II YEAR – III SEMESTER
COURSE CODE: 7MMA3E4

ELECTIVE COURSE- IV- (A) – FUZZY MATHEMATICS

Unit I

Crisp sets and fuzzy sets.

Unit II

Operation on fuzzy sets.

Unit III

Fuzzy relations.

Unit IV

Fuzzy measures.

Unit V

Uncertainty and Information.

Text Books

1. J.Klir and Tina A Folger, Fizzy Sets, Uncertainty and Information, Prentice Hall of India Private Ltd., New Delhi, 2006

Chapters : I, II, III, IV and V upto section 5.5.

Books for Supplementary Reading and Reference:

1. V.Novak, Fuzzy Sets and Their Applications, Adom Hilger, Bristol, 1969.
2. A.Kaufman, Introduction to the Theory of Fuzzy Subsets, Academic Press, 1975.
3. H.J.Zimmermann, Fuzzy Set Theory and its Applications, Allied Publishers, Chennai, 1996.

♣♣♣♣♣♣♣♣♣♣

(1)

unit - I

crisp set :-

A crisp set is defined in such a way us to divide the individuals in some given universe into two groups.

i) members that certainly belongs to the set.

ii) members that certainly do not (or) doesn't belongs to the set.

Ex :-

$Z = \{0, \pm 1, \pm 2, \ldots\}$

R - the set of all real numbers.

$N = \{1, 2, \ldots \infty\}$

$N_0 = \{0, 1, 2, \ldots \infty\}$

$N_n = \{1, 2, \ldots, n\}$

characteristic function (or) Discrimination :-

A set is defined by a function usually called a characteristic function that declares which element of x are members of the set and which are not

(ii) $$\mu_A(x) = \begin{cases} 1 & \text{iff } x \in A \\ 0 & \text{iff } x \notin A \end{cases}$$

which is formally expressed by

$$\mu_A : x \to [0, 1]$$

(ii) This function maps elements of the universal set containing 0 and 1. This can be defined by $\mu_A : x \to [0, 1]$

(i) The universal set x are determined to be either member (or) non members of a set can be defied by a characteristic or determination function for given set A this function a value $\mu_A(x)$ to every $x \in X$ such that

$$\mu_A(x) = \begin{cases} 1 & \text{iff } x \in A \\ 0 & \text{iff } x \notin A \end{cases}$$

(2) Fuzzy Set :-

A Fuzzy set can be defined mathematically by assagning to each possible individuals in the universe of discourse a value representing its grade of membership in the fuzzy set.

Eg :-

1. The class of real members which are much greater than 1.

2. The class of tall man

Universal set :-

The letter x denotes the ~~inver~~ universe of discourse or universal set. This set contains all the possible elements of concern in each particular context (or) application from which sets can be formed unless otherwise stated, x is assumed in this text to contain a finite numbers of elements.

The indicate that an individual objects x is a members (or) element of a set A we write $x \in A$

when ever x is not an element of a set A we write $x \notin A$.

Lest method and Rule method :-

Not Important

A set can be described either by naming all its members (the list method) or by specifying some well-defined

(3) properties satisfied by the member of the set (the rule method).

The list method between can be used only for finite sets the set A whose member are $a_1, a_2, \ldots, a_n$ is usually written as $A = \{a_1, a_2, \ldots, a_n\}$

and the set B whose members satisfy the properties $P_1, P_2, \ldots, P_n$ is usually written as

$$B = \{b \mid b \text{ has properties } P_1, P_2, \ldots, P_n\}$$

where the symbol | denotes the phrase " Such that".

Convex set :-

A set A in R^n is called convex if for every pair of points.

$$r = \{r_i \mid i \in N_n\} \text{ and } s = \{s_i \mid i \in N_n\}$$

in A and every real numbers λ between 0 and 1 exclusively, the point

$$t = (\lambda r_i + (1 - \lambda_i) s_i \mid i \in N_n\}$$

is also in A.

In other words a set A in R^n is convex if for every pair of points r and s in A. all points located on the stright line segment connecting r and s are also in A.

Eg :-

Eg :-

Eg : 2

(ii) $A : [0,2] \cup [3,5]$

$r = 1, s = 1$

$\lambda = 2/5 = 0.4$

$t = \lambda r_i + (1-\lambda) s_i$

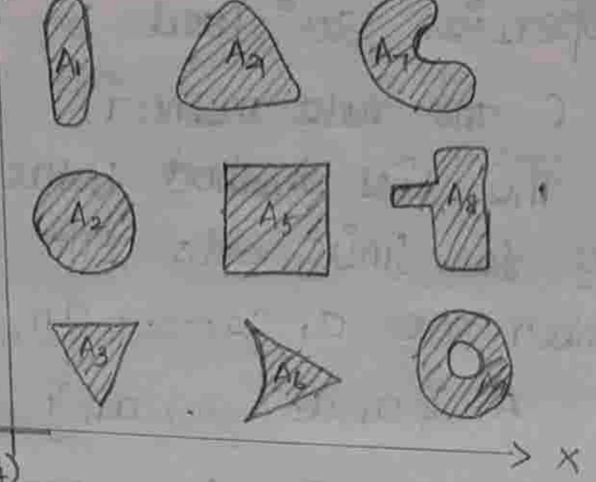
$\Rightarrow (0.4)1 + (1-0.4)(4)$

$\Rightarrow 0.4 + (0.6)(4)$

$\Rightarrow 0.4 + 2.4$

$t = 2.8 \notin A$

$\therefore$ A is non-convex



Example of sets in R^2 that are either convex ($A_1 - A_5$) (or) non convex ($A_6 - A_9$)

Family of sets :-

A set whose elements are themselves sets is often reffered to as a family of sets. It can be defined in the form

$$\{A_i \mid i \in I\}$$

where i and I are called the set identifies and the identification set respectively. Because the index i is used to reference the set Ai, the family of sets is also called an index set.

cardinality of set :- [or] Scalar cardinality

[The number of elements that belong to a set A is called the cardinality of the set and is denoted by |A|] A set that is defined by the rule method may contain an infinite number of elements.

Eg :- $A = \{1, 2, 3, 4, 5, 6\}$ $\quad |A| = \sum_{x \in X} A(x)$

$n(A) = 6 \Rightarrow$ cardinality of A.

5

power set :-

The family of sets consisting of all the subsets of a particular set A is referred to as the power set of A and is indication P(A).

$O(A) = 2^{|A|}$

$$|P(A)| = 2^{|A|}$$

Scalar cordinality :-

The scalar cordinality of For any fuzzy set A defined on a finite universal set X we defined scalar cordinality |A| as follows.

$$|A| = \sum_{x \in X} A(x)$$

Ex :-

$A(x) = 1/x, \quad x = 1, 2, 3, 4, 5$

$A(1) = 1, \quad A(2) = 0.5, \quad A(3) = 0.3, \quad A(4) = 0.25$

$A(5) = 0.2$

$$|A| = 1 + 0.5 + 0.3 + 0.25 + 0.2$$

$$= 2.25$$

Fuzzy cordinality :-

Fuzzy cordinality is defined as a fuzzy number rather that as a real number is the case for the scalar cordinality. when a fuzzy set A has a finite support its fuzzy cordinality $|\tilde{A}|$ is a fuzzy set (fuzzy number) defined on N whose membership function is defined by $\mu_{|\tilde{A}|}(|A_\alpha|) = \alpha$

for all α in the level set of A ($\alpha \in \Lambda_A$)

The fuzzy cordinality of the fuzzy set old from table 1.2 is

$$|\tilde{old}| = \frac{.1}{7} + \frac{.2}{6} + \frac{.4}{5} + \frac{0.6}{4} + \frac{0.8}{3} + \frac{1}{2}$$

(6) Relative complement :-

The relative complement of a set A with respect to a set B is the set containing all the member of B that are not also member of A.

This can be written B-A

$$B-A = \{ x \mid x \in B \text{ ,and } x \notin A \}$$

Law of excluded Middle :-

The elements of the universal set necessarily belong either is a to a set A (or) to its absolute complement $\bar{A}$ the union of A and $\bar{A}$ yields the universal set.

$$\boxed{A \cup \bar{A} = X}$$

This property is usually called the law of contradiction..

Partition :-

A collection of pairwise disjoint non-empty subsets of a set A is called a partition on A. If the union of these subsets yields the original set A.

we denote the partition on A by the symbol $\pi(A)$.

$$\pi(A) = \{ A_i \mid i \in I \text{ , } A_i \subseteq A \}$$

where $A_i \neq \phi$ is a partition on A if and only if

$$A_i \cap A_j = \phi$$

for each pair $i \neq j$, $i, j \in I$ and

7

$$\bigcup_{i \in I} A_i = A$$

thus each element of A belongs to one and only one of the subsets forming the partition.

properties of crisp set operations:-

Involution	$\bar{\bar{A}} = A$
commutativity	$A \cup B = B \cup A$ $A \cap B = B \cap A$
Associativity	$(A \cup B) \cup C = A \cup (B \cup C)$ $(A \cap B) \cap C = A \cap (B \cap C)$
Distributivity	$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
Idempotence	$A \cup A = A$ $A \cap A = A$
Absorption	$A \cup (A \cap B) = A$ $A \cap (A \cup B) = A$
Absorption of complement	$A \cup (\bar{A} \cap B) = A \cup B$ $A \cap (\bar{A} \cup B) = A \cap B$
Absorption by X and ϕ	$A \cup X = X$ $A \cap \phi = \phi$
Identity	$A \cup \phi = A$ $A \cap X = A$

Law of contradiction	$A \cap \bar{A} = \phi$
Law of excluded Middle	$A \cup \bar{A} = X$
Demorgan's laws	$\overline{A \cap B} = \bar{A} \cup \bar{B}$ $\overline{A \cup B} = \bar{A} \cap \bar{B}$

Notation of Fuzzy Set :-

Memership Function can be generalized such that the values assigned to the elements of the universal set fall within a specified range and indicate the membership grade of there elements in the set in question. Larger values denote higher degrees of set membership. such a fn is called a memership fn & the set defined by it a Fuzzy set.

Membership function :- Let X denote a universal set. Then the membership function μ_A by which a Fuzzy set A is usually defined has the form.

$$\mu_A : X \rightarrow [0,1]$$

where [0, 1] denotes the interval of real numbers from 0 to 1. inclusive.

Each Fuzzy set is uniquely defined by one particular memership function.

Eg :- $\mu_A(x) = \dfrac{1}{1+10x^2}$

(a) consider the universal set that consists of 7 levels of education.

0 - no education, 1 - elementary education

2 - high school, 3 - two year college degree

4 - Bachelor degree, 5 - master degree

6 - doctoral degree

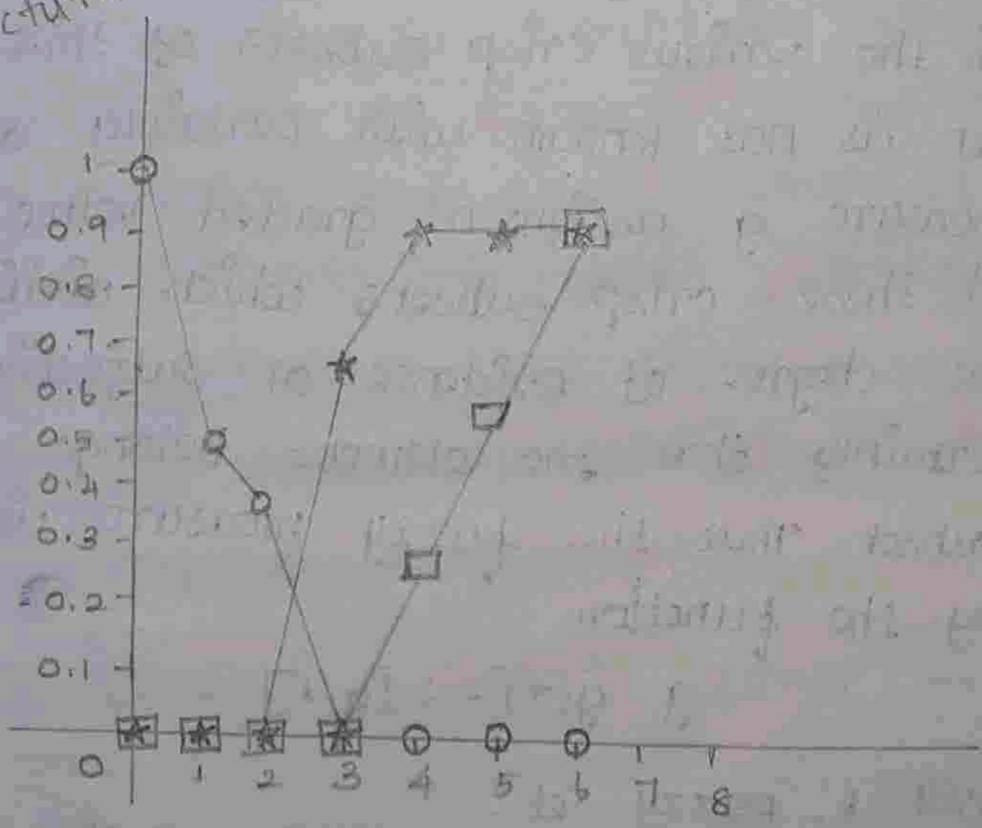
Membership function of three fuzzy set to capture the concepts of

(i) little educator

(ii) Highly educator

(iii) very highly educator people are defined by a symbol o, ★, □.

picture



10

L- Fuzzy Set :-

The generalized membership function has the form by $\mu_A \cdot x \rightarrow L$

where L denoted any set that is atleast partially ordered.

Since L is most frequently a lattice fuzzy sets, Define in the generalized membership grade function are L- Fuzzy sets.

$L = [0,1]^n$ the symbol $[0,1]^n$ is a cartesian product $[0,1] \times [0,1] \times \cdots \times [0,1]$ n times

Defn: Fuzzy measure

Given a particular product element of a universal set of concern whose membership in the various crisp subsets of this universal set is not known with certainly a Fuzzy measure g assigns a graded value to each of these crisp subsets which indicates the degree of evidence or subjective certainly that the element belongs in the subset Thus the fuzzy measure is defined by the function

$$g : g(x) \rightarrow [0,1]$$

Level k- Fuzzy Set :-

A different extension of the Fuzzy set concept involves creating fuzzy subsets of a universal set whose elements are Fuzzy sets. these fuzzy sets are known as level K fuzzy sets where K indicates the depth of nesting.

Eg :-

Given a crisp universal set X, let $\bar{P}(X)$ denote the set of all fuzzy subsets of X and let $\bar{P}^K(X)$ be defined recursively by the eqn

$$\bar{P}^K(X) = \bar{P}(P^{K-1}(X))$$

For all integers $K \geq 2$. Then fuzzy sets of level K are formally defined by membership functions of the form

$$\mu_A : \bar{P}^{K-1}(X) \rightarrow [0,1]$$

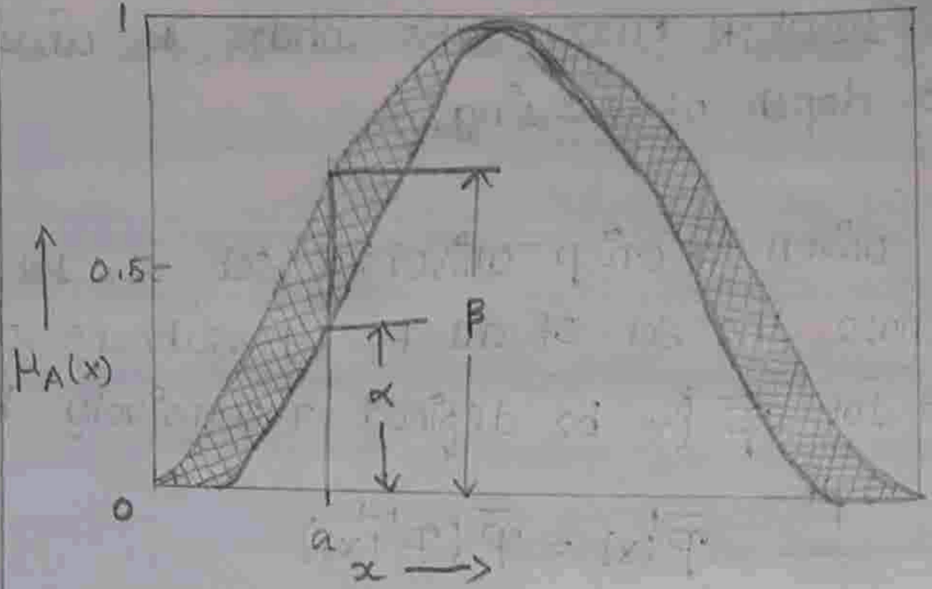
Interval value fuzzy sets :-

The requirement for a precise membership function can also be relaxed by allowing values $\mu_A(x)$ to be intervals of real numbers in $[0,1]$ rather than single numbers. Fuzzy sets of this sort are called interval-valued fuzzy sets. They are formally defined by membership functions of the form

$$\mu_A : X \rightarrow \wp([0,1])$$

where $\mu_A(x)$ is a closed interval in $[0,1]$ for each $x \in X$.

(12)



Basic concepts of Fuzzy sets :-

Defn :- The support of a fuzzy set A in the universal set X is the crisp set that contains all the elements of X that have a non-zero membership grade in A. That is support of fuzzy set in X are obtained by the function

$$\text{supp} : \tilde{P}(X) \rightarrow P(X)$$

where

$$\text{Supp } A = \{ x \in X \mid \mu_A(x) > 0 \}$$

For instance the support of the fuzzy set young from the crisp set

Table 1.2 $\text{supp (young)} = \{5, 10, 20, 30, 40, 50\}$

Elements	Infant	Adult	young	old
5	0	0	1	0
10	0	0	1	0
20	0	.8	.8	.1
30	0	1	.5	.2
40	0	1	.2	.4
50	0	1	.1	.6
60	0	1	0	.8
70	0	1	0	.1

80	0	1	0	1

(13) **Empty Fuzzy set :-**

An empty fuzzy set has an empty support that is the membership function assigns 0 to all elements of the universal set.

(x) **Normalized :-**

The height of a Fuzzy set is the largest membership grade attained by any element in that set $w(A) = 1$. A fuzzy set is called normalized when atleast one of its element attains the maximum possible membership grade.

If membership grade range in the closed interval between 0 and 1 for instance

α-cut :-

A α-cut of a fuzzy set A is a crisp set A_α that contains all the elements of the universal set x that have a membership grade in A greater than or equal to the specified value of α.

This defn can be written as

$$A_\alpha = \{ x \in X \mid \mu_A(x) \geq \alpha \}$$

Eg :-

For $\alpha = 0.2$, the α-cut of the fuzzy set young from table 1.2

$$young_{.2} = \{5, 10, 20, 30, 40\}$$

For $\alpha = 0.8$

$$young_{.8} = \{5, 10, 20\}$$

For $\alpha = 1$

$young_1 = \{5, 10\}$

(12)

Level set of Fuzzy set :-

The set of all levels $\alpha \in [0,1]$ that represent distinct α- cuts of a given fuzzy set A is called a level set of A formally.

$$\Lambda_A = \{\alpha \mid \mu_A(x) = \alpha \text{ for some } x \in X\}$$

where Λ_A denotes the level set of fuzzy set A defined on x.

-X- **Thm :-**

5M 10M A fuzzy set A on R is convex iff $A(\lambda x_1 + (1-\lambda) x_2) \geq \min\{A(x_1) \cdot A(x_2)\}$ $\forall$ $x_1, x_2 \in R$ and for all $\lambda \in [0,1]$

proof :-

Assume that A is convex

To prove : $A(\lambda x_1 + (1-\lambda) x_2) \geq \min\{A(x_1), A(x_2)\}$ $\forall x_1, x_2 \in R$ $\forall \lambda \in [0,1]$

let (W.K.T)

$\alpha = A(x_1) \leq A(x_2)$ ie) $A(x_1) \geq \alpha \Rightarrow x_1 \in {}^{\alpha}A$, $A(x_2) \geq \alpha \Rightarrow x_2 \in {}^{\alpha}A$

Then $x_1, x_2 \in {}^{\alpha}A$

more over $\lambda x_1 + (1-\lambda) x_2 \in {}^{\alpha}A$ (by defn ${}^{\alpha}A$ is convex)

for any $\lambda \in [0,1]$ by convexcity of A.

conversely,

$$A(\lambda x_1 + (1-\lambda) x_2) \leq \alpha_1 = A(x_1)$$

$$= \min\{A(x_1), A(x_2)\}$$

$$A(\lambda x_1 + (1-\lambda) x_2) \geq \min\{A(x_1), A(x_2)\}$$

conversely,

(15) Assume that

$A(\lambda x_1 + (1-\lambda) x_2) \geq \min\{A(x_1), A(x_2)\}$

$\forall x_1, x_2 \in R$

$\forall \lambda \in [0,1]$

To prove A is convex

we need to prove that for any

$\alpha \in [0,1]$, ${}^{\alpha}A$ is convex

Now for any $x_1, x_2 \in {}^{\alpha}A$ [since $A(x_1) \geq \alpha$, $A(x_2) \geq \alpha$]

and for any $\lambda \in [0,1]$

by eqn ①

$A(\lambda x_1 + (1-\lambda) x_2) \geq \min\{A(x_1), A(x_2)\}$

$\geq \min\{\alpha, \alpha\}$

$\geq \alpha$

$A(\lambda x_1 + (1-\lambda) x_2) \in {}^{\alpha}A$

$\therefore$ ${}^{\alpha}A$ is convex $\Rightarrow$ A is convex.

Concept of fuzzy set :-

[α-cut same defn ${}^{\alpha}A = \{x \mid A(x) \geq \alpha\}$]

(*) Strong α-cut :-

2M

A fuzzy set A defined on X and any number $\alpha \in [0,1]$. The strong α-cut ${}^{\alpha+}A$ is the crisp set ${}^{\alpha+}A = \{x \mid A(x) > \alpha\}$

Concave :-

A set A in R^n is called concave if for any every pair of points

$r = \{r_i \mid i \in N_n\}$

$s = \{s_i \mid i \in N_n\}$ in A

and every real number λ between 0 and 1 exclusively the points

16 $$t = \lambda r_i + (1-\lambda) s_i \setminus i \in N_n$$ is not in A.

In other words a set A in R^n is concave. If for every pair of points r and s in A all points not located on the straight line segment connecting r and s are not in A.

Eg:-

Height:-

The height h(A) of a fuzzy set in the largest membership grade obtained by any element in the set

$$h(A) = \sup_{x \in V} A(x)$$

Example: 1

Let $x = \{0, 1, 2, \dots, 10\}$, $A : X \to [0, 1]$ by $A(x) = x/10 \ \forall x \in X$

Soln:-

$A(x) = x/10$

$A(0) = 0/10$

$A(1) = 1/10 = 0.1$

$A(2) = 2/10 = 0.2$

$A(3) = 3/10 = 0.3$

$A(4) = 4/10 = 0.4$

$A(5) = 5/10 = 0.5$

$A(6) = 6/10 = 0.6$

$A(7) = 7/10 = 0.7$

$A(8) = 8/10 = 0.8$

$A(9) = 9/10 = 0.9$

$A(10) = 10/10 = 1$

$h(A) = \sup_{x \in X} A(x) = 1$

2) Let N be the natural number N $A : N \to [0,1]$ defined by $A(x) = 1/2x$.

Soln:- $A(x) = 1/2x$, $A(0) = 1/2(0) = 1/0 = \infty$

$A(1) = 1/2(1) = 1/2 = 0.5$

$A(2) = 1/2(2) = 1/4 = 0.25$

$A(3) = 1/2(3) = 1/6 = 0.16$

$\vdots$

Clearly $h(A) = \sup_{x \in X} A(x) = 0.5$

Normal and Fuzzy normal set:-

A Fuzzy set A is called normal if $h(A) = 1$

Eg:-

Let $X = \{0, 1, 2, \ldots, 10\}$; $A : x \to [0, 1]$ by

$A(x) = x_{A_0}$

Core:-

1-cut A is called the core of A

core of $A' = \{x \mid A(x) \geq 1\}$

Subnormal:-

A Fuzzy set A is called subnormal if $h(A) < 1$

Eg:- Let N be the natural number N

$A : x \to [0,1]$ defined by $A(x) = 1/2x$

Example for α-cut and strong α-cut

x	0	1	2	3	4	5	6	7	8	9	10
$A(x)$	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1

Soln:-

(18)

α - cut

$${}^{\alpha}A = \{x \mid A(x) \geq \alpha\}$$

$\alpha = 0$, ${}^{0}A = \{x \mid A(x) \geq 0\}$

$= \{0, 1, 2, \dots, 10\}$

$\alpha = 0.1$ ${}^{0.1}A = \{x \mid A(x) \geq 0.1\}$

$= \{1, 2, 3, 4, \dots, 10\}$

$\alpha = 0.2$

${}^{0.2}A = \{x \mid A(x) \geq 0.2\}$

$= \{2, 3, 4, \dots, 10\}$

$\alpha = 0.3$

${}^{0.3}A = \{x \mid A(x) \geq 0.3\}$

$= \{3, 4, 5, 6, 7, \dots 10\}$

$\alpha = 0.4$

${}^{0.4}A = \{x \mid A(x) \geq 0.4\}$

$= \{4, 5, \dots, 10\}$

$\alpha = 0.5$, ${}^{0.5}A = \{x \mid A(x) \geq 0.5\}$

$= \{5, 6, 7, 8, 9, 10\}$

$\alpha = 0.6$, ${}^{0.6}A = \{x \mid A(x) \geq 0.6\}$

$= \{6, 7, 8, 9, 10\}$

$\alpha = 0.7$, ${}^{0.7}A = \{x \mid A(x) \geq 0.7\}$

$= \{7, 8, 9, 10\}$

$\alpha = 0.8, \quad {}^{0.8}A = \{x \mid A(x) \geq 0.8\}$
$= \{8, 9, 10\}$

(19) $\alpha = 0.9, \quad {}^{0.9}A = \{x \mid A(x) \geq 0.9\}$
$= \{9, 10\}$

$\alpha = 1, \quad {}^{1}A = \{x \mid A(x) \geq 1\}$
$= \{10\}$

Strong α-cut :-

$${}^{\alpha^+}A = \{x \mid A(x) > \alpha\}$$

$\alpha = 0, \quad {}^{0^+}A = \{x \mid A(x) > 0\} = \{1, 2, \cdots, 10\}$

$\alpha = 0.1, \quad {}^{0.1^+}A = \{x \mid A(x) > 0.1\} = \{2, 3, 4, 5, \ldots, 10\}$

$\alpha = 0.2, \quad {}^{0.2}A^+ = \{x \mid A(x) > 0.2\} = \{3, 4, \cdots, 10\}$

$\vdots$

$\alpha = 1, \quad {}^{1}A^+ = \{x \mid A(x) > 1\} = \{\}$

Basic concept of the fuzzy set:-

consider the three fuzzy set that represent a concept a young, middle and old age persons.

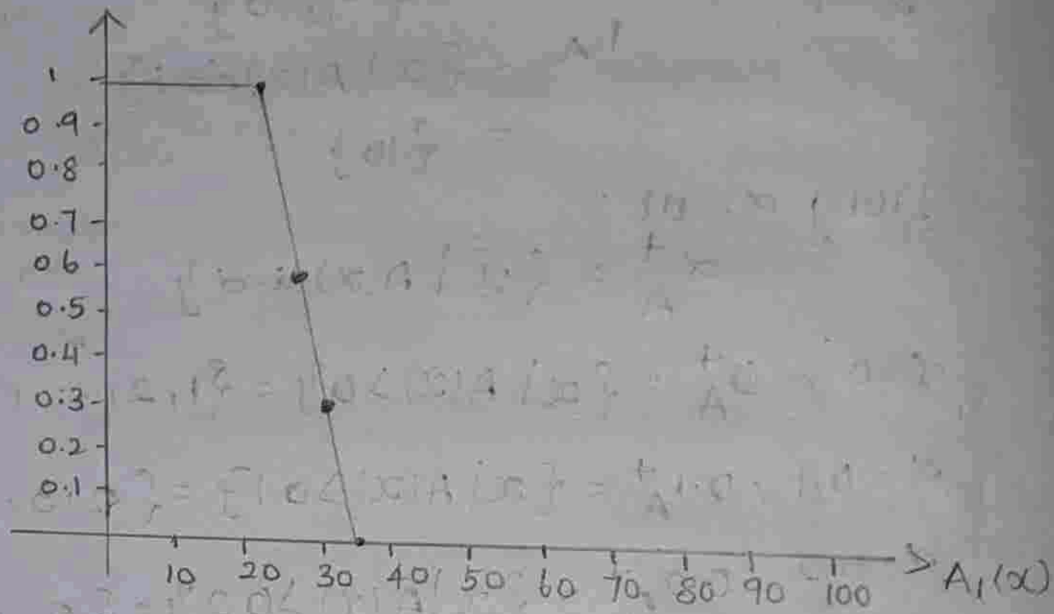
A reasonable expression by this membership function $A_1(x), A_2(x), A_3(x)$ is given below they all defied on $[0, 80]$ as follows,

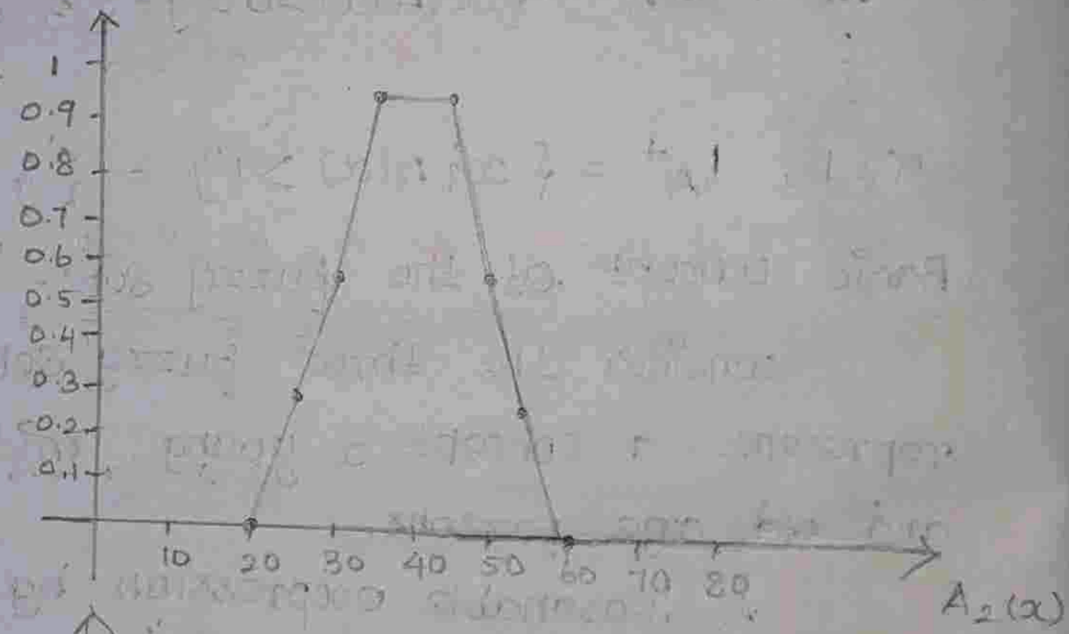
$$A_1(x) = \begin{cases} 1 & \text{if } x \leq 20 \\ \frac{35-x}{15} & \text{if } 20 < x < 35 \\ 0 & \text{if } x > 35 \end{cases}$$

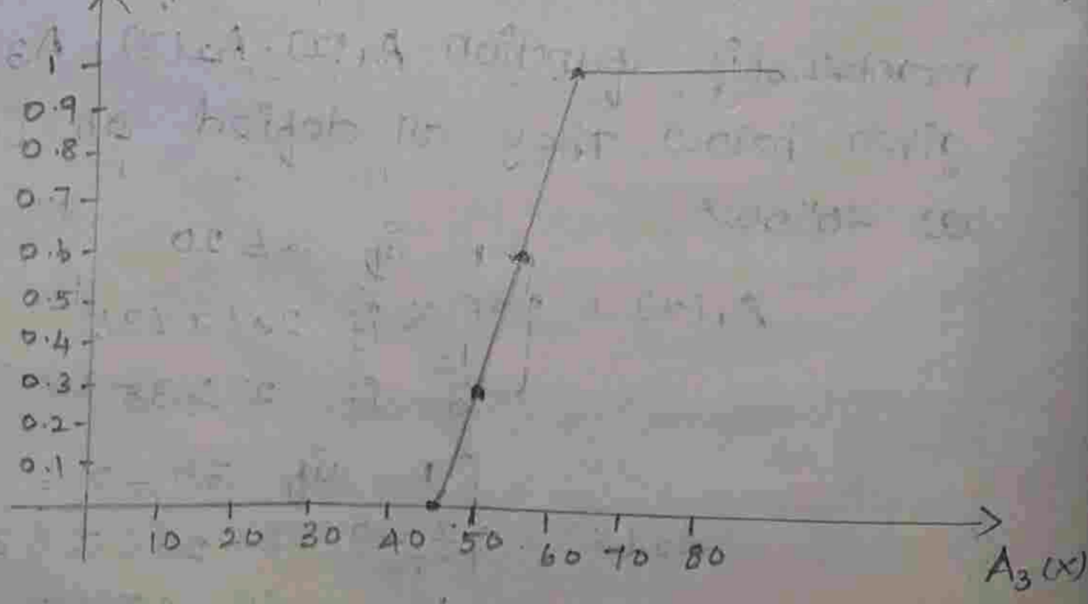
$$A_2(x) = \begin{cases} 1 & \text{if } 35 \leq x \leq 45 \\ \frac{x-20}{15} & \text{if } 20 < x < 35 \\ \frac{60-x}{15} & \text{if } 45 < x < 60 \\ 0 & \text{if } x \leq 20,\ x \geq 60 \end{cases}$$

(29)

$$A_3(x) = \begin{cases} 1 & \text{if } x \geq 60 \\ \frac{x-45}{15} & \text{if } 45 < x < 60 \\ 0 & \text{if } x \leq 45 \end{cases}$$







(21) Convex of the fuzzy set:-

A fuzzy set is said to be convex if every α-cut is convex for all $\alpha \in [0,1]$

Note:

The membership function of a convex fuzzy set is not convex.

Thm: 1.1

(*) A fuzzy set A on R is convex iff $A[\lambda x_1 + (1-\lambda)x_2] \geq \min[A(x_1), A(x_2)]$ $\forall$ x_1, x_2 and all $\lambda \in [0,1]$ where min denotes the minimum operation.

proof:-

Let A be convex and $A(x_1) \leq A(x_2) = \alpha$

To prove:-

$$A[\lambda x_1 + (1-\lambda)x_2] \geq \min[A(x_1), A(x_2)]$$

we know that,

$$^{\alpha}A = \{x \mid A(x) \geq \alpha\}$$

$$A(x) \geq \alpha \Rightarrow x \in {}^{\alpha}A$$

$$A(x_1) \geq \alpha \Rightarrow x_1 \in {}^{\alpha}A$$

$$A(x_2) \geq \alpha \Rightarrow x_2 \in {}^{\alpha}A$$

$$x_1, x_2 \in {}^{\alpha}A$$

since $^{\alpha}A$ is convex.

$[\lambda x_1 + (1-\lambda)x_2] \in {}^{\alpha}A$ for any $\lambda \in [0,1]$

$$\Rightarrow A[\lambda x_1 + (1-\lambda)x_2] \geq \alpha$$

$$\Rightarrow A(x_1)$$

$$= \min[A(x_1), A(x_2)]$$

$$A[\lambda x_1 + (1-\lambda)x_2] \geq \min[A(x_1), A(x_2)]$$

(22)

conversely,

~~To prove~~: Let us assume that

$$A[\lambda x_1 + (1-\lambda)x_2] \geq \min[A(x_1), A(x_2)]$$

To prove: α_A is convex $\forall\, x \in [0,1]$

Let $x_1, x_2 \in \alpha_A$ and $\lambda \in [0,1]$

Then $A(x_1) \geq \alpha$, $A(x_2) \geq \alpha$

By our assumption

$$A[\lambda x_1 + (1-x)x_2] \geq \min[A(x_1), A(x_2)]$$
$$\geq \min[\alpha, \alpha]$$
$$= \alpha$$

$$A[\lambda x_1 + (1-\lambda)x_2] \geq \alpha$$

$$\Rightarrow A[\lambda x_1 + (1-\lambda)x_2] \in \alpha_A$$

$\therefore$ α_A is convex for every $\alpha \in [0,1]$

Hence A is convex.

Standard Fuzzy set operations :-

1) Fuzzy complement :-

$$\bar{A}(x) = 1 - A(x)$$

2) Fuzzy union :-

$$(A \cup B)(x) = \max[A(x), B(x)]$$

3) Fuzzy intersection :-

$$(A \cap B)x = \min[A(x), B(x)]$$

23

Example:-

$$A_1(x) = \begin{cases} 1 & \text{if } x \le 20 \\ \frac{25-x}{15} & \text{if } 20 < x < 35 \\ 0 & \text{if } x \ge 35 \end{cases}$$

$$A_2(x) = \begin{cases} 1 & \text{if } 35 \le x \le 45 \\ \frac{x-20}{15} & \text{if } 20 < x < 35 \\ \frac{60-x}{15} & \text{if } 45 < x < 60 \\ 0 & \text{if } x \ge 60,\ x \ge 20 \end{cases}$$

$$A_3(x) = \begin{cases} 1 & \text{if } x \ge 60 \\ \frac{x-45}{15} & \text{if } 45 < x < 60 \\ 0 & \text{if } x \le 45 \end{cases}$$

$$\overline{A_1}(x) = \begin{cases} 0 & \text{if } x \le 20 \\ \frac{x-20}{15} & \text{if } 20 < x < 35 \\ 1 & \text{if } x \ge 35 \end{cases}$$

$$\overline{A_2}(x) = 1 - A_2(x)$$

$$= \begin{cases} 0 & \text{if } 30 \le x \le 45 \\ \frac{35-x}{15} & \text{if } 20 < x < 35 \\ \frac{x-45}{15} & \text{if } 45 < x < 60 \\ 1 & \text{if } x \le 20,\ x \ge 60 \end{cases}$$

$$\overline{A_3}(x) = 1 - A_3(x)$$

$$\overline{A_3}(x) = \begin{cases} 0 & \text{if } x \ge 60 \\ \frac{60-x}{15} & \text{if } 45 < x < 60 \\ 1 & \text{if } x \le 45 \end{cases}$$

Fuzzy intersection :-

$$(A \cap B)(x) = \min[A(x), B(x)]$$

a) $A_1 \cap A_3 = \{0$ if $0 \leq x \leq 80$

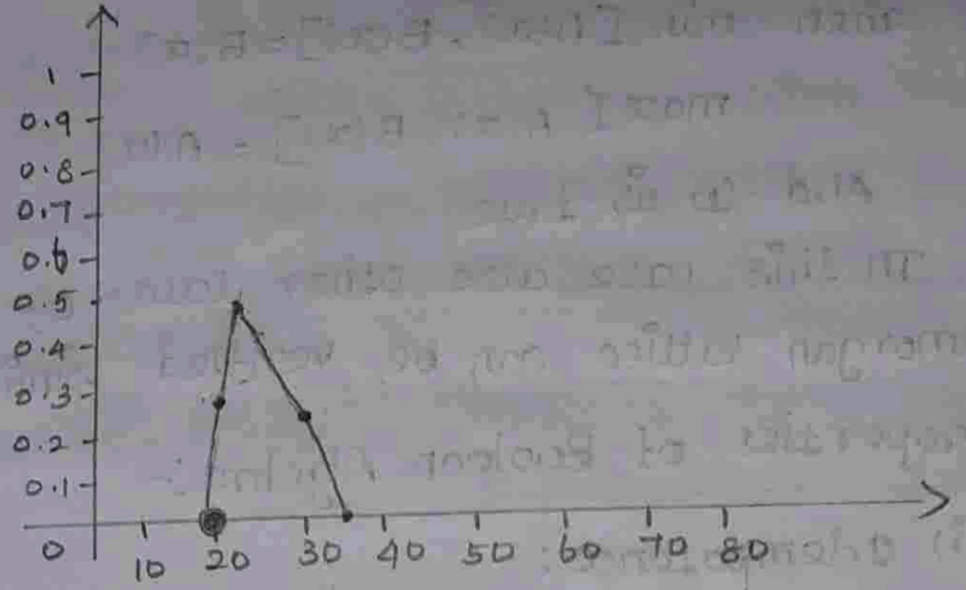
b) $$A_1 \cap A_2 = \begin{cases} 0 & \text{if } x \leq 20,\ x \geq 25 \\ \frac{x-20}{15} & \text{if } 20 < x \leq 27.5 \\ \frac{35-x}{15} & \text{if } 27.5 \leq x < 35 \end{cases}$$

c) $$A_2 \cap A_3 = \begin{cases} 0 & \text{if } x \leq 20 \\ \frac{x-20}{15} & \text{if } 20 < x \leq 27.5 \\ \frac{35-x}{15} & \text{if } 27.5 < x < 35 \\ \frac{x-45}{15} & \text{if } 45 < x < 52.5 \\ \frac{60-x}{15} & \text{if } 52.5 < x < 60 \end{cases}$$

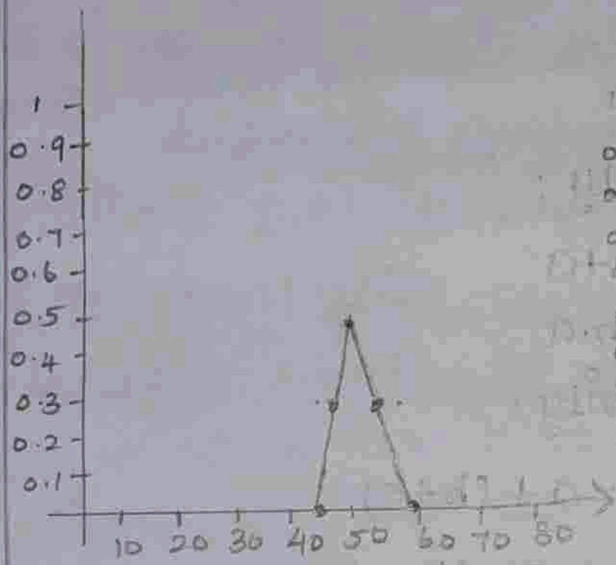
c) $$A_2 \cap A_3 = \begin{cases} 0 & \text{if } x \leq 45,\ x \geq 60 \\ \frac{x-45}{15} & \text{if } 45 < x \leq 52.5 \\ \frac{60-x}{15} & \text{if } 52.5 \leq x < 60 \end{cases}$$

$$B \cup C = \begin{cases} 0 & \text{if } x \leq 20 \\ \frac{x-20}{15} & \text{if } 20 < x \leq 27.5 \\ \frac{35-x}{15} & \text{if } 27.5 < x \leq 35 \\ \frac{x-45}{15} & \text{if } 45 < x < 52.5 \\ \frac{60-x}{15} & \text{if } 52.5 < x < 60 \end{cases}$$

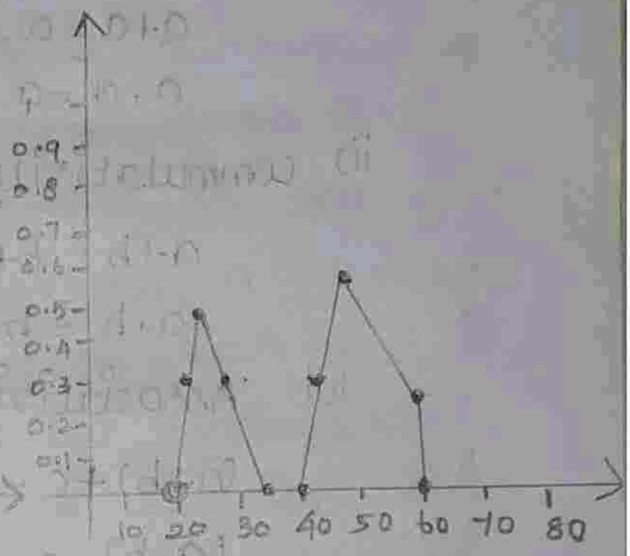
25



(B)



(C)



(BUC)

Note :-

$A \cup (A \cap B) = A$ is true.

To prove:

i.e) $\max[A(x), \min[A(x), B(x)]] = A(x) \; \forall \; x \in X$ — ①

case (i) :-

If $A(x) < B(x)$

then $\min[A(x), B(x)] = A(x)$

$\max[A(x), A(x)] = A(x)$

① is true.

case (ii) :-

If $A(x) > B(x)$

26

Then $\min[A(x), B(x)] = B(x)$

$\max[A(x), B(x)] = A(x)$

And ① is true.

In this case also other laws of Demorgan lattice can be verified similarly.

Properties of Boolean Algebra :-

i) Idempotence :-

$$a+a=a$$
$$a.a=a$$

ii) commutativity :

$$a+b = b+a$$
$$a \cdot b = b \cdot a$$

(iii) Associativity :

$$(a+b)+c = a+(b+c)$$
$$(a \cdot b) \cdot c = a \cdot (b \cdot c)$$

(iv) Absorption :

$$a+(a \cdot b) = a$$
$$a \cdot (a+b) = a$$

(v) Distributively ;-

$$a \cdot (b+c) = (a \cdot b) + (a \cdot c)$$
$$a + (b \cdot c) = (a+b) \cdot (a+c)$$

(vi) Complementary :

$$a+\bar{a} = 1$$
$$a \cdot \bar{a} = 0$$
$$\bar{1} = 0$$

(vii) Involution :-

$$\bar{\bar{a}} = a$$

(viii) universe bound :-

(27)

$$a+0 = a$$
$$a+1 = 1$$
$$a \cdot 0 = 0$$
$$a \cdot 1 = a$$

(ix) Dualization :

$$\overline{a+b} = \overline{a} \cdot \overline{b}$$
$$(\overline{a \cdot b}) = \overline{a} + \overline{b}$$

Existential quantification :-

Existential quantification of a predicate p(x) is expressed by the form (∃(x). P(x)) which represents a sentence if there exists an individuals x (in the universal set X of the individuals variable x) such that x is p(or) the equivalent (f) is called an existential quantifier we have the following equality.

$$((\exists x)\ P(x) = \bigvee_{x \in X} P(x))$$

universal quantification :-

universal quantification of a predicate p(x) is expressed by the form ∀ x, p(x) which represents the sentence for every individuals $x \in P^{\alpha}$ and equalence sentence

" All ∠ x belong to x are p "

The symbol ∀ is called universal qualifies clearly, the following equality holds.

$$[\forall (x)\ P(x) = \bigwedge_{x \in X} P(x)]$$

28

For instants $(\exists x_1)(\forall x_2)(\exists x_3) P(x_1, x_2, x_3)$

There exists an $x \in X$ such that exists $x_3 \in X_3$ such that $P(x_1, x_2, x_3)$

Fuzzy set inclusion:-

consider $[\mathcal{F}(x) \subseteq C]$ where $\subseteq$ denotes a fuzzy set inclusion by which elements of $\mathcal{F}(x)$ are partially ordered.

ie) For $A, B \in \mathcal{F}(x)$ we say that A is the subset of B (write $A \subseteq B$) iff

$A(x) \leq B(x) \; \forall x \in X.$

Then $A \cap B = A$ and $A \cup B = B$

Boolean Algebra:-

A boolean algebra on a set B is defined as the quadrabic $\mathcal{B} = (B, +, \cdot, -)$ where the set B has atleast two elements (bounds) 0 and 1, + and . are binary operation on B and − is uniary operation on B

Note: 2

The law of contradiction is violated for fuzzy set

ie) $\min[A(x), 1 - A(x)] \neq 0$ for atleast one $x \in X$ [infant $A(x) \in [0,1]$) $\min[A(x); 1 - A(x)] = 0$

only when $A(x) \in \{0, 1\}$

The law of contradiction is violated for fuzzy sets, the law of contradiction is satisfies for crisp set.

Types of fuzzy set :-

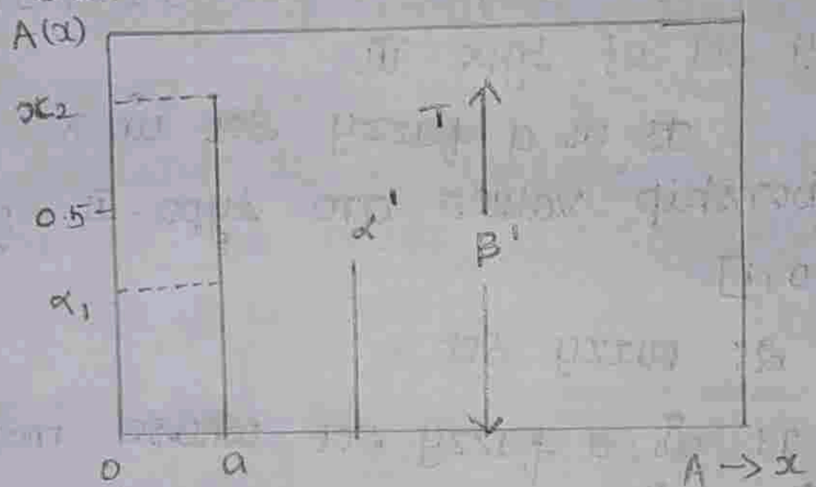
29 **Ordinary fuzzy set :-**

Given an universal set x an ordinary fuzzy set A is defined by a function of the form $A : x \rightarrow [0,1]$

In this each element of the universal set is assign a particular real numbers

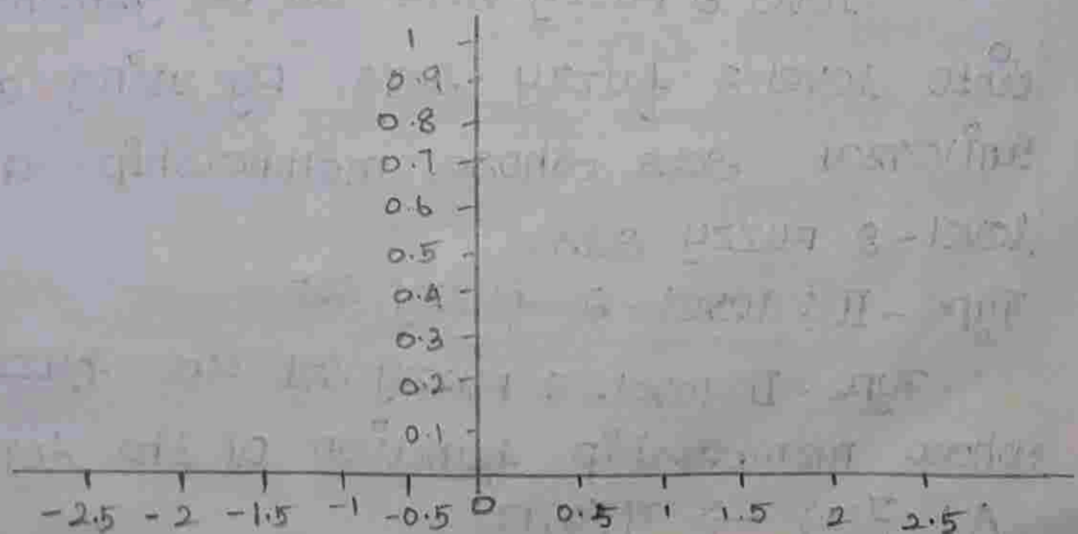
Interval valued fuzzy set :-

Eg :- Example of an interval rate value fuzzy set



We can define a possible membership function for the fuzzy set of real number close to zero as follows

$$\mu_A(x) = \frac{1}{1 + 10x^2}$$



⇒ The number 1 a large grade of 0.9

⇒ The number 0.25 a grade of 0.6

⇒ The number 0 a grade of 1

Fuzzy sets of type - II :-

The membership function of a fuzzy set of type - II have the form $A : x \to \mathcal{F}[(0,1)]$ where $\mathcal{F}[(0,1)]$ denotes the set of all ordinary fuzzy sets can be defined with in the universal sets $[0,1]$ $\mathcal{F}[0,1]$ is also called a fuzzy power set of $[0,1]$

Fuzzy set of type - III :-

It is a fuzzy set in X whose membership values are type - II fuzzy set on $[0,1]$

level 2 : Fuzzy set :-

It is a fuzzy set whose membership function is of the form $A : \mathcal{F}(x) \to [0,1]$ where $\mathcal{F}[x]$ denotes fuzzy power set of X.

ie) $\mathcal{F}(X)$ is the set of all ordinary fuzzy set of X.

Level 2 Fuzzy sets can be generalised into level 3 fuzzy sets. By using a universal sets whose membership are level - 2 fuzzy sets.

Type - II : **level - 2 fuzzy set**

Type - II level - 2 Fuzzy set are fuzzy sets whose membership function of the form $A : \mathcal{F}(x) \to \mathcal{F}[0,1]$.

Unit - II

Operations on fuzzy set

Standard fuzzy union:-

$$\mu_{\bar{A}}(x) = 1 - \mu_A(x)$$

$$\mu_{A \cup B}(x) = max[\mu_A(x), \mu_B(x)]$$

$$\mu_{A \cap B}(x) = min[\mu_A(x), \mu_B(x)]$$

5m **Fuzzy complement:**

A complement of fuzzy set A is specified by a function $c: [0,1] \rightarrow [0,1]$ which assigns a value $c(\mu_A(x))$ to each membership grade $\mu_A(x)$ in order for any function to be consider a fuzzy complement it must satisfy atleast the following 2 axioms.

axiom 1:- (c_1)

$c(0) = 1$ and $c(1) = 0$ (boundary condition)

axiom 2 :-

$\forall a, b \in [0,1]$ if $a < b$, then $c(a) \geq c(b)$

ie) c is monotonic non-increasing.

axiom (c_3):-

c is a continuous function.

axiom (c_4) :-

c is involute which means that

$c(c(a)) = a \;\; \forall a \in [0,1]$

axiom c_1 and c_2 are axiomatic skeleton for fuzzy set

Eg: 1

Equations of general fuzzy complement that satisfy only the axiomatic skeleton or the threshold.

Type: complement defined by

$$C(a) = \begin{cases} 1 & \text{for } a \le t \\ 0 & \text{for } a > t \end{cases} \quad \text{where } a \in [0,1] \text{ and } t \in [0,1)$$

t is called the threshold of C.

Eg: 2

In example of a fuzzy complement that is continuous but not involutive is a function $C(a) = \frac{1}{2}(1 + \cos \pi a)$

put $a = 0.33$, $C(0.33) = \frac{1}{2}[1 + \cos\pi(0.33)]$

$C[C(0.33)] = C(0.75) = 0.75$

$C(a)$ is not an involute

Sugeno class:

2m 5m One class of involute fuzzy complement is the sugeno class defined by $C_\lambda(a) = \frac{1-a}{1+\lambda a}$ where $\lambda \in (-1, \infty)$

Yager class:

2m 5m The class of involute fuzzy complement is defined by $C_w(a) = (1-a^w)^{1/w}$ where $w \in (0, \infty)$. This is called yager class of fuzzy complement.

Theorem 1:

The function $c : [0,1] \to [0,1]$ satisfies the axiom c_2 and c_4. Then c also satisfies axiom c_1 and c_3 moreover c must be a bijective function.

proof:-

Step 1:

Since the range of c is $[0,1]$

$c(0) \le 1$ and $c(1) \ge 0$

By axiom c_2, $c(0) \ge c(1)$

By axiom c_4, $0 = c(c(0)) \ge c(1)$

Hence $c(1) = 1$ (by axiom c_1), $0 \ge c(1)$

Now, again choose axiom c_4, we've

$c(c(1)) \le c(0)$

$1 \le c(0)$

$c(0) = c(c(1)) = 1$

$\therefore c(0) = 1$ by axiom c_1

Step 2:

To prove that c is a bijective function

We observe that $\forall\, a \in [0,1]$, $\exists$ $b = c(a) \in [0,1]$ such that $c(b) = c(c(a)) = a$

$\Rightarrow c(b) = a$

hence c is a onto function.

assume that $c(a_1) = c(a_2)$ by axiom c_4

$a_1 = c(c(a_1)) = c(c(a_2)) = a_2 \Rightarrow a_1 = a_2$

when $c(a_1) = c(a_2)$.

∴ The function c is also 1-1 function

∴ It is a bijective function

Step 3 : continuous function

Since c is a bijective function and satisfies axiom c_2 it can't have any discontinuous points.

Assume that c has a discontinuous at a_0

Then we've $b_0 = \lim_{a \to \bar{a}_0} c(a) > c(a_0)$ and

clearly there must exist $b_1 \in [0,1]$ such that $b_0 > b_1 > c(a_0)$ for which $a \in [0,1]$ exists such that $c(a_1) \geq b_1$ which is a ⇒⇐ to c is a bijective fuction

∴ c is a continuous.

Equilibrium

The equilibrium of a complement c is the degree of membership in fuzzy set A equal in the degree of membership in the component $c(a)$ ie) The equilibrium of the fuzzy complement c which is defined as any value for which $c(a) = a$.

Thm: 2.1

Every fuzzy complement has atmost one equilibrium.

proof:- Let C be a arbitrary fuzzy complement.

An equilibrium of C is a solution of the equation $C(a) - a = 0$ where $a \in [0,1]$

We can demonstrate that any equation

$C(a) - a = v \to (1)$ where v is a real constant must have atmost one solution.

$C(a) - a = v$ such that $a_1 < a_2$

$C(a_2) - a_2 = v$

$C(a_1) - a_1 = v \Rightarrow c(a_1) - a_1 = c(a_2) - a_2$

however C is monotonic non-increasing (by axiom C_2)

$C(a_1) \geq C(a_2)$ and since $a_1 \leq a_2$

$C(a_1) - a_1 > C(a_2) - a_2$

$\Rightarrow\Leftarrow$

Take $v = 0$ in (1), $c(a) - a = 0 \Rightarrow c(a) = a$

~~Then $C(a) \geq a$, $C(a)$.~~

Thus every fuzzy complement has atmost one equilibrium

Thm: 2.3

If c is continuous fuzzy complement then c has a unique equilibrium.

proof:-

The equilibrium e_c of c is the solution of the eqn $c(a) - a = 0$

This is as special case of the eqn $(c(a)) - a = b$ where $b \in [-1, 1]$ is a constant

By axiom C_1

$$c(0) - 0 = 1$$

$$c(1) - 1 = -1$$

write statement

since c is continuous by intermediate value theorem for continuous function

[for each $b \in [-1,1]$ $\exists$ a such that

$c(a) - a > b$]

This demonstrate the necessary existance of an ~~sqpread~~ ~~vatied~~ equilibrium value for a continous function for $b = 0$ $\exists$ a such that

$c(a) - a = b$

ie) $c(a) = a$

c has an equilibrium

Th

Thm 25 For each $a \in [0,1]$, ${}^{d}a = c(a) \Leftrightarrow c(c(a)) = a$ when the complement is involutive.

proof:-

Let ${}^{d}a = c(a)$ sub this in ② ↗ $c({}^{d}a) - {}^{d}a = a - c(a)$

gives $c(c(a)) - c(a) = a - c(a)$

$$c(c(a)) = a$$

conversely,

Let $c(c(a)) = a$.

Then the substitution of $c(c(a))$ for 'a' in dual point eqn

eqn ② becomes $c({}^{d}a) - {}^{d}a = c(c(a)) - c(a)$

because d_a can be substitute for $c(a)$ every where in the eqn to get

$${}^{d}a = c(a).$$

Thm 24 If a complement c has an equilibrium e_c then ${}^{d}e_c = e_c$.

condition for dual point $c({}^{d}a) - {}^{d}a = a - c(a)$

proof:-

If $a = e_c$ then by defn of equilibrium $c(a) = a$, $a - c(a) = 0 \rightarrow$ ①

if ${}^{d}a = e_c$ then $c({}^{d}a) = {}^{d}a$ and $c({}^{d}a) - {}^{d}a = 0 \rightarrow$ ②

From ① & ②

$$c({}^{d}a) - {}^{d}a = a - c(a)$$

This satisfies ② when $a = {}^{d}a = e_c$.

Hence the equilibrium of any component is its own dual point.

Fuzzy union

The union of two fuzzy sets A and B of the form

$$u = [0,1] \times [0,1] \rightarrow [0,1]$$

$$\mu_{A \cup B}(x) = u[\mu_A(x), \mu_B(x)]$$

Axiomatic relation for fuzzy union:

Axiom u_1 : $u(0,0) = 0$; $u(0,1) = u(1,0) = u(1,1) = 1$ (boundary condition)

u_2 : $u(a,b) = u(b,a)$ ie) u is commutative.

u_3 : If $a \le a'$ and $b \le b'$, then $u(a,b) \le u(a',b')$
u is montonic.

u_4 : $u(u(a,b)c) = u(a, u(b,c))$ u is associative.

u_5 : u is continuous function.

u_6 : $u(a,a) = a$ ie) u is idempotent.

Yager class

Yager class is defined by the function

$$u_w(a,b) = \min[1, (a^w + b^w)^{1/w}], \quad w \in (0,\infty)$$

$u_1 \rightarrow u_5$ satisfy.

If $w = 1$, the function becomes

$$u_1(a,b) = \min(1, a+b)$$

$w = 2$,

$$u_2(a,b) = \min[1, \sqrt{a^2+b^2}]$$

Thm: 2.6

$$\lim_{\omega \to \infty} \min [1, (a^{\omega} + b^{\omega})^{1/\omega}] = \max(a, b)$$

proof:-

10m 5m

case (i) If a or b = 0

a = b

The theorem is obvious

lim of $2^{1/\omega}$ as $\omega \to \infty = 1$

case (ii) If $a \neq b$ and the min equals $(a^{\omega} + b^{\omega})^{1/\omega}$

$$\lim_{\omega \to \infty} (a^{\omega} + b^{\omega})^{1/\omega} = \max(a, b)$$

Let us assume that $a < b$, and let $Q = (a^{\omega} + b^{\omega})^{1/\omega}$

$\log(a^{\omega} + b^{\omega})^{1/\omega} = \frac{1}{\omega} \log(a^{\omega} + b^{\omega})$

Then, $\lim_{\omega \to \infty} \log Q = \lim_{\omega \to \infty} \frac{\log(a^{\omega} + b^{\omega})}{\omega}$

Using L-hospitals rule,

$$\underset{\omega\to\infty}{lt} \log Q = \underset{\omega\to\infty}{lt} \frac{1}{\omega}\left(\frac{\omega a^{\omega-1} + \omega b^{\omega-1}}{a^{\omega}+b^{\omega}}\right)$$

$$= \underset{\omega\to\infty}{lt} \frac{a^{\omega} a^{-1} + b^{\omega} b^{-1}}{a^{\omega}+b^{\omega}}$$

$$= \underset{\omega\to\infty}{lt} \frac{a^{\omega}\log a + b^{\omega}\log b}{a^{\omega}+b^{\omega}}$$

$$= \underset{\omega\to\infty}{lt} \frac{\frac{a^{\omega}}{b^{\omega}}\log a + \log b}{\frac{a^{\omega}}{b^{\omega}}+1}$$

$$= \underset{\omega\to\infty}{lt} \frac{(a/b)^{\omega}\log a + \log b}{(a/b)^{\omega}+1}$$

$$= \underset{\omega\to\infty}{lt} \frac{(b/a)^{-\omega}\log a + \log b}{(b/a)^{-\omega}+1}$$

$$= \underset{\omega\to\infty}{lt} \frac{(b/a)^{-\infty}\log a + \log b}{(b/a)^{-\infty}+1}$$

$$= \frac{0\times\log a + \log b}{0+1}$$

$$\underset{\omega\to\infty}{lt} \log Q = \log b$$

Taking exponential on both sides,

$$\underset{\omega\to\infty}{lt} Q = b$$

$$\lim_{\omega\to\infty} Q = \lim_{\omega\to\infty} (a^\omega + b^\omega)^{1/\omega} = b = \max(a,b)$$

It remains to show that the theorem is true when the minimum equals 1.

In this case $(a^\omega + b^\omega)^{1/\omega} \geq 1$

$$(a^\omega + b^\omega) \geq 1$$

$\forall \omega \in (0, \infty)$, when $\omega \to \infty$,

if $a = 1$ or $b = 1$

$$a, b \in [0, 1]$$

Hence the proof.

Thm: 2.7

$$\lim_{\omega\to\infty} i_\omega = \lim_{\omega\to\infty} (1 - \min[1, (1-a)^\omega + (1-b)^\omega]^{1/\omega})$$

$$= \min(a,b).$$

Soln:

First to prove previous thm,

$$\lim_{\omega\to\infty} \min[1, (a^\omega + b^\omega)^{1/\omega}] = \max(a,b)$$

From above we know that

$$\lim_{\omega\to\infty} \min[1, ((1-a)^\omega + (1-b)^\omega)^{1/\omega}] = \max(1-a, 1-b)$$

ie) $i_\omega(a,b) = 1 - \max[1-a, 1-b]$

Let us assume that without loss of generality that $a \leq b$

Then $1-a \geq 1-b$

$$i_\omega(a,b) = 1 - (1-a) = a$$

Hence $i_\omega(a,b) = \min(a,b)$

Hence proved.

Combination of operations

Fuzzy set unions:- that satisfies the axiomatic skeleton (axiom U_1 to U_4) are bounded by the inequalities

$$\max(a,b) \le u(a,b) \le u_{max}(a,b)$$

where $u_{max}(a,b) = \begin{cases} a & \text{when } b=0 \\ b & \text{when } a=0 \\ 1 & \text{otherwise} \end{cases}$

|||ly Fuzzy set intersection that satisfy the axiomatic skeleton (axiom I_1 to I_4) are bounded by an inequalities

$$I_{min}(a,b) \le I(a,b) \le \min(a,b)$$

where $I_{min}(a,b) = \begin{cases} a & \text{when } b=1 \\ b & \text{when } a=1 \\ 0 & \text{otherwise} \end{cases}$

Thm 2.8

$\forall\, a,b \in [0,1]$ $u(a,b) \ge \max(a,b)$

proof:- By associative axiom the equation

$$u(a, u(0,0)) = u(u(a,0),0) \text{ is valid} \quad \to (1)$$

By apply the boundary condition,

$$u(0,0) = 0$$

From (1), $u(a,0) = u(u(a,0),0)$

Assume that $u(a,0) = \alpha \ne a$

Substitute α for $u(a,0)$ in the eqn.

$$u(a,0) = u(u(a,0),0)$$

$$\alpha = u(\alpha, 0)$$

$u(\alpha, 0) = \alpha$

Hence the only solution of the equation is

$u(a, 0) = a$

By monotonicity axiom (u_3)

$u(a, b) \geq u(a, 0) = a$

$u(a, b) = u(b, a) \geq u(b, 0) = b$

Hence $u(a, b) \geq \max(a, b)$

Thm:- 2.9

$\forall\ a, b \in [0, 1]$ $u(a, b) \leq u_{max}(a, b)$

proof:-

By precious thm, when $b = 0$

Then $u(a, b) = a$

similarly by commutativity

when $a = 0$

$u(a, b) = b$

since $u(a, b) \in [0, 1]$

By follows from previous thm

$u(a, 1) = u(1, b) = 1$

Now by monotonicity, we have

$u(a, b) \leq u(a, 1) = u(1, b) = 1$

$\therefore u(a, b) \leq u_{max}(a, b)$.

Thm: 2.10

For all $a, b \in [0,1]$, $i(a,b) \le \min(a,b)$.

proof:-

By associativity axiom

$i(a, i(1,1)) = i(i(a,1), 1) \longrightarrow ①$

Then by using the boundary condition

$i(1,1) = 1$

From ①, we have $i(a,1) = i(i(a,1), 1)$

The only solution of the equation is

$i(a,1) = a$.

Then by monotonicity, we have

$i(a,b) \le i(a,1) = a$.

And by commutativity

$i(a,b) = i(b,a) \le i(b,1) = b$.

$\therefore i(a,b) \le \min(a,b)$.

Thm: 2.11

For all $a, b \in [0,1]$ $i(a,b) \ge i_{min}(a,b)$

proof:-

By using the defn $i_{min}(a,b) = \begin{cases} a & \text{when } b=1 \\ b & \text{when } a=1 \\ 0 & \text{otherwise} \end{cases}$

when $b=1$, $i(a,b) = i(a,1) = a$

and the theorem is satisfied.

III^{ly} by commutativity, that when $a=1$

Then $i_{(a,b)} = b$.

and the theorem is satisfied.

Since $i(a,b) \in [0,1]$ using previous thm

$i(a,0) = i(0,b) = 0$

By monotonicity $i(a,b) \geq i(0,b) = i(a,0) \geq 0$

$i(a,b) \geq 0$

$i(a,b) \geq i_{min}(a,b)$

Hence proved.

Thm: 2.12

$u(a,b) = \max(a,b)$ is the only continuous ~~constant~~ and idempotent fuzzy set union (ie, the only function that satisfies axioms u_1 through u_6).

proof:-

By associativity

$u(a, u(a,b)) = u(u(a,a), b)$

The idempotency property $u(a,a) = a$.

The above eqn becomes

$u(a, u(a,b)) = u(a,b)$.

|||ly

$u(u(a,b), b) = u(a, u(b,b))$

$u(u(a,b), b) = u(a,b)$

Hence $u(a, u(a,b)) = u(u(a,b), b)$

By commutativity

$u(a, u(a,b)) = u(b, u(a,b)) \longrightarrow (1)$

i) when $a=b$ the idempotent is applicable

ie) $u(b, u(b,b)) = u(b,b) = b$

$u(b, u(b,b)) = u(b,b) = b$

and eqn (1) is satisfied.

ii) suppose $a<b$ and assume that $u(a,b) = \alpha$ when $\alpha \neq a$ and $\alpha \neq b$

Hence eqn (1) becomes

$u(a, \alpha) = u(b, \alpha)$

Since u is continuous and monotonic non-decreasing with $u(0, \alpha) = \alpha$ and $u(1, \alpha) = 1$ there exists a pair a, $b \in [0, 1]$ such that $u(a, \alpha) < u(b, \alpha)$

Assume that $u(a, b) = a = \min(a, b)$

This assumption is also unacceptable since it violates the boundary condition when $a = 0$ and $b = 1$.

The final possibility is to consider

$$u(a, b) = b = \max(a, b)$$

In this case, the boundary condition are satisfied and eqn ① becomes

$$u(a, b) = u(b, b)$$

ie) It satisfies for all $a < b$

Because of commutativity the same argument is repeated for $a > b$.

Hence maximum is the only function that satisfies axioms u_1 through u_6.

Thm: 2.14

Fuzzy set operations of union, intersection and continuous complement that satisfy the law of excluded middle and the law of contradiction are not idempotent or distributive.

proof:-

Since the standard operations do not satisfy the two laws of excluded middle and of contradiction.

By thm 2.12 and 2.13 they are the only operations that are idempotent, operations that do satisfy these laws cannot be idempotent.

Next, we must prove that these operations do not satisfy the distributive laws

$$u(a, i(b,d)) = i(u(a,b), u(a,d)) \quad \text{—} ①$$

and

$$i(a, u(b,d)) = u(i(a,b), i(a,d)) \quad \text{—} ②$$

Let e denote the equlibrium of the complement c involved, that is $c(e) = e$.

Then, from the law of excluded middle, we obtain,

$$u(e, c(e)) = u(e, e) = 1$$

similarly, from the law of condraction,

$$i(e, c(e)) = i(e, e) = 0$$

Then, by applying e to the left hand side of ①, we obtain

$$u(e, i(e,e)) = u(e, 0)$$

We observe that e is neither 0 nor 1 because of the requirement that $c(0)=1$ and $c(1)=0$ (axiom c_1).

By thm 2.8 and thm 2.9, whe have $u(e,0)=e$ and consequently

$$u(e, i(e,e)) = e \quad (\neq 1)$$

Now we apply e to the right hand side of ① to obtain

$$i(u(e,e), u(e,e)) = i(1,1) = 1.$$

This demonstrates that the distributive law ① is violated.

Let us now apply e to the second distributive law ②.

By thm 2.10 and 2.11, we obtain

$$i(e, u(e,e)) = i(e,1) = e \quad (\neq 0)$$

and

$$u(i(e,e), i(e,e)) = u(0,0) = 0$$

which demonstrates that on ② is not satisfied.

This completes the proof.

Thm: 2.2

Assume that a given fuzzy complement c has an equilibrium e_c, which by theorem 2.1 is unique. Then

$$a \leq c(a) \text{ if and only if } a \leq e_c$$

and

$$a \geq c(a) \text{ if and only if } a \geq e_c.$$

proof:-

Assume that $a < e_c$, ~~and~~ $a = e_c$ and $a > e_c$.

Since c is monotonic non increasing by axioms,

$c(a) \geq c(e_c)$, $c(a) = c(e_c)$ and $c(a) \leq c(e_c)$

Because $c(e_c) = e_c$, we can rewrite

$c(a) \geq e_c$, $c(a) = e_c$ and $c(a) \leq e_c$.

Due to our initial assumption $a = e_c$,

$c(a) > a$, $c(a) = a$, $c(a) < a$.

Thus, $a \leq e_c$ implies $c(a) \geq a$ and $a \geq e_c$ implies $c(a) \leq a$.

The inverse implication can be shown in a similar manner.

Thm: 2.13

$i(a,b) = \min(a,b)$ is the only continuous and idempotent fuzzy set intersection (ie, the only function that satisfies axioms i1 through i6).

proof:-

This theorem can be proven in exactly the same way as theorem 2.12 by replacing function u with i and by applying axioms i1 through i6 instead of axioms u1 through u6.

By associativity,

$$i(a, i(a,b)) = i(i(a,a), b)$$

By idempotency property $i(a,a) = a$

The above eqn becomes

$$i(a, i(a,b)) = i(a,b)$$

Similarly,

~~$i(a, i(a,b)) = i(a, i(b,b))$~~

~~$i(a, i(a,b)) =$~~

$$i(i(a,b), b)) = i(a, i(b,b))$$

$$i(i(a,b), b) = i(a,b)$$

Hence $i(a, i(a,b)) = i(i(a,b), b)$

By commutativity,

$$i(a, i(a,b)) = i(b, i(a,b)) \rightarrow (1)$$

i) when $a=b$, idempotency is applicable

ie) $i(b, i(b,b)) = i(b,b) = b$

$i(b, i(b,b)) = i(b,b) = b$

and eqn (1) is satisfied

ii) suppose $a<b$ and assume that

$i(a,b) = \alpha$ when $\alpha \neq a$ and $\alpha \neq b$

Hence eqn (1) becomes

$$i(a,\alpha) = i(b,\alpha)$$

Since i is continuous and monotonic non-decreasing with $i(0,\alpha) = 0$ and $i(1,\alpha) = \alpha$ there exist a pair $a,b \in [0,1]$ such that

$i(a,\alpha) < i(b,\alpha)$

Assume that $i(a,b) = b = \max(a,b)$

The assumption is also unacceptable since it violates the boundary condition when $a=0$ and $b=1$

The final possibility is to consider

$$i(a,b) = a = \min(a,b)$$

In this case, the boundary condition are satisfied and eqn ① becomes

$$i(a,b) = i(a,a)$$

ie) It satisfies for all $a<b$

Because of commutativity the same argument is repeated for $a>b$.

Hence minimum is the only function that satisfies axioms i_1 through i_6

General aggregation operation

Defn:

Any aggregation operation is defined by a function $h: [0,1]^n \rightarrow [0,1]$ for some $n \geq 2$. When applied 'n' fuzzy sets $A_1, A_2, \ldots, A_n$ defined on x, h produces an aggregate fuzzy set A by operating on the membership grade of each $x \in X$ in the aggregated sets.

Thus $\mu_A(x) = h(\mu_{A_1}(x), \mu_{A_2}(x), \ldots, \mu_{A_n}(x)]$

Axiom h_1:-

$h(0,0,\ldots,0) = 0$ and

$h(1,1,\ldots,1) = 1$ (boundary condition)

Axiom h_2:-

for any pair $(a_i \mid i \in N_n)$ and $(b_i \mid i \in N_n)$ where $a_i \in [0,1]$ and $b_i \in [0,1]$ if $a_i \geq b_i \; \forall \; i \in N_n$ then $h(a_i \mid i \in N_n) \geq h(b_i \mid i \in N_n)$ that is h is monotonic non-decreasing in all arguments.

Two additional axiom are usually employed to characterize aggregation operations ~~despite~~

Axiom h_3:-

h is a continuous function.

Axiom h_4:-

h is a symmetric function in all its arguments.

ie., $h(a_i \mid i \in N_n) = h(a_{p(i)} \mid i \in N_n)$ for any permutation p on N_n.

Thm: 2.15

Let h_α be given by $h_\alpha(a_1, a_2, \ldots, a_n) = \left(\frac{a_1^\alpha + a_2^\alpha + \ldots + a_n^\alpha}{n}\right)^{1/\alpha}$ Then $\lim_{\alpha \to 0} h_\alpha = (a_1 \cdot a_2 \cdot \ldots \cdot a_n)^{1/n}$

proof:-

Gin $h_\alpha = \left(\frac{a_1^\alpha + a_2^\alpha + \ldots + a_n^\alpha}{n}\right)^{1/\alpha}$

$$\ln h_\alpha = \frac{\ln(a_1^\alpha + a_2^\alpha + \ldots + a_n^\alpha) - \ln n}{\alpha}$$

$$\lim_{\alpha \to 0} \ln h_\alpha = \lim_{\alpha \to 0} \frac{\ln(a_1^\alpha + a_2^\alpha + \ldots + a_n^\alpha) - \ln n}{\alpha}$$

Using L Hospital's rule we have

$$\lim_{\alpha \to 0} \ln h_\alpha = \lim_{\alpha \to 0} \frac{a_1^\alpha \ln a_1 + a_2^\alpha \ln a_2 + \dots + a_n^\alpha \ln a_n}{a_1^\alpha + a_2^\alpha + \dots + a_n^\alpha}$$

$$= \frac{\ln a_1 + \ln a_2 + \dots + \ln a_n}{n}$$

$$= \ln (a_1, a_2, \dots, a_n)^{1/n}$$

Hence $\lim_{\alpha \to 0} h_\alpha = (a_1, a_2, \dots, a_n)^{1/n}$

Defn: Dual point

2m

If we are given a fuzzy complement c and a membership grade whose value is represented by a real number $a \in [0,1]$. Then any membership grade represented by the real number $^d a \in [0,1]$,

$c(^d a) - {^d a} = a - c(a)$ is called the dual point of a with respect to c.

Fuzzy intersection:

The general fuzzy intersection of two fuzzy sets A and B is specified by a function $i : [0,1] \times [0,1] \to [0,1]$

The membership grade of the element in set $A \cap B$

Thus $\mu_{A \cap B}(x) = i[\mu_A(x), \mu_B(x)]$

Axiom i_1 (Boundary condition)

$i(1,1) = 1$; $i(0,1) = i(1,0) = i(0,0) = 0$

ie., i behaves as the classical intersection with crisp sets.

Axiom i_2 : commutativity

$i(a,b) = i(b,a)$

i is commutative

Axiom i_3 Monotonicty

If $a \le a'$ and $b \le b'$ then $i(a,b) \le i(a',b')$

ie., i is monotonic.

Axiom i_4 : Associativity

$i(a, i(b,c)) = i(i(a,b),c)$

ie., i is associative.

This set of axioms are called axiomatic skeletons for fuzzy intersection (or) t-norms

The most important additional rearragement for fuzzy set intersection.

Axiom i_5

i is a continuous function

Axiom i_6

$i(a,a) = a$, i is idempotent.

Unit - III

Fuzzy relation

The carstesian product X and Y denote by X×Y is the crisp set of all order pair such that the first element in each pair is a member of X and the second element the member of Y.

Formally $X \times Y = \{(x, y) / x \in X \text{ and } y \in Y\}$

Note:

If $X \neq Y$ then $X \times Y \neq Y \times X$

Defn of crisp relation

2m (A relation among crisp sets $x_1, x_2, \dots, x_n$ is a subset of the cartesian product $\times_{i \in N_n} X_i$

It is denoted either by $R(X_1, X_2, \dots, X_n)$

or by the apprerriated form $R(X_i ; i \in N_n)$.

Thus $R(X_1, X_2, \ldots, X_n) \subset X_1 \times X_2 \times \ldots \times X_n$

$$\mu_R(x_1, x_2, \ldots, x_n) = \begin{cases} 1 & \text{if } x_1, x_2, \ldots, x_n \\ 0 & \text{otherwise} \end{cases}$$

Ex:

Let R be a fuzzy relation between the two sets $X = \{$paris, New york$\}$ and $Y = \{$paris, America, London$\}$.

That represents the relational concept "very far". This relation can be written in list notation as

$$R(X,Y) = \frac{1}{\text{paris, America}} + \frac{0}{\text{paris, paris}} + \frac{0.6}{\text{paris, Newyork}} + \frac{0.9}{\text{Paris, London}} + \frac{0.7}{\text{NewYork, London}} + \frac{0.3}{\text{NewYork, America}}$$

This relation can also be represented by following two dimensional membership array

	paris	Newyork
paris	0	0.6
America	1	0.3
London	0.9	0.7

Note:

A relation between two sets is called binary if three, four, five sets are involved, the relations are called ternary quarternary and quinary respectively.

In general a relation defined an n-sets is called n-ary or n-dimension.

Ex:

consider the sets $X_1 = \{x, y\}$, $X_2 = \{a, b\}$ and $X_3 = \{*, \$\}$ and the ternary fuzzy relation is

$$R(X_1, X_2, X_3) = \frac{.9}{x,a,*} + \frac{0.4}{x,b,*} + \frac{1}{y,a,*} + \frac{.7}{y,a,\$} + \frac{.8}{y,b,\$}$$

defined on $X_1 \times X_2 \times X_3$

Projection:

Given a relation $R(X_1, X_2, \ldots, X_n)$ Let $[R \downarrow Y]$ denote the projection of R that disregards all variables in X except these in the set $Y = x_j / j \in J \subset N_n$

$$\mu [R \downarrow Y](y) = \max_{x > y} \mu_R(x)$$

Eg:

Let $R_{i,j} = [R \downarrow \{x_i, x_j\}]$ and $R_i = [R \downarrow \{x_i\}]$ for all $(i, j) \in N_3$

Then $R_{1,2} = \frac{.9}{x,a} + \frac{.4}{x,b} + \frac{1}{y,a} + \frac{.8}{y,b}$

$$R_{1,3} = \frac{.9}{x,*} + \frac{1}{y,*} + \frac{.8}{y,\$}$$

$$R_{2,3} = \frac{1}{a,*} + \frac{.4}{b,*} + \frac{.7}{a,\$} + \frac{.8}{b,\$}$$

$$R_1 = \frac{.9}{x} + \frac{1}{y}$$

$$R_2 = \frac{1}{a} + \frac{.8}{b}$$

$$R_3 = \frac{1}{*} + \frac{.8}{\$}$$

Defn:

Another operation on relations, which is in some sense an inverse to the projection is called a cylindric extension.

Let X and Y denote the same families of sets as employed in the defn of projection. Let R be a relation defined on the cartesian product of sets in the family Y and let $[R \uparrow X - Y]$ denote the cylindric extension R into sets X_i $(i \in N_n)$ that are in X but are not in Y.

Then

$$\mu_{[R \uparrow X-Y]}(x) = \mu_R(y)$$

For each x such that $x > y$.

Eg:

cylindric extension of projection calculated in (c)

X_1, X_2, X_3	Membership function of cylindric extension of					
	$R_{1,2}$	$R_{1,3}$	$R_{2,3}$	R_1	R_2	R_3
x, a, *	.9	.9	1	.9	1	1
x, a, $	.9	0	.7	.9	1	.8
x, b, *	.4	.9	.4	.9	.8	1
x, b, $	.4	0	.8	.9	.8	.8
y, a, *	1	1	1	1	1	1
y, a, $	1	.8	.7	1	1	.8
y, b, *	.8	1	.4	1	.8	1
y, b, $	.8	.8	.8	1	.8	.8

The cylindric closure of the first three projections in the previous table is the following fuzzy ternary relation

$$cyl\{R_{1,2}, R_{1,3}, R_{2,3}\} = .9/x,a,* + .4/x,b,* + 1/y,a,* + .7/y,a,\$ + .4/y,b,* + .8/y,b,\$$$

Hence the three projection do not capture the origenal relation fully but approximate is quite well.

Binary relation

[Any relation between two sets X and Y is known as a binary relation. It is usally denoted by R(X, Y)]

When $X \neq Y$, binary relations R(X, Y) are often referred to as bipartite graph

When $X = Y$ they are called directed graph (or) digraph.

In addition to membership matrices another useful representation of binary relation R(X, Y) sagittal diagram.

Defn: Domain

The domain of a crisp binary relation is written as dom R(X, Y) and is defined as the crisp subset of X whose membership participate in the relation.

Formally,

$$\text{dom } R(X, Y) = \{x \mid x \in X, (x, y) \in R \text{ for some } y \in Y\}.$$

~~Defn~~ ~~R~~

Defn: Range

The range of a crisp binary relation $R(X,Y)$ is denoted by $\text{ran } R(X,Y)$ and is defined as the subset of Y whose members participate in the relation. Thus

$$\text{ran } R(X,Y) = \{ y / y \in Y, (x,y) \in R \text{ for some } x \in X \}$$

Note:

i) $\mu_{\text{dom} R}(x) = \max_{y \in Y} \mu_R(x,y)$

ii) $\mu_{\text{ran} R}(y) = \max_{x \in X} \mu_R(x,y)$

Defn: Height

The height of a fuzzy relation R is a number $h(R)$ defined by

$$h(R) = \max_{y \in Y} \max_{x \in X} \mu_R(x,y)$$

If $h(R) = 1$ then the relation is called normal otherwise it is called subnormal.

Defn: completely specified.

If the domain of R is equal to the support of X then the relation is called completely specified.

Otherwise it is called incompletely specified.

Defn:-

If the range of R is equal to the support of set Y, then R is called a relation from X onto Y.

Otherwise it is called a relation from X into Y

Defn:
If each member of the domain of a binary relation R appears exactly once in R the relation is called a mapping or a function.

Defn:
When atleast one member of the domain is related to more that one element of the range the relation is not a mapping and is instead called one to many

Defn:
Each element of the range appears exactly once in the mapping it is called a one-to-one relation

Defn:-
Every fuzzy relation R(X,Y) can be represented by its resolution form

$$R = \bigcup_{\alpha} \alpha R_\alpha , \quad \alpha \in \Lambda_R \text{ (level set of R)}$$

where αR_α is a fuzzy relation defined by

$$\mu_{\alpha R_\alpha}(x, y) = \alpha \, \mu_{R_\alpha}(x, y) \text{ for every } (x, y) \in X \times Y.$$

Note:

$$\mu [R \downarrow Y](y) = \max_{x > y} \mu_R(x)$$

$$\mu_{cyl} [R_i](x) = \min_{i \in I} \mu_{[R \uparrow X - Y]}(x)$$

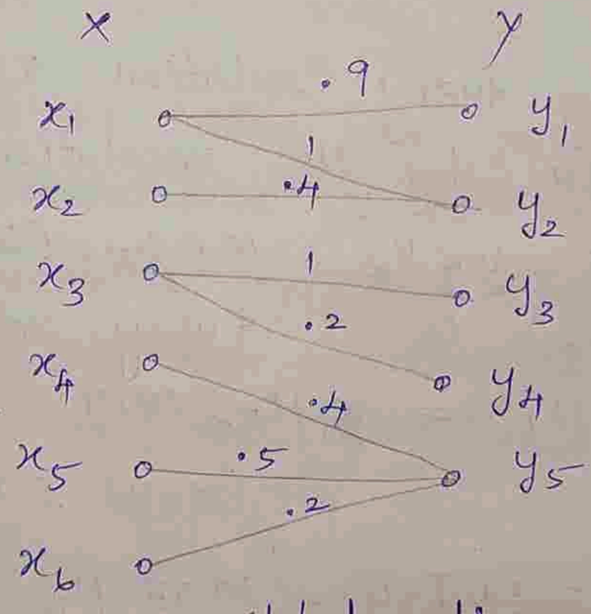
V.a Sagittal diagram:

An useful representation of binary relation R(x, y) (or) sagittal diagrams

Each of the sets (X, Y) is represented by a set of nodes in the diagram. Nodes corresponding to one set are clearly distinguish from nodes representing the other set. Elements of (X × Y) with non-zero membership grades in R(X, Y) are represented in the diagram by lines connecting the respective nodes. These lines are labelled with the value of the membership grade $\mu_{R(X,Y)}(x, y)$

Eg:

An example of the sagittal diagram together with the corresponding membership matrix is shown below.



sagittal diagram

	y_1	y_2	y_3	y_4	y_5
x_1	.9	1	0	0	0
x_2	0	.4	0	0	0
x_3	0	0	1	.2	0
x_4	0	0	0	0	.4
x_5	0	0	0	0	.5
x_6	0	0	0	0	.2

membership matrix

Resolution form

Using the max operator for set unions, every fuzzy relation $R(x, y)$ can be represented by its resolution form $R = \bigcup_{\alpha} \alpha R_\alpha$, $\alpha \in \Lambda_R$ (level set of R) where αR_α is a fuzzy relation defined by $\mu_{\alpha R_\alpha}(x, y) = \alpha \mu_{R_\alpha}(x, y)$ for every $(x, y) \in X \times Y$.

Ex: 5m

Let a binary fuzzy relation R be defined by the following membership matrix

$$M_R = \begin{bmatrix} .7 & .4 & 0 \\ .9 & .1 & .4 \\ 0 & .7 & 1 \\ .7 & .9 & 0 \end{bmatrix}$$

The resolution form $R = \bigcup_{\alpha} \alpha R_\alpha$, $\alpha \in \Lambda_R$ (level set of R)

∴ The level set $\Lambda_R = \{0, .4, .7, .9, 1\}$

i) when $\alpha = 0$.

The term αR_α represents the empty set

ii) when $\alpha = .4$

$$R_{.4} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

$\alpha_A = \{x / A(x) \geq \alpha\}$

$\alpha_{.4} = \{x / A(x) \geq .4\}$

$$.4 R_{.4} = \begin{pmatrix} .4 & .4 & 0 \\ .4 & .4 & .4 \\ 0 & .4 & .4 \\ .4 & .4 & 0 \end{pmatrix}$$

iii) when $\alpha = .7$

$$R_{.7} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

$$.7 R_{.7} = \begin{pmatrix} .7 & 0 & 0 \\ .7 & .7 & 0 \\ 0 & .7 & .7 \\ .7 & .7 & 0 \end{pmatrix}$$

iv) when $\alpha = .9$

$$R_{.9} = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

$$.9 R_{.9} = \begin{pmatrix} 0 & 0 & 0 \\ .9 & .9 & .9 \\ 0 & 0 & .9 \\ 0 & .9 & 0 \end{pmatrix}$$

v) when $\alpha = 1$

$$R_{1} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

$$R = \begin{pmatrix} .4 & .4 & 0 \\ .4 & .4 & .4 \\ 0 & .4 & .4 \\ .4 & .4 & 0 \end{pmatrix} \cup \begin{pmatrix} .7 & 0 & 0 \\ .7 & .7 & 0 \\ 0 & .7 & .7 \\ .7 & .7 & 0 \end{pmatrix} \cup$$

$$\begin{pmatrix} 0 & 0 & 0 \\ .9 & .9 & 0 \\ 0 & 0 & .9 \\ 0 & .9 & 0 \end{pmatrix} \cup \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

Inverse

The inverse of a crisp relation $R(X, Y)$ is written as $R^{-1}(X, Y)$ and is a subset of $Y \times X$ such that

$R^{-1}(X,Y) = \{(y,x) / (x,y) \in R\}$ where $x \in X$, $y \in Y$

For fuzzy relation $R(X,Y)$ the inverse fuzzy relation $R^{-1}(X,Y)$ is defined

$$\mu_{R^{-1}}(X,Y) = \mu_R(x,y) \ \forall (x,y) \in X \times Y$$

Ex:

Let $R(X,Y)$ be a fuzzy relation on $X = \{x, y, z\}$ and $Y = \{a, b\}$ such that

$$M_R = \begin{array}{c} \\ x \\ y \\ z \end{array} \begin{array}{c} \begin{array}{cc} a & b \end{array} \\ \begin{bmatrix} .3 & .2 \\ 0 & 1 \\ .6 & .4 \end{bmatrix} \end{array}$$

Then the inverse of $R(X,Y)$ is specified by the membership matrix

$$M_R^{-1} = M_R^T \begin{array}{c} \\ a \\ b \end{array} \begin{array}{c} \begin{array}{ccc} x & y & z \end{array} \\ \begin{pmatrix} .3 & 0 & .6 \\ .2 & 1 & .4 \end{pmatrix} \end{array}$$ where

M_R^T denotes the transpose of M_R.

Defn:

Consider two crisp binary relation $P(X,Y)$ and $Q(Y,Z)$ defined with a common set Y. The composition of these two relations is denoted by

$$R(X,Z) = P(X,Y) \circ Q(Y,Z)$$

and is defined as a subset $R(X,Z)$ of $X \times Z$ such that $(x,z) \in R$ iff $\exists$ atleast one $y \in Y$ such that $(x,y) \in P$ and $(y,z) \in Q$

$$P \circ Q \neq Q \circ P$$

$$(P \circ Q)^{-1} = Q^{-1} \circ P^{-1}$$

$(P \circ Q) \circ R = P \circ (Q \circ R)$

For fuzzy sets the max min composite denoted by $P(X,Y) \circ Q(Y,Z)$ and is defined by $\mu_{P \circ Q}(x,z) = \max_{y \in Y} \min[\mu_P(x,y), \mu_Q(y,z)]$ for all $x \in X$ and $z \in Z$.

Defn:

The max product composition is denoted by $P(X,Y) \odot Q(Y,Z)$ and defined by

$\mu_{P \odot Q}(x,z) = \max_{y \in Y} [\mu_P(x,y) \cdot \mu_Q(y,z)]$ for $x \in X$, $z \in Z$.

Defn:

The relational join (or) join denoted by $P(X,Y) * Q(Y,Z)$ and defined by

$\mu_{P*Q}(x,y,z) = \min(\mu_P(x,y), \mu_Q(y,z))$ for each $x \in X$, $y \in Y$, $z \in Z$

Ex:

Let $X = \{1,2,3,4\}$, $Y = \{a,b,c\}$, $Z = \{\alpha, \beta\}$ and the relations are

$1 \to a = .7$ (or) $R(1,a) = .7$

$1 \to b = .5$

$2 \to a = 1$

$3 \to b = 1$

$4 \to b = .4$

$4 \to c = .3$

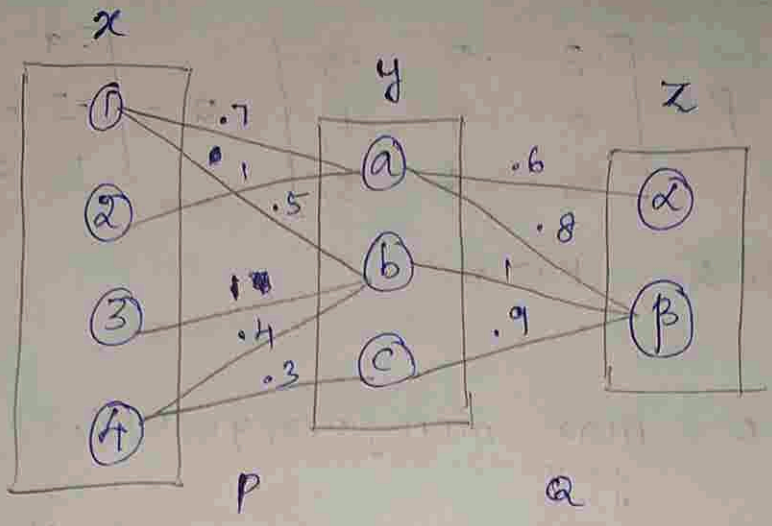
$a \to \alpha = .6$

$a \to \beta = .8$

$b \to \beta = 1$

$c \to \beta = .9$

Find join and composition



join $S = P * Q$ $\mu_S = \min(\mu_P(x,y), \mu_Q(y,z))$

x	y	z	$\mu_S(x,y,z)$
1	a	α	.6
1	a	β	.7
1	b	β	.5
2	a	α	.6
2	a	β	.3
3	b	β	1
4	b	β	.4
4	c	β	.3

composition

$$R = P \circ Q = \mu_R(x,z) = \max_{y \in Y} \min[\mu_P(x,y), \mu_Q(y,z)]$$

x	y	$\mu_R(x,z)$
1	α	.6
1	β	.7
2	α	.6
2	β	.8
3	β	1
4	β	.4

Ex:

Let $P = \begin{bmatrix} .3 & .5 & .8 \\ 0 & .7 & 1 \\ .4 & .6 & .5 \end{bmatrix}$, $Q = \begin{bmatrix} .9 & .5 & .7 & .7 \\ .3 & .2 & 0 & .9 \\ 1 & 0 & .5 & .5 \end{bmatrix}$

Find $P \circ Q$, $P \odot Q$.

Soln:

$$P \circ Q = \max_{y \in Y} \min [P(x, y), Q(y, z)]$$

$$\mu_{P \circ Q}(x, z) = \max_{y \in Y} \min [\mu_P(x, y), \mu_Q(y, z)]$$

Let $R = P \circ Q$

$r_{11} = \max [\min(.3, .9), \min(.5, .3), \min(.8, 1)]$

$= \max[.3, .3, .8]$

$r_{11} = .8$

$r_{12} = \max [.3, .2, 0] = .3$

$r_{13} = \max [.3, 0, .5] = .5$

$r_{14} = \max [.3, .5, .5] = .5$

$r_{21} = \max [0, .3, 1] = 1$

$r_{22} = \max [0, .2, 0] = .2$

$r_{23} = \max [0, 0, .5] = .5$

$r_{24} = \max [0, .7, .5] = .7$

$r_{31} = \max [.4, .3, .5] = .5$

$r_{32} = \max [.4, .2, 0] = .4$

$r_{33} = \max [.4, 0, .5] = .5$

$r_{34} = \max [.4, .6, .5] = .6$

$$R = \begin{bmatrix} .8 & .3 & .5 & .5 \\ 1 & .2 & .5 & .7 \\ .5 & .4 & .5 & .6 \end{bmatrix}$$

$$P \odot Q = \max_{y \in Y} [p(x,y) \cdot Q(x,y)]$$

$$\mu_{P\odot Q} = \max_{y \in Y} [\mu_P(x,y) \cdot \mu_Q(x,y)]$$

Let $T = P \odot Q$

$t_{11} = \max(.27, .15, .8) = .8$

$t_{12} = \max(.15, .10, 0) = .15$

$t_{13} = \max(.21, 0, .40) = .4$

$t_{14} = \max(.21, .45, .40) = .45$

$t_{21} = \max(0, .21, 1) = 1$

$t_{22} = \max(0, .14, 0) = .14$

$t_{23} = \max(0, 0, .5) = .5$

$t_{24} = \max(0, .63, .5) = .63$

$t_{31} = \max(.36, .18, .5) = .5$

$t_{32} = \max(.20, .12, 0) = .2$

$t_{33} = \max(.28, 0, .25) = .28$

$t_{34} = \max(.28, .54, .25) = .54$

$$T = \begin{bmatrix} .8 & .15 & .4 & .45 \\ 1 & .14 & .5 & .63 \\ .5 & .2 & .28 & .54 \end{bmatrix}$$

Home work:

1) Let $M_4 = \begin{bmatrix} 1 & .2 & 0 & 0 \\ 0 & 0 & .4 & .3 \\ 1 & .2 & 0 & 0 \\ 0 & 0 & .4 & .3 \end{bmatrix}$, $M_5 = \begin{bmatrix} .3 & .6 & 0 & 1 \\ .7 & 0 & 1 & .5 \\ .5 & 0 & 0 & .2 \\ 0 & 0 & 1 & 0 \end{bmatrix}$

Find $P \circ Q$, $P \odot Q$

Soln: $P \circ Q = \max_{y \in Y} \min [p(x,y), Q(y,z)]$

$$\mu_{P \circ Q}(x,z) = \max_{y \in Y} \min [\mu_P(x,y), \mu_Q(y,z)]$$

Let $R = P \circ Q$

$r_{11} = \max[\min(1, .3), \min(.2, .7), \min(0, .5), \min(0, 0)]$

$= \max[.3, .2, 0, 0]$

$r_{11} = .3$

$r_{12} = \max[.6, 0, 0, 0] = .6$

$r_{13} = \max[0, .2, 0, 0] = .2$

$r_{14} = \max[1, .2, 0, 0] = 1$

$r_{21} = \max[0, 0, .4, 0] = .4$

$r_{22} = \max[0, 0, 0, 0] = 0$

$r_{23} = \max[0, 0, 0, .3] = .3$

$r_{24} = \max[0, 0, .2, 0] = .2$

$r_{31} = \max[.3, .2, 0, 0] = .3$

$r_{32} = \max[.6, 0, 0, 0] = .6$

$r_{33} = \max[0, .2, 0, 0] = .2$

$r_{34} = \max[1, .2, 0, 0] = 1$

$r_{41} = \max[0, 0, .4, 0] = .4$

$r_{42} = \max[0, 0, 0, 0] = 0$

$r_{43} = \max[0, 0, 0, .3] = .3$

$r_{44} = \max[0, 0, .2, 0] = .2$

$$R = \begin{bmatrix} .3 & .6 & .2 & 1 \\ .4 & 0 & .3 & .2 \\ .3 & .6 & .2 & 1 \\ .4 & 0 & .3 & .2 \end{bmatrix}$$

$$P \odot Q = \max_{y \in Y} [P(x,y) \cdot Q(x,y)]$$

$$\mu_{P \odot Q} = \max_{y \in Y} [\mu_P(x,y) \cdot \mu_Q(x,y)]$$

Let $T = P \odot Q$

$$t_{11} = \max[.3, .14, 0, 0] = .3$$

$$t_{12} = \max[.6, 0, 0, 0] = .6$$

$$t_{13} = \max[0, .2, 0, 0] = .2$$

$$t_{14} = \max[1, 0.10, 0, 0] = 1$$

$$t_{21} = \max[0, 0, 0.20, 0] = 0.20$$

$$t_{22} = \max[0, 0, 0, 0] = 0$$

$$t_{23} = \max[0, 0, 0, .3] = .3$$

$$t_{24} = \max[0, 0, .8, 0] = .8$$

$$t_{31} = \max[.3, .14, 0, 0] = .3$$

$$t_{32} = \max[.6, 0, 0, 0] = .6$$

$$t_{33} = \max[0, .2, 0, 0] = .2$$

$$t_{34} = \max[1, .10, 0, 0] = 1$$

$$t_{41} = \max[0, 0, .20, 0] = .20$$

$$t_{42} = \max[0, 0, 0, 0] = 0$$

$$t_{43} = \max[0, 0, 0, .3] = .3$$

$$t_{44} = \max[0, 0, .8, 0] = .8$$

$$T = \begin{bmatrix} .3 & .6 & .2 & 1 \\ .2 & 0 & .3 & .8 \\ .3 & .6 & .2 & 1 \\ .2 & 0 & .3 & .8 \end{bmatrix}$$

Binary relations on a single set

A binary relation on a single set can be denoted by $R(X,X)$ or $R(X^2)$ and is a subset of $X \times X = X^2$ these relations are referred to as directed graphs (or) digraphs.

Binary relations $R(X,X)$ can be expressed by the same forms as central binary relations (matrices, sagittal diagram, tables)

It can be conveniently expressed in terms of simple diagram with the following properties

1) each element of the set X is represented by a single node in the diagram

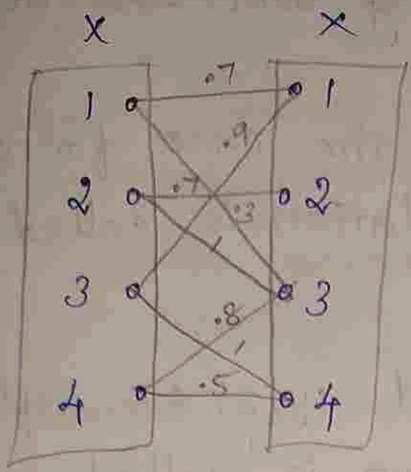
2) directed connections between nodes indicate pairs of elements of X for which the grade of membership in R is non-zero

3) Each connection in the diagram is labelled by the actual membership grade of the corresponding pair in R.

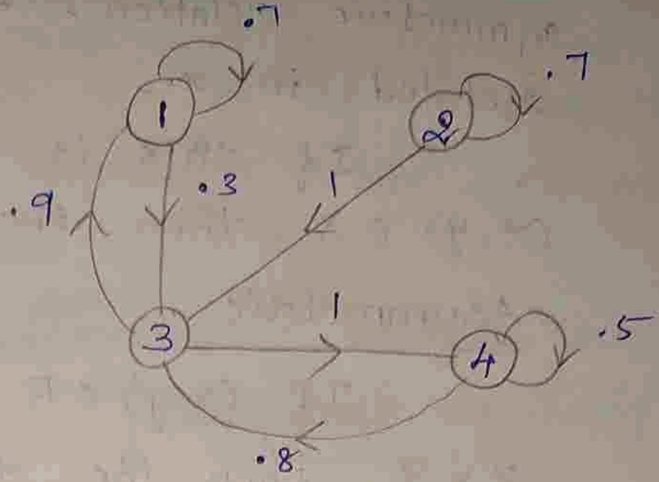
An example of this diagram for a relation $R(X,X)$ defined on X. $X = \{1,2,3,4\}$ is shown where it can be compared with the other forms of represented of binary relation

X \ X	1	2	3	4
1	.7	0	.3	0
2	0	.7	1	0
3	.9	0	0	1
4	0	0	.8	.5

membership matrix.



sagittal diagram



simple diagram

x	y	$\mu_R(x,y)$
1	1	.7
1	3	.3
2	2	.7
2	3	1
3	1	.9
3	4	1
4	4	.8
4	5	.5

Table

Reflexive:

(A crisp relation $R(X,X)$ is __reflexive__ iff $(x,x) \in R$ $\forall$ $x \in X$), (ie) if every elements of X is related to itself.

Otherwise, $R(x,x)$ is called __irreflexive__

(If $(x,x) \notin R$ for every x in X, the relation is called __anti reflexive__)

Symmetric

$(x,y) \in R \Leftrightarrow (y,x) \in R$

A crisp relation $R(x,y)$ is __symmetric__ iff for every $(x,y) \in R$ it is also the case that $(y,x) \in R$ where $x, y \in X$.

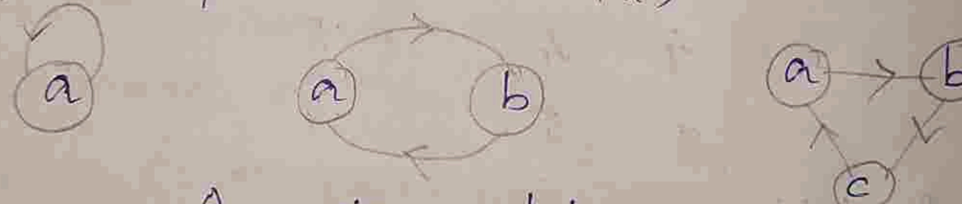
Thus whenever an element x is related to an element y through a symmetric relation, then y will also be related to x.

If this is not the case for each $(x, y) \in X$, then the relation is called <u>Asymmetric</u>

(If $(x,y) \in R$ and $(y,x) \in R$ implies $x = y$ then the relation is called <u>antisymmetric</u>)

(If either $(y,x) \in R$ or $(x,y) \in R$ whenever $x \neq y$ then the relation is called <u>strictly anti symmetric</u>)

The properties of relaxivity, symmetry and transitivity are illustrated for crisp relation $R(x, x)$



A crisp relation $R(X, X)$ is called <u>transitive</u> iff $(x, z) \in R$ whenever both $(x,y) \in R$ and $(y,z) \in R$ for atleast one $y \in X$.

In otherwoods, the relation of x to y and of y to z implies the relation of x to z in a transitive relation

(A relation that does not satisfy this property is called <u>non-transitive</u>)

(If $(x,z) \notin R$ whenever both $(x,y) \in R$ and $(y,z) \in R$ then the relation is called ~~antisyd~~ <u>anti transitive</u>.)

Ex:

Let R be a crisp relation defined on $X \times X$, where X is the set of all university courses and R represents the relation "is a prerequisite of".

R is antireflexive because a course is never a prerequisite of itself

Further, if one course is a prerequisite of another, the reverse will never be true

∴ R is antisymmetric

Finally if a course is a prerequisite for a second course which is itself a prerequisite for a third, then the first course is also a prerequisite for the third course

Then the relation R is transistive

Fuzzy relation

5m (i) For fuzzy relation $R(X, X)$ is reflexive iff $\mu_R(x, x) = 1$ for all $x \in X$

If this is not the case for some $x \in X$ then the relation is called irreflexive

If $\mu_R(x, x) \neq 1$ for every $x \in X$ it is called antireflexive.

If $\mu_R(x, x) \geq \varepsilon$, where $0 < \varepsilon < 1$ $\forall x \in X$ it is called ε-reflexive.

ii) A fuzzy relation $R(x, y)$ is called symmetric if $\mu_R(x, y) = \mu_R(y, x)$ for all $x, y \in X$

If it is not satisfied for some $x \in X$, the relation is called asymmetric

When $\mu_R(x,y) > 0$ and $\mu_R(y,x) > 0$ implies that $x = y$ for all $x, y \in X$ then the relation R is called antisymmetric.

iii) $R(X, X)$ is called transitive or max min transitive if

$$\mu_R(x,z) \geq \max_{y \in Y} \min[\mu_R(x,y), \mu_R(y,z)]$$

for every point $(x,z) \in X^2 = X \times X$

If it not satisfy for some members of X it is called non-transitive

If $\mu_R(x,z) < \max_{y \in Y} \min[\mu_R(x,y), \mu_R(y,z)]$ for all $(x,z) \in X^2$ then the relation is called anti transitive.

Note:

Max product transitivity

$\mu_R(x,z) \geq \max_{y \in Y} [\mu_R(x,y), \mu_R(y,z)]$ for all $(x,z) \in X^2$

Ex:

i) Let R be the fuzzy relation defined on the set of cities and representing the concept very near we may assume that a city is certainly (ie) to a degree of 1) very near to itself

∴ The relation is reflexive.

ii) Further if city a is very near to city B then B is certainly very near to A to the same degree.

∴ The relation is symmetric

iii) Finally if city A is very near to city B to some degree say ·7 and city B is very near city c to some degree say ·8

It is possible (although not necessary) that city A is very near to city c to a smaller degree say ·5

The relation is non-transitive.

Transitive closure

Given a relation $R(x, x)$ its transitive closure $R_T(x, x)$ can be determined by a simple algorithm that consists of following three steps.

i) $R' = R \cup (R \circ R)$

ii) If $R' \neq R$, make $R' = R$ and go step (i)

iii) stop $R' = R_T$

Eg:

Using the algorithm given we can determine the transitive max min closure $R_T(x, x)$ for a fuzzy relation $R(x, x)$ defined by the membership matrix.

$$R = \begin{bmatrix} .7 & .5 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & .4 & 0 & 0 \\ 0 & 0 & .8 & 0 \end{bmatrix}$$

Soln:-

Applying step ① we get

$$R \circ R = \begin{bmatrix} .7 & .5 & 0 & .5 \\ 0 & 0 & .8 & 0 \\ 0 & 0 & 0 & .4 \\ 0 & .4 & 0 & 0 \end{bmatrix}$$

$$R \cup (R \circ R) = \begin{bmatrix} .7 & .5 & 0 & .5 \\ 0 & 0 & .8 & 1 \\ 0 & .4 & 0 & .4 \\ 0 & .4 & .8 & 0 \end{bmatrix} = R'$$

since $R' = R$ we take R' as a new matrix R and respectively the previous procedure we get

$$R \circ R = \begin{bmatrix} .7 & .5 & .5 & .5 \\ 0 & .4 & .8 & .4 \\ 0 & .4 & .4 & .4 \\ 0 & .4 & .4 & .4 \end{bmatrix}$$

$$R \cup (R \circ R) = \begin{bmatrix} .7 & .5 & .5 & .5 \\ 0 & .4 & .8 & 1 \\ 0 & .4 & .4 & .4 \\ 0 & .4 & .8 & .4 \end{bmatrix} = R'$$

since $R' \neq R$ at this stage again repeat the procedure with the new relation we get

$$R \cup (R \circ R) = \begin{bmatrix} .7 & .5 & .5 & .5 \\ 0 & .4 & .8 & 1 \\ 0 & .4 & .4 & .4 \\ 0 & .4 & .8 & .4 \end{bmatrix} = R$$

$$R_T = \begin{bmatrix} .7 & .5 & .5 & .5 \\ 0 & .4 & .8 & 1 \\ 0 & .4 & .4 & .4 \\ 0 & .4 & .8 & .4 \end{bmatrix}$$ is the membership matrix of the transitive closure R_T (corresponding to the given relation $R(X, X)$.

Equivalence and similarity relation

(A crisp binary relation $R(X,X)$ ie., reflexive, symmetric and transistive is called an equivalence relation for each element $x \in X$.)

We can define a crisp set A_x, which contains all the elements of X that are related to x. By the equivalence relation

($A_x = \{y / (x, y) \in R(X, X)\}$ A_x is called the equivalence class of $R(X,X)$ with respect to x)

A_x clearly a subset of X. the element x is itself contained in A_x due to the reflexivity of R because R is transitive and symmetric each member of A_x is related to all the other members of A_x.

Furthermore no member of A_x is related to any element of X not included in A_x.

(This set A_x is referred be as an equivalence class of $R(X, X)$ with respect to x)

The family of all such equal classes define by the relation which is usually denoted by X/R forms a partition on X.

Ex:

Let $X = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$.

The cartesian product $X \times X$ contains 100 members $(1,1), \ldots (10,10)$

Let $R(X, X) = \{(x, y) / x$ and y have the same remainder when divided by $3\}$

The relation is easily shown to be reflexive, symmetric and transitive and is therefore an equivalence relation on X. The three equivalence classes defined by this relation are $A_1 = A_4 = A_7 = A_{10} = \{1, 4, 7, 10\}$

$A_2 = A_5 = A_8 = \{2, 5, 8\}$

$A_3 = A_6 = A_9 = \{3, 6, 9\}$

Hence $X/R = \{\{1, 4, 7, 10\}, \{2, 5, 8\}, \{3, 6, 9\}\}$

(A fuzzy binary relation (ie) reflexive, symmetric and transitive is known as a <u>similarity relation</u>.)

For each $x \in X$, a similarity class can be defined as a fuzzy set in which the membership grade of any particular element represents the similarity of that element to the element x. If all the elements in the class are similar to x to the degree of 1 and similar to all elements outside the set to the degree of 0 then the grouping again becomes an equivalence class

Defn: partition tree.

(Let $\pi(R_\alpha)$ denote the partition corresponding to the equivalence relation. Clearly two elements x and y belong to same block of partition $\pi(R_\beta)$ iff $\mu_R(x,y) \geq \alpha$, each similarity relation is associated with the set $\pi(R) = \{\pi(R_\alpha) / \alpha \in \Lambda_\alpha\}$ of partitions). These partitions are nested in the sense that $\pi(R_\alpha)$ is a refinement of $\pi(R_\beta)$ iff $\alpha \geq \beta$.

Ex:

The fuzzy relation $R(x,x)$ represented by a membership matrix

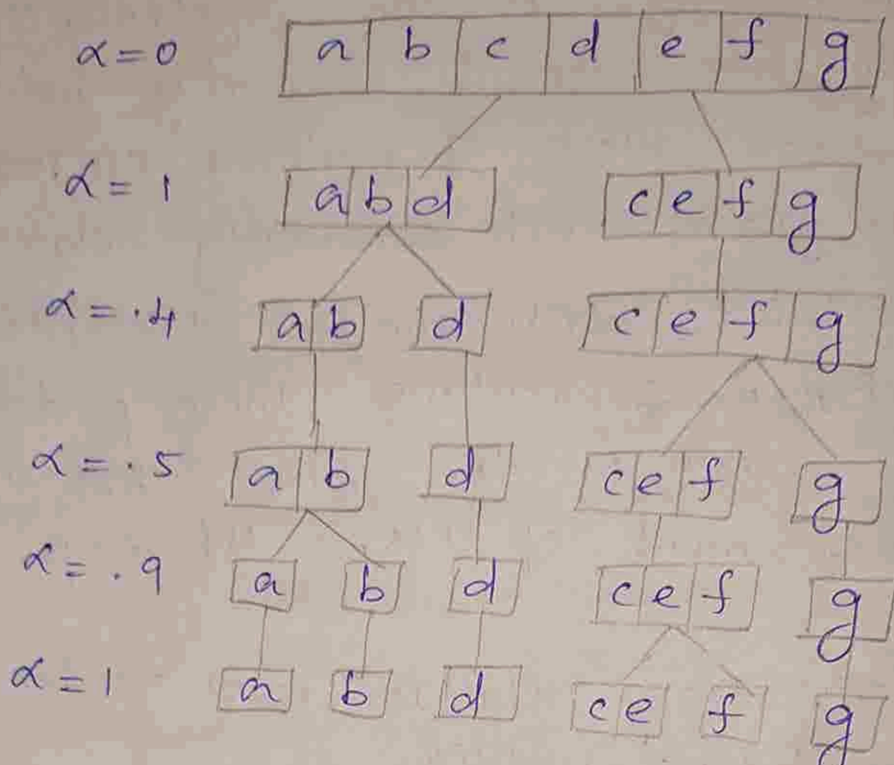
	a	b	c	d	e	f	g
a	1	.8	0	.4	0	0	0
b	.8	1	0	.4	0	0	0
c	0	0	1	0	1	.9	.5
d	.4	.4	0	1	0	0	0
e	0	0	1	0	1	.9	.5
f	0	0	.9	0	.9	1	.5
g	0	0	.5	0	.5	.5	1

is a similarity relation on $X = \{a, b, c, d, e, f, g\}$

The level set of R is $(\Lambda_R) = \{0.4, .5, .8, .9, 1\}$

R is associated with the sequence of five nested partitions $\pi(R_\alpha)$, for $\alpha \in \Lambda_R$ and $\alpha > 0$

Their refinement relationship can be convinently diagrammed by a partition tree



Compatibility (or) Tolerance relation

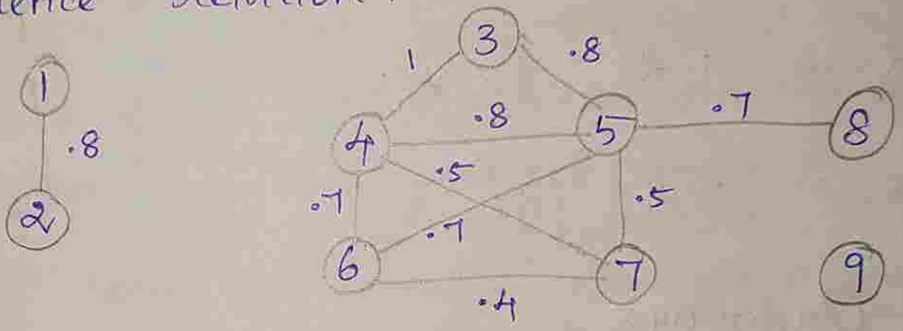
(A binary relation $R(X, X)$ that is reflexive and symmetric is usually called a compatibility relation or a tolerance relation) (When $R(X, X)$ is a reflexive and symmetric fuzzy relation is called a proximity relation.)

An important concept associated with compatibility relation are compatibility classes (also called tolerance classes). Given a crisp compatibility relation $R(X, X)$ a compatibility class is a subset A of X such that $x R y$ for all $x, y \in A$. A maximal compatibility class or maximal compatible is a compatibility class that is not properly contained within any other compatibility class.

The family consisting of all the maximal compatibles induced by R on X is called a complete cover of X with respect to R.

Quasi equivalence relation:

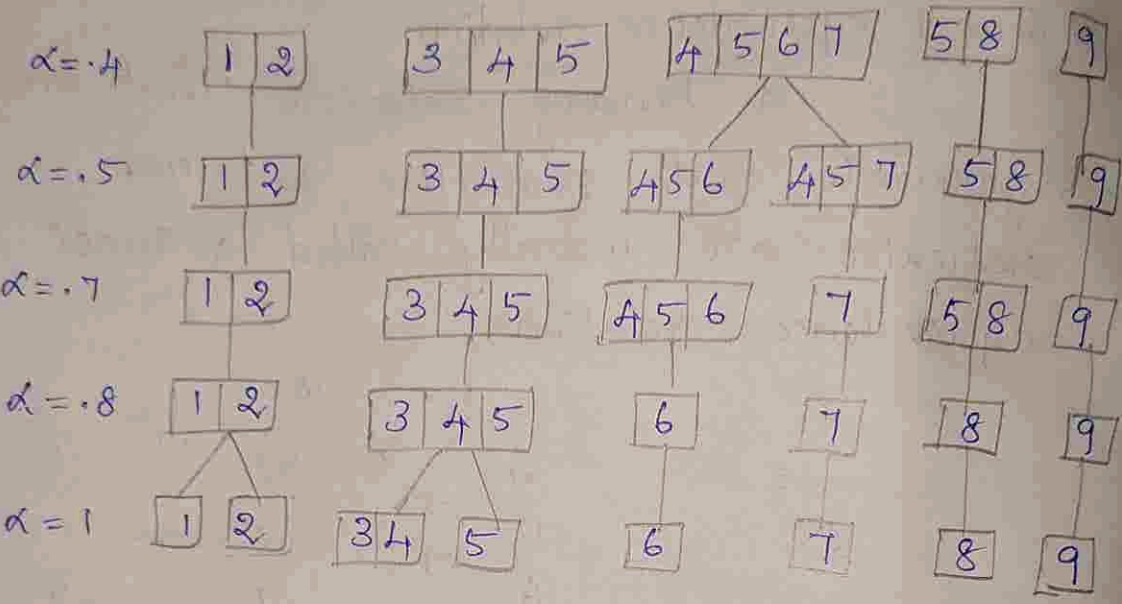
A binary relation that are symmetric and transitive but not reflexive are usually called a Quasi equivalence relation.



Eg: consider a fuzzy relation $R(X,X)$ defined on $X = N_9$ by the following membership matrix.

	1	2	3	4	5	6	7	8	9
1	1	.8	0	0	0	0	0	0	0
2	.8	1	0	0	0	0	0	0	0
3	0	0	1	1	.8	0	0	0	0
4	0	0	1	1	.8	.7	.5	0	0
5	0	0	.8	.8	1	.7	.5	.7	0
6	0	0	0	.7	.7	1	.4	0	0
7	0	0	0	.5	.5	.4	1	0	0
8	0	0	0	0	.7	0	0	1	0
9	0	0	0	0	0	0	0	0	1

Since the matrix is symmetric and all entries on the main diagonal are equal to 1 the relation represented is reflexive and symmetric it is therefore a compatibility relation.



Orderings

A crisp binary relation $R(X, X)$ that is reflexive, antisymmetric and transitive is called a partial ordering. The common symbol $\leq$ is suggestive of the properties of this class of relations.

Thus $x \leq y$ denotes $(x, y) \in R$ and signifies that x precedes y.

The inverse partial ordering $R^{-1}(X, X)$ is suggested by the symbol $\geq$.

If $y \geq x$, indicating that $(y, x) \in R^{-1}$, then we say that y succeeds x.

When $x \leq y$, x is also referred to as a predeccesor of y, while y is called a successor of x.

When $x \leq y$ and there is no z such that $x \leq z$ and $z \leq y$, then x is called an immediate predecessor of y and y is called an immediate successor of x.

A partial ordering $\leq$ on X does not guarantee that all pairs of elements x, y in X are comparable in the sense that either $x \leq y$ or $y \leq x$.

Thus for some $x, y \in X$, it is possible that x is neither a predeccessor nor a successor of y. Such pairs are called non comparable with respect to $\leq$.

Defn of some fundamental concepts associated with partial orderings

(* If $x \in X$ and $x \leq y$ for every $y \in X$, then x is called the first member of X with respect to the relation denoted by $\leq$.

* If $x \in X$ and $y \leq x$ for every $y \in X$, then x is called the last member of X with respect to the partial ordering relation.

* If $x \in X$ and $y \leq x$ implies $x = y$ then x is called a minimal member of X with respect to the relation.

* If $x \in X$ and $x \leq y$ implies $x = y$ then x is called a maximal member of X with respect to the relation.)

Properties of partial ordering:

1) There exists atmost one first member and atmost one last member.

2) There exists atleast one maximal member and atleast one minimal member.

3) If a first member exists, then only one minimal member exists and it is identical with the first member.

4) If a last member exists, then only one maximal member exists and it is identical with the last member.

5) The first and last members of a partial ordering relation correspond to the last and first members of the inverse partial ordering respectively.

Let X again be a set on which a partial ordering is defined and let A be a subset of X $(A \subseteq X)$. If $x \in X$ and $x \leq y$ for every $y \in A$, then x is called a lower bound of A on X with respect to the partial ordering.

If $x \in X$ and $y \leq x$ for every $y \in A$, then x is called an upper bound of A on X with respect to the relation.

If a particular lower bound suceeds every other lower bound of A, then it is called the greatest lower bound, or infimum, of A.

If a particular upper bound precedes every other upper bound of A, then it is called the least upper bound, or supremum of A

A partial ordering on a set X that contains a greatest lower bound and a least upper bound for every subset of X is called a lattices.

A partial ordering $\leq$ on X is said to be connected if and only if for all $x, y \in X, x \neq y$ implies either $x \leq y$ or $y \leq x$.

When partial ordering is connected, then all pairs of elements of X are comparable by the ordering. Such an ordering is usually called a linear ordering (or) total ordering (or) simple ordering (or) complete ordering.

Every partial ordering on a set X can be conveniently represented by a diagram in which each element of X is expressed by a single node that is connected only to the nodes representing its immediate predecessors and immediate successors. The connections are directed in order to distinguish predecessors from successors: the arrow $\leftarrow$ indicates the inequality $\leq$. Diagrams of this sort are called Hasse diagrams.

Eg: 3.18

Three crisp partial orderings P, Q and R on the set $X = \{a, b, c, d, e\}$ are defined by their membership matrices (crisp) and their Hasse diagrams.

The underlined entries in each matrix indicate the relationship of immediate predecessor and successor that is employed in the corresponding Hasse diagram. P has no special properties, Q is a lattices, and R is an example of a lattices that represents a linear ordering.

P

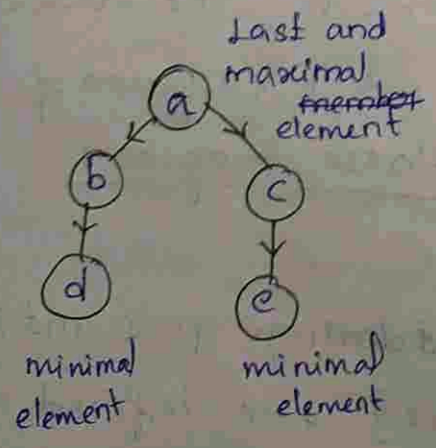
	a	b	c	d	e
a	1	0	0	0	0
b	$\underline{1}$	1	0	0	0
c	$\underline{1}$	0	1	0	0
d	1	$\underline{1}$	0	1	0
e	1	0	$\underline{1}$	0	1

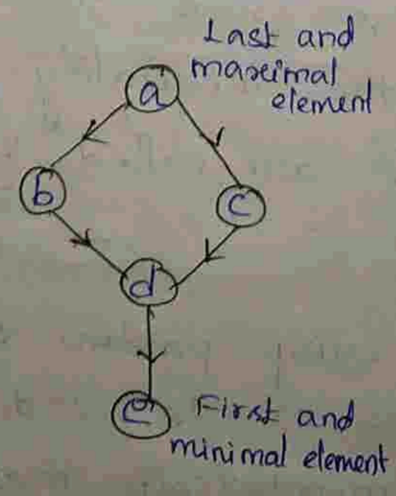
Q

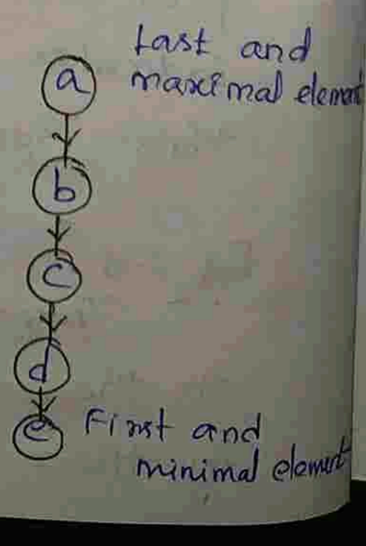
	a	b	c	d	e
a	1	0	0	0	0
b	$\underline{1}$	1	0	0	0
c	$\underline{1}$	0	1	0	0
d	1	$\underline{1}$	$\underline{1}$	1	0
e	1	1	1	$\underline{1}$	1

R

	a	b	c	d	e
a	1	0	0	0	0
b	$\underline{1}$	1	0	0	0
c	1	$\underline{1}$	1	0	0
d	1	1	$\underline{1}$	1	0
e	1	1	1	$\underline{1}$	1







(A fuzzy binary relation R on a set X is a fuzzy partial ordering if and only if it is reflexive, antisymmetric and transitive under some form of fuzzy transitivity.)

(When a fuzzy partial ordering is defined on a set X, then two fuzzy set are associated with each element x in X. The first is called the dominating class of x. It is denoted by $R \geq \{x\}$ and is defined by

$$\mu_{R \geq \{x\}} = \mu_R(x, y),$$

where $y \in X$

In other words, the dominating class of x contains the members of X to the degree to which they dominate x.)

The second fuzzy set of concern is the class dominated by x, which is denoted by $R_{\leq \{x\}}$ and defined by

$$\mu_{R \leq \{x\}}(y) = \mu_R(y, x),$$

where $y \in X$.

The class dominated by x contains the elements of X to the degree to which they are dominated by x.

An element $x \in X$ is undominated if and only if

$$\mu_R(x, y) = 0$$

for all $y \in X$ and $y \neq x$.

For a crisp subset A of a set X on which a fuzzy partial ordering R is defined, the fuzzy upper bound for A is the fuzzy set denoted by U(R, A) and defined by

$$U(R,A) = \bigcap_{x \in A} R_{\geq}\{x\}$$

where $\cap$ denotes an appropriate fuzzy intersection.

If a least upper bound of the set A exists, it is the unique element x in U(R, A) such that

$$\mu_{U(R,A)}(x) > 0 \quad \text{and} \quad \mu_R(x, y) > 0$$

for all elements y in the support of U(R, A)

Eg: 3.19

The following membership matrix defines a fuzzy partial ordering R on the set X = {a, b, c, d, e}

	a	b	c	d	e
a	1	.7	0	1	.7
b	0	1	0	.9	0
c	.5	.7	1	1	.8
d	0	0	0	1	0
e	0	1	0	.9	1

rows - dominating
column - class dominated by x

The dominating class for each element is given by the row of the matrix corresponding to that element.

The columns of the matrix ~~as~~ given the dominated class for each element.

Under this ordering, the element d is undominated and the element c is undominating. For the subset $A = \{a, b\}$, the upper bound is the fuzzy set produced by the intersection of the dominating classes for a and b. Employing the min operator for fuzzy intersection, we obtain

$$U(R, \{a, b\}) = .7/b + .9/d$$

The unique least upper bound for the set A is the element b.

A fuzzy preordering is a fuzzy relation that is reflexive and transistive.

A fuzzy weak ordering R is an ordering satisfying all the properties of a fuzzy linear ordering except antisymmetry.

Alternatively, it can be thought of as a fuzzy preordering in which either $\mu_R(x, y) > 0$ or $\mu_R(y, x) > 0$ for all $x \neq y$.

A fuzzy strict ordering is antireflexive, antisymmetric and transitive. It can clearly be derived from any partial ordering R by replacing the values $\mu_R(x, x) = 1$ with zeroes for all $x \in X$.

Morphism:

If two crisp binary relation $R(X,X)$ and $Q(Y,Y)$ are defined on set X and Y respectively then a function $h: X \to Y$ is said to be homomorphism from (X,R) to (Y,Q) if $(x_1, x_2) \in R$ implies $h(x_1), h(x_2) \in Q$ for all $x_1, x_2 \in X$.

In other words, a homomorphism implies that for every two elements of set x that are related under the relation R, their homomorphic images $h(x_1), h(x_2)$ in set Y are related under the relation Q.

When $R(X,X)$ and $Q(Y,Y)$ are fuzzy binary relations, this implication can be generalized by $\mu_R(x_1, x_2) \le \mu_Q(h(x_1), h(x_2))$ $\forall x_1, x_2 \in X$ and their images $h(x_1), h(x_2) \in Y$

Eg:-

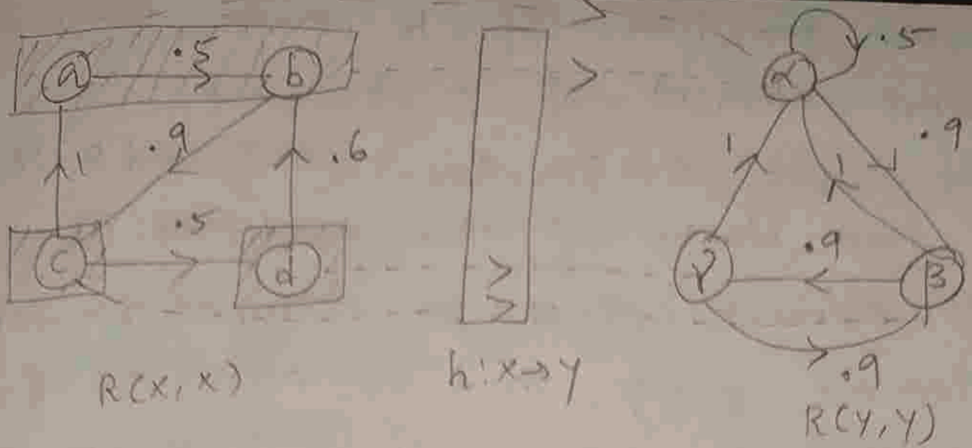
The membership matrices represent fuzzy relations $R(X,X)$ and $Q(Y,Y)$ defined on sets $X = \{a, b, c, d\}$ and $Y = \{\alpha, \beta, \gamma\}$ respectively.

	a	b	c	d
a	0	.5	0	0
b	0	0	.9	0
c	1	0	0	.5
d	0	.6	0	0

	α	β	γ
α	.5	.9	0
β	1	0	.9
γ	1	.9	0

$h: a, b \to \alpha$
$c \to \beta$
$d \to \gamma$

$h: a, b \to \alpha$
$c \to \beta$



$$\mu_R(x_1, x_2) \le \mu_Q[h(x_1), h(x_2)]$$

Let $x_1 = a$, $x_2 = b$

$$\mu_R(a, b) \le \mu_Q(\alpha, \alpha)$$

$$.5 \le .5$$

Let $x_1 = b$, $x_2 = d$

$$\mu_R(b, d) \le \mu_Q(\alpha, \gamma)$$

$$.6 < 1$$

Let $x_1 = c$, $x_2 = d$

$$\mu_R(c, d) \le \mu_Q(\beta, \gamma)$$

$$.5 \le .9$$

Figure illustrates this fuzzy homomorphism

Strong homomorphism:

A function h is called a strong homomorphism for a crisp sets. It satisfies the two implications $(x_1, x_2) \in R$ implies $(h(x_1), h(x_2)) \in Q$ for all $x_1, x_2 \in X$ and $(y_1, y_2) \in Q$ implies $(x_1, x_2) \in R$ for all $y_1, y_2 \in Y$ where $x_1 \in h^{-1}(y_1)$ and $x_2 \in h^{-1}(y_2)$

The function h imposes a partition π_h on the set X such that any two elements $x_1, x_2 \in X$ belong to the same block of the partition iff h maps them to the same element of Y.

Let $A = \{a_1, a_2, \dots, a_n\}$ and $B = \{b_1, b_2, \dots, b_n\}$ be two blocks of this partition π_h and let all elements of A be mapped to some element $y_1 \in Y$ and all elements of B be mapped to some element $y_2 \in Y$ then the function h is said to be strong homomorphism from (X, R) to (Y, Q) iff the degree the strongest relation between any element of A and any element of B in the fuzzy relation R equals the strength of the relation between y_1 and y_2 in the fuzzy relation Q.

Formally

$$\max_{i,j} \mu_R(a_i, b_j) = \mu_Q(y_1, y_2)$$

This equality must be satisfied for each pair of blocks of the partition π_h.

Ex:

consider two relation $R(X, X)$ and $Q(Y, Y)$ defined on $X = \{a, b, c, d, e, f\}$ and $Y = \{\alpha, \beta, \gamma\}$ respectively which are represented by the following membership matrices.

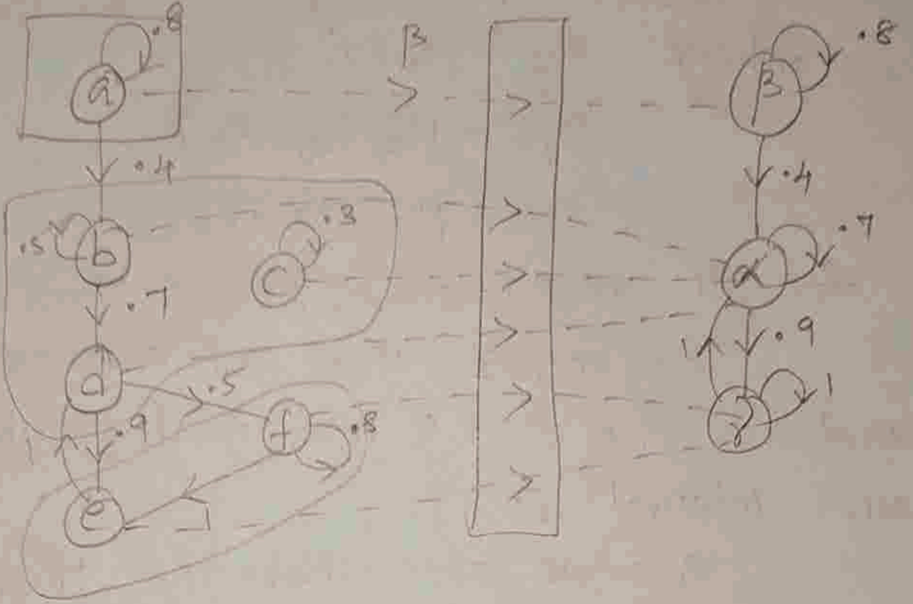
	a	b	c	d	e	f
a	.8	.4	0	0	0	0
b	0	.5	0	.7	0	0
c	0	0	.3	0	0	0
d	0	.5	0	0	.9	.5
e	0	0	0	1	0	0
f	0	0	0	0	1	.8

	α	β	γ
α	.7	0	.9
β	.4	.8	0
γ	1	0	1

$h: a \to \beta$

$b, c, d \to \alpha$

$e, f \to \gamma$



$$\max_{i,j} \mu_R(a_i, b_j) = \mu_Q(y_1, y_2)$$

i) $\mu_R(a,b) = .4 \quad h(a) = \beta \quad h(b) = \alpha$

$\mu_Q(\beta, \alpha) = .4$

ii) $\mu_R(d,f) = .5 \quad \mu_R(d,e) = .9$

$\max[\mu_R(d,f), \mu_R(d,e)] = \max(.5, .9)$

$= .9$

$\mu_Q(\alpha, \gamma) = .9$

Figure depicts this strong homomorphism.

Fuzzy relation equations

consider three fuzzy binary relations $P(X,Y)$, $Q(Y,Z)$ and $R(X,Z)$ which are defined on the sets

$X = \{x_i \mid i \in I\}$, $Y = \{y_j \mid j \in J\}$

$Z = \{z_k \mid k \in K\}$ where we assume that

$I = N_n$, $J = N_m$, $K = N_s$

Let the membership matrices of P, Q and R we denoted by $P = [p_{ij}]$, $Q = [q_{jk}]$ and $R = [r_{ik}]$ respectively.

where $p_{ij} = \mu_P(x_i, y_j)$

$q_{jk} = \mu_Q(y_j, z_k)$

$r_{ik} = \mu_R(x_i, z_k)$

for all $i \in I (= N_n)$, $j \in J (= N_m)$ and $k \in K (= N_s)$

This means that all entries in the matrices P, Q and R are real numbers in the unit interval $[0, 1]$.

Assume now that the three relations constrained each other in such a way that $P \circ Q = R \rightarrow ①$ where $\circ$ denotes the max min composition.

This means that

$$\max_{j \in J} \min(p_{ij}, q_{jk}) = r_{ik} \rightarrow ②$$

for all $i \in J$, $k \in K$

ie) The matrix equation ① encompasses $n \times s$; simultaneous equation of the form of equation ②

When two in each of the eqns are given and one is unknown. This equations are referred to as fuzzy relation equations.

The problem of determining P from R and Q as a decomposition of R with respect to Q

Let each particular matrix P that satisfies eqn ① we called its solution and let

$S(Q,R) = \{P \mid P \circ Q = R\}$ denote the set of all solutions (the solution set)

Let the matrix equations $P_i \circ Q = r_i$ for all $i \in I$ where

$$P_i = [p_{ij} \mid j \in J] \text{ and } r_i = [r_{ik} \mid k \in K]$$

If $\max_{j \in J} q_{jk} < \max_{i \in I} r_{ik}$ for some k in K then $S(Q,R) = \phi$

Ex:
consider the matrix equation

$$\begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \end{bmatrix} \circ \begin{bmatrix} .9 & .5 \\ .7 & .8 \\ 1 & .4 \end{bmatrix} = \begin{bmatrix} .6 & .3 \\ .2 & 1 \end{bmatrix}$$

where general form is

$$[p_{ij}] \circ [q_{jk}] = [r_{ik}] \text{ where } i \in N_2, j \in N_3 \text{ and } k \in N_2$$

The given matrix represents the following four equations of the form of eqn (2)

$\max\{\min(p_{11}, .9), \min(p_{12}, .7), \min(p_{13}, 1)\} = .6 \rightarrow$ (a)

$\max\{\min(p_{11}, .5), \min(p_{12}, .8), \min(p_{13}, .4)\} = .3 \rightarrow$ (b)

$\max\{\min(p_{21}, .9), \min(p_{22}, .7), \min(p_{23}, 1)\} = .2 \rightarrow$ (c)

$\max\{\min(p_{21}, .5), \min(p_{22}, .8), \min(p_{23}, .4)\} = 1 \rightarrow$ (d)

Equation a and b contain only unknowns p_{11}, p_{12} and p_{13} where as eqn (c) and (d) contain only unknown p_{21}, p_{22} and p_{23}

The given matrix eqn can be decompose into two singular matrix eqn

$$[P_{11}\ P_{12}\ P_{13}] \circ \begin{bmatrix} .9 & .5 \\ .7 & .8 \\ 1 & .4 \end{bmatrix} = [.6\ \ .3] \rightarrow (I)$$

and $$[P_{21}\ P_{22}\ P_{23}] \circ \begin{bmatrix} .9 & .5 \\ .7 & .8 \\ 1 & .4 \end{bmatrix} = [.2\ \ 1] \rightarrow (II)$$

Let us take the eqn (II) here $i = 2$ and $k = 2$

If $\max_{j \in J} q_{jk} < \max_{i \in I} r_{ik}$

then $S(Q, R) = \phi$

L.H.S

$$\max_{j \in J} q_{jk} = \max_{j=1,2,3} q_{jk} = \max\{q_{12}, q_{22}, q_{32}\}$$
$$= \max\{.5, .8, .4\}$$
$$= .8$$

R.H.S

$$\max_{i \in I} r_{i,k} = \max_{i=2} r_{22}$$
$$= 1$$

$$.8 < 1$$

$$S(Q, R) = \phi$$

Thus the given matrix eqn has no Solution.

The procedure for solving finite max-min fuzzy relation equation.

Basic procedure :

i) decompose $P \circ Q = R$ into eqns of the form $p \circ Q = r$ — (A)

One for each row in P and R (p is associated with index j, Q with indices j and k and r with index k)

ii) For each question ⓐ, if $\max_{j \in J} q_{jk} < \max r_k$ for some k then the eqn has no soln: $S(Q, r) = \phi$ and the procedure terminates; otherwise proceed to step ③

iii) determine $\hat{P}$ by procedure ①

iv) If $\hat{P}$ is not a solution of eqn ⓐ, then the eqn has no solution; $S(Q, r) = \phi$ and the procedure terminates; otherwise proceed to step ⑤

v) For each $\hat{P}_j = 0$ and $r_K = 0$, exclude these components as well as the corresponding rows j and columns k from matrix Q in eqn ⓐ : This results $P \circ Q = r \longrightarrow$ ⓑ where we assume $j \in J$, $k \in K$

vi) determine all minimal solution of the reduced eqn ⓑ by procedure ②; This results in $\check{S}(Q, r)$

vii) determine the soln set of the reduced eqn ⓑ

$$S(Q, r) = \bigcup_{\check{P}} \langle \check{P}, \hat{P} \rangle$$ where the union is taken over all $\check{P} \in \check{S}(Q, r)$

viii) Extend all solutions in $S(Q, r)$ by zeros that where excluded in step ⑤ : This results in the solution set $S(Q, r)$ of eqn ⓐ

ix) repeat steps ② - ⑧ for all equations of type in the eqn ⓐ that are embedded in eqn ① : This results in all matrices p that satisfy eqn ①.

procedure: 1

Form the vector $\hat{p} = [\hat{p}_j \mid j \in J]$ in which $\hat{p}_j = \min_{x} \sigma(q_{j,k}, r_k)$

where $\sigma(q_{j,k}, r_k) = \begin{cases} r_k & \text{if } q_{j,k} > r_k \\ 1 & \text{otherwise} \end{cases}$

procedure: 2

i) permute elements of r and the corresponding columns of Q appropriately to arrange them decreasing order

ii) determine sets

$$J_k(\hat{p}) = \{ j \in J \mid \min(p_j, q_{j,k}) = r_k \}$$

for all $k \in K$ and form

$$J(\hat{p}) = \underset{k \in K}{\times} J_k(\hat{p})$$

iii) For each $\beta \in J(\hat{p})$ and each $j \in J$, determine the set $K(\beta, j) = \{ k \in K \mid \beta_k = j \}$

iv) For each $\beta \in J(\hat{\beta})$, generate the touple

$g(\beta) = \{ g_j(\beta) \mid j \in J \}$ by taking

$$g_j(\beta) = \begin{cases} \max r_k & \text{if } K(\beta, j) \neq \phi, k \in K(\beta, j) \\ 0 & \text{otherwise} \end{cases}$$

v) From all touples $g(\beta)$ generated in step (iv), select only the minimal points: This results in $\check{S}(Q, r)$

Given $Q = \begin{bmatrix} .1 & .4 & .5 & .1 \\ .9 & .7 & .2 & 0 \\ .8 & 1 & .5 & 0 \\ .1 & .3 & .6 & 0 \end{bmatrix}$ and $r = [.8\ .7\ .5\ 0]$

determines all solutions of $P \circ Q = r$

Soln:-

Step : 1

Let us write in the form $P \circ Q = r$.

$$P \circ \begin{bmatrix} .1 & .4 & .5 & .1 \\ .9 & .7 & .2 & 0 \\ .8 & 1 & .5 & 0 \\ .1 & .3 & .6 & 0 \end{bmatrix} = [.8\ .7\ .5\ 0] \quad \text{--- (1)}$$

Step 2

check $\max_{j \in J} q_{j,k} < \max r_k$

i) $k = 1 \Rightarrow$ L.H.S

$$\Rightarrow \max_{j \in J} q_{j,k} = \max(q_{11}, q_{21}, q_{31}, q_{41})$$

$$= \max(.1, .9, .8, .1)$$

$$= .9$$

R.H.S $\max r_k = r_1 = .8$

$$.9 \not< .8$$

ii) $k = 2$

L.H.S $\Rightarrow \max(q_{12}, q_{22}, q_{32}, q_{42})$

$$= \max(.4, .7, 1, .3) = 1$$

R.H.S $\max r_k = r_2 = .7$

$$1 \not< .7$$

iii) $k = 3$

$$\max(q_{13}, q_{23}, q_{33}, q_{43}) = .6$$

$$r_3 = .5$$

$$.6 \not< .5$$

iv) $K = 4$

$\max(q_{14}, q_{24}, q_{34}, q_{44}) = .1$

$r_4 = 0$

$.1 \neq 0$

$\max_{j \in J} q_{j,k} \neq \max r_k$

① has a solution

ie) $S(Q, r) \neq \phi$

Step : 3

$\hat{P} = [\hat{p}_j / j \in J]$

$\hat{p}_j = \min_k \sigma(q_{jk}, r_k)$

$$\sigma(q_{jk}, r_k) = \begin{cases} r_k & \text{if } q_{j,k} > r_k \\ 1 & \text{otherwise} \end{cases}$$

$j=1 \Rightarrow \sigma(q_{jk}, r_k) \Rightarrow \sigma(q_{11}, r_1) = \sigma(.1, .8) = 1$

$\sigma(q_{12}, r_2) \Rightarrow \sigma(.4, .7) = 1$

$\sigma(q_{13}, r_3) \Rightarrow \sigma(.5, .5) = 1$

$\sigma(q_{14}, r_4) \Rightarrow \sigma(1, 0) = 0$

$j=2$

$\Rightarrow \sigma(q_{2k}, r_k)$

$\Rightarrow \sigma(q_{21}, r_1) \Rightarrow \sigma(.9, .8) = .8$

$\Rightarrow \sigma(q_{22}, r_2) \Rightarrow \sigma(.7, .7) = 1$

$\Rightarrow \sigma(q_{23}, r_3) \Rightarrow \sigma(.2, .5) = 1$

$\Rightarrow \sigma(q_{24}, r_4) \Rightarrow \sigma(0, 0) = 1$

$\hat{p}_j = \hat{p}_2 = \min[.8, 1, 1, 1] = .8$

$j=3$

$\Rightarrow \sigma(q_{3k}, r_k)$

$\Rightarrow \sigma(q_{31}, r_1) = 1$

$\Rightarrow \sigma(q_{32}, r_2) = .7$

$\Rightarrow \sigma(q_{33}, r_3) = 1$

$\Rightarrow \sigma(q_{34}, r_4) = 1$

$\hat{p}_3 = .7$

$j = 4 \Rightarrow \sigma(q_{4k}, r_4)$

$\Rightarrow \sigma(q_{41}, r_1) = 1$

$\Rightarrow \sigma(q_{42}, r_2) = 1$

$\Rightarrow \sigma(q_{43}, r_3) = .5$

$\Rightarrow \sigma(q_{44}, r_4) = 1$

$\hat{p}_4 = .5$

Thus $\hat{p} = [0, .8, .7, .5]$

Step : 4

$\hat{p} \in S(Q, r)$, $S(Q, r) \neq \phi$

Step : 5

since $\hat{p} = 0$ and $r_4 = 0$

we exclude the 1st row and 4th column of Q

$\hat{p} = [.8, .7, .5]$

The reduced equation has the form

$$[P_1 \; P_2 \; P_3] \circ \begin{bmatrix} .9 & .7 & .2 \\ .8 & 1. & .5 \\ .1 & .3 & .6 \end{bmatrix} = [.8 \; .7 \; .5]$$

P_1, P_2, P_3 represents P_2, P_3, P_4 respectively.

Step : 6

To determine the minimal solution let us find

$$J_K(\hat{p}) = \{ j \in J \,/\, \min(\hat{p}_j, q_{j,k}) \} = r_k$$

$K=1$, To find $J_1(\hat{p})$

$$\min(\hat{p}_j, q_{j,k}) = r_K$$

$j=1$, $\min(\hat{p}_1, q_{11}) = \min(.8, .9) = .8 = r_1$

$j=2$, $\min(\hat{p}_2, q_{21}) = \min(.7, .8) = .7 \neq r_1$

$j=3$, $\min(.5, .1) = .1 \neq r_1$

$$J_1(\hat{p}) = \{1\}$$

$K=2$,

$j=1$, $\min(.8, .7) = .7 = r_2$

$j=2$, $\min(.7, .1) = .7 = r_2$

$j=3$, $\min(.5, .3) = .3 \neq r_2$

$$J_2(\hat{p}) = \{1, 2\}$$

$K=3$,

$j=1$, $\min(.8, .2) = .2 \neq r_3$

$j=2$, $\min(.7, .5) = .5 = r_3$

$j=3$, $\min(.5, .6) = .5 = r_3$

$$J_3(\hat{p}) = \{2, 3\}$$

$$J(\hat{p}) = \underset{k \in K}{\times} J_K(\hat{p})$$

$$= J_1(\hat{p})\, J_2(\hat{p})\, J_3(\hat{p})$$

$$= \{1\} \times \{1,2\} \times \{2,3\}$$

$K(\beta, j)$	j = 1	2	3	$g(\beta)$
$\beta_1\ \beta_2\ \beta_3$				
$\beta = 1\ 1\ 2$	$\{1,2\}$	$\{3\}$	ϕ	$(.8, .5, 0)$
$1\ 1\ 3$	$\{1,2\}$	ϕ	$\{3\}$	$(.8, 0, .5)$
$1\ 2\ 2$	$\{1\}$	$\{2,3\}$	ϕ	$(.8, .7, 0)$
$1\ 2\ 3$	$\{1\}$	$\{2\}$	$\{3\}$	$(.8, .7, .5)$

$$g(\beta) = \{g_j(\beta) / j \in J\}$$

$$g_j(\beta) = \begin{cases} \max r_k & \text{if } K(\beta, j) \neq \phi \\ 0 & \text{otherwise} \quad k \in K(\beta, j) \end{cases}$$

$$\check{S}(Q, r) = \{ {}^1\check{p} = \{.8, .5, 0\}, {}^2\check{p} = \{.8, 0, .5\} \}$$

$${}^1\check{p} = (0, .8, .5, 0), \quad {}^2\check{p} = (0, .8, 0, .5)$$

$$\check{S}(Q, r) = \{ {}^1\check{p} = (0, .8, .5, 0), {}^2\check{p} = (0, .8, 0, .5) \}$$

i) The maximal soln : $\hat{p} = (0, .8, .7, .5)$

ii) The minimal soln : ${}^1\check{p} = (0, .8, .5, 0)$

$${}^2\check{p} = (0, .8, 0, .5)$$

$$S(Q, r) = \{ p \in P \mid {}^1\check{p} \leq p \leq \hat{p} \} \cup \{ {}^2\check{p} \leq p \leq \hat{p} \}$$

is the solution set

Unit - IV

Fuzzy measures

A fuzzy measure is defined by a function $g: \mathcal{P}(x) \to [0,1]$ which assigns to each crisp subset of x a number in the unit interval [0,1]. The following are axioms of fuzzy measures.

Axiom g_1 boundary condition.

$g(\phi) = 0$ and $g(X) = 1$.

Axiom g_2 monotonicity

For every $A, B \in \mathcal{P}(x)$ if $A \subseteq B$ then $g(A) \leq g(B)$

Continuity Axiom g_3

For every sequence $(A_i \in \mathcal{P}(x) \mid i \in N)$ of subsets of x if either $A_1 \subseteq A_2 \subseteq \cdots$ or

$A_1 \supseteq A_2 \supseteq \cdots$

(iii) the sequence of A_i is monotonic)

(iv) $\lim_{i \to \infty} g(A_i) = g(\lim_{i \to \infty} A_i)$

Borel field (or) σ-field

A fuzzy measure is defined more generally as a function $g : \mathcal{B} \to [0,1]$ where $\mathcal{B} \subset \mathcal{P}(X)$ is a family of subsets of X such that

1. $\phi \in \mathcal{B}$ and $X \in \mathcal{B}$
2. If $A \in \mathcal{B}$ then $\bar{A} \in \mathcal{B}$
3. $\mathcal{B}$ is closed under the operation of set union, (i.e.) if $A \in \mathcal{B}$ and $B \in \mathcal{B}$ then $A \cup B \in \mathcal{B}$

The set $\mathcal{B}$ is called a Borel field (or) σ field

(2) **Belief measure**

A belief measure is a function $Bel : \mathcal{P}(X) \to [0,1]$ that satisfies the axioms g_1 through g_3 of fuzzy measure axioms and the following additional axiom

$Bel(A_1 \cup A_2 \cup \ldots \cup A_n) \geq \sum_i Bel(A_i) - \sum_{i<j} Bel(A_i \cap A_j) \ldots + (-1)^{n+1} Bel(A_1 \cap \ldots)$

for every $n \in N$ and every collection of subsets of X

For $n = 2$

$Bel(A_1 \cup A_2) \geq Bel(A_1) + Bel(A_2) - Bel(A_1 \cap A_2)$

For $n = 3$

$$B(A_1 \cup A_2 \cup A_3) \geq Bel(A_1) + Bel(A_2) + Bel(A_3) - Bel(A_1 \cap A_2) - Bel(A_2 \cap A_3) - Bel(A_1 \cap A_3) + Bel(A_1 \cap A_2 \cap A_3)$$

For $n = 4$

$$B(A_1 \cup A_2 \cup A_3 \cup A_4) \geq Bel(A_1) + Bel(A_2) + Bel(A_3) + Bel(A_4) - Bel(A_1 \cap A_2) - Bel(A_1 \cap A_3) - Bel(A_1 \cap A_4) - Bel(A_2 \cap A_3) - Bel(A_2 \cap A_4) - Bel(A_3 \cap A_4) \cdot Bel(A_1 \cap A_2 \cap A_3 \cap A_4)$$

Plausibility measure.

A plausibility measure is a function $Pl : P(X) \rightarrow [0,1]$ that satisfies three axioms g_1 through g_3 of fuzzy measures and the following the additional axiom.

$$Pl(A_1 \cap A_2 \cap \cdots \cap A_n) \leq \sum_i Pl(A_i) - \sum_{i<j} Pl(A_i \cup A_j) + \cdots + (-1)^{n+1} Pl(A_1 \cup A_2 \cup \cdots \cup A_n)$$

for every $n \in N$ and every collection of subsets of X.

For $n = 2$

$$Pl(A_1 \cap A_2) \leq Pl(A_1) + Pl(A_2) - Pl(A_1 \cup A_2)$$

(1) If $A \subset B$ then P.T $Bel(A) \leq Bel(B)$

Proof: Let $A \subset B$ and let $C = B - A$

Then $A \cup C = B$ and $A \cap C = \phi$

For $n = 2$ W.K.T

$Bel(A_1 \cup A_2) \geq Bel(A_1) + Bel(A_2) - Bel(A_1 \cap A_2)$ — (*)

Apply A and C in thm

$Bel(A \cup C) \geq Bel(A) + Bel(C) - Bel(A \cap C)$

$Bel(B) \geq Bel(A) + Bel(C) - Bel(\phi)$

$\Rightarrow Bel(B) \geq Bel(A) + Bel(C)$

and Consequently

$Bel(B) \geq Bel(A)$

(2) P.T $Bel(A) + Bel(\bar{A}) \leq 1$ (or) Derive the fundamental property of belief measure

Let $A_1 = A$ and $A_2 = \bar{A}$ in (*)

$Bel(A \cup \bar{A}) \geq Bel(A) + Bel(\bar{A}) - Bel(A \cap \bar{A})$

$\Rightarrow Bel(X) \geq Bel(A) + Bel(\bar{A}) - Bel(\phi)$

$1 \geq Bel(A) + Bel(\bar{A}) - 0$

$Bel(A) + Bel(\bar{A}) \leq 1$

Note (1) $Pl(A) = 1 - Bel(\bar{A})$ $\forall A \in P(X)$

11y (2) $Bel(A) = 1 - Pl(\bar{A})$

(3) P.T $Pl(A) + Pl(\bar{A}) \geq 1$ (or) Derive the basic inequality of plausibility measure

Proof For $n = 2$

W.K.T

$Pl(A_1 \cap A_2) \le Pl(A_1) + Pl(A_2) - Pl(A_1 \cup A_2)$

Let $A_1 = A$ and $A_2 = \bar{A}$

Then $Pl(A \cap \bar{A}) \le Pl(A) + Pl(\bar{A}) - Pl(A \cup \bar{A})$

$\Rightarrow Pl(\phi) \le Pl(A) + Pl(\bar{A}) - Pl(X)$

$0 \le Pl(A) + Pl(\bar{A}) - 1$

$\Rightarrow Pl(A) + Pl(\bar{A}) \ge 1$

Hence Proof.

b) P.T plausibility is the counterpart for belief.

Proof W.K.T For $n = 3$

$Bel(A_1 \cup A_2 \cup A_3) \ge Bel(A_1) + Bel(A_2) + Bel(A_3) - Bel(A_1 \cap A_2) - Bel(A_1 \cap A_3) - Bel(A_2 \cap A_3) + Bel(A_1 \cap A_2 \cap A_3)$ —①

and $Bel(A) = 1 - P(\bar{A})$ —②

Apply ② in ①, we get

$1 - Pl(\overline{A_1 \cup A_2 \cup A_3}) \ge 1 - Pl(\bar{A_1}) + 1 - P(\bar{A_2}) + 1 - Pl(\bar{A_3}) - (1 - P(\overline{A_1 \cap A_2})) - (1 - P(\overline{A_1 \cap A_3})) - (1 - P(\overline{A_2 \cap A_3})) + 1 - P(\overline{A_1 \cap A_2 \cap A_3})$

$\Rightarrow 1 - Pl(\bar{A}_1 \cap \bar{A}_2 \cap \bar{A}_3) \geq 1 - Pl(\bar{A}_1) + 1 - Pl(\bar{A}_2) - 1 - Pl(\bar{A}_3) + 1 - Pl(\bar{A}_3) - (1 - Pl(\bar{A}_1 \cup \bar{A}_2)) - (1 - Pl(\bar{A}_1 \cup \bar{A}_3)) - (1 - Pl(\bar{A}_2 \cup \bar{A}_3)) + 1 - Pl(\bar{A}_1 \cup \bar{A}_2 \cup \bar{A}_3)$

Cancelling the 1's and multiply the inequality by -1.

$\Rightarrow Pl(\bar{A}_1 \cap \bar{A}_2 \cap \bar{A}_3) \leq Pl(\bar{A}_1) + Pl(\bar{A}_2) + Pl(\bar{A}_3) - Pl(\bar{A}_1 \cup \bar{A}_2) - Pl(\bar{A}_1 \cup \bar{A}_3) - Pl(\bar{A}_2 \cup \bar{A}_3) + Pl(\bar{A}_1 \cup \bar{A}_2 \cup \bar{A}_3)$

This is the counterpart of ①

(5) Every belief measure and its dual plausibility measure can be expressed interms of a function $m: \mathcal{P}(X) \to [0,1]$ such that $m(\phi) = 0$ and $\sum_{A \in \mathcal{P}(X)} m(A) = 1$

Where $m(A)$ is interpreted as the degree of evidence supporting the claim that a specific element of X belongs to the set A but not to any special subset of A (or) as the degree to which we believe that such a claim is warranted.

Q9 Explain Dempster's rule of Combination

Soln Evidence obtained in the same Context from two independent Sources and expressed by two basic assignments m1 and m_2 on some power set $P(X)$ must be appropriately combined to obtain a joint basic assignment $m_{1,2}$

In general evidence can be combined in various ways some of which may take into Consideration the reliability of the Sources and other relevant aspects

The Standard way of Combining evidence is expressed by the formula

$$m_{1,2}(A) = \frac{\sum_{B \cap C = A} m_1(B) \cdot m_2(C)}{1 - K} \rightarrow (1)$$

for $A \neq \phi$ where

$$K = \sum_{B \cap C = \phi} m_1(B) \cdot m_2(C) \rightarrow (2)$$

$$m_{1,2}(\phi) = 0$$

Formula (1) is referred to as Dempster's rule of Combination

Q9 Explain Dempster's rule of Combination

Soln Evidence obtained in the same context from two independent sources and expressed by two basic assignments m_1 and m_2 on some power set $P(X)$ must be appropriately combined to obtain a joint basic assignment $m_{1,2}$

In general evidence can be combined in various ways some of which may take into consideration the reliability of the sources and other relevant aspects

The standard way of combining evidence is expressed by the formula

$$m_{1,2}(A) = \frac{\sum_{B \cap C = A} m_1(B) \cdot m_2(C)}{1-K} \rightarrow (1)$$

for $A \neq \phi$ where

$$K = \sum_{B \cap C = \phi} m_1(B) \cdot m_2(C) \rightarrow (2)$$

$$m_{1,2}(\phi) = 0$$

Formula (1) is referred to as Dempster's rule of Combination

(i) According to the rule, the degree of evidence $m_1(B)$ from the first source that focuses on set $B \in P(X)$ and the degree of evidence $m_2(C)$ from the second source that focuses on set $C \in P(X)$ are combined by taking the product $m_1(B) \cdot m_2(C)$ which focuses on the intersection $B \cap C$.

ii) This is exactly the same way in which the joint probability distribution is calculated from two independent marginal distributions and

iii) Since, However, some intersections of focal elements from the first and second source may result in the same set A we must add the corresponding products to obtain $m_{1,2}(A)$

iv) moreover, some of the intersections may be empty since it is required that $m_{1,2}(\phi) = 0$ the value K expressed by (2) is not included in the definition of the joint basic assignment $m_{1,2}$. This means that the sum of products $m_1(B) \cdot m_2(C)$ for all focal elements B of m_1 and all focal elements C of m_2 such that $B \cap C \neq \phi$ is equal to $1-K$

v) To obtain a normalized basic assignment $m_{1,2}$ as required.

Σ m(A) = 1 we must divide m(A)

each of these products by the factor 1 - k as indicated in (3).

Ex Assume that an old painting was discovered that strongly resembles paintings by Raphael. Such a discovery is likely to generate various questions regarding the status of the painting. Assume the following three questions:

i) Is the discovered painting a genuine painting by Raphael?

ii) Is the discovered painting a product of one of Raphael's many disciples?

iii) Is the discovered painting a counterfeit?

Soln

let R, D and C denote subsets of our universal set X – the set of paintings – that contain the set of paintings by Raphael, the set of all paintings by disciples of Raphael

$\sum m(A) = 1$ we must divide $m(A)$ each of these products by the factor $1 - K$ as indicated in (3).

Ex Assume that an old painting was discovered that strongly resembles paintings by Raphael. Such a discovery is likely to generate various questions regarding the status of the painting. Assume the following three questions:

i) Is the discovered painting a genuine painting by Raphael?

ii) Is the discovered painting a product of one of Raphael's many disciples?

iii) Is the discovered painting a counterfeit?

Soln

Let R, D and C denote subsets of our universal set X – the set of paintings – that contain the set of paintings by Raphael, the set of all paintings by disciples of Raphael and the set of all Contemporary Radhas's painting respectively

Assume that two experts Performed Careful examination of the Painting and Subsequently Provided us with basic assignments m_1 and m_2 specified in Table

Combination of Degrees of evidence from two Independent sources

Focal element	Expert 1		Expert 2		Combined evidence	
	m_1	Bel_1	m_2	Bel_2	$m_{1,2}$	$Bel_{1,2}$
R	.05	.05	.15	.15	.21	.21
D	0	0	0	0	.09	0
C	.05	.05	.05	.05	.12	.09
RUD	.15	.2	.05	.2	.2	.34
RUC	.1	.2	.2	.4	.05	.5
DUC	.05	.1	.05	.1	.31	.16
RUDUC	.6	1	.5	1		1

W.K.T.

$$Bel(A) = \sum_{B \subseteq A} m(B) \text{ and}$$

$$m_{1,2}(A) = \frac{\sum_{B \cap C = A} m_1(B) \cdot m_2(C)}{1 - K}$$

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

$K = m_1(R) \cdot m_2(D) + m_1(R)\, m_2(C) +$
$m_1(R)\, m_2(D \cup C) + \{m_1(D)\, m_2(R)$
$+ m_1(D)\, m_2(C) + m_1(D)\, m_2(R \cup C)$
$+ m_1(C)\, m_2(R) + m_1(C)\, m_2(D)$
$+ m_1(C)\, m_2(R \cup D) + m_1(R \cup D)\, m_2(C)$
$+ m_1(R \cup C)\, m_2(D) + m_1(D \cup C) \cdot m_2(R)$

$= 0.03$

The normalised factor $1 - K =$

$1 - 0.03 = 0.97$

$m_{12}(R) = [m_1(R)\, m_2(R) + m_1(R)\, m_2(R \cup D)$
$+ m_1(R)\, m_2(R \cup C) + m_1(R)\, m_2(R \cup D \cup C) + m_1(R \cup D)\, m_2(R)$
$+ m_1(R \cup D) \cdot m_2(R \cup C) + m_1(R \cup C)\, m_2(R \cup D)$
$+ m_1(R \cup C)\, m_2(R)$
$+ m_1(R \cup D \cup C) \cdot m_2(R)] / 0.97$

$= 0.21$.

$m_{12}(D \cup C) = m_1(D \cup C)\, m_2(D \cup C) + m_1(D \cup C)\, m_2(R \cup D \cup C) + m_1(R \cup D \cup C)\, m_2(D \cup C) / .97$

<u>PROBABILITY MEASURES</u>

When Axiom (A_3) for Belief measures is replaced with a stronger Axiom

$Bel(A \cup B) = Bel(A) + Bel(B)$ when ever $(A \cap B) = \phi$.

Obtain a special type of belief measures (classical probability measures (or) Bayesian belief measures)

(or)

State & prove the most fundamental property of probability measures as special types of belief measures.

Thm:

A belief measure Bel on a finite power set $P(X)$ is a probability measure iff its basic assignment m is given by $m(\{x\}) = Bel\{\{x\}\}$ and $m(A) = 0$ for all subsets of X that are not singletons.

Proof

Assume that Bel is a probability measure

Case (i) For empty set ϕ, the theorem trivially holds.

Since $m(\phi) = 0$ by defn of m

Case (ii) let $A \neq \phi$ and

Assume $A = \{x_1, x_2, \ldots x_n\}$,

then by repeated application of ①

we obtain $Bel(A) = Bel(\{x_1\} + Bel\{x_2, x_3, \ldots x_n\})$

$= Bel(\{x_1\}) + Bel(\{x_2\}) + \dots + Bel\{x_n\}$

WKT:

$$Bel(A) = \sum_{B \subseteq A} m(B)$$

$\therefore Bel(\{x\}) = m\{x\}$ for any $x \in X$

we have, $Bel(A) = m(\{x_1\} + m\{x_2\} + \dots + m\{x_n\})$

$$= \sum_{i=1}^{n} m(\{x_i\})$$

Hence Bel is defined in terms of a basic assignment that focuses only on singletons.

Conversely.

Assume that a basic assignments m is given such that $\sum_{x \in X} m(\{x\}) = 1$

T.P.T Bel is a possibility measure.

(ie) T.P.T $Bel(A \cup B) = Bel(A) + Bel(B)$

For any sets $A, B \subset \mathcal{P}(X)$ such that $A \cap B = \phi$

we have, $Bel(A) + Bel(B) = \sum_{x \in A} m(\{x\}) + \sum_{x \in B} m(\{x\})$

$$= \sum_{x \in A \cup B} m(\{x\}) = Bel(A \cup B)$$

And consequently Bel is a possibility measure Hence the Proof

Defn

The focal elements of a body of evidence ($\mathcal{F}$, m) be nested, then the associated belief and plausibility measures are called consonant.

* **Thm** Given a consonant belief of evidence ($\mathcal{F}$, m) the associated consonant belief and plausibility measures possess the following properties

i) $Bel(A \cap B) = \min[Bel(A), Bel(B)]$ for all $A, B \in \mathcal{P}(X)$

ii) $Pl(A \cup B) = \max[Pl(A), Pl(B)]$ for all A, B ∈ P(X)

Proof Since the focal elements in $\mathcal{F}$ are nested they may be linearly ordered by the subset relationship.

Let $\mathcal{F} = \{A_1, A_2, \ldots A_n\}$ and

Assume that $A_i \subset A_j$ whenever $i < j$

Consider arbitrary subset A and B of X

Let i_1 be the largest integer i

$A_i \subset A$ and

Let i_2 be the largest integer i such that $A_i \subset B$

Then $A_i \subset A$ and $A_i \subset B$ iff $i \leq i_1$ and $i \leq i_2$ respectively

Moreover $A_i \subset A \cap B$ iff $i \leq \min(i_1, i_2)$

Hence

$$Bel(A\cap B) = \sum_{i=1}^{\min(i,j)} m(A_i) *$$

$$= \min\left\{\sum_{i=1}^{i} m(A_i),\ \sum_{i=1}^{j} m(A_i)\right]$$

$$= \min\left[Bel(A),\ Bel(B)\right]$$

ii) Assume that (i) holds

W.K.T $Pl(A) = 1 - Bel(\bar{A})$

$$\therefore Pl(A\cup B) = 1 - Bel(\overline{A\cup B})$$

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

$$= 1 - \min\left[Bel(\bar{A}),\ Bel(\bar{B})\right]$$

$$= \max\left[1 - Bel(\bar{A}),\ 1 - Bel(\bar{B})\right]$$

$$= \max\left[Pl(A),\ Pl(B)\right]$$

$\forall A, B \in P(X)$

Hence Proved

Defn

Consonant belief and plausibility measures are referred to as necessity measures and possibility measures respectively.

Let η and π denote a necessity measure and possibility measure on $P(X)$ respectively. Then

$$\eta(A\cap B) = \min\left[\eta(A),\ \eta(B)\right]$$

and $$\pi(A\cup B) = \max\left[\pi(A),\ \pi(B)\right]$$

$\forall A, B \in P(X)$

Note

$\eta(A) = 1 - \pi(\bar{A}) \ \forall A \in P(X)$

Thm 3

Every Possibility measure π on $P(X)$ can be uniquely determined by possibility distribution function $r: X \to [0,1]$ via the formula $\pi(A) = \max_{x \in A} r(x)$ — (1) for each $A \in P(X)$

Proof We prove the theorem by induction on the cardinality of the set A

Step 1 Let $|A| = 1$ then $A = \{x\}$ where $x \in X$ and eqn (1) is trivially satisfied

Step 2 Assume that eqn (1) is satisfied for $|A| = n-1$

Step 3 Let $|A| = n$ that is $A = \{x_1, x_2, \ldots x_n\}$

We know that $\pi(A \cup B) = \max[\pi(A), \pi(B)]$

$\therefore \pi(A) = \max[\pi(\{x_1, x_2, \ldots x_{n-1}\}), \pi\{x_n\}]$

$= \max[\max(\pi\{x_1\}, \pi\{x_2\}, \ldots \pi(\{x_{n-1}\})), \pi\{x_n\}]$

$= \max[\pi(\{x_1\}), \pi(\{x_2\}), \ldots \pi(\{x_n\})]$

$= \max_{x \in A} r(x)$

Defn

Consider a basic assignments m defined on the Cartesian Product

$Z = X \times Y$ (i.e.) $m: \mathcal{P}(X \times Y) \to [0,1]$

Each focal element of m is a binary relation R on $X \times Y$

Let R_x denote the projection of R on x

(i.e.) $R_x = \{x \in X \mid (x,y) \in R$ for some $y \in Y\}$

Similarly

$R_y = \{y \in Y \mid (x,y) \in R$ for some $x \in X\}$

Define the projection of R on y

The projection m_x of m on X is defined as $m_x(A) = \sum_{R: A = R_x} m(R)$ for $A \in \mathcal{P}(X)$

Similarly,

$m_y(B) = \sum_{R: B = R_y} m(R)$ $\forall B \in \mathcal{P}(Y)$

Even defines the projection m of Y

Let m_x and m_y be called marginal basic assignments and let $(\mathcal{F}_x, m_x)$ and $(\mathcal{F}_y, m_y)$ be the associated marginal bodies of evidence.

Two marginal bodies of evidence $(\mathcal{F}_x, m_x)$ and $(\mathcal{F}_y, m_y)$ are said to be non interactive iff for all $A \in \mathcal{F}_x$ and $B \in \mathcal{F}_y$

$m(A \times B) = m_x(A) \; m_y(B)$

and $m(R) = 0$ for all $R \neq A \times B$

Defn

A probability distribution p is defined on the Cartesian product $X \times Y$. it is called a joint probability distribution

Defn

Projections p_x and p_y of p on X and Y respectively are called marginal probability distributions, they defined by the formulas.

$$p_x(x) = \sum_{y \in Y} p(x,y) \text{ for each } x \in X$$

$$p_y(y) = \sum_{x \in X} p(x,y) \text{ for each } y \in Y.$$

Defn

Sets X and Y are called noninteractive w.r.t to p if

$$p(x,y) = p_x(x) \cdot p_y(y) \text{ for each } x \in X \text{ \& all } y \in Y$$

Defn Two conditional probability distributions $p_{X/Y}$ and $p_{Y/X}$ are defined in terms of a joint distribution p by the formulas.

$$p_{(X/Y)}(x/y) = \frac{p(x,y)}{p_Y(y)} \text{ and}$$

$$p_{(Y/X)}(y/x) = \frac{p(x,y)}{p_X(x)} \text{ for all } x \in X \text{ \& } y \in Y$$

Defn

Set X is called independent of Y if $p_{X/Y}(x/y) = p_X(x) \ \forall x \in X \text{ \& all } y \in Y$

Set Y is called independent of X if $p_{Y/X}(y/x) = p_Y(y) \ \forall x \in X \text{ \& all } y \in Y$

5

Uncertainty and Information

5.1 Types of Uncertainty

Upon consulting a common dictionary about the term *uncertain*, we find that the word has a broad semantic content. For example, *Webster's New Twentieth Century Dictionary* gives six clusters of meanings for the term:

1. Not certainly known; questionable; problematical.
2. Vague; not definite or determined.
3. Doubtful; not having certain knowledge; not sure.
4. Ambiguous.
5. Not steady or constant; varying.
6. Liable to change or vary; not dependable or reliable.

When we further examine these various meanings, again using the dictionary, two categories of uncertainty emerge quite naturally; they are captured quite well by the terms *vagueness* and *ambiguity*, respectively.

In general, *vagueness* is associated with the difficulty of making sharp or precise distinctions in the world; that is, some domain of interest is vague if it cannot be delimited by sharp boundaries. *Ambiguity*, on the other hand, is associated with one-to-many relations, that is, situations in which the choice between two or more alternatives is left unspecified.

Each of these two distinct forms of uncertainty—vagueness and ambiguity—is connected with a set of kindred concepts. Some of the concepts connected with vagueness are fuzziness, haziness, cloudiness, unclearness, indistinctiveness, and sharplessness; some of the concepts connected with ambiguity are nonspecificity, one-to-many relation, variety, generality, diversity, and divergence.

It is easy to see that the concept of a fuzzy set provides a basic mathematical framework for dealing with vagueness. The concept of a fuzzy measure, on the other hand, provides a general framework for dealing with ambiguity. Indeed, a fuzzy measure specifies the degrees to which an arbitrary element of the universal set X (which is a priori nonlocated) belongs to the individual crisp subsets of X. That is, the measure specifies a set of alternative subsets of X that are associated with any given element of X to various degrees according to the available evidence.

Three types of ambiguity are readily recognizable within this framework. The first is connected with the size of the subsets that are designated by a fuzzy measure as prospective locations of the element in question. The larger the subsets, the less specific the characterization. This type of ambiguity therefore has the meaning of *nonspecificity in evidence*, as expressed by the given fuzzy measure.

The second type of ambiguity is exhibited by disjoint subsets of X that are designated by the given fuzzy measure as prospective locations of the element of concern. In this case, evidence focusing on one subset conflicts with evidence focusing on the other subsets. This type of ambiguity thus has the meaning of *conflict or dissonance in evidence*.

The third type of ambiguity is associated with the number of subsets of X that are designated by a fuzzy measure as prospective locations of the element under consideration and that do not overlap or overlap only partially. The multitude of partially or totally conflicting evidence is a source of confusion. This type of ambiguity therefore characterizes *confusion in evidence*.

It follows from this preliminary discussion that fuzzy sets and fuzzy measures reflect two fundamentally different types of uncertainty—vagueness and ambiguity. Consequently, each of them constitutes a distinct framework within which appropriate measures of the corresponding type of uncertainty must be formulated. In accordance with a current terminological trend in the literature, measures of uncertainty related to vagueness are referred to in this text as *measures of fuzziness*. They are discussed in Sec. 5.2. Measures related to ambiguity are further divided into three types referred to as *measures of nonspecificity*, *measures of dissonance*, and *measures of confusion* in evidence. These are discussed in Secs. 5.4 through 5.6.

Prior to the entry of fuzzy set theory into our mathematical repertory, the only well-developed mathematical apparatus for dealing with uncertainty was *probability theory*. Although useful and successful in many applications, probability theory is, in fact, appropriate for dealing with only a very special type of uncertainty. Its limitations, some of which are discussed later in this chapter, have increasingly been recognized.

It is well known that a measure of uncertainty can also be used for measuring information. That is, the amount of uncertainty regarding some situation represents the total amount of potential information in this situation. According to this view, the reduction of uncertainty by a certain amount (due to new evidence from, for instance, the outcome of an experiment or a received message) indicates the gain of an equal amount of information.

A measure of uncertainty, when adopted as a *measure of information*, does not include semantic and pragmatic aspects of information. As such, it is not adequate for dealing with information in human communication. However, when we are dealing with structural (syntactic) aspects of systems, such a measure is not only adequate but highly desirable. It can be used for measuring the degree of constraint among variables of interest, thus comprising a powerful tool for dealing with systems problems such as systems modeling, analysis, or design.

2 MEASURES OF FUZZINESS

The question of how to measure vagueness or fuzziness has been one of the concepts associated with the development of the theory of fuzzy sets. In general, a measure of fuzziness is a function

$$f: \tilde{\mathcal{P}}(X) \to \mathbb{R},$$

where $\tilde{\mathcal{P}}(X)$ denotes the set of all fuzzy subsets of X. That is, the function f assigns a value $f(A)$ to each fuzzy subset A of X that characterizes the degree of fuzziness of A.

In order to qualify as a meaningful measure of fuzziness, f must satisfy certain axiomatic requirements. Although not necessarily unique, these requirements must fully capture the meaning of an intuitively acceptable characterization of the concept *degree of fuzziness*.

There are three axiomatic requirements that every meaningful measure of fuzziness must satisfy. Only one of them is unique; the remaining two depend on the meaning given to the concept of the degree of fuzziness.

The unique requirement states that the degree of fuzziness must be zero for all crisp sets in $\tilde{\mathcal{P}}(X)$ and only for these. Formally,

Axiom f1. $f(A) = 0$ if and only if A is a crisp set.

It is obvious that any function f violating this requirement would be totally unacceptable as a measure of fuzziness.

The second requirement is based upon a particular definition of the relation "sharper than" (i.e., "less fuzzy than") on the set $\tilde{\mathcal{P}}(X)$. If, according to a particular meaning given to the concept of the degree of fuzziness, set A is viewed (perceived) as sharper (less fuzzy) than set B, it is required that $f(A) \leq f(B)$. Formally,

Axiom f2. If $A < B$, then $f(A) \leq f(B)$.

here $A < B$ denotes that A is sharper than B. Clearly, each particular definition of the sharpness relation must properly capture the underlying conception of the degree of fuzziness.

The third requirement states that the degree of fuzziness must attain the maximum value only for a fuzzy set in $\tilde{\mathcal{P}}(X)$ that is viewed (perceived) as maximally fuzzy. Formally,

Axiom f3. $f(A)$ assumes the maximum value if and only if A is maximally fuzzy.

Of course, for each particular conception of the degree of fuzziness, the term *maximally fuzzy* attains a unique meaning.

Several measures of fuzziness have been proposed in the literature. One of them, perhaps the best known, is based on the following concepts:

1. The sharpness relation $A \prec B$ in Axiom f2 is defined by

$$\mu_A(x) \le \mu_B(x) \quad \text{for } \mu_B(x) \le \tfrac{1}{2}$$

and

$$\mu_A(x) \ge \mu_B(x) \quad \text{for } \mu_B(x) \ge \tfrac{1}{2}$$

for all $x \in X$.

2. The term maximally fuzzy in Axiom f3 is defined by the membership grade $\frac{1}{2}$ for all $x \in X$.

This measure of fuzziness is defined by the function

$$f(A) = -\sum_{x \in X} \{\mu_A(x)\log_2 \mu_A(x) + [1 - \mu_A(x)]\log_2[1 - \mu_A(x)]\}. \tag{5.1}$$

Its normalized version, $\hat{f}(A)$, for which

$$0 \le \hat{f}(A) \le 1,$$

is clearly given by

$$\hat{f}(A) = \frac{f(A)}{|X|},$$

where $|X|$ denotes the cardinality of the universal set X.

Another measure of fuzziness, referred to as an *index of fuzziness*, is defined in terms of a metric distance (Hamming or Euclidean) of A from any of the nearest crisp sets, say crisp set C, for which

$$\mu_C(x) = 0 \quad \text{if } \mu_A(x) \le \tfrac{1}{2}$$

and

$$\mu_C(x) = 1 \quad \text{if } \mu_A(x) > \tfrac{1}{2}.$$

When the Hamming distance is used, the measure of fuzziness is expressed by the function

$$f(A) = \sum_{x \in X} |\mu_A(x) - \mu_C(x)|, \tag{5.2}$$

for the Euclidean distance,

$$f(A) = \left(\sum_{x \in X} [\mu_A(x) - \mu_C(x)]^2\right)^{1/2}. \tag{5.3}$$

It is clear that other metric distances may be used as well. For example, the Minkowski class of distances yields a class of fuzzy measures

$$f_w(A) = \left(\sum_{x \in X} |\mu_A(x) - \mu_C(x)|^w\right)^{1/w}, \tag{5.4}$$

where $w \in [1, \infty]$; obviously, (5.2) and (5.3) are special cases of (5.4) for $w = 1$ and $w = 2$, respectively. It follows directly from Theorem 2.6 that

$$f_\infty(A) = \max_{x \in X} |\mu_A(x) - \mu_C(x)|. \tag{5.5}$$

It is easy to see that (5.4), for every $w \in [1, \infty]$, is based on the same conception of sharpness in Axiom f2 and maximum fuzziness in Axiom f3 as (5.1). In fact, it is known that (5.1) and (5.4) are only special cases of a larger class of measures of fuzziness, which are all based on these meanings given to requirements Axiom f2 and Axiom f3. This larger class can be expressed by the form

$$f(A) = h\left(\sum_{x \in X} g_x(\mu_A(x))\right), \tag{5.6}$$

where g_x $(x \in X)$ are functions

$$g_x : [0, 1] \to \mathbb{R}^+$$

(different, in general, for different x), which are all monotonically increasing in $[0, \frac{1}{2}]$, monotonically decreasing in $[\frac{1}{2}, 1]$, and satisfy the requirements that $g_x(0) = g_x(1) = 0$ and that $g_x(\frac{1}{2})$ be a unique maximum of g_x; h is a monotonically increasing function from $\mathbb{R}^+$ to $\mathbb{R}^+$.

For example, when

$$g_x(\mu_A(x)) = -\mu_A(x)\log_2 \mu_A(x) - [1 - \mu_A(x)]\log_2[1 - \mu_A(x)]$$

for all $x \in X$ and h is an identity function on $\mathbb{R}^+$, measure (5.1) is obtained. When, for a given $w \in [1, \infty]$,

$$g_x(\mu_A(x)) = \begin{cases} \mu_A^w(x) & \text{for } \mu_A(x) \in [0, \frac{1}{2}] \\ [1 - \mu_A(x)]^w & \text{for } \mu_A(x) \in [\frac{1}{2}, 1] \end{cases}$$

for all $x \in X$ and

$$h(a) = a^{1/w},$$

measures of the class (5.4) are obtained.

It has been more recently argued that the degree of fuzziness of a fuzzy set can be expressed, in the most natural way, in terms of the lack of distinction between the set and its complement. Indeed, it is precisely the lack of distinction between sets and their complements that distinguishes fuzzy sets from crisp sets. The less a set differs from its complement, the fuzzier it is.

The formulation of measures of fuzziness based upon this approach depends on the fuzzy complement selected. When we employ this approach within the general class of fuzzy complements c, whose properties are discussed in Sec. 2.2, the term *sharpness* in Axiom f2 is defined by

$$A \prec B \quad \text{if and only if} \quad |\mu_A(x) - c(\mu_A(x))| \geq |\mu_B(x) - c(\mu_B(x))|$$

for all $x \in X$, and the term *maximum fuzziness* in Axiom f3 means that $\mu_A(x) = e$, for all $x \in X$, provided that the complement employed has an equilibrium, e.

It has been established that a general class of measures of fuzziness based upon the lack of distinction between a set and its complement is exactly the same as the class of measures of fuzziness in which this lack of distinction is expressed in terms of a metric distance that is based on some form of aggregating the individual differences

$$|\mu_A(x) - c(\mu_A(x))| = \delta_{c,A}(x)$$

for all $x \in X$. For example, using any metric distance from the Minkowski class, we obtain for each $w \in [1, \infty]$ one particular distance

$$D_{c,w}(A, A') = \Big[\sum_{x \in X} \delta^w_{c,A}(x)\Big]^{1/w}, \tag{5.7}$$

where A' denotes the complement of A produced by function c.

For a general metric distance D, the measure of fuzziness has the form

$$f_{c,D}(A) = D_c(Z, Z') - D_c(A, A'), \tag{5.8}$$

where Z denotes any arbitrary crisp subset of X so that $D_c(Z, Z')$ is the largest possible distance in $\mathcal{P}(X)$ for a given c. Clearly,

$$0 \le f_{c,D}(A) \le D_c(Z, Z'). \tag{5.9}$$

The normalized version $\hat{f}_{c,D}$ of this measure of fuzziness is given by the formula

$$\hat{f}_{c,D}(A) = 1 - \frac{D_c(A, A')}{D_c(Z, Z')}, \tag{5.10}$$

that is,

$$0 \le \hat{f}_{c,D}(A) \le 1.$$

When the Minkowski class of metric distances, expressed by Eq. (5.7), is used, let $f_{c,w}$ denote the measure of fuzziness for the distance $D_{c,w}$. Then

$$D_{c,w}(Z, Z') = |X|^{1/w}, \tag{5.11}$$

Eq. (5.8) becomes

$$f_{c,w}(A) = |X|^{1/w} - D_{c,w}(A, A'), \tag{5.12}$$

and Eq. (5.10) becomes

$$\hat{f}_{c,w} = 1 - \frac{D_{c,w}(A, A')}{|X|^{1/w}}. \tag{5.13}$$

Example 5.1

Consider the fuzzy set A on $X = \mathbb{N}_8$ specified in Table 5.1. To illustrate the effect of the complement and distance employed, Table 5.1 shows calculations of the measure of fuzziness (regular and normalized) for two complements of the Yager class and for four distances of the Minkowski class ($w = 1, 2, 3, 5$). For each case, we first calculate the local differences $\delta_{c,A}(x)$ for all $x \in X$. Then, using Eq. (5.7), we calculate for each case the distance $D_w(A, A^c)$ between the given set A and its complement A^c, as well as the distance $D_w(Z, Z^c)$ between an arbitrary crisp set on X and its complement, by Eq. (5.11); we have $D_w(Z, Z^c) = |X|^{1/w} = 9^{1/w}$, which is independent of the complement employed. Finally, using Eqs. (5.12) and (5.10), we calculate the measure of fuzziness $f_w(A)$ and its normalized version $\hat{f}_w(A)$, respectively.

Observe that the measure of fuzziness (regular as well as normalized) decreases with increasing w for both of the complements. The rate of decrease, however, is smaller for the classical fuzzy complement.

Measures of fuzziness defined in terms of different distance functions are based upon different measurement units. Although the choice of a unit is not a critical issue, it is often desirable to use a unit that is intuitively appealing in the sense that it has a simple interpretation in terms of some significant canonical situation. For instance, it seems intuitively pleasing for one unit of fuzziness to indicate that the membership in the fuzzy set of one element of the universal set is maximally uncertain. A measure of fuzziness equal to two of these units would therefore indicate that we cannot determine set membership or nonmembership at all for two elements of the universal set. Ultimately, the maximally fuzzy set is one for which we cannot make this determination for any of the elements of the universal set; the fuzziness in this case is equal to the cardinality of the (finite) universal set. To define this unit formally, we require that the degree of fuzziness is 1 for every set A (defined on a finite universal set X) for which $\mu_A(x)$ is equal to the equilibrium of the complement employed for one particular $x \in X$ and that

TABLE 5.1 THE EFFECT OF THE COMPLEMENT AND DISTANCE (VALUE OF w) EMPLOYED ON THE MEASURE OF FUZZINESS EXPRESSED BY EQS. (5.8) AND (5.10).

		$c(a) = 1 - a$				$c(a) = (1 - a^2)^{1/2}$			
x	$\mu_A(x)$	$\delta^1_A(x)$	$\delta^2_A(x)$	$\delta^3_A(x)$	$\delta^4_A(x)$	$\delta^1_A(x)$	$\delta^2_A(x)$	$\delta^3_A(x)$	$\delta^4_A(x)$
1	[illegible]	.8	.56	.27	.36	.79	.67	.85	.29
2	[illegible]	.8	.0	.6	.01	.37	.54	.08	.03
3	[illegible]	.4	.16	.06	.16	.01	.0	.1	.8
4	[illegible]	.6	.35	.22	.08	.29	.81	.9	.3
5	[illegible]	.8	.64	.54	.11	.46	.25	.89	.02
6	[illegible]	1	1	1	1	1	1	1	1
7	[illegible]	.2	.08	.03	0	.20	.98	.65	.0
8	[illegible]	.2	.08	.01	.3	.52	.27	.54	.04
9	[illegible]	.8	.63	.55	.31	.89	.79	.76	.56
		$D_1(A, A^c)$ 4.6	$D_2(A, A^c)$ 1.8	$D_3(A, A^c)$ 1.35	$D_4(A, A^c)$ 1.0	$D_1(A, A^c)$ 3.13	$D_2(A, A^c)$ 1.76	$D_3(A, A^c)$ 1.28	$D_4(A, A^c)$ 1.11
		$\lvert X\rvert$ 9.0	$\lvert X\rvert^{1/2}$ 3.0	$\lvert X\rvert^{1/3}$ 2.08	$\lvert X\rvert^{1/4}$ 1.73	$\lvert X\rvert$ 9.0	$\lvert X\rvert^{1/2}$ 3.0	$\lvert X\rvert^{1/3}$ 2.08	$\lvert X\rvert^{1/4}$ 1.73
		$f_1(A)$ 4.4	$f_2(A)$ 1.2	$f_3(A)$ [illegible]	$f_4(A)$ [illegible]	$f_1(A)$ 4.57	$f_2(A)$ 1.24	$f_3(A)$ .78	$f_4(A)$ .43
		$\hat{f}_1(A)$ .49	$\hat{f}_2(A)$ .4	$\hat{f}_3(A)$ .35	$\hat{f}_4(A)$ .32	$\hat{f}_1(A)$ .51	$\hat{f}_2(A)$ .41	$\hat{f}_3(A)$ .38	$\hat{f}_4(A)$ .36

$\mu_A(x) \in \{0, 1\}$ for all $x \in X$ that is set different from c. (Note that the condition on c is the membership value which is the same in a fuzzy set and its complement; it therefore indicates maximum uncertainty concerning set membership.) Then

$$\delta_{c,A}(x) = \begin{cases} c & \text{when } x = c \\ 1 & \text{otherwise} \end{cases}$$

and our requirement leads to the equation

$$f_{c,w}(A) = |X| - (|X| \cdot 1)^{1/w} = 0.$$

This equation is satisfied only for $w = 1$, thus implying that our requirement is satisfied only by the Hamming distance.

This unit of fuzziness characterizes the maximum fuzziness or uncertainty regarding the truth value of a single proposition, namely, the proposition that an element x belongs to fuzzy set A. Since truth values of propositions are often denoted by binary digits (0 and 1), the unit that characterizes full uncertainty represented in our case by fuzziness for a single proposition is usually called a bit, which is an abbreviation for binary digit. The same reasoning is used to develop the notion of a bit in the context of the Shannon entropy discussed in Sec. 5.5, as well as any other measure of uncertainty.

When we accept bits as units of fuzziness, Eqs. (5.12) and (5.13) become more specific:

$$f(A) = |X| - \sum_{x \in X} |\mu_A(x) - c(\mu_A(x))| \quad \text{(in bits)} \tag{5.14}$$

$$\hat{f}(A) = \frac{f(A)}{|X|} \tag{5.15}$$

The two values of $f_{c,1}$ in Table 5.1 provide us with the degrees of fuzziness, as measured in bits, of the set A for the two complements.

The previous definitions of measures of fuzziness are based on the assumption that the universal set X is finite. These definitions can be readily extended to infinite sets. Consider, for example, that $X = [a, b]$. Then, Eq. (5.7) becomes

$$D_{c,w}(A, A^c) = \left(\int_a^b \delta^w_{c,A}(x)\, dx \right)^{1/w} \tag{5.16}$$

Since

$$D_w(Z, Z^c) = \left(\int_a^b dx \right)^{1/w} = (b - a)^{1/w}, \tag{5.17}$$

we obtain

$$f_{c,w}(A) = (b - a)^{1/w} - \left(\int_a^b \delta^w_{c,A}(x)\, dx \right)^{1/w} \tag{5.18}$$

[illegible]

$$[illegible]$$

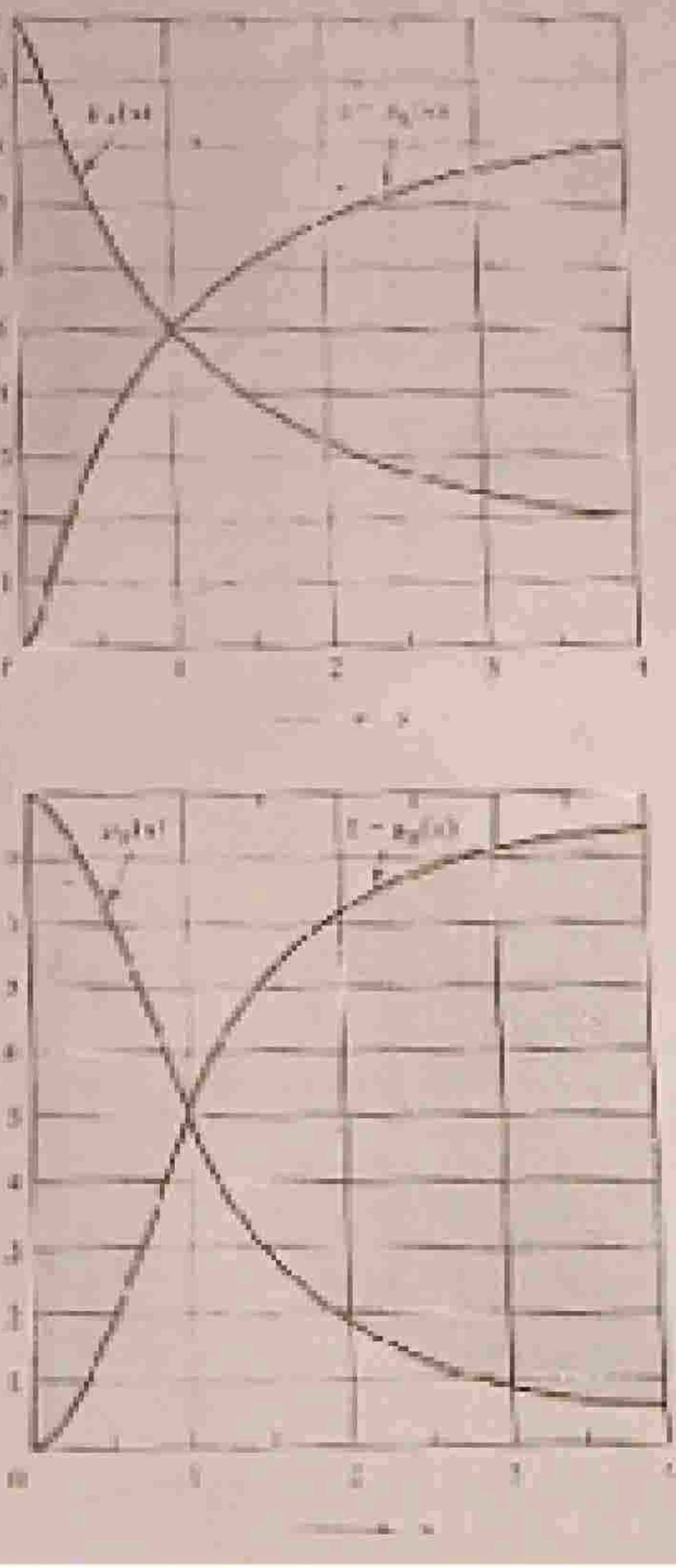


Figure 5.1. Comparison of the [illegible] of [illegible] (Example [illegible])

Example 5.2

Consider two fuzzy sets, A and B, defined on the set of real numbers $[0, 4]$ with the membership grade functions

$$\mu_A(x) = \frac{1}{1+x}$$

and

$$\mu_B(x) = \frac{1}{1+x^2}.$$

Plots of these functions and their standard classical complements are shown in Fig. 5.1. By inspecting these plots, we observe that set A is less distinct from its complement than is set B. To calculate the actual degrees of fuzziness of these sets, we use the Hamming distance ($w = 1$). Then,

$$\delta_A(x) = |2\mu_A(x) - 1| = \left|\frac{2}{1+x} - 1\right|$$

and

$$D_1(A, A') = \int_0^4 \left|\frac{2}{1+x} - 1\right| dx$$

$$= \int_0^1 \left(\frac{2}{1+x} - 1\right) dx + \int_1^4 \left(1 - \frac{2}{1+x}\right) dx$$

$$= [2\ln(1+x) - x]_0^1 + [x - 2\ln(1+x)]_1^4$$

$$= 1.55$$

Now applying Eq. (5.18), we obtain

$$f_{c_1}(A) = 4 - 1.55 = 2.45;$$

after normalization, expressed by Eq. (5.19), we have

$$f_{c_1}(A) = \frac{2.45}{4} = .61.$$

The following is the same calculation for the set B:

$$\delta_B(x) = |2\mu_B(x) - 1| = \left|\frac{2}{1+x^2} - 1\right|,$$

$$D_1(B, B') = \int_0^4 \left|\frac{2}{1+x^2} - 1\right| dx$$

$$= \int_0^1 \left(\frac{2}{1+x^2} - 1\right) dx + \int_1^4 \left(1 - \frac{2}{1+x^2}\right) dx$$

$$= [2\tan^{-1}x - x]_0^1 + [x - 2\tan^{-1}x]_1^4$$

$$= 4\tan^{-1}1 - 2\tan^{-1}4 + 2 = 2.49,$$

$$f_{c_1}(B) = 4 - 2.49 = 1.51,$$

$$f_{c_1}(B) = \frac{1.51}{4} = .38.$$

Thus, as we anticipated, the degree of fuzziness of the set B is smaller than that of the set A. The difference between the two degrees of fuzziness is .95 bits.

5.3 CLASSICAL MEASURES OF UNCERTAINTY

Prior to the theory of fuzzy sets, two principal measures of uncertainty were recognized. One of them, proposed by Hartley [1928], is based solely on the classical set theory. The other, introduced by Shannon [1948], is formulated in terms of probability theory. Both of these measures pertain to some aspects of ambiguity, as opposed to vagueness or fuzziness. As shown later in this chapter, however, each measures a different aspect of ambiguity: Hartley's measure pertains to nonspecificity, Shannon's measure to conflict or discordance in evidence.

Both Hartley and Shannon introduced their measures for the purpose of measuring information in terms of uncertainty. Therefore, these measures are often referred to as measures of information. It has been more common, however, to refer to the measure invented by Shannon as the Shannon *entropy*. The name entropy was suggested by Shannon himself, presumably because of a similarity in the mathematical form between his measure and that of physical entropy as defined in certain formulations of statistical mechanics.

In this section, we introduce these classical measures of uncertainty and information and review their most fundamental properties. We refer to them by names that are predominant in the literature: the *Hartley information* and *Shannon entropy*.

Hartley Information

Consider a finite set X of n elements that are viewed as symbols that convey certain meanings to the parties involved in a certain context. For example, the set may consist of all possible measurements of a physical variable (defined in a specific way and subject to a specific accuracy), all possible states of an investigated system, all possible primitive messages, and the like.

Sequences can be formed from elements of set X by successive selections. These are determined, for example, by the variable measured, by the dynamic properties of the system under investigation, or by the sender of a message. At each selection, all possible elements that might have been chosen are eliminated except one. Similarly, as the selection proceeds, all possible sequences of elements of each particular length are eliminated except one.

Since all elements of X represent possible alternatives prior to a selection, we experience ambiguity, the amount of which is proportional to the number of alternatives. This ambiguity is totally resolved when one of the alternatives is selected. The amount of information conveyed (to the experimenter, observer, message receiver, and the like) can thus be meaningfully defined as the amount of ambiguity eliminated by the selection.

The number of all possible sequences of s selections (each from set X) is n^s, where $n = |X|$. The amount of information $I(n^s)$ associated with s selections from X should be proportional to s; that is,

$$I(n^s) = K(n)\,s, \tag{5.20}$$

where $K(n)$ is a constant, for set with n, that 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 sets X_1 and X_2, respectively, are such that they yield the same number of sequences, then the amount of information associated with these sequences should also be the same. Formally, when

$$n_1^{s_1} = n_2^{s_2} \tag{5.21}$$

then

$$K(n_1)\,s_1 = K(n_2)\,s_2. \tag{5.22}$$

From Eqs. (5.21) and (5.22) we obtain

$$\frac{s_2}{s_1} = \frac{\log_b n_1}{\log_b n_2}$$

and

$$\frac{s_2}{s_1} = \frac{K(n_1)}{K(n_2)},$$

respectively. Hence,

$$\frac{\log_b n_1}{\log_b n_2} = \frac{K(n_1)}{K(n_2)}.$$

This equation 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 a set with n elements is thus given by the formula

$$I(n^s) = K_0\, s \log_b n. \tag{5.23}$$

By making a particular choice of values for K_0 and b in this formula, we define a unit by which information is measured. When we choose $K_0 = 1$ and $b = 2$, the information is measured in bits.* Then,

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

or

$$I(N) = \log_2 N, \tag{5.24}$$

where N denotes the total number of alternatives regardless of whether they are realized by one selection from a set or by a sequence of selections. One bit of information is obtained when one of two possible alternatives ($N = 2$) is eliminated. This is equivalent to learning the truth value of a single proposition, which was not known prior to the observation or receipt of the message. Because of from X should be proportional to s; that is,

$$I(n^s) = K(n)\,s, \tag{5.20}$$

where $K(n)$ is a constant of proportionality that 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 sets X_1 and X_2, respectively, are such that they yield the same number of sequences, then the amount of information associated with these sequences should also be the same. Formally, when

$$n_1^{s_1} = n_2^{s_2} \tag{5.21}$$

then

$$K(n_1)s_1 = K(n_2)s_2. \tag{5.22}$$

From Eqs. (5.21) and (5.22) we obtain

$$\frac{s_1}{s_2} = \frac{\log_b n_2}{\log_b n_1}$$

and

$$\frac{s_1}{s_2} = \frac{K(n_2)}{K(n_1)},$$

respectively. Hence,

$$\frac{\log_b n_2}{\log_b n_1} = \frac{K(n_2)}{K(n_1)}.$$

This equation 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 a set with n elements is thus given by the formula

$$I(n^s) = K_0 s \log_b n. \tag{5.23}$$

By making a particular choice of values for K_0 and b in this formula, we define a unit by which information is measured. When we choose $K_0 = 1$ and $b = 2$, the information is measured in bits.[*] Then,

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

or

$$I(N) = \log_2 N, \tag{5.24}$$

where N denotes the total number of alternatives regardless of whether they are realized by one selection from a set or by a sequence of selections. One bit of information is obtained when one of two possible alternatives ($N = 2$) is designated. This is equivalent to ascertaining the truth value of a single proposition, which was not known prior to the observation or receipt of the message. Because of this intuitively appealing interpretation of the amount of information measured in bits, Hartley information is usually presented in the form (5.24).

For each $N \in \mathbb{N}$, the value of $I(N)$ can also be viewed as the amount of information needed to characterize one of N alternatives. When measured in bits, $I(N)$ expresses the number of single propositions whose truth values must be determined in order to characterize one of the alternatives.

Hartley information can also be characterized by the following axioms:

Axiom I1 (*additivity*). $I(N \cdot M) = I(N) + I(M)$ for all $N, M \in \mathbb{N}$.

Axiom I2 (*monotonicity*). $I(N) \leq I(N + 1)$ for all $N \in \mathbb{N}$.

Axiom I3 (*normalization*). $I(2) = 1$.

Axiom I1 involves a set with $N \cdot M$ elements, which can be partitioned into N subsets each with M elements. A characterization of an element from the full set requires the amount $I(N \cdot M)$ of information. However, we can also proceed in two steps to characterize the elements by taking advantage of the partition of the set. First, we characterize the subset to which the element belongs; the required amount of information is $I(N)$. Then, we characterize the element within the subset; here the required amount of information is $I(M)$. These two amounts of information completely characterize an element of the full set and, hence, their sum should be equal to $I(N \cdot M)$. This is exactly what the axiom requires.

Axiom I2 represents an essential and rather obvious requirement: the larger the number of alternatives, the more information is gained by eliminating all of them except one. Axiom I3 is needed only to define the unit of information. In our case, the defined unit is the bit.

As expressed by the following uniqueness theorem, the Hartley information (5.24) is the only function that satisfies these axioms.

Theorem 5.1. Function $I(N) = \log_2 N$ is the only function that satisfies Axioms I1 through I3.

Proof: Let N be an integer greater than 2. For every integer i, define the integer $q(i)$ such that

$$2^{q(i)} \leq N^i < 2^{q(i)+1} \tag{5.25}$$

These inequalities can be written as

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

When we divide the inequalities by i and replace $\log_2 2$ with 1, we obtain

$$\frac{q(i)}{i} \leq \log_2 N < \frac{q(i) + 1}{i} \tag{5.26}$$

and, consequently,

$$\lim_{i \to \infty} \frac{q(i)}{i} = \log_2 N. \tag{5.27}$$

Let I denote a function that satisfies Axioms I1 through I3. Then, by Axiom I2,

$$I(a) \leq I(b) \tag{5.28}$$

for $a < b$. Combining (5.28) and (5.25), we obtain

$$I(2^{q(l)}) \leq I(N^l) \leq I(2^{q(l)+1}). \tag{5.29}$$

By Axiom I1, we have

$$I(a^k) = kI(a). \tag{5.30}$$

Hence,

$$I(N^l) = lI(N), \tag{5.31}$$

$$I(2^{q(l)}) = q(l)I(2),$$

and

$$I(2^{q(l)+1}) = (q(l) + 1)I(2).$$

Since $I(2) = 1$ by Axiom I3, we can rewrite the last two equations as

$$I(2^{q(l)}) = q(l)$$

and

$$I(2^{q(l)+1}) = q(l) + 1.$$

Applying these two equations and (5.31) to (5.29), we obtain

$$q(l) \leq lI(N) \leq q(l) + 1$$

and, consequently,

$$\lim_{l \to \infty} \frac{q(l)}{l} = I(N). \tag{5.32}$$

Comparing (5.32) with (5.27), we conclude that $I(N) = \log_2 N$ for $N \geq 2$. Since $\log_2 2 = 1$ and $\log_2 1 = 0$, function $\log_2 N$ clearly satisfies the axioms for $N = 1, 2$ as well. This concludes the proof. ■

Consider now two sets X and Y that are interrelated in the sense that selections from one of the sets are constrained by selections from the other. Assume that the constraint is expressed by a relation $R \subset X \times Y$. Then, three types of Hartley information can be defined on these sets:

- *Simple information*

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

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

- *Joint information*

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

* *Conditional information*

$$H(X \mid Y) = \log_2 \frac{|R|}{|Y|} = \log_2 |R| - \log_2 |Y|$$

$$H(Y \mid X) = \log_2 \frac{|R|}{|X|} = \log_2 |R| - \log_2 |X|$$

Observe that $|R|/|Y|$ in $H(X \mid Y)$ represents the average number of elements of X that can be selected under the condition that an element of Y has already been selected. Similarly, $|R|/|X|$ in $H(Y \mid X)$ characterizes the average number of elements of Y that can be selected provided that an element of X has been selected. Observe also that

$$H(X \mid Y) = H(X, Y) - H(Y) \tag{5.10}$$

and

$$H(Y \mid X) = H(X, Y) - H(X). \tag{5.11}$$

If selections from X do not depend on selections from Y, then sets X and Y are called *noninteractive*. Then, $R = X \times Y$ and we can readily obtain the following equations:

$$H(X, Y) = \log_2 |X \times Y| = \log_2 (|X| \cdot |Y|)$$
$$= \log_2 |X| + \log_2 |Y| = H(X) + H(Y),$$
$$H(X \mid Y) = H(X),$$
$$H(Y \mid X) = H(Y).$$

The following symmetric function, which is usually referred to as *information transmission*, is a useful indicator of the strength of constraint between sets X and Y:

$$T(X, Y) = H(X) + H(Y) - H(X, Y). \tag{5.12}$$

When the sets are noninteractive, we have $T(X, Y) = 0$; otherwise, $T(X, Y) > 0$.

Information transmission can be generalized to express the constraint among more than two sets. It is always expressed as the difference between the total information based on the individual sets and the joint information. Formally,

$$T(X_1, X_2, \ldots, X_n) = \sum_{i=1}^{n} H(X_i) - H(X_1, X_2, \ldots, X_n). \tag{5.13}$$

Example 5.3

Consider two variables x and y whose values are taken from sets $X = \{low, medium, high\}$ and $Y = \{1, 2, 3, 4\}$, respectively. It is known that the variables are constrained by the relation R expressed by the matrix

	1	2	3	4
Low	1	1	1	1
Medium	1	0	1	0
High	0	1	0	0

We can see that the low value of x does not constrain y at all, the medium value of x constrains y partially, and the high value constrains it totally. The following types of Hartley information can be calculated in this example:

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

$$I(Y) = \log_2 |Y| = \log_2 4 = 2,$$

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

$$I(X \mid Y) = I(X, Y) - I(Y) = 2.8 - 2 = .8,$$

$$I(Y \mid X) = I(X, Y) - I(X) = 2.8 - 1.6 = 1.2,$$

$$T(X, Y) = I(X) + I(Y) - I(X, Y) = 1.6 + 2 - 2.8 = .8.$$

The Hartley information is based on uncertainty associated with a choice among a certain number of alternatives. The larger the number of alternatives, the larger the uncertainty and, consequently, the more information is conveyed by the uncertainty. The degree of uncertainty in this case can be viewed, most naturally, as the degree of nonspecificity. Indeed, the fewer alternatives we have, the more specific is our choice, and vice versa.

The Hartley information can thus be viewed as a simple *measure of nonspecificity* based solely on classical set theory. As discussed in Sec. 5.6, it is a special case of more general nonspecificity measures based upon fuzzy set theory.

Shannon Entropy

The Shannon entropy, which is a measure of uncertainty and information formulated in terms of probability theory, is expressed by the function

$$H(p(x) \mid x \in X) = -\sum_{x \in X} p(x) \log_2 p(x), \tag{5.37}$$

where $(p(x) \mid x \in X)$ is a probability distribution on a finite set X. It is thus a function of the form

$$H : \mathcal{P} \to [0, \infty),$$

where $\mathcal{P}$ denotes the set of all probability distributions on finite sets.

Shannon entropy was considered for many years to be the only feasible basis for information theory. It has certainly dominated the literature on information theory since it was proposed by Shannon in 1948. Hartley information, which is in fact a predecessor of Shannon entropy, is rarely mentioned in the current literature. When it is mentioned, it is almost always given one of two probabilistic interpretations. In the first, it is viewed as a measure that only distinguishes between zero and nonzero probabilities in the given probability distribution, that is, a measure that is totally insensitive to the actual values of the probabilities. It is derived from Shannon entropy by replacing any nonzero probability in the probability distribution with one.

The second probabilistic interpretation views Hartley information as equivalent to Shannon entropy under the assumption that all elements of the set X are equally probable. In this case, the equal probabilities are $1/|X|$. When we substitute them for $p(x)$ in formula (5.37), we readily obtain the Hartley information, $\log_2 |X|$.

Although the probabilistic interpretations try to subsume the Hartley information under the framework of information theory based on the Shannon entropy, such attempts are ill conceived. This should be obvious from the axiomatic treatment of the Hartley information earlier in this section, which is totally independent of any probabilistic assumptions. In fact, the Shannon entropy and Hartley information measure quite different aspects of uncertainty and information, as discussed in Secs. 5.4 through 5.6.

Why is the Shannon entropy significant as a measure of uncertainty and information? First, let us justify it on simple intuitive grounds. Suppose a particular element x of our universal set X occurs with the probability $p(x)$. When the probability of x is very high, say $p(x) = .99$, then the actual occurrence of x is taken almost for granted and, consequently, its occurrence does not surprise us very much. That is, our uncertainty in anticipating x is quite small and, therefore, our observation that x has actually occurred contains very little information content. When the probability is very small, on the other hand, say $p(x) = .01$, then we are greatly surprised by the occurrence of x. This means that we are highly uncertain in our anticipation of x and, hence, the observation of x has a very large information content. The information content of observing x, expressed by our anticipatory uncertainty prior to the observation, should therefore be characterized by a decreasing function of the probability $p(x)$: the more likely the occurrence of x, the less information the observation of x actually contains.

Let u denote a function that for each $x \in X$ with probability $p(x)$ characterizes the anticipatory uncertainty of x. Since $p(x) \in [0, 1]$, we have

$$u : [0, 1] \rightarrow [0, \infty),$$

where

$$u(a) > u(b) \quad \text{for } a < b.$$

In addition, function u should be additive with respect to joint observations of elements from two sets that are independent in the probabilistic sense. That is, for each $x \in X$ and each $y \in Y$, if

$$p(x, y) = p(x) \cdot p(y),$$

then u should satisfy the equation

$$u(p(x) \cdot p(y)) = u(p(x)) + u(p(y)).$$

This functional equation is known as the Cauchy equation. Its solution is

$$u(a) = K \log_b a,$$

where K is a constant ($K \in \mathbb{R}$). Since u is required to be a decreasing function on $[0, 1]$ and the logarithmic function is increasing, K must be negative. When

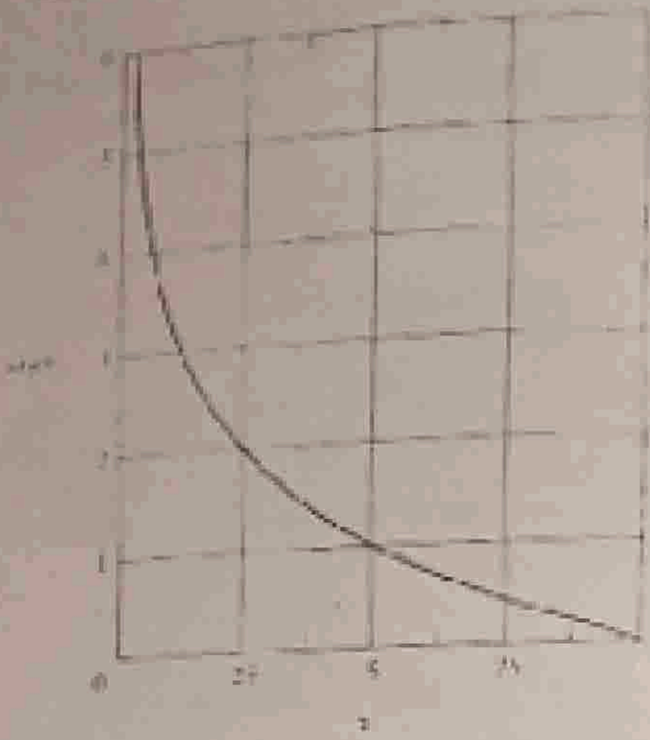


Figure 5.2 Graph of function u(a) = −log₂ a.

we take $b = 2$ and add a normalization requirement that $u(\frac{1}{2}) = 1$ [illegible] the anticipatory uncertainty in bits), we obtain $k = 1$ and

$$u(a) = -\log_2 a.$$

A graph of this function is shown in Fig. 5.2.

Consider now a set X of alternatives with probabilities $p(x)$ for each [illegible] It is certain that exactly one of them must occur under some relevant [illegible] for example, exactly one must be observed as an outcome of an experim[illegible] be received as a message. When we know that x occurs (by actually obse[illegible] as the outcome of our experiment or by receiving it as a message), the inform[illegible] content of this fact is $-\log_2 p(x)$ bits. Prior to the occurrence of the outcom[illegible] this information content is not known. However, we can calculate the expe[illegible] information content as the weighted arithmetic mean

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

which is exactly the Shannon entropy.

When only two alternatives whose probabilities are a and $1 - a$ are availa[illegible] the expected information $H(a, 1 - a)$ depends on a as illustrated in Fig. 5.[illegible] graphs of its components $-a \log_2 a$ and $-(1 - a) \log_2 (1 - a)$ are shown in Fi[illegible] 5.3(b).

Probabilistic measures of uncertainty and information have also been tre[illegible] axiomatically in various ways. To illustrate this more rigorous treatment, [illegible] that

$$X = \{x_1, x_2, \ldots, x_n\},$$

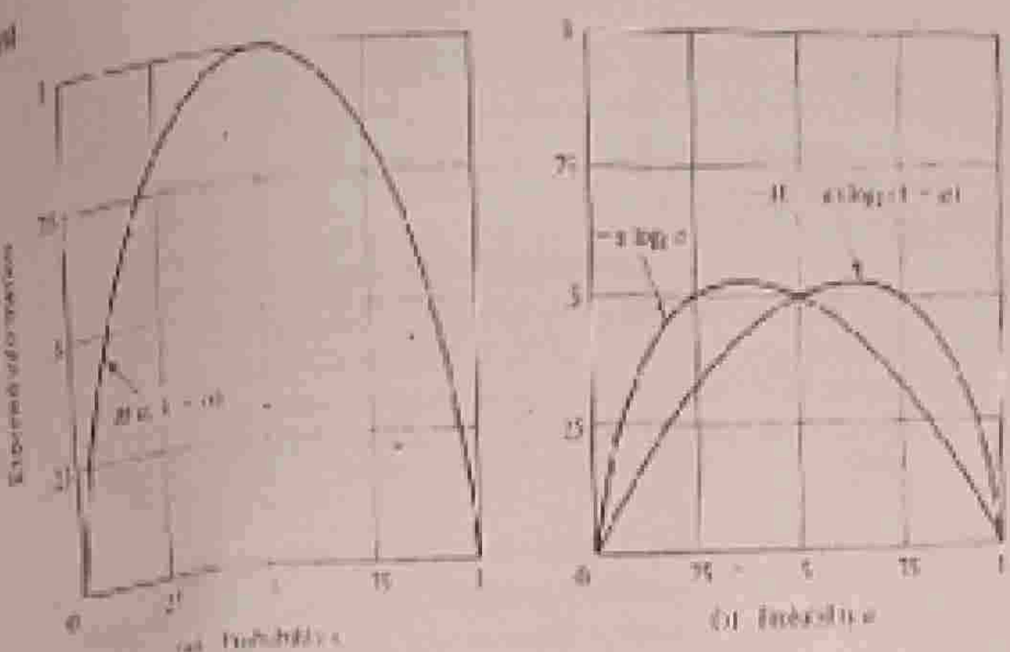


Figure 5.3 Graphs of $H(a, 1-a)$ and its components $-a \log_2 a$ and $-(1-a)\log_2(1-a)$.

and let p_i denote the probability of $x_i \in X$ for all $i \in \mathbb{N}_n$. In addition, let

$${}^n\mathcal{P} = \left\{(p_1, p_2, \ldots, p_n) \mid p_i \geq 0 \text{ for all } i \in \mathbb{N}_n \text{ and } \sum_{i=1}^{n} p_i = 1\right\}$$

denote the set of all probability distributions with n components and let

$$\mathcal{P} = \bigcup_{n \in \mathbb{N}} {}^n\mathcal{P}.$$

Then, every measure of uncertainty based on probability distributions is characterized by a function

$$H : \mathcal{P} \to [0, \infty)$$

that satisfies some requirements considered desirable for such a measure.

Different subsets of the following requirements, which are universally considered essential for a probabilistic measure of uncertainty and information, are usually taken as axioms of probabilistic information theory:

(H1) Expansibility—when a component with zero probability is added to a probability distribution, the uncertainty should not change; formally,

$$H(p_1, p_2, \ldots, p_n) = H(p_1, p_2, \ldots, p_n, 0)$$

for all $(p_1, p_2, \ldots, p_n) \in \mathcal{P}$.

(H2) Symmetry—the uncertainty should be invariant with respect to permutations of probabilities of a given probability distribution; formally,

$$H(p_1, p_2, \ldots, p_n) = H(\pi(p_1, p_2, \ldots, p_n))$$

for all $n \in \mathbb{N}$, this requirement can be expressed in a simpler way as

$$f(nm) = f(n) + f(m) \tag{5.40}$$

for all $n, m \in \mathbb{N}$. Let this requirement be called a *weak additivity*.

(H7) Monotonicity—for probability distributions with equal probabilities $1/n$ $(n \in \mathbb{N})$, the uncertainty should increase with increasing n. Formally, the function

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

should be a monotonically increasing function of n; that is,

$$\text{if } n < m, \text{ then } f(n) < f(m)$$

for all $n, m \in \mathbb{N}$.

(H8) Branching—given a probability distribution on a universal set X, we should be able to measure the uncertainty associated with it either directly or indirectly by adding uncertainties involved in a two-stage measurement process: in the first stage, we measure the uncertainty of two disjoint subsets of X that form a partition of X; in the second stage, we measure the uncertainty associated with the conditional probability distributions on these subsets and weight each of them by the probability of the subset. To formalize this requirement, let $A = \{x_1, x_2, \ldots, x_s\}$ and $B = \{x_{s+1}, x_{s+2}, \ldots, x_n\}$ be two disjoint subsets of the universal set $X = \{x_1, x_2, \ldots, x_n\}$. Furthermore, given a probability distribution $(p_1, p_2, \ldots, p_n)$ on X, where $p_i = p(x_i)$ for all $i \in \mathbb{N}_n$, let

$$p_A = \sum_{i=1}^{s} p_i \quad \text{and} \quad p_B = \sum_{i=s+1}^{n} p_i$$

denote the probabilities of A and B, respectively. Then, the branching requirement means that

$$H(p_1, p_2, \ldots, p_n) = H(p_A, p_B) + p_A H\left(\frac{p_1}{p_A}, \frac{p_2}{p_A}, \ldots, \frac{p_s}{p_A}\right) + p_B H\left(\frac{p_{s+1}}{p_B}, \frac{p_{s+2}}{p_B}, \ldots, \frac{p_n}{p_B}\right) \tag{5.41}$$

should be satisfied for any probability distribution $(p_1, p_2, \ldots, p_n) \in \mathcal{P}$ and any subsets A and B of X that form a partition on X. This requirement, which is also called a *grouping requirement*, is sometimes presented in various alternative forms. For example, one of its weaker forms is described by the equation

$$H(p_1, p_2, p_3) = H(p_1 + p_2, p_3) + (p_1 + p_2) H\left(\frac{p_1}{p_1 + p_2}, \frac{p_2}{p_1 + p_2}\right).$$

It matters little which of these forms is adopted since they can be derived from each other.

(H9) Normalization—To ensure (if desirable) that the uncertainty is measured in bits, we require that

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

The listed requirements for probabilistic measures of uncertainty are extensively discussed and well justified in the plentiful literature on classical information theory. The following subsets of these requirements are the best known examples of axiomatic characterizations of the probabilistic measure of uncertainty:

1. Continuity, weak additivity, monotonicity, branching, and normalization.
2. Expansibility, continuity, maximum, branching, and normalization.
3. Symmetry, continuity, branching, and normalization.
4. Expansibility, symmetry, continuity, subadditivity, additivity, and normalization.

Any of these collections of requirements (as well as some additional collections) is, when taken as a set of axioms, sufficient to characterize the Shannon entropy uniquely. That is, it has been proven that the Shannon entropy is the only function that satisfies any of these sets of axioms. To illustrate in detail the important issue of uniqueness, which gives the Shannon entropy its great significance, we present the uniqueness proof for the first of the listed sets of axioms, that is, for continuity, weak additivity, monotonicity, branching, and normalization. Since the proof is rather lengthy, it is placed in Appendix A.1.

The literature dealing with information theory based on the Shannon entropy is extensive. We do not attempt to give a comprehensive coverage of the theory in this book. In the rest of this section, however, we briefly overview the most fundamental properties of the Shannon entropy. In addition, some notes at the end of this chapter provide the reader with key literature resources for a deeper study of various aspects of this classical information theory.

First, let us show that

$$0 \le H(p_1, p_2, \ldots, p_n) \le \log_2 n. \tag{5.42}$$

To derive the lower bound, we observe that $-p_i \log_2 p_i \ge 0$ for all $p_i \in (0, 1]$. For $p_i = 0$, the function $-p_i \log_2 p_i$ is not defined. However, employing l'Hospital's rule for indeterminate forms, we can calculate its limit for $p_i \to 0$:

$$\lim_{p_i \to 0} -p_i \log_2 p_i = \lim_{p_i \to 0} \frac{-\log_2 p_i}{\frac{1}{p_i}} = \lim_{p_i \to 0} \frac{\frac{-1}{p_i \ln 2}}{\frac{-1}{p_i^2}} = \lim_{p_i \to 0} \frac{p_i}{\ln 2} = 0$$

Clearly, the lower bound of (5.42) is obtained only when $p_i = 1$ for some particular i and $p_j = 0$ for all $j \neq i$ ($i, j \in \mathbb{N}_n$); this probability distribution indeed represents no uncertainty.

To derive the upper bound of (5.42), we must determine the maximum of function H for all probability distributions in ${}^n\mathcal{P}$. In this case, H is a function of n variables, but one of these is dependent on the other variables due to the requirement that the probabilities must add to one. Without any loss of generality, let p_n be the dependent variable. Then,

$$p_n = 1 - (p_1 + p_2 + \cdots + p_{n-1})$$

and the necessary conditions for an extremal value of H are

$$\frac{\partial H}{\partial p_i} = 0 \quad \text{for all } i \in \mathbb{N}_{n-1}.$$

The partial derivatives are

$$\frac{\partial H}{\partial p_i} = \sum_{k=1}^{n} \frac{\partial H}{\partial p_k}\frac{\partial p_k}{\partial p_i} = \frac{\partial(-p_i \log_2 p_i)}{\partial p_i} + \frac{\partial(-p_n \log_2 p_n)}{\partial p_n}\frac{\partial p_n}{\partial p_i}$$

since

$$\frac{\partial H}{\partial p_k}\frac{\partial p_k}{\partial p_i} = 0 \quad \text{for all } k \neq i, n.$$

Clearly,

$$\frac{\partial p_n}{\partial p_i} = \frac{\partial[1 - (p_1 + p_2 + \cdots + p_i + \cdots + p_{n-1})]}{\partial p_i} = -1$$

and, consequently,

$$\frac{\partial H}{\partial p_i} = -\log_2 p_i - \frac{1}{\ln 2} + \log_2 p_n + \frac{1}{\ln 2}$$

$$= -\log_2 p_i + \log_2 p_n.$$

By setting the derivatives to zero, we obtain

$$p_i = p_n \quad \text{for all } i \in \mathbb{N}_{n-1}.$$

Hence, an extremal value of H exists for the distribution with equal probabilities. Since

$$\left.\frac{\partial^2 H}{\partial p_i^2}\right|_{p_1 = \cdots = p_n = 1/n} = \frac{-2n}{\ln 2} < 0$$

for all $i \in \mathbb{N}_{n-1}$,

$$H\left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right) = \log_2 n$$

is the maximum of H in ${}^n\mathcal{P}$.

In some applications, it is desirable to define the ratio

$$\hat{H}(p_1, p_2, \ldots, p_n) = \frac{H(p_1, p_2, \ldots, p_n)}{\log_2 n}$$

of the actual entropy and its upper bound, which is called a *normalized Shannon entropy*. Clearly,

$$0 \le \hat{H}(p_1, p_2, \ldots, p_n) \le 1.$$

Before examining additional properties of the Shannon entropy, let us illustrate the meaning of its branching property, which is employed as one of the axioms in our uniqueness proof (Appendix A.1). In principle, the branching property allows us to calculate uncertainty either directly, in terms of the probability distribution on the universal set, or in stages, using probability distributions on various subsets of the universal set. These alternative ways of calculating uncertainty can be best illustrated by an example.

Example 5.6

Let the set $X = \{x_1, x_2, x_3, x_4\}$ with the probability distribution

$$p = (p_1 = .25, p_2 = .5, p_3 = .125, p_4 = .125)$$

be given, where p_i denotes the probability of x_i for all $i \in N_4$. Consider the four branching schemes specified in Fig. 5.4 for calculating the uncertainty of this probability distribution. Employing the branching property of Shannon entropy, the resulting uncertainty should be the same regardless of which of the branching schemes we use. Let us perform and compare the four schemes of calculating the uncertainty.

Scheme I: According to this scheme, we calculate the uncertainty directly: $H(p) = -.25 \log_2 .25 - .5 \log_2 .5 - 2 \times .125 \log_2 .125 = .5 + .5 + .375 + .375 = 1.75$.

Scheme II: $H(p) = H(p_1, p_a) + p_a H(p_2/p_a, p_b/p_a) + p_b H(p_3/p_b, p_4/p_b) = H(.25, .75) + .75H(.667, .333) + .25H(.5, .5) = .811 + .689 + .25 = 1.75$.

Scheme III: $H(p) = H(p_1, p_a) + p_a H(p_2/p_a, p_3/p_a, p_4/p_a) = H(.25, .75) + .75H(.667, .167, .167) = .811 + .939 = 1.75$.

Scheme IV: $H(p) = H(p_1, p_a) + p_a H(p_2/p_a, p_b/p_a) + p_b H(p_3/p_b, p_4/p_b) = H(.25, .75) + .75H(.667, .333) + .25H(.5, .5) = .811 + .689 + .25 = 1.75$.

These results thus demonstrate that the uncertainty can be calculated in terms of any branching scheme. There are, of course, many additional branching schemes in this example, each of which can be employed for calculating the uncertainty and each of which must lead to the same result.

We now present a theorem that plays an important role in classical information theory. This theorem is essential for proving some basic properties of Shannon entropy as well as for introducing some additional important concepts of information theory.

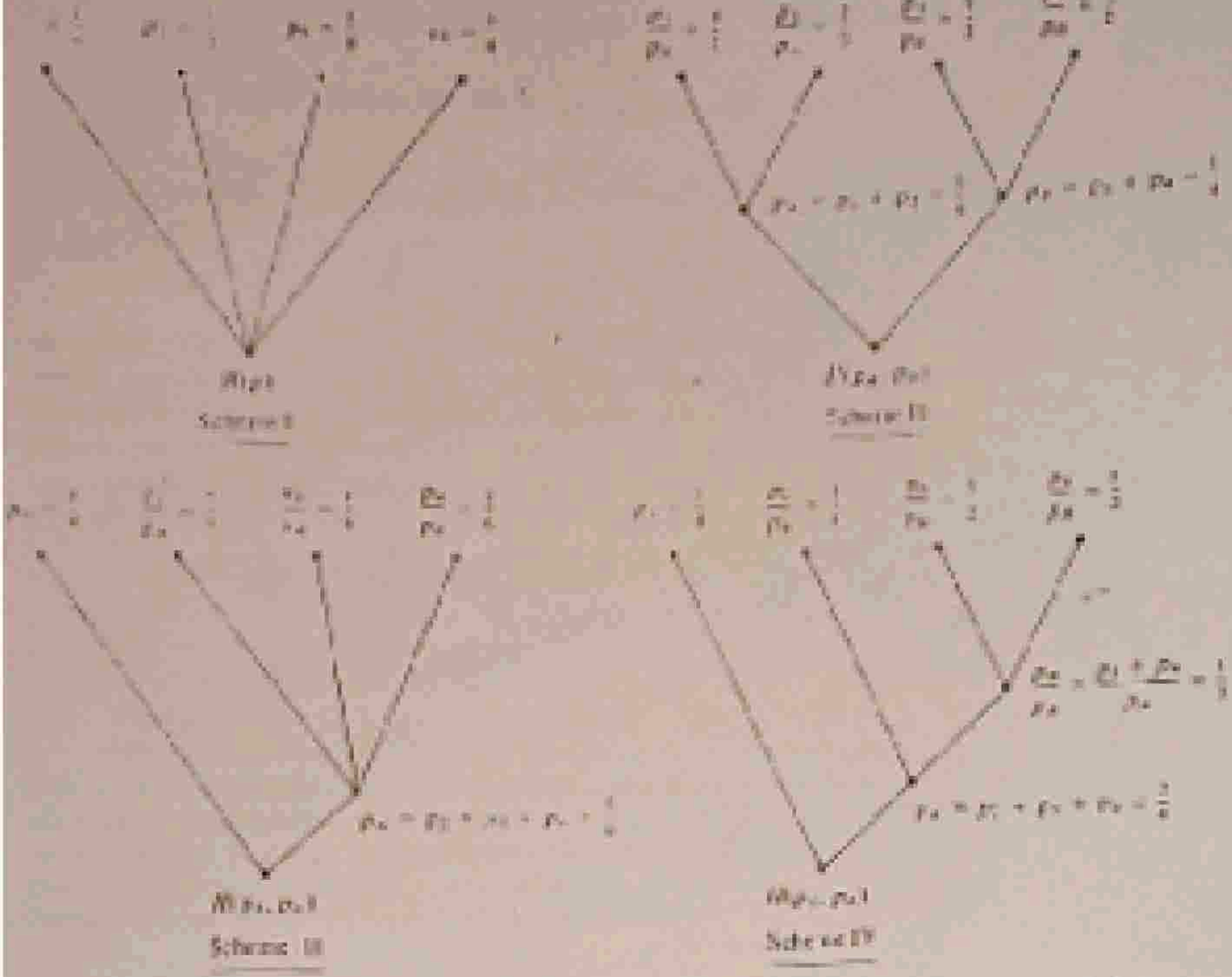


Figure 5.4 Application of the branching property of Shannon entropy (Example 5.4).

Theorem 5.2. The inequality

$$-\sum_{i=1}^{n} p_i \log_2 p_i \le -\sum_{i=1}^{n} p_i \log_2 q_i \tag{5.43}$$

is satisfied for all probability distributions $\langle p_i \mid i \in \mathbb{N}_n \rangle \in {}^n\mathcal{P}$ and $\langle q_i \mid i \in \mathbb{N}_n \rangle \in {}^n\mathcal{P}$ and for all $n \in \mathbb{N}$; the equality in (5.43) holds if and only if $p_i = q_i$ for all $i \in \mathbb{N}_n$.

Proof. Consider the function

$$s(p_i, q_i) = p_i(\ln p_i - \ln q_i) - p_i + q_i$$

for $p_i, q_i \in [0, 1]$. This function is finite and differentiable for all values of p_i and q_i except the pair $q_i = 0$ and $p_i \ne 0$. For each fixed $q_i \ne 0$, the partial derivative with respect to p_i is

$$\frac{\partial s(p_i, q_i)}{\partial p_i} = \ln p_i - \ln q_i.$$

That is,

$$\frac{\partial h(p_i, q_i)}{\partial p_i}\begin{cases} < 0 & \text{for } p_i < q_i \\ = 0 & \text{for } p_i = q_i \\ > 0 & \text{for } p_i > q_i \end{cases}$$

and, consequently, h is a convex function of p_i with its minimum at $p_i = q_i$. Hence, for any given i, we have

$$p_i \ln p_i - p_i \ln q_i - p_i + q_i \geq 0,$$

where the equality holds if and only if $p_i = q_i$. This inequality is also satisfied for $q_i = 0$, since the expression on its left-hand side is $+\infty$ if $p_i \neq 0$ and $q_i = 0$, and it is zero if $p_i = 0$ and $q_i = 0$. Taking the sum of this inequality for all $i \in \mathbb{N}_n$, we obtain

$$\sum_{i=1}^{n} [p_i \ln p_i - p_i \ln q_i - p_i + q_i] \geq 0,$$

which can also be written as

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

The last two terms on the left-hand side of this inequality cancel each other out, since each of the sums is equal to one. Hence,

$$\sum_{i=1}^{n} p_i \ln p_i - \sum_{i=1}^{n} p_i \ln q_i \geq 0,$$

which is equivalent to (5.43) when multiplied by $1/\ln 2$. ■

The inequality (5.43) is usually referred to as the *Gibbs' inequality* or *Gibbs' theorem*.

Let us examine now Shannon entropies of joint, marginal, and conditional probability distributions defined on two sets X and Y. In agreement with a common practice in the literature dealing with Shannon entropy, we simplify the notation in the rest of this section by using $H(X)$ instead of $H(p(x) \mid x \in X)$ or $H(p_1, p_2, \ldots, p_n)$. Furthermore, assuming that $x \in X$ and $y \in Y$, we use the symbols $p(x)$ and $p(y)$ to denote marginal probabilities on sets X and Y, respectively, the symbol $p(x, y)$ for joint probabilities on $X \times Y$, and the symbols $p(x \mid y)$ and $p(y \mid x)$ for the corresponding conditional probabilities. In this simplified notation, the meaning of each symbol is uniquely determined by the argument shown in the parentheses.

Given two sets X and Y, we can recognize three types of entropies:

1. Two *simple entropies* based on the marginal probability distributions:

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

and

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

2. A *joint entropy* defined in terms of the joint probability distribution on $X \times Y$,

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

3. Two *conditional entropies* defined in terms of weighted averages of local conditional entropies as

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

and

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

In addition to these entropies, the function

$$T(X, Y) = H(X) + H(Y) - H(X, Y) \tag{5.44}$$

is often used in the literature as a measure of the strength of relationship (in the probabilistic sense) between elements of sets X and Y. The function is called *information transmission*. It is analogous to the function defined by Eq. (5.35) for Hartley information; it can be generalized to more than two sets in the same way.

Our next subject is an examination of the relationship among the various types of entropies and the information transmission. The key properties of this relationship are expressed by the next several theorems.

Theorem 5.3.

$$H(X \mid Y) = H(X, Y) - H(Y) \tag{5.45}$$

Proof:

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

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

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

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

* The following useful formula for the conditional entropy follows immediately from this equation:

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

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

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

$$= H(X, Y) - H(Y). \quad \blacksquare$$

The same theorem can obviously be proven for the other conditional entropy as well:

$$H(Y \mid X) = H(X, Y) - H(X). \tag{5.46}$$

The theorem can also be generalized to more than two sets. The general form, which is derived from either

$$H(X, Y) = H(X \mid Y) + H(Y)$$

or

$$H(X, Y) = H(Y \mid X) + H(X)$$

is

$$H(X_1, X_2, \ldots, X_n) = H(X_1) + H(X_2 \mid X_1) + H(X_3 \mid X_1, X_2)$$
$$+ \cdots + H(X_n \mid X_1, X_2, \ldots, X_{n-1}). \tag{5.47}$$

This equation is valid for any permutation of the sets involved.

Theorem 5.4.

$$H(X, Y) \le H(X) + H(Y). \tag{5.48}$$

Proof:

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

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

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

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

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

$$+ \log_2 \sum_{x \in X} p(x, y)]$$

$$= -\sum_{\langle x, y \rangle \in X \times Y} p(x, y) [\log_2 p(x) + \log_2 p(y)]$$

$$= -\sum_{\langle x, y \rangle \in X \times Y} p(x, y) \log_2 [p(x) \cdot p(y)]$$

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

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

$$= H(X,Y) - H(Y). \quad \blacksquare$$

The same theorem can obviously be proven for the other conditional entropy as well:

$$H(Y \mid X) = H(X,Y) - H(X). \tag{5.46}$$

The theorem can also be generalized to more than two sets. The general form, which is derived from either

$$H(X,Y) = H(X \mid Y) + H(Y)$$

or

$$H(X,Y) = H(Y \mid X) + H(X)$$

is

$$H(X_1, X_2, \ldots, X_n) = H(X_1) + H(X_2 \mid X_1) + H(X_3 \mid X_1, X_2)$$
$$+ \cdots + H(X_n \mid X_1, X_2, \ldots, X_{n-1}). \tag{5.47}$$

This equation is valid for any permutation of the sets involved.

Theorem 5.4.

$$H(X,Y) \leq H(X) + H(Y). \tag{5.48}$$

Proof:

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

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

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

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

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

$$+ \log_2 \sum_{x \in X} p(x,y)]$$

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

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

By Gibbs' inequality (Theorem 5.2), we have

$$H(X, Y) = -\sum_{(x,y)\in X\times Y} p(x, y)\log_2 p(x, y)$$
$$\leq -\sum_{(x,y)\in X\times Y} p(x, y)\log_2[p_X(x) \cdot p_Y(y)] = H(X) + H(Y).$$

Hence, $H(X, Y) \leq H(X) + H(Y)$; furthermore (again by Gibbs' inequality), the equality holds if and only if

$$p(x, y) = p_X(x) \cdot p_Y(y),$$

that is, only if the sets X and Y are noninteractive in the probabilistic sense. ■

Theorem 5.4 can be easily generalized to more than two sets. Its general form is

$$H(X_1, X_2, \ldots, X_n) \leq \sum_{i=1}^{n} H(X_i), \quad (5.49)$$

which holds for every $n \in \mathbb{N}$.

Theorem 5.5.

$$H(X) \geq H(X \mid Y). \quad (5.50)$$

Proof: From Theorem 5.3,

$$H(X, Y) = H(X \mid Y) + H(Y),$$

and from Theorem 5.4,

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

Hence,

$$H(X \mid Y) + H(Y) \leq H(X) + H(Y)$$

and the inequality

$$H(X \mid Y) \leq H(X)$$

follows immediately. ■

Exchanging X and Y in Theorem 5.5, we obtain

$$H(Y) \geq H(Y \mid X). \quad (5.51)$$

Additional equations expressing the relationships among the various entropies and the information transmission can be obtained by simple formula manipulations with the aid of the key properties contained in Theorems 5.2 through 5.5. For example, when we substitute for $H(X, Y)$ from Eq. (5.45) into Eq. (5.44), we obtain

$$T(X, Y) = H(X) - H(X \mid Y); \quad (5.52)$$

By Gibbs' inequality (Theorem 5.2), we have

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

$$\leq -\sum_{(x,y)\in X\times Y} p(x, y)\log_2[p(x) \cdot p(y)] = H(X) + H(Y).$$

Hence, $H(X, Y) \leq H(X) + H(Y)$; furthermore (again by Gibbs' inequality), the equality holds if and only if

$$p(x, y) = p(x) \cdot p(y),$$

that is, only if the sets X and Y are noninteractive in the probabilistic sense. ■

Theorem 5.4 can be easily generalized to more than two sets. Its general form is

$$H(X_1, X_2, \ldots, X_n) \leq \sum_{i=1}^{n} H(X_i), \tag{5.49}$$

which holds for every $n \in \mathbb{N}$.

Theorem 5.5.

$$H(X) \geq H(X \mid Y). \tag{5.50}$$

Proof: From Theorem 5.3,

$$H(X, Y) = H(X \mid Y) + H(Y),$$

and from Theorem 5.4,

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

Hence,

$$H(X \mid Y) + H(Y) \leq H(X) + H(Y)$$

and the inequality

$$H(X \mid Y) \leq H(X)$$

follows immediately. ■

Exchanging X and Y in Theorem 5.5, we obtain

$$H(Y) \geq H(Y \mid X). \tag{5.51}$$

Additional equations expressing the relationship among the various entropies and the information transmission can be obtained by simple formula manipulations with the aid of the key properties contained in Theorems 5.2 through 5.5. For example, when we substitute for $H(X, Y)$ from Eq. (5.45) into Eq. (5.44), we obtain

$$T(X, Y) = H(X) - H(X \mid Y); \tag{5.52}$$

similarly, by substituting for $H(X, Y)$ from Eq. (5.46) into Eq. (5.48), we obtain [illegible]

$$T(X, Y) = H(Y) - H(Y \mid X).$$

By comparing Eqs. (5.52) and (5.53), we also have

$$H(X) - H(Y) = H(X \mid Y) - H(Y \mid X).$$

We leave additional simple derivations of this kind to the reader as [illegible]

Boltzmann Entropy

One important aspect of Shannon entropy remains to be discussed [illegible] connected with its restriction to finite sets. Is this restriction necessary? It [illegible] that the formula

$$B(q(x) \mid x \in [a, b]) = -\int_a^b q(x) \log_2 q(x)\, dx,$$

where q denotes a probability density function on the real interval $[a, b]$, [illegible] to formula (5.33) for Shannon entropy and could thus be viewed [illegible] extension of Shannon entropy to the domain of real numbers. Moreover, [illegible] B defined by Eq. (5.55) is usually referred to as the *Boltzmann entropy*. [illegible] the analogy between the two functions is suggestive, the following question [illegible] be avoided: is the Boltzmann entropy a genuine extension of the Shannon entropy? To answer this nontrivial question, we must establish a connection between the two functions.

Let q be a probability density function on the interval $[a, b]$ of real numbers. That is, $q(x) \geq 0$ for all $x \in [a, b]$ and

$$\int_a^b q(x)\, dx = 1.$$

Consider a sequence of probability distributions ${}^n\mathbf{p} = ({}^n p_1, {}^n p_2, \ldots$ [illegible] that

$${}^n p_i = \int_{x_{i-1}}^{x_i} q(x)\, dx$$

for every $n \in \mathbb{N}$ and every $i \in \mathbb{N}_n$, where

$$x_i = a + i\frac{b - a}{n}$$

for each $i \in \mathbb{N}_n$. For convenience, let

$$\Delta_n = \frac{b - a}{n}$$

so that

$$x_i = a + i\Delta_n.$$

For each probability distribution ${}^n\mathbf{p} = \langle {}^np_1, {}^np_2, \ldots, {}^np_n \rangle$, let ${}^n\mathbf{d}(x)$ denote a probability density function on $[a, b]$ such that

$${}^n\mathbf{d}(x) = \langle {}^nd_i(x) \mid i \in \mathbb{N}_n \rangle,$$

where

$${}^nd_i(x) = \frac{{}^np_i}{\Delta_n} \quad \text{for} \quad x \in [x_{i-1}, x_i) \tag{5.58}$$

and for all $i \in \mathbb{N}_n$. Then, due to continuity of $\varphi(x)$, the sequence $\langle {}^n\mathbf{d}(x) \mid n \in \mathbb{N} \rangle$ converges to $\varphi(x)$ uniformly on $[a, b]$.

Given the probability distribution ${}^n\mathbf{p}$ for some $n \in \mathbb{N}$, its Shannon entropy is

$$H({}^n\mathbf{p}) = -\sum_{i=1}^{n} {}^np_i \log_2 {}^np_i$$

or, using the introduced probability density function ${}^n\mathbf{d}$,

$$H({}^n\mathbf{p}) = -\sum_{i=1}^{n} {}^nd_i(x)\Delta_n \log_2[{}^nd_i(x)\Delta_n].$$

This equation can be modified as follows:

$$H({}^n\mathbf{p}) = -\sum_{i=1}^{n} {}^nd_i(x)\Delta_n \log_2 {}^nd_i(x) - \sum_{i=1}^{n} {}^nd_i(x)\Delta_n \log_2 \Delta_n$$

$$= -\sum_{i=1}^{n} [{}^nd_i(x) \log_2 {}^nd_i(x)]\Delta_n - \log_2 \Delta_n \sum_{i=1}^{n} {}^np_i.$$

Since probabilities ${}^np_i$ of the distribution ${}^n\mathbf{p}$ must add to one and by the definition of Δ_n, we obtain

$$H({}^n\mathbf{p}) = -\sum_{i=1}^{n} [{}^nd_i(x) \log_2 {}^nd_i(x)]\Delta_n + \log_2 \frac{n}{b-a}. \tag{5.59}$$

When $n \to \infty$ (or $\Delta_n \to 0$), we have

$$\lim_{n\to\infty} -\sum_{i=1}^{n} [{}^nd_i(x) \log_2 {}^nd_i(x)]\Delta_n = -\int_a^b \varphi(x) \log_2 \varphi(x)\, dx$$

according to the introduced relation among ${}^n\mathbf{p}$, $\varphi(x)$, and ${}^n\mathbf{d}(x)$, in particular Eqs. (5.57) and (5.58). Equation (5.59) can thus be rewritten for $n \to \infty$ as

$$\lim_{n\to\infty} H({}^n\mathbf{p}) = B(\varphi(x)) + \lim_{n\to\infty} \log_2 \frac{n}{b-a}. \tag{5.60}$$

The last term in this equation clearly diverges. This means that the Boltzmann entropy is not a limit of the Shannon entropy for $n \to \infty$ and, consequently, it is not a measure of uncertainty and information.

For each probability distribution ${}^n\mathbf{p} = ({}^np_1, {}^np_2, \ldots, {}^np_n)$, let ${}^n\mathbf{d}(x)$ denote a probability density function on $[a, b]$ such that

$${}^n\mathbf{d}(x) = \langle {}^nd_i(x) \mid i \in \mathbb{N}_n \rangle,$$

where

$${}^nd_i(x) = \frac{{}^np_i}{\Delta_n} \quad \text{for} \quad x \in [x_{i-1}, x_i) \tag{5.58}$$

and for all $i \in \mathbb{N}_n$. Then, due to continuity of $q(x)$, the sequence $\{{}^n\mathbf{d}(x) \mid n \in \mathbb{N}\}$ converges to $q(x)$ uniformly on $[a, b]$.

Given the probability distribution ${}^n\mathbf{p}$ for some $n \in \mathbb{N}$, its Shannon entropy is

$$H({}^n\mathbf{p}) = -\sum_{i=1}^{n} {}^np_i \log_2 {}^np_i$$

or, using the introduced probability density function ${}^n\mathbf{d}$,

$$H({}^n\mathbf{p}) = -\sum_{i=1}^{n} {}^nd_i(x)\Delta_n \log_2[{}^nd_i(x)\Delta_n].$$

This equation can be modified as follows:

$$H({}^n\mathbf{p}) = -\sum_{i=1}^{n} {}^nd_i(x)\Delta_n \log_2 {}^nd_i(x) - \sum_{i=1}^{n} {}^nd_i(x)\Delta_n \log_2 \Delta_n$$

$$= -\sum_{i=1}^{n} [{}^nd_i(x)\log_2 {}^nd_i(x)]\Delta_n - \log_2 \Delta_n \sum_{i=1}^{n} {}^np_i.$$

Since probabilities ${}^np_i$ of the distribution ${}^n\mathbf{p}$ must add to one and by the definition of Δ_n, we obtain

$$H({}^n\mathbf{p}) = -\sum_{i=1}^{n} [{}^nd_i(x)\log_2 {}^nd_i(x)]\Delta_n + \log_2 \frac{n}{b-a}. \tag{5.59}$$

When $n \to \infty$ (or $\Delta_n \to 0$), we have

$$\lim_{n\to\infty} -\sum_{i=1}^{n} [{}^nd_i(x)\log_2 {}^nd_i(x)]\Delta_n = -\int_a^b q(x)\log_2 q(x)\,dx$$

according to the introduced relations among ${}^n\mathbf{p}$, $q(x)$, and ${}^n\mathbf{d}(x)$, in particular Eqs. (5.57) and (5.58). Equation (5.59) can thus be rewritten for $n \to \infty$ as

$$\lim_{n\to\infty} H({}^n\mathbf{p}) = B(q(x)) + \lim_{n\to\infty} \log_2 \frac{n}{b-a}. \tag{5.60}$$

The last term in this equation clearly diverges. This means that the Boltzmann entropy is not a limit of the Shannon entropy for $n \to \infty$ and, consequently, it is not a measure of uncertainty and information.

The discrepancy between Shannon and Boltzmann entropies can be reconciled, for example, when we take the difference of entropies for two probability distributions. Then, the last term of Eq. (5.60), which is the same for both entropies, is canceled and the difference of the Shannon entropies converges for $n \to \infty$ to the difference of the corresponding Boltzmann entropies. One implication of this fact is that

$$T_B(q(x,y) \mid x \in [a,b], y \in [c,d]) = -\int_a^b q(x) \log_2 q(x)\, dx$$

$$-\int_c^d q(y) \log_2 q(y)\, dy + \int_a^b \int_c^d q(x,y) \log_2 q(x,y)\, dx\, dy \quad (5.61)$$

is a genuine Boltzmann counterpart of the information transmission T defined by Eq. (5.48). Also

$$B(q(x), q'(x) \mid x \in [a,b]) = -\int_a^b q(x) \log_2 q'(x)\, dx + \int_a^b q(x) \log_2 q(x)\, dx \quad (5.62)$$

is a genuine Boltzmann counterpart of the function

$$D(p(x), p'(x) \mid x \in X) = -\sum_{x \in X} p(x) \log_2 p'(x) + \sum_{x \in X} p(x) \log_2 p(x) \quad (5.63)$$

which is known in information theory as the Shannon cross-entropy or relative entropy (often also called a *directed divergence*). This function is usually employed as a measure of the degree to which an estimated probability distribution p' approximates the distribution p. Observe that the function is always nonnegative, due to the Gibbs inequality (Theorem 5.2), but it is applicable only when $p(x) \neq 0$ implies $p'(x) \neq 0$.

We return to the cross-entropy again in Sec. 5.8 and describe some of its applications in Chap. 5.

MEASURES OF DISSONANCE

Let us begin now to investigate the various aspects of uncertainty within the framework of belief and plausibility measures. In this section, we focus on one aspect of uncertainty that is connected with conflict or dissonance in evidence.

At the common-sense level, we associate the terms conflict and dissonance with properties such as sharp disagreement (between claims, beliefs, interests, etc.), opposition, incompatibility, contradiction, or inconsistency. When we apply the common-sense meaning of these terms to the domain of belief and plausibility measures (within which we deal with degrees of evidence regarding subsets of a universal set), we encounter conflict or dissonance in evidence whenever nonzero degrees of evidence are allocated to disjoint subsets of the universal set. This follows directly from the fact that the element of concern (a priori unknown) can belong to only one of several disjoint subsets of the universal set. Whenever our evidence suggests that the element may belong to either of two disjoint subsets, then we clearly have a conflict in the evidence.

In order to derive a measure of conflict or dissonance in evidence within the framework of belief and plausibility measures, let us consider two bodies of evidence,* $(\mathcal{F}_1, m_1)$ and $(\mathcal{F}_2, m_2)$. Assume that these bodies of evidence, which are obtained from two independent sources, are defined on the same universal set X and for the same purpose. They conflict with each other whenever $m_1(A) \neq 0$, $m_2(B) \neq 0$, and $A \cap B = \varnothing$. Hence, the total amount of their conflict or dissonance should be monotonic increasing with

$$K = \sum_{A \cap B = \varnothing} m_1(A) \cdot m_2(B), \tag{5.64}$$

where $A \in \mathcal{F}_1$ and $B \in \mathcal{F}_2$. Observe that K is also the factor employed in normalizing the joint basic assignment obtained by Dempster's rule (4.34).

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

$$\text{con}(m_1, m_2) = -\log_2(1 - K). \tag{5.65}$$

This function takes values from 0 to ∞ and it is clearly monotonic increasing with K. $\text{con}(m_1, m_2) = 0$ only if m_1 and m_2 do not conflict at all ($K = 0$), and $\text{con}(m_1, m_2) = \infty$ only if they conflict totally ($K = 1$). The choice of the logarithmic function of K is motivated by similar arguments encountered in the discussion of Shannon entropy in Sec. 5.3; logarithm base 2 is used here solely for consistency with our formulation of the Hartley information and Shannon entropy in Sec. 5.2.

Our aim in this section is to derive a function

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

such that $E(m)$ can be justified as a meaningful measure of conflict or dissonance in evidence represented by the basic assignment m; $\mathcal{M}$ denotes here the set of all basic assignments on power sets with 2^n elements for any $n \in \mathbb{N}$. For further reference, let E be called a *measure of dissonance*.

Function (5.65) is obviously relevant for defining a meaningful measure of dissonance, but it must be adjusted to express a conflict within a single body of evidence. To make an appropriate adjustment of the function, let m_A denote a basic assignment on X such that

$$m_A(B) = \begin{cases} 1 & \text{if } B = A \\ 0 & \text{otherwise} \end{cases}$$

for all $B \in \mathcal{P}(X)$. Then, given a body of evidence $(\mathcal{F}, m)$, the function

$$\text{con}(m, m_A) = -\log_2[1 - \sum_{B \cap C = \varnothing} m(B) \cdot m_A(C)],$$

where $A, B, C \in \mathcal{F}$, expresses the amount of conflict of the focal element A with all the other focal elements in $\mathcal{F}$. Since $m_A(C) = 0$ for all $C \neq A$, this function

* The term *body of evidence* is used here in exactly the same sense as defined in Sec. 4.5.

can be defined by a simpler form

$$\mathrm{con}(m, m_A) = -\log_2\left[1 - \sum_{B \cap A = \emptyset} m(B)\right]. \qquad (5.66)$$

Furthermore, observe that

$$1 - \sum_{B \cap A = \emptyset} m(B) = \sum_{B \cap A \neq \emptyset} m(B)$$

and, by Eq. (4.11),

$$\sum_{B \cap A \neq \emptyset} m(B) = Pl(A).$$

Hence,

$$\mathrm{con}(m, m_A) = -\log_2 Pl(A). \qquad (5.67)$$

It is now easy to go one step further and define $E(m)$ as a weighted average of $\mathrm{con}(m, m_A)$ for all $A \in \mathcal{F}$. That is, we define the *measure of dissonance* as

$$E(m) = \sum_{A \in \mathcal{F}} m(A)\,\mathrm{con}(m, m_A)$$

or, using Eq. (5.67), as

$$E(m) = -\sum_{A \in \mathcal{F}} m(A)\log_2 Pl(A). \qquad (5.68)$$

Example 5.5

Let $m(\{x_1, x_2\}) = .4$, $m(\{x_3\}) = .1$, $m(\{x_1, x_3\}) = .3$, and $m(\{x_1, x_2, x_3\}) = .2$ be a basic assignment representing a body of evidence with four focal elements. In order to calculate $E(m)$, we must determine the degrees of plausibility for the focal elements. We have

$$Pl(\{x_1, x_2\}) = m(\{x_1, x_2\}) + m(\{x_1, x_3\}) + m(\{x_1, x_2, x_3\}) = .9,$$

$$Pl(\{x_3\}) = m(\{x_3\}) + m(\{x_1, x_3\}) + m(\{x_1, x_2, x_3\}) = .6,$$

$$Pl(\{x_1, x_3\}) = Pl(\{x_1, x_2, x_3\}) = 1.$$

Applying formula (5.68), we obtain

$$E(m) = -.4\log_2 .9 - .1\log_2 .6 - .3\log_2 1 - .2\log_2 1 = .06 + .07 + 0 + 0 = .13$$

Before we investigate various mathematical properties of function E defined by Eq. (5.68), let us demonstrate that it reduces to the Shannon entropy when m represents a probability distribution.

Theorem 5.5. *If m represents a probability distribution on X, then the measure of dissonance in evidence, as defined by Eq. (5.68), becomes equal to the Shannon entropy (Eq. (5.37)).*

Proof: For probability measures defined on X, $m(A) = 0$ for $A \neq \{x\}$, $x \in X$. Hence,

$$m(A)\log_2 Pl(A) = 0$$

for all $A \neq \{x\}$, $x \in X$. This implies that

$$E(m) = -\sum_{x \in X} m(\{x\}) \log_2 Pl(\{x\}).$$

Since

$$m(\{x\}) = Pl(\{x\}) = p(x),$$

where $p(x)$ denotes the probability of x, we have

$$E(m) = -\sum_{x \in X} p(x) \log_2 p(x) = H(p(x) \mid x \in X). \quad ■$$

The significance of Theorem 5.6 lies in the new insight it gives us into the meaning of Shannon entropy. It is quite clear from the theorem that Shannon entropy measures one specific type of uncertainty, that of dissonance in evidence.

Some properties of the measure of dissonance E are rather obvious. Since $m(A) \in [0, 1]$ and $Pl(A) \in [0, 1]$ for all $A \in \mathcal{P}(X)$, clearly $E(m) \geq 0$ for all basic assignments m; since $m(A) \neq 0$ implies $Pl(A) \neq 0$, it is guaranteed that $E(m)$ is finite for every $m \in \mathcal{M}$. Moreover, it is also clear that E is expansible, continuous, and properly normalized so that its unit of measurement is the bit. Minima and maxima of E are characterized by the following three theorems. Proofs of these theorems are rather simple and we leave them to the reader as an exercise.

Theorem 5.7. If $(\mathcal{F}, m)$ is a consonant body of evidence, then $E(m) = 0$.

Theorem 5.8. $E(m) = 0$ if and only if m has the following property: for all $A, B \in \mathcal{P}(X)$, if $m(A) \neq 0$ and $m(B) \neq 0$, then $A \cap B \neq \emptyset$.

Theorem 5.9. For all basic assignments defined on any universal set X with n elements, $\log_2 n$ is the only maximum of function E, which is obtained for the uniformly distributed probability measure, that is, for $m(\{x\}) = 1/n$ for each $x \in X$.

Given a universal set X, function E thus has the range

$$0 \leq E(m) \leq \log_2 |X|. \quad (5.49)$$

The maximum, which is unique, is identical with the maximum of the Shannon entropy on X. The minimum is not unique; it consists of all possibility and necessity measures as well as some other belief and plausibility measures (as characterized by Theorem 5.8).

It is not surprising that possibility and necessity measures have no dissonance in evidence. These measures are based on nested focal elements and, consequently, no conflict in evidence is involved. This is even reflected in their name—consonant measures.

Function E is also additive with respect to marginal bodies of evidence that are noninteractive in the sense of Eq. (4.18). This is expressed by the next theorem.

for all $A \neq \{x\}$, $x \in X$. This implies that

$$E(m) = -\sum_{x \in X} m(\{x\}) \log_2 Pl(\{x\}).$$

Since

$$m(\{x\}) = Pl(\{x\}) = p(x),$$

where $p(x)$ denotes the probability of x, we have

$$E(m) = -\sum_{x \in X} p(x) \log_2 p(x) = H(p(x) \mid x \in X). \quad \blacksquare$$

The significance of Theorem 5.6 is in the new insight it gives us into the meaning of Shannon entropy. It is quite clear from the theorem that Shannon entropy measures one specific type of uncertainty, that of dissonance in evidence.

Some properties of the measure of dissonance E are rather obvious. Since $m(A) \in [0, 1]$ and $Pl(A) \in [0, 1]$ for all $A \in \mathcal{P}(X)$, clearly $E(m) \geq 0$ for all basic assignments m; since $m(A) \neq 0$ implies $Pl(A) \neq 0$, it is guaranteed that $E(m)$ is finite for every $m \in \mathcal{M}$. Moreover, it is also clear that E is expansible, continuous, and properly normalized so that its unit of measurement is the bit. Minima and maxima of E are characterized by the following three theorems. Proofs of these theorems are rather simple and we leave them to the reader as an exercise.

Theorem 5.7. If $(\mathcal{F}, m)$ is a consonant body of evidence, then $E(m) = 0$.

Theorem 5.8. $E(m) = 0$ if and only if m has the following property: for all $A, B \in \mathcal{P}(X)$ if $m(A) \neq 0$ and $m(B) \neq 0$, then $A \cap B \neq \emptyset$.

Theorem 5.9. For all basic assignments defined on any universal set X with n elements, $\log_2 n$ is the only maximum of function E, which is obtained for the uniformly distributed probability measure, that is, for $m(\{x\}) = 1/n$ for each $x \in X$.

Given a universal set X, function E thus has the range

$$0 \leq E(m) \leq \log_2 |X|. \tag{5.49}$$

The maximum, which is unique, is identical with the maximum of the Shannon entropy on X. The minimum is not unique; it consists of all possibility and necessity measures as well as some other belief and plausibility measures (as characterized by Theorem 5.8).

It is not surprising that possibility and necessity measures have no dissonance in evidence. These measures are based on nested focal elements and, consequently, no conflict in evidence is involved. This is even reflected in their name—consonant measures.

Function E is also additive with respect to marginal bodies of evidence that are noninteractive in the sense of Eq. (4.18). This is expressed by the next theorem.

Theorem 5.10. Let m_X and m_Y be marginal basic assignments on sets X and Y, respectively, and let m be a joint basic assignment on $X \times Y$ such that

$$m(A \times B) = m_X(A) \cdot m_Y(B)$$

for all $A \in \mathcal{P}(X)$ and $B \in \mathcal{P}(Y)$. Then,

$$E(m) = E(m_X) + E(m_Y). \tag{5.70}$$

Proof: First, we must determine the meaning of $Pl(A \times B)$ in this case. By definition,

$$Pl(A \times B) = \sum m(C \times D),$$

where the sum is taken over all sets C and D such that $(C \times D) \cap (A \times B) \neq \varnothing$. This equation can be rewritten, according to the assumption of the theorem, as

$$Pl(A \times B) = \sum m_X(C) \cdot m_Y(D),$$

where the sum is taken over the same sets C and D as before. This equation is clearly identical to

$$Pl(A \times B) = \sum_{C \cap A \neq \varnothing} m_X(C) \cdot \sum_{D \cap B \neq \varnothing} m_Y(D).$$

Hence,

$$Pl(A \times B) = Pl_X(A) \cdot Pl_Y(B). \tag{5.71}$$

Using this result, the rest of the proof consists of the following simple manipulations of the formula for $E(m)$, where all the sums are taken over $A \in \mathcal{P}(X)$ and $B \in \mathcal{P}(Y)$ or just over the corresponding focal elements:

$$\begin{aligned}
E(m) &= -\sum_{A,B} m(A \times B)\log_2 Pl(A \times B) \\
&= -\sum_{A,B} m_X(A) \cdot m_Y(B)\log_2[Pl_X(A) \cdot Pl_Y(B)] \\
&= -\sum_{A,B} m_X(A) \cdot m_Y(B)\log_2 Pl_X(A) - \sum_{A,B} m_X(A) \cdot m_Y(B)\log_2 Pl_Y(B) \\
&= -\sum_{A} m_X(A)\log_2 Pl_X(A) \sum_{B} m_Y(B) - \sum_{B} m_Y(B)\log_2 Pl_Y(B) \sum_{A} m_X(A) \\
&= -\sum_{A} m_X(A)\log_2 Pl_X(A) - \sum_{B} m_Y(B)\log_2 Pl_Y(B) \\
&= E(m_X) + E(m_Y). \quad \blacksquare
\end{aligned}$$

Example 5.6

To illustrate the meaning of Theorem 5.10, let us consider the noninteractive marginal basic assignments m_X and m_Y on sets $X = \{x_1, x_2, x_3\}$ and $Y = \{y_1, y_2, y_3\}$, respectively, that are specified in Table 5.2(a). Their joint basic assignment m, which is defined by

$$m(A \times B) = m_X(A) \cdot m_Y(B)$$

TABLE 5.2. NONINTERACTIVE MARGINAL AND JOINT BASIC ASSIGNMENTS, PLAUSIBILITIES, AND BELIEFS (EXAMPLES 5.6 AND 5.7)

(a) $m(A \times B) = m_X(A) \cdot m_Y(B)$

$m(A \times B)$	$m_Y(\{y_1\}) = .2$	$m_Y(\{y_2, y_3\}) = .5$	$m_Y(\{y_1, y_2, y_3\}) = .3$
$m_X(\{x_1, x_2\}) = .4$	.08	.2	.12
$m_X(\{x_1\}) = .1$	.02	.05	.03
$m_X(\{x_2, x_3\}) = .3$	.06	.15	.09
$m_X(\{x_1, x_2, x_3\}) = .2$	.04	.1	.06

(b) $Pl(A \times B) = Pl_X(A) \cdot Pl_Y(B)$

$Pl(A \times B)$	$Pl_Y(\{y_1\}) = .2$	$Pl_Y(\{y_2, y_3\}) = .8$	$Pl_Y(\{y_1, y_2, y_3\}) = .8$
$Pl_X(\{x_1, x_2\}) = .8$	.16	.72	.72
$Pl_X(\{x_1\}) = .6$	.12	.48	.48
$Pl_X(\{x_2, x_3\}) = 1$	.2	.8	.8
$Pl_X(\{x_1, x_2, x_3\}) = 1$	.2	.8	.8

(c) $Bel(A \times B) = Bel_X(A) \cdot Bel_Y(B)$

$Bel(A \times B)$	$Bel_Y(\{y_1\}) = .2$	$Bel_Y(\{y_2, y_3\}) = .8$	$Bel_Y(\{y_1, y_2, y_3\}) = 1$
$Bel_X(\{x_1, x_2\}) = .4$	.08	.32	.4
$Bel_X(\{x_1\}) = .1$	.02	.08	.1
$Bel_X(\{x_2, x_3\}) = .4$	.08	.32	.4
$Bel_X(\{x_1, x_2, x_3\}) = 1$	.2	.8	1

for all $A \in \mathcal{P}(X)$ and $B \in \mathcal{P}(Y)$, is also given in Table 5.2(c). To illustrate the additivity of the measure of dissonance E, we need the corresponding marginal and joint plausibilities for all focal elements; these are given in Table 5.2(b). The marginal plausibilities are calculated from their basic assignment counterparts by Eq. (4.11); the joint plausibilities can then be calculated by Eq. (5.71). We can now calculate the amount of dissonance in the marginal and joint bodies of evidence:

$E(m_X) = .14$ (from Example 5.5),

$$E(m_Y) = -.2 \log_2 .2 - .5 \log_2 .8 - .3 \log_2 .8$$
$$= .46 + .16 + .1 = .72,$$

$$E(m) = -.08 \log_2 .16 - .2 \log_2 .72 - .12 \log_2 .72 - .02 \log_2 .12$$
$$- .05 \log_2 .48 - .03 \log_2 .48 - .06 \log_2 .2 - .15 \log_2 .8$$
$$- .09 \log_2 .8 - .04 \log_2 .2 - .1 \log_2 .8 - .06 \log_2 .8$$
$$= .2 + .1 + .06 + .06 + .05 + .03 + .14 + .05$$
$$+ .03 + .09 + .03 + .02 = .86.$$

We can see that $E(m_X) + E(m_Y) = .14 + .72 = .86 = E(m)$.

5.5 MEASURES OF CONFUSION

As explained in Sec. 4.2, plausibility and belief measures are dual in the sense that

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

It is thus reasonable to define a natural counterpart of the dissonance measure by replacing $Pl(A)$ in the definition of E (Eq. (5.68)) with $Bel(A)$. This results in the function

$$C(m) = -\sum_{A \in \mathcal{F}} m(A) \log_2 Bel(A),$$

where $\mathcal{F}$ is the set of focal elements of the basic assignment m. This function is usually called in the literature a *measure of confusion*.

Function C is clearly expansible, continuous, and normalized in the sense [illegible] of measurement is the bit. Since $Bel(A) \leq Pl(A)$, it follows from [illegible] Eqs. (5.68) and (5.72) that

$$C(m) \geq E(m)$$

for any basic assignment m. It is also obvious that $C(m) \geq 0$. In addition, [illegible] $m(A) > 0$ implies $Bel(A) > 0$, $C(m)$ is always finite.

The minimum of function C, $C(m) = 0$, can be attained if and only [illegible] $= 1$ for all focal elements. According to Eq. (4.10), this is possible only [illegible] $= 1$ for one particular $A \in \mathcal{P}(X)$ and $m(B) = 0$ for all $B \in \mathcal{P}(X)$ [illegible] A. Unlike the dissonance measure E, the confusion measure C is not [illegible] consonant bodies of evidence that have at least two focal elements.

Since $Bel(A) \geq m(A)$, it follows from Eq. (5.72) that

$$C(m) \leq -\sum_{A \in \mathcal{F}} m(A) \log_2 m(A).$$

Hence, C attains its maximum (not necessarily unique) if and only if [illegible] distributed among the largest possible number of subsets of X such that none of these subsets is included in any other. Families of subsets, none of which is included in any other, are known in combinatorial theory as *antichains*. It [illegible] established that the largest antichains consist of all subsets with $n/2$ elements when n is even and with either $\lfloor n/2 \rfloor$ or $\lfloor n/2 \rfloor + 1$ elements when n is odd, where $\lfloor n/2 \rfloor$ denotes the integer part of $n/2$. Hence, C has one maximum when n is even and two maxima when n is odd. For any n, the largest antichains contain

$$\binom{n}{\left\lfloor \frac{n}{2} \right\rfloor}$$

subsets. The maximum of C is thus attained when

$$m(A) = \frac{1}{\binom{n}{\lfloor \frac{n}{2} \rfloor}}$$

for all subsets in one of the largest antichains. This maximum is equal to

$$\log_2 \binom{n}{\lfloor \frac{n}{2} \rfloor}$$

Hence,

$$0 \le C(m) \le \log_2 \binom{n}{\lfloor \frac{n}{2} \rfloor} \qquad (5.75)$$

where $n = |X|$.

We can now see that the name *measure of confusion* captures quite well the nature of the function C defined by Eq. (5.72). Indeed, $C(m)$ characterizes the multitude of subsets supported by evidence as well as the uniformity of the distribution of strength of evidence among the subsets. Clearly, the greater the number of subsets involved and the more uniform the distribution, the more we tend to be confused by the presentation of evidence.

When m represents a probability measure, then $m(A) = Pl(A) = Bel(A)$ for all focal elements (singletons) and, consequently, the confusion measure becomes the Shannon entropy. The generalization of the Shannon entropy from probability measures to evidence measures is thus not unique. Functions E and C, both of which are generalizations of Shannon entropy, are sometimes called *entropy-like measures*.

Like the dissonance measure E, the confusion measure is also additive with respect to marginal bodies of evidence that are noninteractive. This property is formally stated by the following theorem.

Theorem 5.11. Let m_X and m_Y be marginal basic assignments on sets X and Y, respectively, and let m be a joint basic assignment on $X \times Y$ such that

$$m(A \times B) = m_X(A) \cdot m_Y(B)$$

for all $A \in \mathcal{P}(X)$ and $B \in \mathcal{P}(Y)$. Then,

$$C(m) = C(m_X) + C(m_Y). \qquad (5.76)$$

Proof: The proof is analogous to the proof of Theorem 5.10. ■