$ which associates with each x in X a value called membership in the interval $[0, 1]$. In other words $\mu_A(\cdot)$ represents the degree of membership of x in A, that is, the closer is the value of $\mu_A(\cdot)$ to 1, the higher is the degree of membership of x in A. It is clear that if the membership function can only take the value either 0 or 1, then the fuzzy set, A, will reduce to the ordinary or crisp set. Also it is clear that the fuzzy set can deal with both precise and imprecise information.

A fuzzy relation R on the set $X \times Y$ is a fuzzy subset such that for every pair (x, y), $R\,(x, y)$ measures the membership grade of the pair (x, y) in R and it takes value in a unit interval $[0, 1]$. In other words $R\,(x, y)$ can be considered as the degree of relationship between x and y. Now we can define the concept of a similarity relation as follows:

Defintion 1. A similarity relation S on X is a fuzzy subset on $X \times X$ that satisfies the following properties:

1. S is reflexive: $S(x,x) = 1$ for all $x \in X$.

2. S is symmetric: $S(x,y) = S(y,x)$.

3. S is transitive: $S(x,z) \geq Max_y\,[S(x,y) \wedge S(y,z)]$, where $(\wedge = \min]$.

It is easy to see that an equivalence relation is a special case of a similarity relation. (In the case of equivalence relation $S(x,y) \in \{0,1\}$).

Also we can define a similarity class as follows.

Defintion 2. Let S be a similarity relation defined on X. With each $x \in X$ we can associate a fuzzy subset of X. This fuzzy subset is called the similarity class of X, denoted as $S_{[x]}$ and its membership function is defined as

$$S_{[x]}(y) = S\left(x,y\right).$$

Note that the different similarity classes do not have the nice crisp characteristic of being either disjoint or equal. In fact they can overlap. However, the similarity classes do cover X. It means that

$$X = \cup_{x \in X} S_{[x]}.$$

The following facts are true about similarity classes.

1. If x and y are such that $S\left(x,y\right) = 1$, then their similarity classes are equal.

2. If x and z are such that $S\left(x,z\right) = 0$, then their similarity classes are disjoint.

Thus we can use $1 - S\left(x,y\right)$ as a measure of the degree of distinction between two similarity classes $S_{[x]}$ and $S_{[y]}$.

Defintion 3. A fuzzy subset A is called normal if there exists at least one element having membership value equal to one.

Because for any simililarity class $S_{[x]}$, $S_{[x]}(x) = 1$, therefore all the similarity classes are normal.

Defintion 4. Let A be a fuzzy subset of X. Then the cardinality of A, denoted $CardA$, is defined as $CardA = \sum_{x \in X} A(x)$.

It is to be noted that $Card\left(S_{[x]}\right) \geq 1$.

Let A be a fuzzy subset of X and $P = \{p_1, p_2, \ldots, p_n\}$ be a probability distribution on X. Then the probability of A, $\Pr(A)$, is defined as

$$\Pr(A) = \sum_{i=1}^{n} p_i A(x_i).$$

In other words, the probability of A is the expected membership grade of A [13].

In the next section we shall consider the measure of uncertainty of type β under similarity relation and prove some interesting results.

2. Uncertainty of Type β Under Similarity Relation

We defined the measure of uncertainty of type $\beta, H_\beta(P)$, in the previous section for any probability distribution. It is known that this measure takes the minimum value 0 if $p_i = 1$ for some x_i. Also we know that it takes its maximum value when $p_i = \frac{1}{n}$ for all x_i. The maximum value is $\dfrac{n^{1-\beta}-1}{e^{1-\beta}-1}$.

Now consider the problem of comparing elements or objects of X with respect to some attribute or characteristic (variable) V. Also let there be a similarity relation on X generated by this attribute V. Then consider the problem of selecting an object or element from X based on the given probability distribution on X. Our interest is in determining the uncertainty of an element chosen rather than its identity with respect to this attribute V. In order to handle this situation, we shall consider a generalization of the measure of uncertainty of type β. This generalization is defined as

Defintion 5. The measure of uncertainty of type β of the probability distribution P with respect to the similarity relation S, $H_\beta\left(P/S\right)$, is defined as

$$H_\beta\left(P/S\right) = A_\beta[\sum p_i(a_i^{\beta-1} - 1)], \qquad \beta > 0,\ \beta \neq 1\,,$$

where

$$a_i = \sum_{j=1}^{n} p_j S_{[x_i]}(x_j) = \sum_{j=1}^{n} p_j S(x_i, x_j).$$

It is to be noted that

$$\lim_{\beta \to 1} H_\beta\left(P/S\right) = H\left(P/S\right)\,,$$

where

$$H\left(P/S\right) = -\sum p_i \ln a_i$$

is the Shannon entropy of the probability distribution P with respect to the similarity relation S.

Now we shall consider the usual situation in which we are interested in the uncertainty with respect to the identity of the element selected. In other words, consider the similarity relation S_I defined as

$$\begin{aligned} S_I(x,x) &= 1 \qquad \text{for all } x \text{ in } X\,, \\ S_I(x,y) &= 0 \qquad \text{for } x \neq y\,. \end{aligned}$$

In other words, two elements of S_I are similar iff they are same or identical. Thus

$$a_i = \sum_{j=1}^{n} p_j S(x_i, x_j) = p_i.$$

Therefore

$$H_\beta\,(P/S) = A_\beta[\sum p_i(a_i^{\beta-1} - 1)] = A_\beta[\sum p_i(p_i^{\beta-1} - 1)] = H_\beta(P).$$

In the following theorem, we will consider the situation when S is an equivalence relation.

Theorem 1. *Let S be an equivalence relation with q distinct equivalence classes $E_1, E_2, \ldots, E_q$. Then.*

$$H_\beta\,(P/S) = A_\beta[\sum_{i=1}^{q} b_i(b_i^{\beta-1} - 1)], \qquad \beta > 0,\ \beta \neq 1\,,$$

where b_i is the sum of the probabilities of the elements in class E_i.

Proof. We have

$$H_\beta\,(P/S) = A_\beta[\sum_{i=1}^{n} p_i(a_i^{\beta-1} - 1)]\,,$$

where

$$a_i = \sum_{j=1}^{n} p_j S(x_i, x_j).$$

Also if x_i and x_j are in the same equivalence class, then $S\,(x_i, x_j) = 1$ otherwise $S\,(x_i, x_j) = 0$. Thus

$$a_i = b_i = \sum_j p_j\,,$$

where summation is over those j such that $S(x_i, x_j) = 1$.

In other words, the original probability distribution P is replaced by another probability distribution $B = (b_1, b_2, \ldots\ldots, b_q)$, where b_i is the probability of the i-th equivalence class and is equal to the sum of the probabilities of the elements in that class. Thus $H_\beta(P/S)$ is essentially the uncertainty measure of type β, $H_\beta(B)$, where B is the probability distribution of the equivalence class. □

Corollary. *For any equivalence relation S, we have*

$$H_\beta(P/S) \le H_\beta(P).$$

Proof. We know that

$$H_\beta(P) = A_\beta[\sum_{i=1}^{n} p_i(p_i^{\beta-1} - 1)]\,,$$

and

$$H_\beta(P/S) = A_\beta[\sum_{j=1}^{q} b_j(b_j^{\beta-1} - 1)].$$

Let u_i be the value of b_j such that $x_i \in E_j$. Also b_j is the sum of the probabilities of the elements in the equivalence class E_j. Thus

$$H_\beta(P/S) = A_\beta[\sum_{i=1}^{n} p_i(u_i^{\beta-1} - 1)].$$

Since $u_i \ge p_i$, thus we have

$$\begin{aligned} & H_\beta(P) - H_\beta(P/S) \\ = \; & A_\beta[\sum_{i=1}^{n} p_i(p_i^{\beta-1} - 1)] - A_\beta[\sum_{i=1}^{n} p_i(u_i^{\beta-1} - 1)] \\ = \; & A_\beta[\sum_{i=1}^{n} p_i(p_i^{\beta-1} - u_i^{\beta-1})] \ge 0. \end{aligned}$$

This completes the proof. □

Now we shall consider the general situation when S is a similarity relation.

Theorem 2. *Let P be a probability distribution and S be a similarity relation on the set X. Then we have*

$$H_\beta(P/S) \le H_\beta(P).$$

Proof. Since $a_i = \sum_{j=1}^{n} p_j S(x_i, x_j)$, we have $a_i \geq p_i$. Thus

$$\begin{aligned} & H_\beta(P) - H_\beta(P/S) \\ = \; & A_\beta[\sum_{i=1}^{n} p_i(p_i^{\beta-1} - 1)] - A_\beta[\sum_{i=1}^{n} p_i(a_i^{\beta-1} - 1)] \\ = \; & A_\beta[\sum_{i=1}^{n} p_i(p_i^{\beta-1} - a_i^{\beta-1})] \geq 0. \end{aligned}$$

In other words, the uncertainty measure of type β is still an upper bound on the uncertainty, when S is a similarity relation. Also we note that

$$H_\beta(P) - H_\beta(P/S) = A_\beta[\sum_{i=1}^{n} p_i(p_i^{\beta-1} - a_i^{\beta-1})].$$

This means that each similarity class causes a reduction in uncertainty in accordance with the term $A_\beta(p_i^{\beta-1} - a_i^{\beta-1})$.

If $a_i = p_i$, then there is no contribution of the similarity class. Also we note that if the similarity class is larger, then the reduction in uncertainty is large. Thus the reduction in uncertainty occurs due to the introduction of the similarity class and for any probability distribution, uncertainty measure of type β provides an upper bound on the uncertainty. □

In the following, we shall define the least generous and the most generous similarity relation and prove some interesting facts about these similarity relations.

Defintion 6. Let S and S' be two similarity relations on X such that for every x, y in X

$$S(x,y) \geq S'(x,y).$$

Then S is a more generous similarity relation and we denote this as $S \geq S'$.

Defintion 7. Let S_I be the similarity relation such that

$$\begin{aligned} S_I(x,x) &= 1 \text{ for all } x \text{ in } X \\ S_I(x,y) &= 0 \qquad x \neq y. \end{aligned}$$

Then S_I is called the least generous similarity relation.

Defintion 8. Let S^0 be the most generous similarity relation defined by $S^0(x,y) = 1$ for all x.

Corollary. For any similarity relation S

$$H_\beta(P/S_I) \geq H_\beta(P/S)\,.$$

Proof. It is clear that

$$H_\beta\left(P/S_I\right) = H_\beta(P).$$

Thus

$$H_\beta(P) = H_\beta(P/S_I) \geq H_\beta(P/S).$$

Therefore, the least generous similarity relation gives us the most uncertainty of type β. □

Theorem 3. *For any probability distribution P on X, we have*

$$H_\beta\left(P/S^0\right) = 0.$$

Proof.

$$H_\beta\left(P/S^0\right) = A_\beta[\sum_{i=1}^{n} p_i(a_i^{\beta-1} - 1)]\,,$$

where

$$a_i = \sum_{j=1}^{n} p_j S^0(x_i, x_j).$$

But $S^0\left(x_i, x_j\right) = 1$. Therefore $a_i = \sum_j p_j = 1$. Thus

$$H_\beta\left(P/S^0\right) = A_\beta[\sum_{i=1}^{n} p_i(1^{\beta-1} - 1)] = 0\,.$$

This completes the proof. □

Also we can easily check that for any probability distribution P and similarity relation S, the relation $H_\beta\left(P/S\right) \geq 0$ is always true. This is due to the fact that

$$a_i = \sum_{j=1} p_i S\left(x_i, x_j\right) \leq 1.$$

From the above discussion, we can conclude that for any probability distribution P and similarity relation S, the following is true

$$0 = H_\beta\left(P/S^0\right) \leq H_\beta(P/S) \leq H_\beta(P/S_I) = H_\beta(P), \quad \beta > 0,\ \beta \neq 1.$$

In the following theorem, we shall obtain a result with regards to the ordering of the uncertainty measure of type β and the similarity relationship.

Theorem 4. *Let S and S' be two similarity relations on the set X and let P be any probability distribution on X. Also let S be more generous than S' $(S \geq S^{'})$. Then we have*

$$H_\beta\left(P/S'\right) \geq H_\beta\left(P/S\right).$$

In other words, the more generous is the similarity relationship, the less is the uncertainty of type β in the environment.

Proof.

$$H_\beta\left(P/S\right) = A_\beta[\sum_{i=1}^{n} p_i(a_i^{\beta-1} - 1)],$$

$$H_\beta\left(P/S^{'}\right) = A_\beta[\sum_{i=1}^{n} p_i((a_i')^{\beta-1} - 1)],$$

where

$$a_i = \sum_{j=1} p_j S\left(x_i, x_j\right),$$

and

$$a_i' = \sum_{j=1} p_j S\left(x_i, x_j\right).$$

Since $S \geq S'$, then $S(x_i, x_j) \geq S'\left(x_i, x_j\right)$. This means that $a_i \geq a_i'$. Therefore,

$$\begin{aligned} & H_\beta\left(P/S'\right) - H_\beta\left(P/S\right) \\ = \; & A_\beta[\sum_{i=1}^{n} p_i(a_i^{\beta-1} - 1)] - A_\beta[\sum_{i=1}^{n} p_i((a_i')^{\beta-1} - 1)] \\ = \; & A_\beta[\sum_{i=1}^{n} p_i(a_i^{\beta-1} - (a_i')^{\beta-1}] \geq 0. \end{aligned}$$

This completes the proof. □

What is the effect of the probability distribution on the uncertainty measure of type β under a given similarity environment? In the next theorem, we will show that the certainty in regards to a probability distribution gives us certainty under any S.

Theorem 5. *Let us assume that S be any similarity relation on X and P be a probability distribution on X such that $p_i = 1$ for some x_i. Then*

$$H_\beta\,(P/S) = 0.$$

Proof. Let us assume that $p_k = 1$ for some x_k. Then

$$H_\beta(P/S) = A_\beta[\sum_{i=1}^{n} p_i(a_i^{\beta-1} - 1)] = A_\beta(a_k^{\beta-1} - 1).$$

Also $a_k = \sum_{j=1} p_j S_k(x_j) = p_k S_k(x_k) = 1$. Thus $H_\beta(P/S) = 0$. This completes the proof. □

Theorem 6. *Let all the elements of X of cardinality n be equally likely. Then*

$$H_\beta(P/S) \leq \frac{n^{1-\beta} - 1}{e^{1-\beta} - 1}.$$

Proof. Assume that the elements of X are equally likely. Then $p_i = \frac{1}{n}$ for each i. Furthermore

$$H_\beta(P/S) = A_\beta[\sum_{i=1}^{n} \frac{1}{n}(a_i^{\beta-1} - 1)]\,,$$

where

$$a_i = \sum_{j=1}^{n} p_j S_i(x_j) = \frac{1}{n} Card[S_{[i]}].$$

Thus

$$\begin{aligned} H_\beta(P/S) &= A_\beta[\sum_{i=1}^{n} \frac{1}{n}((\frac{1}{n}Card[S_{[i]}])^{\beta-1} - 1)] \\ &= A_\beta[\sum_{i=1}^{n} n^{-\beta} Card[S_{[i]}]^{\beta-1} - 1] \end{aligned}$$

Since $Card(S_{[i]}) \geq 1$, therefore $H_\beta(P/S) \leq A_\beta[n^{-\beta} - 1]$. This completes the proof. □

Now we would like to know that for any similarity relation S, does the equally likely probability give the maximum entropy? In the following example, we will see that it is not always the case.

Example. Let $X = \{x_1, x_2, x_3\}$ and

$$S = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

a) Assume that $p_1 = p_2 = p_3 = 1/3$ and $\beta = 2$.

$$H_\beta(P/S) = \frac{4e}{9(e-1)}.$$

b) Assume that $p_1 = p_3 = 1/2$, $p_2 = 0$ and $\beta = 2$.

$$H_\beta(P/S) = \frac{e}{(e-1)}.$$

This example shows that the equally likely elements of a set do not give us the maximum entropy under all similarity relations. □

In the following theorem, an interesting relation is proved for equivalence relations.

Theorem 7. *Let X be a set with n elements. Let S be an equivalence relation on X which has m equivalence classes. Also assume that P^* is a probability distribution on X such that the sum of the probabilities of the elements in each equivalence class is $\frac{1}{m}$. Then, we have*

$$H_\beta(P^*/S) \geq H_\beta(P/S) \qquad \text{for any } P.$$

Proof.

$$H_\beta(P/S) = A_\beta[\sum_{i=1}^{n} p_i(a_i^{\beta-1} - 1)],$$

where

$$a_i = \sum_{j=1}^{n} p_j S(x_i, x_j).$$

We have $S(x_i, x_j) = 1$ or 0 because S is an equivalence relation. Therefore, $a_i = \sum_{j \in S_i} p_j$. Let $E_1, E_2, \ldots, E_m$ be the equivalence classes and let q_i be the sum of the probabilities in each class i. Then

$$H_\beta(P/S) = A_\beta[\sum_{i=1}^{m} q_i(q_i^{\beta-1} - 1)].$$

We know that this attains maximum value when $q_i = 1/m$. □

3. Conclusion

We have studied the generalized form of uncertainty measure of type β under similarity relations. The results, proved here, extend the results proved by Yager [11]. When $\beta \to 1$, the results proved in this paper reduce to Shannon's entropy, proved by Yager [11].

References

[1] J. Aczel, Z. Daroczy, *On Measures of Information and their Characterization*, Academic Press, New York (1975).

[2] D. Bhandari, N.R. Pal, Some new information measures for fuzzy sets, *Information Sciences*, **67** (1993), 209-228.

[3] A. DeLuca, S. Termini, A definition of nonprobabilistic entropy in the setting of fuzzy sets theory, *Information and Control*, **20** (1972), 301-312.

[4] B. Ebanks, On measures of fuzziness and their representations, *J. Math. Anal. Appl.*, **94** (1983), 24-37.

[5] J.H. Havrda, F. Charvat, Quantification method of classification processes: Concepts of structural α entropy, *Kybernetica*, **3** (1976), 149-172.

[6] J.N. Kapur, H.K. Kesavan, *Entropy Optimization Principles with Applications*, Academic Press, New York (1992).

[7] A. Kauffman, *Introduction to the Theory of Fuzzy Subsets*, Academic Press, New York, Vol. **1** (1975).

[8] N.R. Pal, James C. Bezdek, Measuring fuzzy entropy, *IEEE Trans. Fuzzy Systems*, **2**, No. 2 (1994), 107-118.

[9] A. Renyi, On measures of entropy and information, In: *Proc. 4-th Berkeley Symp. Maths. Stat. Prob.*, **1** (1961), 547-561.

[10] C.L. Shannon, The mathematical theory of communication, *Bell System Technical Journal*, **48** (1948), 379-423.

[11] Ronald R. Yager, Entropy measures under similarity relations, *Technical Report MII-1113*, Iona College, Machine Intelligence Institute, New Rochelle, New York 10801.

[12] L.A. Zadeh, Similarity relations and fuzzy orderings, *Information Sciences,* **3** (1971), 177-200.

[13] L.A. Zadeh, Probability measures of fuzzy events, *Journal of Math. Anal. and Appl.*, **10** (1968), 421-427.

[14] Guy Jumarie, *Information theory: Theory and Applications*, Springer-Verlag, Berlin (1987).

[15] George Klir, Tina Folger, *Fuzzy Sets and Applications*, Prentice Hall, New York (1987).

[16] George Klir, Bo Yuan, *Fuzzy Sets and Fuzzy Logic*, Prentice Hall, New York (1995).