CS 215- Data Interpretation and Analysis (Post Midsem)

Chi-square distribution, Fuzzy Logic cntd.

Pushpak Bhattacharyya
Computer Science and Engineering Department
IIT Bombay

Lecture-6

30oct23

Recap

Ch-Square Distribution Definition

Sheldon Edition 3, Pp 185

Statement of Chi-Square distribution

$$X = Z_1^{\;2} + Z_2^{\;2} + Z_3^{\;2} + \ldots Z_n^{\;2}$$

Each Z_i is a standard normal variable ($Z_i \sim N(0,1)$)

$$X \sim \chi^2$$

X is said to have a chi-square distribution with ‘n’ degrees of freedom

Fitting Distribution to Data

Motivation Slide

- Scarcity of data
- Have to use transfer learning
- Transfer learning is essentially borrowing the distribution of another domain and/or data for the purpose at hand
- For example, sentiment analysis in the movie domain- in case data is absent- can be attempted through data in book domain (sentiment about the book ***that*** was picturized will have bearing on the sentiment on the picture)

Transfer Learning = Distribution Adaptation

- Distribution adaptation needs distribution fitting
- Distribution fitting needs testing the goodness of fit
- A well established classical area
- HYPOTHESIS TESTING: Distribution ‘D’ fits the Data ‘d’

coin

Fitting Binomial Distribution

Die Tossing: χ^2 Test

Toss a Die 120 times

- Observe the no. of times each face appears
- Test the hypothesis that the Dice is FAIR

Face	Frequency
1	25
2	17
3	15
4	23
5	24
6	16

Toss a Die 120 times

- Observe the no. of times each face appears
- Test the hypothesis that the Dice is FAIR

Face	Frequency
1	25
2	17
3	15
4	23
5	24
6	16

Condition for FAIRNESS

- If the dice was fair, we would get 20 times for each face

Face	Frequency	Expected
1	25	20
2	17	20
3	15	20
4	23	20
5	24	20
6	16	20
Total	120	120

Compute $\chi^2_{observed}$

- Take $(O-E)^2/E$ for each observation
- Sum them; that gives $\chi^2_{observed}$

Face	Frequency (O)	Expected (E)	O-E	(O-E)^2	(O-E)^2/E
1	25	20	5	25	1.25
2	17	20	-3	9	0.45
3	15	20	-5	25	1.25
4	23	20	3	9	0.45
5	24	20	4	16	0.8
6	16	20	-4	16	0.8
Total	120	120		$\chi^2_{observed}$	5

Find $\chi^2_{critical}$

DoF=6-1=5; Significance level α=0.05; $\chi^2_{critical}$= 11.1

Critical values of the Chi-square distribution with *d* degrees of freedom

	Probability of exceeding the critical value						
d	0.05	0.01	0.001	*d*	0.05	0.01	0.001
1	3.841	6.635	10.828	11	19.675	24.725	31.264
2	5.991	9.210	13.816	12	21.026	26.217	32.910
3	7.815	11.345	16.266	13	22.362	27.688	34.528
4	9.488	13.277	18.467	14	23.685	29.141	36.123
5	11.070	15.086	20.515	15	24.996	30.578	37.697
6	12.592	16.812	22.458	16	26.296	32.000	39.252
7	14.067	18.475	24.322	17	27.587	33.409	40.790
8	15.507	20.090	26.125	18	28.869	34.805	42.312
9	16.919	21.666	27.877	19	30.144	36.191	43.820
10	18.307	23.209	29.588	20	31.410	37.566	45.315

***INTRODUCTION TO POPULATION GENETICS,* Table D.1**

Compare $\chi^2_{observed}$ and $\chi^2_{critical}$

- $\chi^2_{observed} < \chi^2_{critical}$
- So cannot reject NULL Hypothesis
- H_0: the dice is FAIR

Fitting Poisson Distribution

Die Tossing: χ^2 Test

Proverb Data

- The table below shows the number of times proverbs occur in a set of 50 documents.

X (num proverbs)	F(num docs)
0	21
1	18
2	7
3	3
4	1
	50

Poisson Formula

$$P(X = x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$$

P(X=x) is the probability of the random variable X taking the value x.

In our case X is the r.v denoting the #proverbs in a document

λ is the parameter of the distribution, equal to the mean and standard deviation (can be shown by MGF)

Mean of Poisson for the example

Mean=λ=

$(0*21+1*18+2*7+3*3+4*1)/50=45/50$

$= 0.9$

X (num proverbs)	F(num docs)
0	21
1	18
2	7
3	3
4	1
	50

Calculate the Expected No. of proverbs

$$P(X = 0) = \frac{e^{-0.9}\lambda^{0}}{0!} = 0.4$$

Similarly find P(X=1), P(X=2), P(X=3), P(X=4)

P(X=1)= 0.37,

P(X=2)=0.17,

P(X=3)=0.05,

P(X=4)=0.01

Get expected values, P(X=x)*50

X (#Proverbs)	F(#docs)	Exp #docs, after rounding off
0	21	20
1	18	18
2	7	8
3	3	2
4	1	1

$(Exp\text{-}obs)^2/Exp$

X (#Proverbs)	F(Observed #docs)	Exp #docs	(obs-exp)^2/exp
0	21	20	0.05
1	18	18	0
2	7	8	0.125
3	3	2	0.5
4	1	1	0

The fit looks good at the first impression!

Get ChiSquare Observed

$$\chi^2_{obs} = \sum_{i \varepsilon categories} \frac{(\exp_i - obs_i)^2}{\exp}$$

$$= 0.675$$

and Compare with ChiSq Critical

$$\chi^2_{critical,\, dof=5-1=4,\, \alpha=0.05}$$

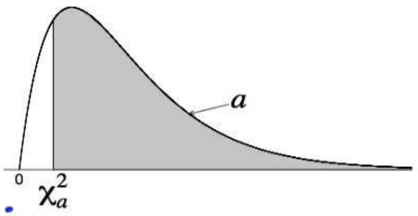
$$=9.48$$

$ChiSq_{obseved} < ChiSq_{critical}$

No reason to reject null hypothesis

H_0= Data follows Poisson distribution

Chi-Square table



df	$\chi^2_{0.9995}$	$\chi^2_{0.999}$	$\chi^2_{0.995}$	$\chi^2_{0.990}$	$\chi^2_{0.975}$	$\chi^2_{0.95}$	$\chi^2_{0.90}$	$\chi^2_{0.85}$	$\chi^2_{0.80}$
1	0.000	0.000	0.000	0.000	0.001	0.004	0.016	0.036	0.064
2	0.001	0.002	0.010	0.020	0.051	0.103	0.211	0.325	0.446
3	0.015	0.024	0.072	0.115	0.216	0.352	0.584	0.798	1.005
4	0.064	0.091	0.207	0.297	0.484	0.711	1.064	1.366	1.649
5	0.158	0.210	0.412	0.554	0.831	1.145	1.610	1.994	2.343
6	0.299	0.381	0.676	0.872	1.237	1.635	2.204	2.661	3.070
7	0.485	0.598	0.989	1.239	1.690	2.167	2.833	3.358	3.822
8	0.710	0.857	1.344	1.646	2.180	2.733	3.490	4.078	4.594
9	0.972	1.152	1.735	2.088	2.700	3.325	4.168	4.817	5.380
10	1.265	1.479	2.156	2.558	3.247	3.940	4.865	5.570	6.179
11	1.587	1.834	2.603	3.053	3.816	4.575	5.578	6.336	6.989
12	1.934	2.214	3.074	3.571	4.404	5.226	6.304	7.114	7.807
13	2.305	2.617	3.565	4.107	5.009	5.892	7.042	7.901	8.634
14	2.697	3.041	4.075	4.660	5.629	6.571	7.790	8.696	9.467

Fitting Normal Distribution

Cricket Score problem

Range	Midpoint (MP)	#innings (I)
0-20	10	10
21-40	30	20
41-60	50	40
61-80	70	20
81-100	90	10
		100

Compute Mean

Range	Midpoint (MP)	#innings (I)	MP X I
0-20	10	10	100
21-40	30	20	600
41-60	50	40	2000
61-80	70	20	1400
81-100	90	10	900
		100	5000
		AV	50

Compute Standard Deviation

Midpoint (MP), Xi	#innings (I)	MP X I		(Xi-av)	(Xi-av)^2	sqr X I
10	10	100		-40	1600	16000
30	20	600		-20	400	8000
50	40	2000		0	0	0
70	20	1400		20	400	8000
90	10	900		40	1600	16000
	100	5000			4000	48000
	AV	50				
	var	484.848				
	std	22.0193				

Making the ranges continuous, Computing low and high values

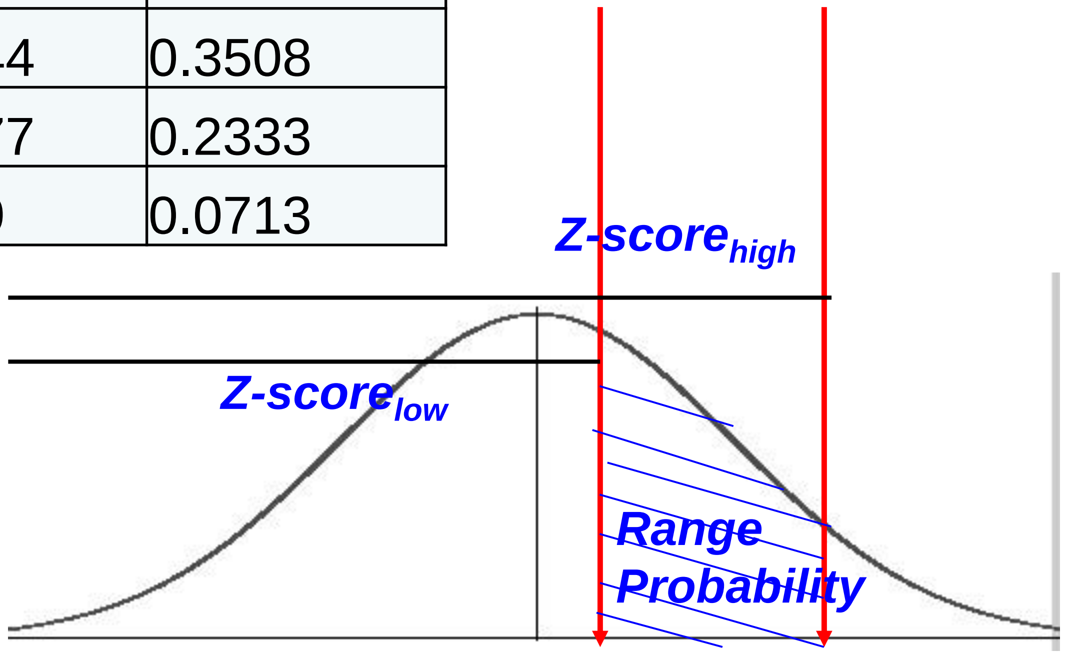
Range	Low range	Hi range	X_low-mu	X_high-mu
0-20	-0.5	20.5	-50.5	-29.5
21-40	20.5	40.5	-29.5	-9.5
41-60	40.5	60.5	-9.5	10.5
61-80	60.5	80.5	10.5	30.5
81-100	80.5	100.5	30.5	50.5
	Mean=50			

Z_{low} [=(X_{low}-μ)/σ] and Z_{high} [=(X_{high}-μ)/σ)

Low range	Hi range	X_low-mu	X_high-mu	Zlow	Zhigh
-0.5	20.5	-50.5	-29.5	-2.29337	-1.33969
20.5	40.5	-29.5	-9.5	-1.33969	-0.43143
40.5	60.5	-9.5	10.5	-0.43143	0.47684
60.5	80.5	10.5	30.5	0.47684	1.3851
80.5	100.5	30.5	50.5	1.3851	2.29337
				Mean=50	
				std 22.02	

Compute *Z-score* and the range probability

Z-score,low	Zscore, high	Range Probability (P)
0.011	0.0901	0.0791
0.0901	0.3336	0.2435
0.3336	0.6844	0.3508
0.6844	0.9177	0.2333
0.9177	0.989	0.0713



Compute Expected Frequency, Range Probability X Midpoint

Range	Midpoint (MP)	Range Probability (P)	Expected freq (MP X P)
0-20	10	0.0791	7.91
21-40	30	0.2435	24.35
41-60	50	0.3508	35.08
61-80	70	0.2333	23.33
81-100	90	0.0713	9.33
			100

Compare Observed and Expected

Range	Midpoint (MP)	Observed Frequency #innings (I)	Expected freq (MP X P)
0-20	10	10	7.91
21-40	30	20	24.35
41-60	50	40	35.08
61-80	70	20	23.33
81-100	90	10	9.33
		100	100

Seems like from Normal Distribution!

Compute $\chi^2_{observed}$

Observed Frequency #innings (I)	Expected freq (MP X P)	obs-expected	(obs-exp)^2	(obs-exp)^2/exp
10	7.91	2.09	4.3681	0.552225032
20	24.35	-4.35	18.9225	0.777104723
40	35.08	4.92	24.2064	0.690034208
20	23.33	-3.33	11.0889	0.475306472
10	9.33	0.67	0.4489	0.048113612
100	100			

$X^2_{observed}$=2.54 (sum of last col)

Compare $\chi^2_{observed}$ and $\chi^2_{critical}$

- $\chi^2_{observed} = 2.54$
- $\chi^2_{critical} = 9.48$ (DoF: 4, α=0.05)

- Cannot reject the null hypothesis

H_0: The data comes from a normal distribution with μ=50 and σ=22.01

End Recap

Modeling Human Reasoning

Fuzzy Logic

Fuzzy Logic tries to capture the human ability of reasoning with imprecise information

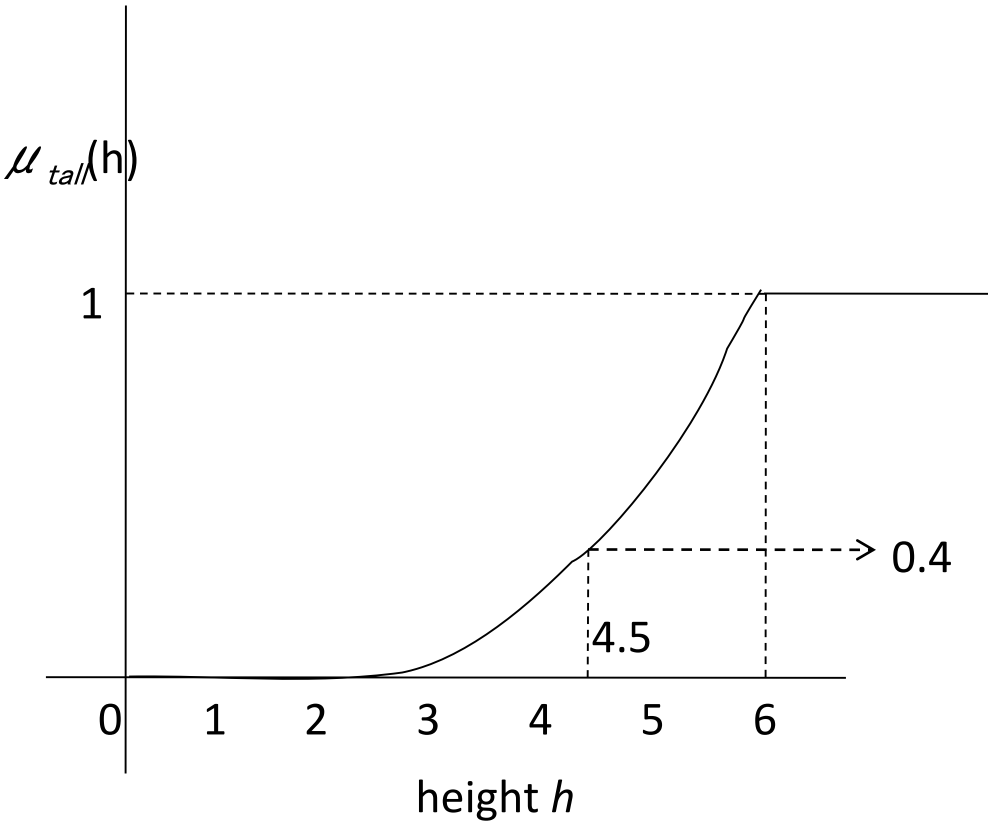
- Works with imprecise statements such as:

  In a process control situation, "*If* the temperature is <u>moderate</u> and the pressure is <u>high</u>, *then* turn the knob <u>slightly right</u>"
- The rules have "Linguistic Variables", typically adjectives qualified by adverbs (adverbs are <u>hedges</u>).

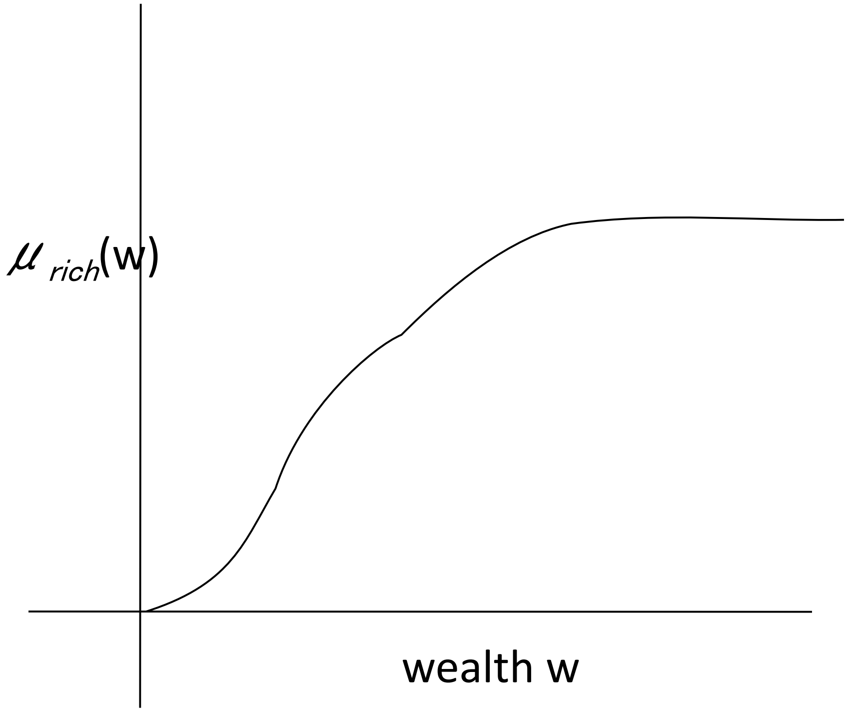
Linguistic Variables

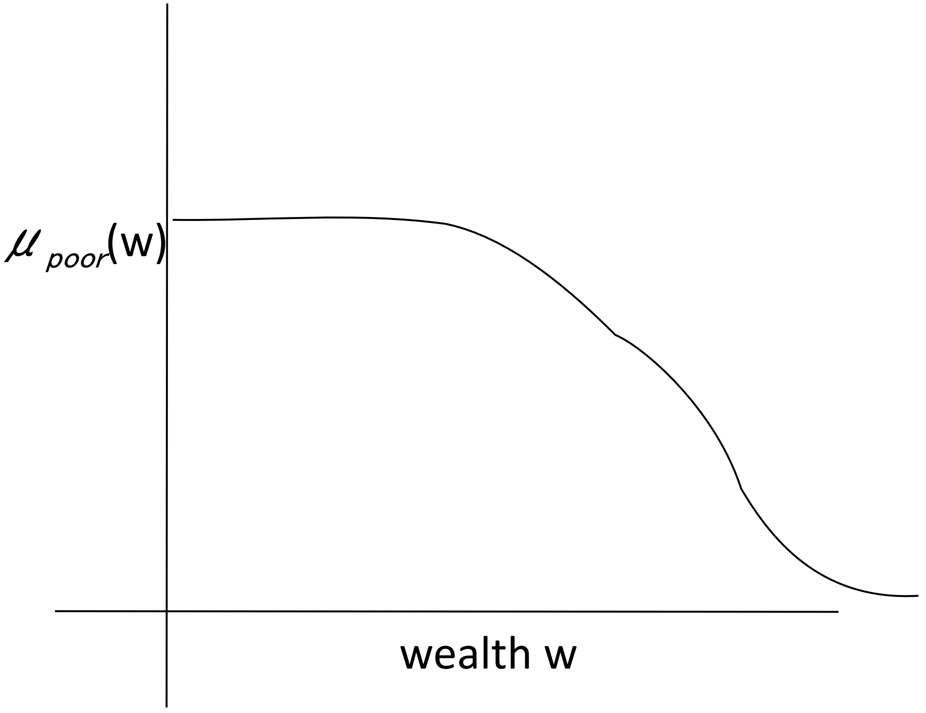
- Fuzzy sets are named by Linguistic Variables (typically adjectives).
- Underlying the LV is a numerical quantity

  E.g. For ‘tall’ (LV), ‘height’ is numerical quantity.
- Profile of a LV is the plot shown in the figure shown alongside.

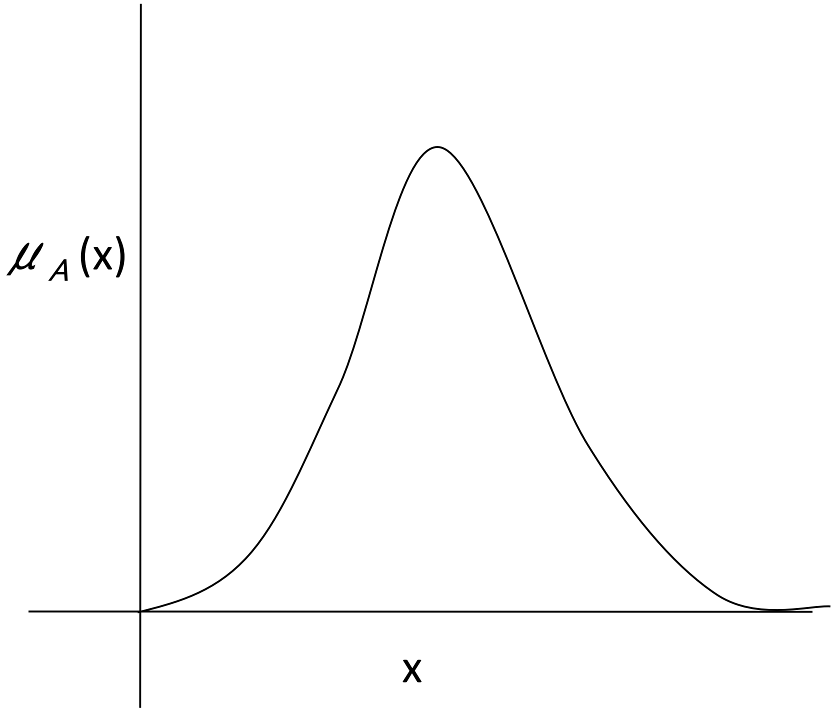


Example Profiles

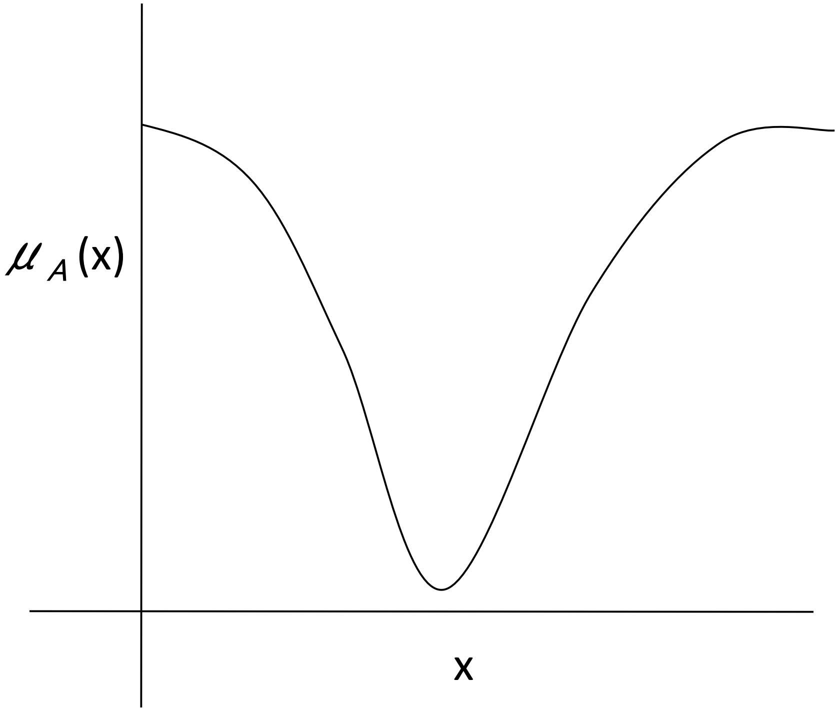




Example Profiles



Profile representing
moderate (*e.g.* moderately rich)



Profile representing
extreme

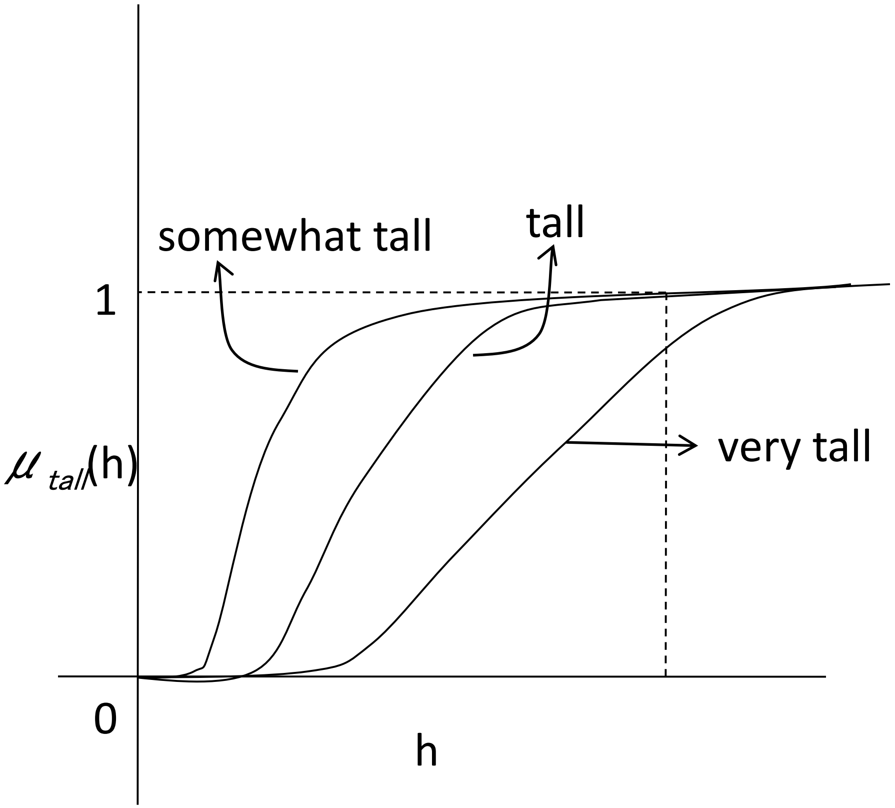
Concept of Hedge

- Hedge is an intensifier
- Example:

  LV = tall, LV_1 = very tall,
  LV_2 = somewhat tall
- 'very' operation:

  $\mu_{very\ tall}(x) = \mu^2_{tall}(x)$
- 'somewhat' operation:

$\mu_{somewhat\ tall}(x) =$
$\sqrt{(\mu_{tall}(x))}$



Fuzzy Sets

Theory of Fuzzy Sets

- Intimate connection between logic and set theory.
- Given any set ‘S’ and an element ‘e’, there is a very natural predicate, $\mu_s(e)$ called as the *belongingness predicate*.
- The predicate is such that,

$$\mu_s(e) = 1, \qquad \textit{iff } e \in S$$
$$= 0, \qquad \textit{otherwise}$$

- For example, $S = \{1, 2, 3, 4\}$, $\mu_s(1) = 1$ and $\mu_s(5) = 0$
- A predicate *P(x)* also defines a set naturally.

  $S = \{x \mid P(x) \text{ is } \textit{true}\}$

  For example, *even(x)* defines $S = \{x \mid x \text{ is even}\}$

Fuzzy Set Theory (contd.)

- Fuzzy set theory starts by questioning the fundamental assumptions of set theory *viz.*, the belongingness predicate, μ, value is 0 or 1.
- Instead in Fuzzy theory it is assumed that,

$$\mu_s(e) = [0, 1]$$

- Fuzzy set theory is a generalization of classical set theory *aka* called Crisp Set Theory.
- In real life, *belongingness* is a fuzzy concept.

Example: Let, T = "tallness"

$$\mu_T(\text{height=6.0ft}) = 1.0$$

$$\mu_T(\text{height=3.5ft}) = 0.2$$

An individual with height 3.5ft is "tall" with a degree *0.2*

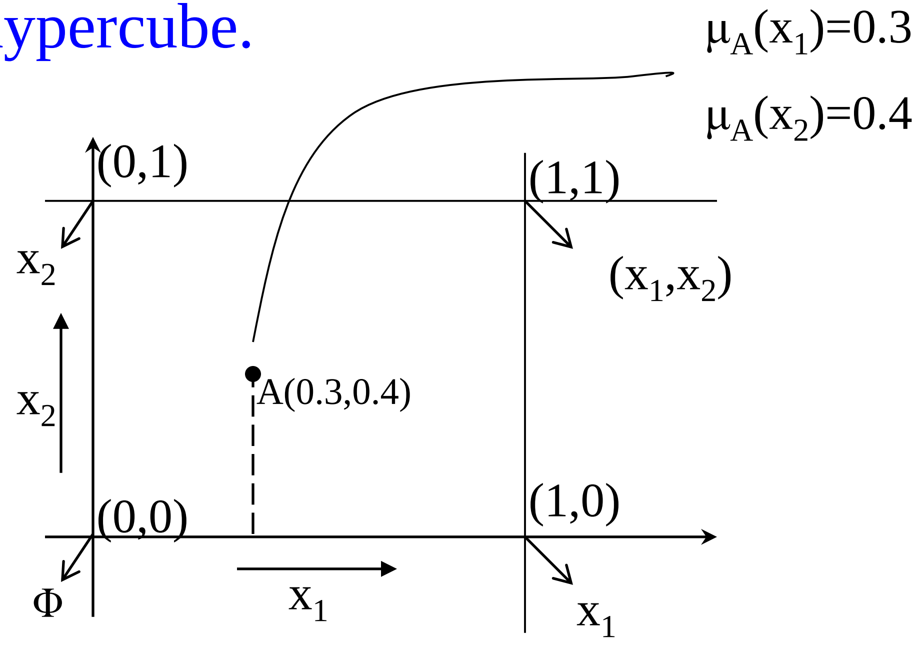
Representation of Fuzzy sets

Let $U = \{x_1, x_2, \ldots\ldots, x_n\}$

$|U| = n$

The various sets composed of elements from U are presented as points on and inside the n-dimensional hypercube. The crisp sets are the corners of the hypercube.

$U=\{x_1,x_2\}$



A fuzzy set A is represented by a point in the n-dimensional space as the point $\{\mu_A(x_1), \mu_A(x_2), \ldots\ldots \mu_A(x_n)\}$

Degree of fuzziness

The centre of the hypercube is the *most fuzzy* set. Fuzziness decreases as one nears the corners

Measure of fuzziness

Called the entropy of a fuzzy set

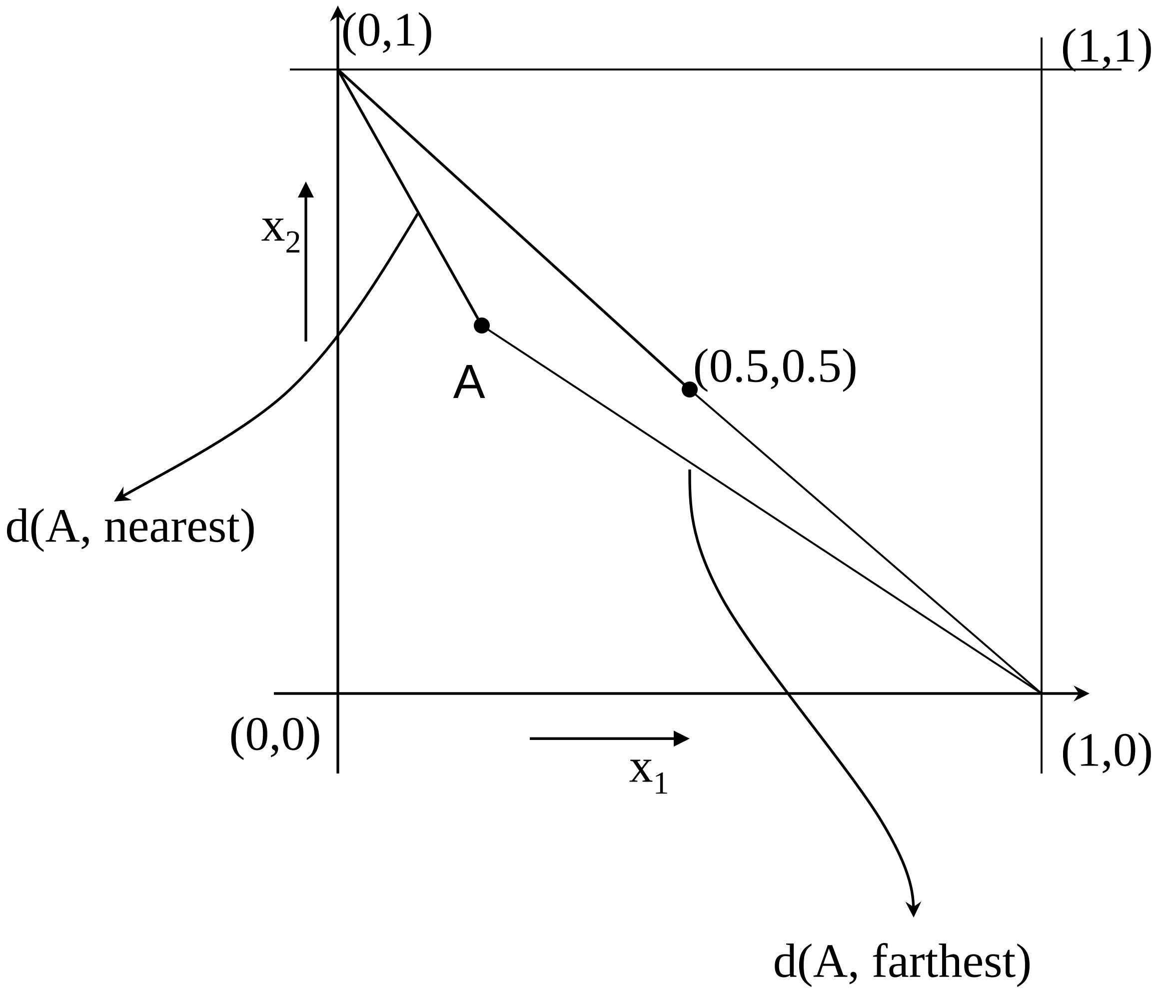
Fuzzy set

Farthest corner

$$E(S) = d(S, nearest) / d(S, farthest)$$

Entropy

Nearest corner


(0,1)
(1,1)
x_2
A
(0.5,0.5)
d(A, nearest)
(0,0)
x_1
(1,0)
d(A, farthest)

Definition

Distance between two fuzzy sets

$$d(S_1, S_2) = \sum_{i=1}^{n} \underbrace{| \mu_{s_1}(x_i) - \mu_{s_2}(x_i) |}_{L_1 \text{ - norm}}$$

Let C = fuzzy set represented by the centre point

d(c,nearest) = |0.5-1.0| + |0.5 – 0.0|

= 1

= d(C,farthest)

=> E(C) = 1

Definition

Cardinality of a fuzzy set

$$m(s) = \sum_{i=1}^{n} \mu_s(x_i)$$ (generalization of cardinality of classical sets)

Union, Intersection, complementation, subset hood

$$\mu_{s_1 \cup s_2}(x) = \max(\mu_{s_1}(x), \mu_{s_2}(x)), \forall x \in U$$

$$\mu_{s_1 \cap s_2}(x) = \min(\mu_{s_1}(x), \mu_{s_2}(x)), \forall x \in U$$

$$\mu_{s^c}(x) = 1 - \mu_s(x)$$

Example of Operations on Fuzzy Set

- Let us define the following:
  - Universe $U=\{X_1, X_2, X_3\}$
  - Fuzzy sets
    - $A=\{0.2/X_1, 0.7/X_2, 0.6/X_3\}$ and
    - $B=\{0.7/X_1, 0.3/X_2, 0.5/X_3\}$

Then Cardinality of A and B are computed as follows:

Cardinality of $A=|A|=0.2+0.7+0.6=1.5$

Cardinality of $B=|B|=0.7+0.3+0.5=1.5$

While distance between A and B

$d(A,B)=|0.2-0.7)+|0.7-0.3|+|0.6-0.5|=1.0$

What does the cardinality of a fuzzy set mean? In crisp sets it means the number of elements in the set.

Example of Operations on Fuzzy Set (cntd.)

Universe U={X_1 ,X_2 ,X_3}

Fuzzy sets A={0.2/X_1 ,0.7/X_2 ,0.6/X_3} and B={0.7/X_1 ,0.3/X_2 ,0.5/X_3}

A U B= {0.7/X_1, 0.7/X_2, 0.6/X_3}

A ∩ B= {0.2/X1, 0.3/X2, 0.5/X3}

A^c = {0.8/X_1, 0.3/X_2, 0.4/X_3}

Laws of Set Theory

- The laws of Crisp set theory also holds for fuzzy set theory (verify them)
- These laws are listed below:
  - Commutativity: $A \cup B = B \cup A$
  - Associativity: $A \cup (B \cup C) = (A \cup B) \cup C$
  - Distributivity: $A \cup (B \cap C) = (A \cap C) \cup (B \cap C)$

    $A \cap (B \cup C) = (A \cup C) \cap (B \cup C)$
  - De Morgan's Law: $(A \cup B)^{C} = A^{C} \cap B^{C}$

    $(A \cap B)^{C} = A^{C} \cup B^{C}$

Distributivity Property Proof

- Let Universe $U=\{x_1,x_2,\ldots x_n\}$

$p_i = \mu_{A\cup(B\cap C)}(x_i)$

$= \max[\mu_A(x_i), \mu_{(B\cap C)}(x_i)]$

$= \max[\mu_A(x_i), \min(\mu_B(x_i),\mu_C(x_i))]$

$q_i = \mu_{(A\cup B)\cap(A\cup C)}(x_i)$

$= \min[\max(\mu_A(x_i), \mu_B(x_i)), \max(\mu_A(x_i), \mu_C(x_i))]$

Distributivity Property Proof

- Case I: $0<\mu_C<\mu_B<\mu_A<1$

  $p_i = \max[\mu_A(x_i), \min(\mu_B(x_i),\mu_C(x_i))]$

  $= \max[\mu_A(x_i), \mu_C(x_i)]=\mu_A(x_i)$

  $q_i = \min[\max(\mu_A(x_i), \mu_B(x_i)), \max(\mu_A(x_i), \mu_C(x_i))]$

  $= \min[\mu_A(x_i), \mu_A(x_i)]=\mu_A(x_i)$

- Case II: $0<\mu_C<\mu_A<\mu_B<1$

  $p_i = \max[\mu_A(x_i), \min(\mu_B(x_i),\mu_C(x_i))]$

  $= \max[\mu_A(x_i), \mu_C(x_i)]=\mu_A(x_i)$

  $q_i = \min[\max(\mu_A(x_i), \mu_B(x_i)), \max(\mu_A(x_i), \mu_C(x_i))]$

  $= \min[\mu_B(x_i), \mu_A(x_i)]=\mu_A(x_i)$

  Prove it for rest of the 4 cases.