UNIT- IV

1

UNCERTAINITY AND INFORMATION

TYPES OF UNCERTAINITY:

There are many words or phrasee to define the word Uncertainity. Dictionary gives 6 clusters of meanings for the term, Uncertainity.

i) Not certainly known
ii) Vague
iii) Doubtful
iv) Ambiguous
v) Not - Steady
vi) Liable to change

We categorize the term uncertainity into two terms namely

* Vagueness and
* Ambiguity

The term vagueness is associated with the difficulty of making sharp.

Ambiguity is associated with one to many relations.

i.e., Situations in which the choice between 2 or more alternatives is left unspecified

Measures of uncertainity related to vagueness are referred to measures of fuzzyness.

Measures related to ambiguity are divided into three types 2

i) Measure of non-specificity
ii) Measure of Dissonance
iii) Measure of confusion in evidence.

Measure of fuzziness:

In general, a measure of fuzziness is a function $f: \tilde{P}(X) \longrightarrow R$, where $\tilde{P}(X)$ denotes the set of all fuzzy subsets of X

i.e., the function "f" assigns a value $f(A)$ to each fuzzy subset A of X that characterizes the degree of fuzzyness of A.

There are 3 axiomatic requirements, that every meaningful measure of fuzzyness must satisfy

axiom f_1:

$f(A) = 0 \iff A$ is a crisp set

axiom f_2:

If $A < B$, then $f(A) \leq f(B)$.

axiom f_3:

$f(A)$ assumes the maximum value iff A is maximally fuzzy.

i) The sharpness relation:

$A < B$ in axiom f_2 is defined by $\mu_A(x) \leq \mu_B(x)$, for $\mu_B(x) \leq 1/2$

$\mu_A(x) \geq \mu_B(x)$, for $\mu_B(x) \geq 1/2$, $\forall x \in X$

ii) Maximally fuzzy:

This measure of fuzziness is defined by the function

$$f(A) = -\sum_{x \in X}\left[\mu_A(x)\log_2 \mu_A(x) + (1-\mu_A(x))\log_2(1-\mu_A(x))\right]$$

iii) Index of fuzziness:

Index of fuzzyness is defined interms of metric distance (Hamming or Euclidean) of A from any of the nearest crisp sets C for which

$$\boxed{\begin{aligned} \mu_C(x) &= 0, \quad \text{if } \mu_A(x) \leq 1/2 \\ \mu_C(x) &= 1, \quad \text{if } \mu_A(x) > 1/2 \end{aligned}}$$

When the hamming distance is used, the measure of fuzzyness is expressed by the function

$$\boxed{f(A) = \sum_{x \in X} |\mu_A(x) - \mu_C(x)|} \quad \text{——} \quad (1)$$

For the Euclidean distance,

$$\boxed{f(A) = \left[\sum_{x \in X}[\mu_A(x) - \mu_C(x)]^2\right]^{1/2}} \quad \text{——} \quad (2)$$

The other metric distances may be used as well

For example,

The Minkowski class of distances yields a class of fuzzy measure

$$f_{\omega}(A) = \left[\sum_{x \in X} \left[\mu_A(x) - \mu_C(x)\right]^{\omega} \right]^{1/\omega} \quad \text{(3)}$$

where, $\omega \in [1, \infty]$

For $\omega = 1$, eqn. (3) reduces to eqn. (1)

and for $\omega = 2$, eqn (3) reduces to eqn. (2)

Classical measure of Uncertainty

(1) Hartley and Shannon introduced measure of cencertainty in different views. These measure are referred as Hartley information and Shannon entropy

Hartley information:

Consider a finite set X of n elements, sequence can be formed from elements of X by successive selections. At each selection, all possible elements that might have been chosen or eliminated except one. The number of all possible sequences of $s \in N$ selections from the set X is n^s, where $n = |X|$

Now, we defined the amount of information $I(n^s)$ associated with s selections from the set X should be proportional to s.

(2)

i.e., $I(n^s) = k(n) \cdot s$

where, $k(n)$ is a constant depends on n.

Consider two sets X_1 and X_2 such that $|X_1| = n_1$ and $|X_2| = n_2$, when the numbers s_1 and s_2 of selections from the sets X_1 and X_2 respectively are such that they yield same number of sequences, then the amount of information associated with the sequences should be the same.

i.e., $n_1^{s_1} = n_2^{s_2}$ — (1) $\Rightarrow I(n_1^{s_1}) = I(n_2^{s_2})$

$\Rightarrow k(n_1) \cdot s_1 = k(n_2) \cdot s_2$ —— (2)

taking log on both sides in (1)

$$\log n_1^{s_1} = \log n_2^{s_2}$$

$$\Rightarrow s_1 \log n_1 = s_2 \log n_2$$

$$\Rightarrow \frac{s_2}{s_1} = \frac{\log n_1}{\log n_2} \quad \text{—— (3)}$$

from (2),

$$\frac{s_2}{s_1} = \frac{k(n_1)}{k(n_2)} \quad \text{—— (4)}$$

from (5) and (4)

$$\frac{\log_b n_1}{\log_b n_2} = \frac{k(n_1)}{k(n_2)}$$

This eqn can be satisfied only by

$$k(n) = k_0 \log_b n$$

where, k_0 is a common constant

The amount of information conveyed by a sequence of s - selections from
3 the set X with n - elements is given by the formula

$$I(n^s) = k(n) \cdot s$$
$$= k_0 \, s \log_b n$$

Particular case:

Choose $k_0 = 1$ and $b = 2$

$$I(n^s) = s \log_2 n$$
$$= \log_2 n^s$$
$$I(N) = \log_2 N$$

where N denotes the total no. of alternatives

Information function:

$I(N)$ can also be viewed as the amount of information needed to characterize one of N alternatives. Now, Hartley information can also be charactersed by the following axioms:

Axiom I_1: (Additivity)

$$I(NM) = I(N) + I(M)$$

where $N, M \in$ Natural numbers

Axiom I_2: (Monotonicity)

$$I(N) \leq I(N+1) \quad , \forall \; N \in \text{natural numbers}$$

Axiom I_3: (Normalization)

$$I(2) = 1$$

THEOREM:

4 Hartley information is the only function that satisfies all the three axioms of Information function.

Proof:

Let $N > 2$, for each integer i, define $q(i)$ such that

$$2^{q(i)} \leq N^i \leq 2^{q(i)+1} \quad \text{——} \quad (1)$$

This eqn. can be written in the form

$$\log_2 2^{q(i)} \leq \log_2 N^i \leq \log_2 2^{q(i)+1}$$

$$q(i) \log_2 2 \leq i \log_2 N \leq (q(i)+1) \log_2 2$$

$\div i$ and replace $\log_2 2 = 1$

$$\frac{q(i)}{i} \leq \log_2 N \leq \frac{q(i)+1}{i}$$

Taking the limit $i \to \infty$

$$\lim_{i\to\infty} \frac{q(i)}{i} \leq \log_2 N \leq \lim_{i\to\infty} \left[\frac{q(i)}{i}\right]$$

$$\Rightarrow \lim_{i\to\infty} \frac{q(i)}{i} = \log_2 N \quad \text{——— (2)}$$

Let I be a function that satisfies all the three axioms of the information function.

If $a < b$, then $I(a) \leq I(b)$ [from axiom I_2

from (1),

$$I(2^{q(i)}) \leq I(N^i) \leq I(2^{q(i)+1}) \quad \text{——— (2)}$$

using axiom - I_1,

$$I(a^k) = I(a \cdot a \cdot a \ldots a) \quad (k \text{ times})$$

$$= I(a) + I(a) + \cdots + I(a)$$

$$= k \cdot I(a)$$

$$\therefore I(N^i) = i \cdot I(N)$$

$$I(2^{q(i)}) = q(i) \cdot I(2)$$

$$= q(i) \quad [\because I(2) = 1, \text{ by axiom } 3$$

$$I(2^{q(i)+1}) = (q(i)+1)\, I(2)$$

$$= q(i)+1 \quad [\because \text{ by axiom } 3$$

(3) $\Rightarrow$ $$q(i) \leq i \cdot I(N) \leq q(i)+1$$

$\div$ by i & taking $\lim_{i\to\infty}$ q

$$\lim_{i\to\infty} \frac{q(i)}{i} \le I(N) \le \lim_{i\to q} \frac{q(i)}{i} \quad \left[\because \lim_{i\to\infty} \frac{1}{i} = 0\right.$$

(6)

$$\Rightarrow \quad \lim_{i\to\infty} \frac{q(i)}{i} = I(N) \quad \text{———} \quad (7)$$

from (2) and (4),

$$I(N) = \log_2 N \quad , \quad \text{for } N > 2$$

when, $N = 2$,

$$I(2) = \log_2 2 = 1$$

and

$$I(N) = I(N \cdot 1)$$
$$I(N) = I(N) + I(1)$$

$$\Rightarrow \quad I(1) = 0$$

Hence, the Hartley information function is the only function that satisfies all the three axioms of information function

Types of information:

Consider the 2 sets X and Y, that are inter-related in the sense that selections from one of the sets constrained by selections from the other.

Assume that the constraint is expressed by the relation $R \subset X \times Y$, then 3 types of Hartley information can be defined on these sets.

1. Simple information:

$$I(X) = \log_2 |X|$$

$$I(Y) = \log_2 |Y|$$

2. Joint information:

$$I(X,Y) = \log_2 |R|$$

3. Conditional information:

$$I(X|Y) = \cancel{\frac{\log}{I(Y)}}$$

$$= \log_2 \left(\frac{|R|}{|Y|}\right)$$

$$= \log_2 |R| - \log_2 |Y|$$

$$\therefore I(X|Y) = I(X,Y) - I(Y)$$

and

$$I(Y|X) = \log_2 \left(\frac{|R|}{|X|}\right)$$

$$= \log_2 |R| - \log_2 |X|$$

$$\therefore I(Y|X) = I(X,Y) - I(X)$$

If selections from X do not depend on selections from Y, then sets X and Y are called non-interactive. Then

$$I(X,Y) = I(X) + I(Y)$$

$$I(X|Y) = I(X,Y) - I(Y)$$

$$= I(X) + I(Y) - I(Y)$$

$$= I(X)$$

$$I(Y|X) = I(X,Y) - I(X)$$

$$= I(X) + I(Y) - I(X)$$

$$= I(Y)$$

Information transmission function:

Definition:

A symmetric function which is usually referred to as information transmission is a useful indicator of the strength of constraint between the sets X and Y defined by

$$T(X,Y) = I(X) + I(Y) - I(X,Y)$$

when the sets are non-interactive

$$T(X,Y) = 0$$

Generalized information transformation is defined by

8

$$T(X_1, X_2, \ldots X_n) = \sum_{i=1}^{n} I(X_i) - I(X_1, X_2, \ldots X_n)$$

Example:

Consider 2 variables X, Y whose values are taken from the sets

$$X = \{low,\ medium,\ high\}$$

$$Y = \{1, 2, 3, 4\}$$

It is known that the variable are constrained by the relation R by the matrix

$$\begin{array}{c c} & \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{array}{c} low \\ medium \\ high \end{array} & \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \end{array}$$

We can see that, the low values of X doesnot contain Y at all. The medium value of X contains

Y partially and the high value

Y totally 12

The following types of Hartley information can be calculated in this example.

Soln:

1. Simple information:

$$I(X) = \log_2 |X| = \log_2 3 = 1.6$$

$$I(Y) = \log_2 |Y| = \log_2 4 = \log_2 2^2 = 2\log_2 2 = 2.1 = 2$$

2. Joint information:

$$I(X,Y) = \log_2 |R| = \log_2 7 = 2.8$$

3. Conditional information:

$$I(X|Y) = I(X,Y) - I(Y)$$

$$= 2.8 - 2$$

$$= 0.8$$

$$I(Y|X) = I(X,Y) - I(X)$$

$$= 2.8 - 1.6$$

$$= 1.2$$

4. Information transmission

$$T(X,Y) = I(X) + I(Y) - I(X,Y)$$

$$= 1.6 + 2 - 2.8$$

$$= 0.8$$

Shannon entropy:

Shannon entropy is a measure of uncertainty formulated interms of Probability theory is given by the function.

$$H(p(x) \mid x \in X) = -\sum_{x \in X} p(x) \log_2 p(x)$$

where, $(p(x) \mid x \in X)$ is a ~~class~~ probability distribution on a finite set X.

It is thus, a function of the form

$$H : \wp \longrightarrow [0, \infty)$$

where $\wp$ denotes the set of all probability distributions on finite sets.

Axioms which are essential for a probabilistic measure of uncertainty.

Axiom - H_1 : (Expansibility)

When a component with zero probability is added to a probability distribution, the uncertainty should not change.

i.e., $H(P_1, P_2, \ldots P_n) = H(P_1, P_2, \ldots P_n, 0)$, $\forall P_1, P_2, \ldots P_n \in \wp$

Axiom - H_2 : (Symmetry)

The uncertainty should be invariant with respect to permutations of probabilities of a given probability distribution.

i.e., $H(P_1, P_2, \ldots P_n) = H(\text{Perm}(P_1, P_2, \ldots P_n))$, $\forall P_1, P_2, \ldots P_n \in \wp$

Axiom - H_3 : (Continuity)

Function H should be continuous in

all its arguments $p_1, p_2 \dots p_n$, $n \in N$

Axiom - H_4: (Maximum)

For every $n \in N$.

$$H(p_1, p_2 \dots p_n) \le H\left(\frac{1}{n}, \frac{1}{n} \dots \frac{1}{n}\right)$$

Axiom - H_5: (Sub-additivity)

The uncertainity of a joint probability distribution should not be the greater the sum of the uncertainty of the corresponding marginal probability distributions.

i.e., $H(p_{11}, p_{12}, \dots p_{1n}, p_{21}, p_{22} \dots p_{2n}, \dots, p_{s1}, p_{s2}, \dots p_{sn})$

$$\le H\left(\sum_{i=1}^{s} p_{i1}, \sum_{i=1}^{s} p_{i2} \dots \sum_{i=1}^{s} p_{in}\right) + H\left(\sum_{j=1}^{n} p_{1j}, \sum_{j=1}^{n} p_{2j}, \dots \sum_{j=1}^{n} p_{sj}\right)$$

for any joint probability distribution in $\wp$.

Axiom - H_6: (Additivity)

For any two marginal probability distribution that are non-interactive, the uncertainty of the associated joint distribution should be equal to the sum of uncertainties of the marginal distributions.

i.e., $H(p_1q_1, p_1q_2 \dots p_1q_s, p_2q_1, p_2q_2 \dots p_2q_s, \dots$
$\dots, p_nq_1, p_nq_2 \dots p_nq_s)$

$$= H(p_1, p_2, \dots, p_n) + H(q_1, q_2, \dots, q_s)$$

for any 2 probability distributions $p_1, p_2, \dots p_n \in \wp$ & $q_1, q_2 \dots q_s \in \wp$.

Weaker additivity:

This requirement is sometimes replaced with a weaker additivity requirement defined only for marginal distribution with equal probabilities $\frac{1}{n}$ and $\frac{1}{s}$

i.e., $H\left(\frac{1}{ns}, \frac{1}{ns}, \dots \frac{1}{ns}\right) = H\left(\frac{1}{n}, \frac{1}{n}, \dots \frac{1}{n}\right) + H\left(\frac{1}{s}, \frac{1}{s}, \dots \frac{1}{s}\right)$

for all $n, s \in N$

Note:

$f(n) = H\left(\frac{1}{n}, \frac{1}{n}, \dots \frac{1}{n}\right)$

$f(ns) = f(n) + f(s)$

Axiom H7: (Monotonicity)

For probability distribution with equal probabilities $\frac{1}{n}$, $n \in N$, the uncertainty should increase with increasing n.

i.e., If $n < s$, then $f(n) < f(s)$, $\forall\, n, s \in N$

Axiom H8: (Branching)

Let $A = \{x_1, x_2, \dots x_s\}$

$B = \{x_{s+1}, x_{s+2}, \dots x_n\}$

be two disjoint subsets of X. Further more given a probability distribution, $p_1, p_2, \dots p_n$ on X, where $p_i = P(x_i)$. $\forall\, i \in N$

Let $P_A = \sum_{i=1}^{s} p_i$, $P_B = \sum_{i=s+1}^{n} p_i$

denote the probabilities of A and B is

Then, the branching requirement means that,

16

$$H(p_1, p_2, \dots p_n) = H(p_A, p_B) + p_A H\left(\frac{p_1}{p_A}, \frac{p_2}{p_A}, \dots \frac{p_s}{p_A}\right) + p_B H\left(\frac{p_{s+1}}{p_B}, \frac{p_{s+2}}{p_B}, \dots \frac{p_n}{p_B}\right)$$

$$\forall \; p_1, p_2 \dots \in \wp.$$

Axiom H9: (Normalization)

$$H\left(\tfrac{1}{2}, \tfrac{1}{2}\right) = 1$$

THEOREM-1:

Shanon entropy is the only function that satisfies all the 9 axioms of probabilistic measure of uncertainty. i.e., s.t.

$$0 \leq H(p_1, p_2 \dots p_n) \leq \log_2 n$$

Proof:

First let us prove that

$$H(p_1, p_2, \dots p_n) \geq 0$$

w.k.T.

$$-p_i \log_2 p_i \geq 0, \quad \forall \; p_i \in (0, 1]$$

$\log_2 p_i = \log$

for $p_i = 0$, $\log p_i$ is not defined

$$\lim_{p_i \to 0} -p_i \log_2 p_i = -\lim_{p_i \to 0} \frac{\log_2 p_i}{1/p_i}$$

$$= -\lim_{p_i \to 0} \frac{\frac{1}{p_i \log_e 2}}{\frac{-1}{p_i^2}}$$

$$= \lim_{p_i \to 0} \frac{p_i^2}{p_i \log_e^2}$$

$$= \lim_{p_i \to 0} \left[\frac{p_i}{\log_e^2} \right]$$

$$= 0$$

$\therefore H(p_1, p_2 \dots p_n) \geq 0$

Let p_n be the dependent variable, then $p_n = 1 - (p_1 + p_2 + \dots + p_{n-1})$

and the necessary condition for an external value of H are

$$\frac{\partial H}{\partial p_i} = 0 \quad , \quad \forall i$$

The partial derivatives are

$$\frac{\partial H}{\partial p_i} = \frac{\partial H}{\partial p_1} \cdot \frac{\partial p_1}{\partial p_i} + \dots + \frac{\partial H}{\partial p_n} \cdot \frac{\partial p_n}{\partial p_i} , \quad i = 1, 2 \dots n$$

Since,

$$\frac{\partial H}{\partial p_k} \cdot \frac{\partial p_k}{\partial p_i} = 0 \quad , \quad \forall k \neq i, n$$

$$\frac{\partial H}{\partial p_i} = \frac{d(-p_i \log_2 p_i)}{dp_i} \frac{\partial p_i}{\partial p_i} + \frac{d(-p_n \log_2 p_n)}{dp_n} \cdot \frac{\partial p_n}{\partial p_i}$$

$$= -\left[\log_2 p_i + p_i \frac{1}{p_i \log_e^2} \times \frac{\partial p_i}{\partial p_i} \right] (1)$$

$$- \left[\log_2 p_n + p_n \frac{1}{p_n \log_e^2} \times \frac{\partial p_n}{\partial p_n} \right] (-1)$$

$$= -\log_2 p_i - \frac{1}{\log_e^2} + \log_2 p_n + \frac{1}{\log_e^2}$$

$$\frac{\partial H}{\partial p_i} = -\log_2 p_i + \log_2 p_n$$

18

$$\frac{\partial H}{\partial p_i} = 0$$

$\Rightarrow \quad \log_2 p_n = \log_2 p_i$

$\Rightarrow \quad p_i = p_n$

Hence $\quad p_1 = p_2 = \dots = p_n = 1/n$

Hence an extremal value of H exist for the distribution with equal probabilities $1/n$.

now,

$$\frac{\partial^2 H}{\partial p_i^2} = -\frac{1}{p_i \log_e^2} \times \frac{\partial p_i}{\partial p_i} + \frac{1}{p_n \log_e^2} \times \frac{\partial p_n}{\partial p_i}$$

$$= -\frac{1}{p_i \log_e^2} + \frac{1}{p_n \log_e^2}(-1)$$

$$= -\frac{1}{p_i \log_e^2} - \frac{1}{p_n \log_e^2}$$

$$= -\frac{1}{\left(\frac{1}{n}\right)\log_e^2} - \frac{1}{\left(\frac{1}{n}\right)\log_e^2}$$

$$= \frac{-2n}{\log_e^2} < 0 \quad , \quad \forall i$$

Hence, H is maximum.

Now,

$$H\left(\frac{1}{n}, \frac{1}{n}, \dots \frac{1}{n}\right) = -\sum \frac{1}{n} \log_2\left(\frac{1}{n}\right)$$

$$= -\sum \frac{1}{n} \log_2 (n)^{-1}$$

$$= \frac{1}{n} \log_2 n \geq 1$$

$$= \frac{1}{n} \log_2 n(n)$$

$$= \log_2 n$$

Hence the inequalities

i.e., $0 \leq H(p_1, p_2, \dots p_n) \leq \log_2 n$

$\therefore H\left(\frac{1}{n}, \frac{1}{n}, \dots \frac{1}{n}\right) = \log_2 n$ is the maximum of H.

Hence the inequalities is true.

<u>Gibb's inequality (or) Gibb's theorem - 2:</u>

The inequality $-\sum_{i=1}^{n} p_i \log_2 p_i \leq -\sum_{i=1}^{n} p_i \log_2 q_i$

It satisfied for all probability distributions p_i and q_i, $i \in N_n$, $n \in N$; the inequality holds iff $p_i = q_i$, $\forall i \in N_n$

Proof:

Consider the function,

$$h(p_i, q_i) = p_i (\log_e p_i - \log_e q_i) - p_i + q_i, \quad \forall p_i, q_i \in [0,1]$$

This function is finite and differentiable for all values of p_i and q_i except the pair $q_i = 0$ and $p_i \neq 0$

for each fixed $q_i \neq 0$

$$\frac{\partial h}{\partial p_i} = p_i \left(\frac{1}{p_i} - 0\right) + (\log_e p_i - \log_e q_i) 1 - 1$$

$$= \log_e p_i - \log_e q_i$$

i.e., $\frac{\partial h}{\partial p_i} = \begin{cases} < 0, & p_i < q_i \\ = 0, & p_i = q_i \\ > 0, & p_i > q_i \end{cases}$ 20

Hence 'h' is a convex function of p_i with its minimum at $p_i = q_i$

Hence, $p_i(\log_e p_i - \log_e q_i) - p_i + q_i \geq 0$ —— (1)

taking summation for $i = 1, 2, \dots n$ in (1), we have

$$\sum_{i=1}^{n} p_i \log_e p_i - \sum_{i=1}^{n} p_i \log_e q_i - \sum_{i=1}^{n} p_i + \sum_{i=1}^{n} q_i \geq 0$$

$$\Rightarrow \sum_{i=1}^{n} p_i \log_e p_i \geq \sum_{i=1}^{n} p_i \log_e q_i \quad \left[\because \sum_{i=1}^{n} p_i = \sum_{i=1}^{n} q_i = 1 \text{ tot. prob.} = 1\right.$$

$$\Rightarrow -\sum_{i=1}^{n} p_i \log_e p_i \leq -\sum_{i=1}^{n} p_i \log_e q_i$$

$\div$ by $\log_e 2$

$$-\sum_{i=1}^{n} p_i \log_2 p_i \leq -\sum_{i=1}^{n} p_i \log_2 q_i \quad \left[\because \log_a b = \frac{\log_c b}{\log_c a}\right.$$

Let us examine now, shannon entropy of marginal, Joint and conditional probability distributions on 2 sets X and Y.

i) Simple entropy s based on the marginal probability distribution

$$H(X) = -\sum_{x \in X} p(x) \log_2 p(x)$$

$$H(Y) = -\sum_{y \in Y} p(y) \log_2 p(y)$$

ii) Joint entropy :

$$H(X,Y) = -\sum_{(x,y) \in (X \times Y)} p(x,y) \log_2 p(x,y)$$

iii) Conditional entropies:

$$H(X|Y) = -\sum_{y \in Y} p(y) \sum_{x \in X} p(x|y) \log_2 p(x|y)$$

$$H(Y|X) = -\sum_{x \in X} p(x) \sum_{y \in Y} p(y|x) \log_2 p(y|x)$$

Information transmission function:

$$T(x, y) = H(x) + H(y) - H(x, y)$$

Joint entropies:

$$H(X, Y) = -\sum_{x \in X} \sum_{y \in Y} p(x, y) \log_2 p(x, y)$$

THEOREM - 3:

Prove that $H(X|Y) = H(X,Y) - H(Y)$

Proof:

$$L.H.S = H(X|Y)$$

$$= -\sum_{y\in Y} p(y) \sum_{x\in X} p(x|y) \log_2 p(x|y)$$

$$= -\sum_{y\in Y} p(y) \sum_{x\in X} \frac{p(x,y)}{p(y)} \log_2\left(\frac{p(x,y)}{p(y)}\right)$$

$$= -\sum_{y\in Y} \sum_{x\in X} p(x,y)\left[\log_2(p(x,y)) - \log_2 p(y)\right]$$

$$= -\sum_{y\in Y} \sum_{x\in X} p(x,y) \log_2 p(x,y) + \sum_{y\in Y} \sum_{x\in X} p(x,y) \log_2 p(y)$$

$$= H(X,Y) + \sum_{y\in Y} \log_2 p(y) \left(\underline{\sum_{x\in X} p(x,y)}\right)$$

$$= H(X,Y) + \sum_{y\in Y} p(y) \log_2 p(y)$$

$$= H(X,Y) - H(Y)$$

$$= RHS$$

$|||^{rly}$,

$$H(Y|X) = H(X,Y) - H(X)$$

THEOREM - 4:

Prove that $H(X,Y) \leq H(X) + H(Y)$

Proof:

$$H(X) = -\sum_{x\in X} p(x) \log_2 p(x) = -\sum_{x\in X} \sum_{y\in Y} p(x,y) \sum_{y\in Y} \log_2 p(x,y)$$

$$H(Y) = -\sum_{y\in Y} p(y) \log_2 p(y) = -\sum_{x\in X} \sum_{y\in Y} p(x,y) \sum_{x\in X} \log_2 p(x,y)$$

$$H(X) + H(Y) = -\sum_{x\in X} \sum_{y\in Y} p(x,y)\left[\log_2 \sum_{x\in X} p(x,y) + \log_2 \sum_{y\in Y} p(x,y)\right]$$

$$H(X) + H(Y) = -\sum_{x \in X} \sum_{y \in Y} p(x,y) \left[\log_2 p(y) + \log_2 p(x) \right]$$

$$\therefore H(X) + H(Y) = -\sum_{x \in X} \sum_{y \in Y} p(x,y) \log_2 \left(p(x) \cdot p(y) \right).$$

~~From Gibb's theorem~~

$$H(X,Y) = -\sum_{x \in X} \sum_{y \in Y} p(x,y) \log_2 p(x,y)$$

From Gibb's theorem,

$$\leq -\sum_{x \in X} \sum_{y \in Y} p(x,y) \log_2 \left(p(x) \cdot p(y) \right)$$

$$= H(X) + H(Y)$$

$$\therefore H(X,Y) \leq H(X) + H(Y)$$

THEOREM - 5:

Prove that $H(X) \geq H(X|Y)$ & $H(Y) \geq H(Y|X)$

Proof:

From theorem - 3,

$$H(X|Y) = H(X,Y) - H(Y) \quad \text{—— } ①$$

$$H(Y|X) = H(X,Y) - H(X) \quad \text{—— } ②$$

From theorem - 4,

$$H(X,Y) \leq H(X) + H(Y) \quad \text{—— } ③$$

Substitute ① in ③, we get

$$H(X|Y) + \cancel{H(Y)} \leq H(X) + \cancel{H(Y)}.$$

$$\Rightarrow \quad H(X|Y) \leq H(X)$$

Substitute ② in ③, we get

$$H(Y|X) + \cancel{H(X)} \leq \cancel{H(X)} + H(Y)$$

$$\Rightarrow \quad H(Y|X) \leq H(Y)$$

THEOREM-6:

Prove that $H(X) - H(Y) = H(X|Y) - H(Y|X)$

Proof:

Consider the information transition function

$$T(X,Y) = H(X) + H(Y) - H(X,Y)$$

$$= H(X) - (H(X,Y) - H(Y))$$

$$= H(X) - H(X|Y) \quad \text{—— } (1)$$

[from thrm 3

and

$$T(X,Y) = H(Y) - (H(X,Y) - H(X))$$

$$= H(Y) - H(Y|X) \quad \text{—— } (2)$$

[from thrm 3

$(1) = (2) \Rightarrow$

$$H(X) - H(X|Y) = H(Y) - H(Y|X)$$

$$\Rightarrow H(X) - H(Y) = H(X|Y) - H(Y|X)$$

Measure of dissonance:

Let us consider two bodies of evidence (f_1, m_1) and (f_2, m_2). Assume that they are defined on the same Universal set X and obtained from independent sources.

The conflict which each other

$m_1(A) \neq 0$, $m_2(B) \neq 0$ and $A \cap B = \phi$

The total amount of conflict of dissonance should be monotonic increasing with

$$K = \sum_{A \cap B = \phi} m_1(A) \cdot m_2(B)$$

where,

$A \in f_1$, $B \in f_2$

The total amount of conflict between two bodies of evidence is expressed by the function

$$Con(m_1, m_2) = -\log_2 (1-k)$$

where $k \in [0, \infty)$

This function takes values, from 0 to ∞

If $k=0$, $Con(m_1, m_2) = -\log_2 1 = 0$

If $k=1$, $Con(m_1, m_2) = -\log_2 0 = \infty$ (not defined)

$\therefore$ It is clearly monotonic increasing with k; $Con(m_1, m_2) = 0$, only if m_1 & m_2 donot conflict at all $(k=0)$.

and $Con(m_1, m_2) = \infty$, only if they conflict totally, $(k=1)$.

Defn: (Measure of dissonance)

$$E : \mathcal{M} \longrightarrow [0, \infty)$$

where $\mathcal{M}$ denotes the set of all basic assignments defined on the power sets with 2^n elements, for any $n \in N$

THEOREM: 7

$E(m)$ can be justified as a measure of conflict ~~of~~ or dissonance. 26

To derive $E(m)$, Let $m_A(B)$ denote a basic assignment defined on the Universal set X, such that

$$m_A(B) = \begin{cases} 1, & B = A \\ 0, & \text{otherwise}, \forall\ B \in \wp(X) \end{cases}$$

The given body of evidence is given by the family of focal elements (f, m)

$$Con(m, m_A) = -\log_2 (1-k)$$
$$= -\log_2 \left(1 - \sum_{B \cap C = \phi} m(B)\, m_A(C)\right)$$

Since, $m_A(c) = 0$, $\forall\ C \neq A$ and $m_A(c) = 1$, for $C = A$

$$Con(m, m_A) = -\log_2 \left(1 - \sum_{B \cap A = \phi} m(B)\right)$$
$$= -\log_2 \left(\sum_{B \cap A \neq \phi} m(B)\right)$$
$$= -\log_2 Pl(A)$$

now, we define $E(m)$ as a weighted average of conflict $Con(m, m_A)$

$$E(m) = \sum_{A \in f_1} m(A)\, Con(m, m_A)$$

$$\therefore E(m) = -\sum_{A \in f_1} m(A) \log_2 Pl(A)$$

THEOREM : 8

If m represents a probability distribution on X, then the measure (m) of dissonance in evidence is equivalent to shannon entropy.

Proof:

For probability measures defined on X,

Case(i): $m(A) = 0$, for $A \neq \{x\}$, $x \in X$

Hence, $m(A) \cdot \log_2 pl(A) = 0$, $\forall A \neq \{x\}$, $x \in X$

Case(ii): For $A = \{x\}$,

$$E(m) = -\sum m(\{x\}) \log_2 pl(\{x\})$$

Since,

$$m(\{x\}) = pl(\{x\})$$

$= P(x)$, where $P(x)$ is the probability of x.

$$E(m) = -\sum_{x \in X} P(x) \cdot \log_2 P(x)$$

$$\therefore E(m) = H(P(x) \mid x \in X)$$

THEOREM - 9:

Let m_x and m_y be marginal basic assignments on set X and Y respectively, and let 'm' be a joint basic assignment on $X \times Y$, such that

$$m(A \times B) = m_x(A) \cdot m_y(B), \quad \forall A \in \wp(X), B \in \wp(Y)$$

Then,

$$E(m) = E(m_x) + E(m_y)$$

Proof: W.K.T

$$pl(A \times B) = \sum_{(C \times D) \cap (A \times B) \neq \phi} m(C \times D)$$

$$Pl(A \times B) = \sum_{(C \times D) \cap (A \times B) \neq \phi} m_x(C)\, m_y(D)$$

$$= \sum_{C \cap A \neq \phi} m_x(C) \sum_{D \cap B \neq \phi} m_y(D)$$

$$\therefore Pl(A \times B) = Pl(A) \cdot Pl(B)$$

$$E(m) = -\sum m(A \times B) \log_2 Pl(A \times B)$$

$$= -\sum_{A,B} m_x(A) \cdot m_y(B) \cdot \log_2 (Pl(A) \cdot Pl(B))$$

$$= -\sum_{A,B} m_x(A) \cdot m_y(B) \cdot [\log_2 Pl(A) + \log_2 Pl(B)]$$

$$= -\sum_{A,B} m_x(A)\, m_y(B) \log_2 Pl(A) - \sum_{A,B} m_x(A)\, m_y(B) \log_2 Pl(B)$$

$$= -\sum_{A \subset X} m_x(A) \cdot \log_2 Pl(A) - \sum_{B \subset Y} m_y(B) \log_2 Pl(B)$$

$$\therefore E(m) = E(m_x) + E(m_y)$$

Measure of confusion:

Plausibility measure and Belief measure are dual in sense that

$$Pl(A) = 1 - Bel(\bar{A})$$

We can define a measure of confusion as

$$C(m) = -\sum_{A \in f} m(A) \log_2 Bel(A)$$

where f is the set of all local elements of the basic assignment. This function $C(m)$ is called measure of confusion

Since, $Bel(A) \le Pl(A)$, $C(m) \ge E(m)$

THEOREM - 10:

Let m_x and m_y be marginal basic assignments on the set X and Y respectively and let 'm' be a joint basic assignment on $X \times Y$, such that

$$m(A \times B) = m_x(A) \cdot m_y(B), \quad \forall A \in \mathcal{P}(X), B \in \mathcal{P}(Y)$$

then,

$$C(m) = C(m_x) + C(m_y)$$

Proof:

W.k.T.

$$Bel(A \times B) = \sum_{(C \times D) \subseteq (A \times B) \neq \phi} m(C \times D)$$

By assumption of the theorem,

$$Bel(A \times B) = \sum_{(C \times D) \subseteq (A \times B) \neq \phi} m_x(C) \cdot m_y(D)$$

$$= \sum_{C \cap A \neq \phi} m_x(C) \cdot \sum_{D \cap B \neq \phi} m_y(D).$$

$$Bel(A \times B) = Bel(A) \cdot Bel(B)$$

now,

$$C(m) = -\sum_{A, B \subset X, Y} m(A \times B) \log_2 Bel(A \times B)$$

$$= -\sum_{A, B \subset X, Y} m_x(A) \cdot m_y(B) \cdot \log_2 (Bel(A) \cdot Bel(B))$$

$$= -\sum_{A, B \subset X, Y} m_x(A) \cdot m_y(B) \left[\log_2 Bel(A) + \log_2 Bel(B)\right.$$

$$C(m) = -\sum_{A,B \subset X,Y} m_x(A) \cdot m_y(B) \log_2 Bel(A)$$

$$- \sum_{A,B \subset X,Y} m_x(A) \cdot m_y(B) \log_2 Bel(B)$$

30 $$= -\sum_{A \subset X} m_x(A) \log_2 Bel(A)$$

$$- \sum_{B \subset Y} m_y(B) \log_2 Bel(B)$$

$$\therefore C(m) = C(m_x) + C(m_y)$$

Hence the proof.

Example - 5.6 :

To illustrate the meaning of thrm 9, let us consider the non-interactive marginal basic assignments m_x and m_y on set $X = \{x_1, x_2, x_3\}$ and $Y = \{y_1, y_2, y_3\}$ respectively that are specified in table I(a). Their joint basic assignment m, which is defined by

$$m_i(A \times B) = m_x(A) \cdot m_y(B)$$

Table I(a): $m(A \times B) = m_x(A) \cdot m_Y(B)$

$m(A \times B)$	$m_y(\{y_1\}) = .2$	$m_y(\{y_2, y_3\}) = .5$	$m_Y(\{y_2\}) = .3$
$m_x(\{x_1, x_2\}) = 0.4$	0.08	0.02	0.12
$m_x(\{x_3\}) = 0.1$	0.02	0.05	0.03
$m_x(\{x_1, x_3\}) = 0.3$	0.06	0.15	0.09
$m_x(\{x_1, x_2, x_3\}) = 0.2$	0.04	0.1	0.06

$$Pl(A \times B) = Pl_x(A) \cdot Pl_y(B)$$

31

$Pl(A \times B)$	$Pl_y\{y_1\} = 0.2$	$Pl_Y(\{y_2, y_3\}) = 0.8$	$Pl_y(\{y_3\}) = 0.8$
$Pl_x(\{x_1, x_2\}) = 0.9$	0.18	0.72	0.72
$Pl_x(\{x_3\}) = 0.6$	0.12	0.48	0.48
$Pl\{x_1, x_3\} = 1$	0.2	0.8	0.8
$Pl\{x_1, x_2, x_3\} = 1$	0.2	0.8	0.8

Solution:

Verify $E(m) = E(m_x) + E(m_y)$

W.K.T $E(m_x) = -\sum_{A \in F} m(A) \log_2 Pl(A)$

and $Pl(A) = \sum_{B \cap A \neq \phi} m_x(B)$

$\therefore Pl(\{x_1, x_2\}) = m\{x_1, x_2\} + m\{x_1, x_3\} + m_1\{x_1, x_2, x_3\}$

$= 0.4 + 0.3 + 0.2$

$= 0.9$

$Pl(\{x_3\}) = m(\{x_3\}) + m\{x_1, x_3\} + m\{x_1, x_2, x_3\}$

$= 0.1 + 0.3 + 0.2$

$= 0.6$

$Pl(\{x_1, x_3\}) = m\{x_1, x_2\} + m\{x_3\} + m\{x_1, x_3\} + m\{x_1, x_2, x_3\}$

$= 0.4 + 0.1 + 0.3 + 0.2$

$= 1$

$Pl(\{x_1, x_2, x_3\}) = m\{x_1, x_2, x_3\} + m\{x_3\} + m\{x_1, x_3\} + m\{x_1, x_2\}$

$= 0.2 + 0.1 + 0.3 + 0.4$

$= 1$

$$E(m_x) = -\sum m(A) \log_2 Pl(A)$$

$$= -m\{x_1, x_2\} \log_2 Pl(\{x_1, x_2\})$$

0.022

32 $$= -.4 \log_2 .9 - 0.1 \log_2 .6 - 0.3 \log_2 1.0 - 0.2 \log_2 1$$

$+0.04$

$$E(m_x) = 0.14$$

III^rly,

$$E(m_y) = -\sum m(B) \log_2 Pl(B)$$

$$Pl(\{y_1\}) = 0.2$$

$$Pl(\{y_2, y_3\}) = 0.8$$

$$Pl(\{y_2\}) = 0.8$$

$$E(m_y) = -0.2 \log_2 0.2 - 0.5 \log_2 0.8 - 0.3 \log_2 0.8$$

$$E(m_y) = 0.72$$

$\log_{10} 0.2$

$\log_{10} 2$

$$\therefore E(m_x) + E(m_y) = 0.14 + 0.72$$

$$E(m_x) + E(m_y) = 0.86 \quad \text{—— } (1)$$

also,

$$E(m) = -\sum m(A \times B) \log_2 Pl(A \times B)$$

$$= -0.08 \log_2 0.18 - 0.2 \log_2 0.72 - 0.12 \log_2 0.72 - 0.02 \log_2 0.12$$
$$- 0.05 \log_2 0.48 - 0.03 \log_2 0.48$$
$$- 0.6 \log_2 0.2 - 0.15 \log_2 0.8$$
$$- 0.09 \log_2 0.8 - 0.04 \log_2 0.2$$
$$- 0.1 \log_2 0.8 - 0.06 \log_2 0.8$$

$$= 0.2 + 0.1 + 0.06 + 0.06 + 0.05 + 0.03$$
$$+ 0.14 + 0.05 + 0.03 + 0.09 + 0.03 + 0.02$$

$$E(m) = 0.86 \quad \text{—— } (2)$$

(1), (2) $\Rightarrow$ $$E(m) = E(m_x) + E(m_y)$$

Measures of Non-specificity:

The Hartley measure captures that aspect of uncertainty that is well characterized by the term non-specificity. It is therefore, reasonable to assume that the possibilistic measure of uncertainty derived from the Hartley measure is also connected with non-specificity.

In addition, of course, it must also satisfy appropriate generalizations of the requirements upon which the Hartley measure is founded.

Let $U : \mathcal{R} \longrightarrow [0, \infty)$

where $\mathcal{R}$ denotes the set of all ordered possibility distributions, be a function such that $U(r)$ is supposed to characterize the amount of uncertainty associated with the possibility distribution r.

The function U should satisfy all properties of the Shannon entropy.

U_1) Expansibility:

When components of zero are added to a given possibility distribution, the value of U should not change.

U_2) Subadditivity:

When r_X and r_Y are marginal possibility distributions that are calculated from a joint possibility distribution r by the eqn's

$$r_X(x) = \max_{y \in Y} [r(x, y)], \text{ for each } x \in X$$

34

and

$$r_Y(y) = \max_{x \in X} [r(x, y)], \text{ for each } y \in Y$$

then we have

$$U(r) \leq U(r_X) + U(r_Y)$$

U3) Additivity:

If r_X and r_Y in U2 are non-interactive (in the possibilistic sense) that is

$$r(x, y) = \min [r_X(x), r_Y(y)], \forall \begin{matrix} x \in X \\ y \in Y \end{matrix}$$

then

$$U(r) = U(r_X) + U(r_Y)$$

U4) Continuity:

U should be a continuous function

U5) Monotonicity:

For any pair r_1, r_2 of possibility distributions of the same length ($r_1, r_2 \in R$ for some $n \in N$), if $r_1 \leq r_2$, then

$$U(r_1) \leq U(r_2)$$

U6) Minimum:

For all possibility distributions, $U(r) = 0$ iff exactly one component of r is equal to 1 and all of its remaining components are 0s.

U7) Maximum:

Among all possibility distributions of the same length n ($n \in N$), function U should attain its maximum for the distribution for which all elements are 1s.

U8) Branching:

for every possibility distribution $r = (p_1, p_2, \ldots p_n)$ of length n $(n \in N)$

$$U(p_1, p_2, \ldots p_n) = U(p_1, p_2 \ldots p_{k-2}, p_k, p_k, p_{k+1}, \ldots p_n)$$
$$+ (p_{k-2} - p_k)\, U\Big(\underbrace{1, 1, \ldots 1}_{k-2}, \frac{p_{k-1} \cdots p_k}{p_{k-2} \cdots p_k}, \underbrace{0, 0, \ldots 0}_{n-k+1}\Big)$$
$$- (p_{k-2} - p_k)\, U\Big(\underbrace{1, 1, \ldots 1}_{k-2}, \underbrace{0, 0, \ldots 0}_{n-k+2}\Big)$$

for each $k \in N_{3,n}$

U9) Normalization:

$$U(1, 1) = 1$$

It has been established that the only function that satisfies the requirements U_1 through U_9 is

$$U(r) = \sum_{i=1}^{n} (p_i - p_{i+1}) \log_2 |A_i| \qquad (1)$$

where $p_{n+1} = 0$ by convention and

$$A_i = \{x \in X \mid r(x) \geq p_i\} \qquad (2)$$

The set A_i in (2) can be viewed as the p_i-cut of a fuzzy set A defined on X by

$$\mu_A(x) = r(x), \quad \forall\, x \in X$$

In this sense, $U(r)$ can be viewed as a weighted average of the Hartley information of the ρ_i - cuts A_i of the fuzzy set A, since $\rho_i - \rho_{i+1} = m(A_i)$ by

$$\mu_i = \rho_i - \rho_{i+1}, \quad \text{—} \quad (*)$$

36 the weights of sets A_i are the values $m(A_i)$ of the basic assignment corresponding to the possibility distribution r.

Function U defined by (1) is usually called a U-uncertainty.

Since possibility distributions on which U is defined are assumed to be ordered and the sets A_i are nested, it is obvious that $|A_i| = i$, $\forall\ i \in N$.

(1) can thus be re-written in a simpler form,

$$U(r) = \sum_{i=1}^{n} (\rho_i - \rho_{i+1}) \log_2 i \quad \text{—} \quad (3)$$

Furthermore, using the 1-1 correspondance between possibility distributions and basic distributions given by $\rho_i = \sum_{k=1}^{n} \mu_k$ and $\mu_i = \rho_i - \rho_{i+1}$, we can also express the U-uncertainty in terms of the basic distributions. One form, based upon (1) is

$$U(r) = U(t^{-1}(m)) = \sum_{i=1}^{n} m(A_i) \log_2 |A_i| \quad \text{—} \quad (4)$$

Another form, based upon (2) is

$$U(r) = U(t^{-1}(m)) = \sum_{i=1}^{n} \mu_i \log_2 i \quad \text{—} \quad (5)$$

where, $(\mu_i \mid i \in N_n) = (t(r_i) \mid i \in N_n)$

Example :

Calculate $U(r)$ for the possibility distribution

$$r = \{1, 1, 0.8, 0.7, 0.7, 0.7, 0.4, 0.3, 0.2, 0.2\}$$

37

Soln :

Given possibity distribution,

$$r = \{\rho_1, \rho_2, \rho_3, \rho_4, \rho_5, \rho_6, \rho_7, \rho_8, \rho_9, \rho_{10}\}$$

$$= \{1, 1, 0.8, 0.7, 0.7, 0.7, 0.4, 0.3, 0.2, 0.2\}$$

We defined

$$t : R \rightarrow M \ni$$

$$t(r) = m$$

W.K.T.

$$m = (\mu_1, \mu_2, \mu_3, \mu_4, \mu_5, \mu_6, \mu_7, \mu_8, \mu_9, \mu_{10})$$

(4) $\Rightarrow$

$$m = (0, 0.2, 0.1, 0, 0, 0.3, 0.1, 0.1, 0, 0.2)$$

(5) $\Rightarrow$

$$U(r) = \sum_{i=1}^{n} \mu_i \log_2 i$$

$$= 0.\log_2 1 + 0.2 \log_2 2 + 0.1 \log_2 3 + 0 \log_2 4 + 0.\log_2 5 + 0.3 \log_2 6 + 0.1 \log_2 7 + 0.1 \log_2 8 + \overset{0.\log_2 9}{} 0.2 \log_2 11$$

$$= 0 + 0.2 \frac{\log_{10} 2}{\log_{10} 2} + 0.1 \frac{\log_{10} 3}{\log_{10} 2} + 0 + 0.\frac{\log_{10} 5}{\log_{10} 2} + 0.3 \frac{\log_{10} 6}{\log_{10} 2} + 0.1 \frac{\log_{10} 7}{\log_{10} 2} + 0.1 \frac{\log_{10} 8}{\log_{10} 2} + 0 + 0.2 \frac{\log_{10} 11}{\log_{10} 2}$$

$$= 0 + 0.2 + 0.1(1.58) + 0.3(2.58) + 0.1(2.81) + 0.1(3) + 0.2(3.32)$$

$$\therefore U(r) = 2.38$$