CS 215- Data Interpretation and Analysis
(Post Midsem)

Closure, Fuzzy Implication

Pushpak Bhattacharyya
Computer Science and Engineering Department
IIT Bombay

Lecture-6

9nov23

Recap

ML and Hypothesis Testing

Bias detection

	English				
	Muril	**XLMR**	**Mbert**	**Bernice**	**IndicBERT**
Gender	47.08	**50.26**	47.61	56.61	52.38
Socioeconomic	58.77	64.03	**53.50**	54.38	63.15
Age	**49.15**	44.06	47.45	45.76	54.23
Physical-appearance	58.13	58.13	**53.48**	62.79	62.79
Disability	65.51	68.96	58.62	72.41	**51.72**
Crows Pair	52.87	55.17	**50.75**	56.09	56.55

From our AAAI submitted paper on Bias detection

Hypothesis testing on bias

- H_0: $\mu_1 = \mu_2$ (there is no diff in mean biasedness)
- H_A: $\mu_1 <> \mu_2$
- Under H_0, $\mu_{\theta1-\theta2}=0$, θ_1 and θ_2 are sample means
- $\sigma_{\theta1-\theta2}= [37.84/6+ 21.56/6]^{1/2}=3.15$
- Z

$=[\{(\theta_1-\theta_2)-0\}/\ \sigma_{\theta1-\theta2}]$

$=[(54.83-56.33)/3.15]=-0.48$

Examination of H_0

test-type (col)			
vs.	Two-Tail	1 sided to +inf	1 sided from -inf
Confidence Interval (significance level)			
90% (0.10)	(- and +) 1.65	-1.28 to +inf	-inf to +1.28
95% (0.05)	(- and +) 1.96	-1.65 to +inf	-inf to +1.65
99% (0.01)	(- and +) 2.58	-2.33 to +inf	-inf to 2.33

- Cannot reject H_0: for 95% CI, since -1.96<-0.48<+1.96
- Nor for 99%, nor for 90%

Conclusion for relative biasedness of MURIL and INDICBERT

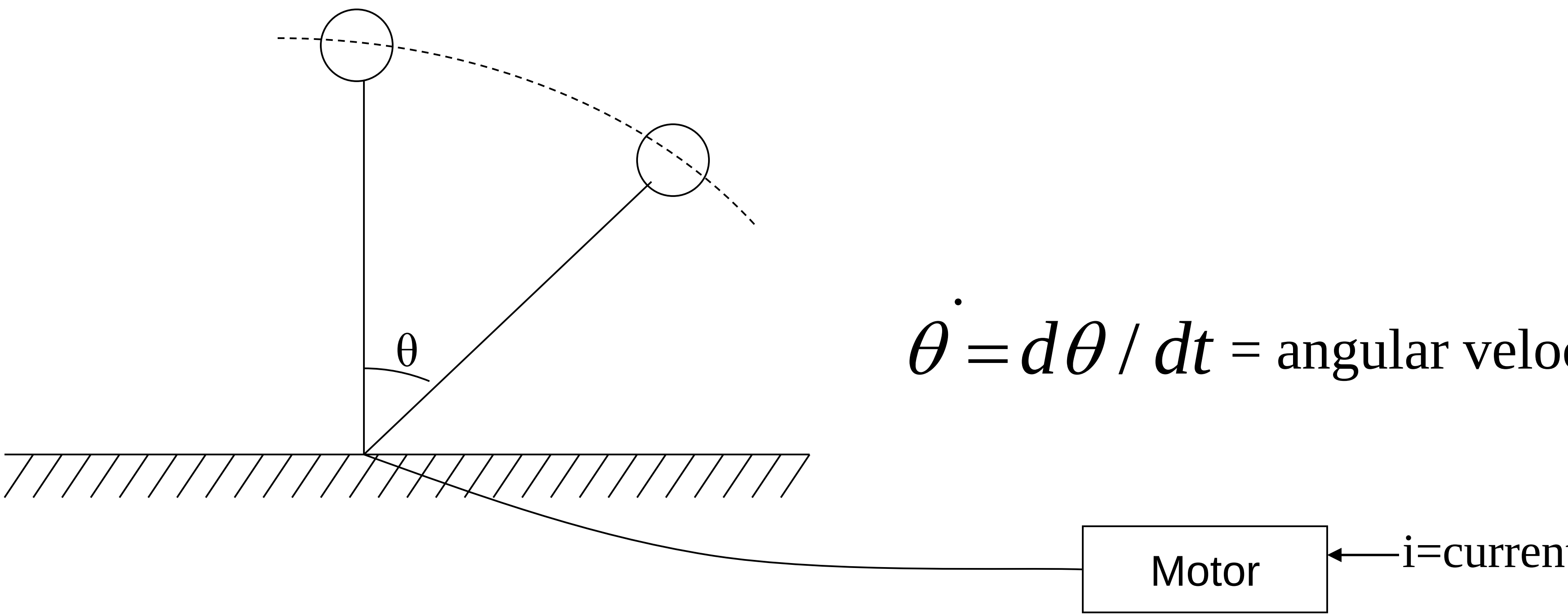
- There is 95% probability that MURIL and INDICBERT are equally biased

	English				
	Muril	**XLMR**	**Mbert**	**Bernice**	**IndicBERT**
Gender	47.08	**50.26**	47.61	56.61	52.38
Socioeconomic	58.77	64.03	**53.50**	54.38	63.15
Age	**49.15**	44.06	47.45	45.76	54.23
Physical-appearance	58.13	58.13	**53.48**	62.79	62.79
Disability	65.51	68.96	58.62	72.41	**51.72**
Crows Pair	52.87	55.17	**50.75**	56.09	56.55

Fuzzification and Defuzzification- Control of inverted pendulum

An Example

Controlling an inverted pendulum:



$$\dot{\theta} = d\theta / dt$$ = angular velocity

Control Matrix

$\dot{\theta}$ \ θ	-ve med	-ve small	Zero	+ve small	+ve med
-ve med					
-ve small		+ve med	+ve small	Zero	
Zero		+ve small	Zero	-ve small	
+ve small		Zero	-ve small	-ve med	
+ve med					

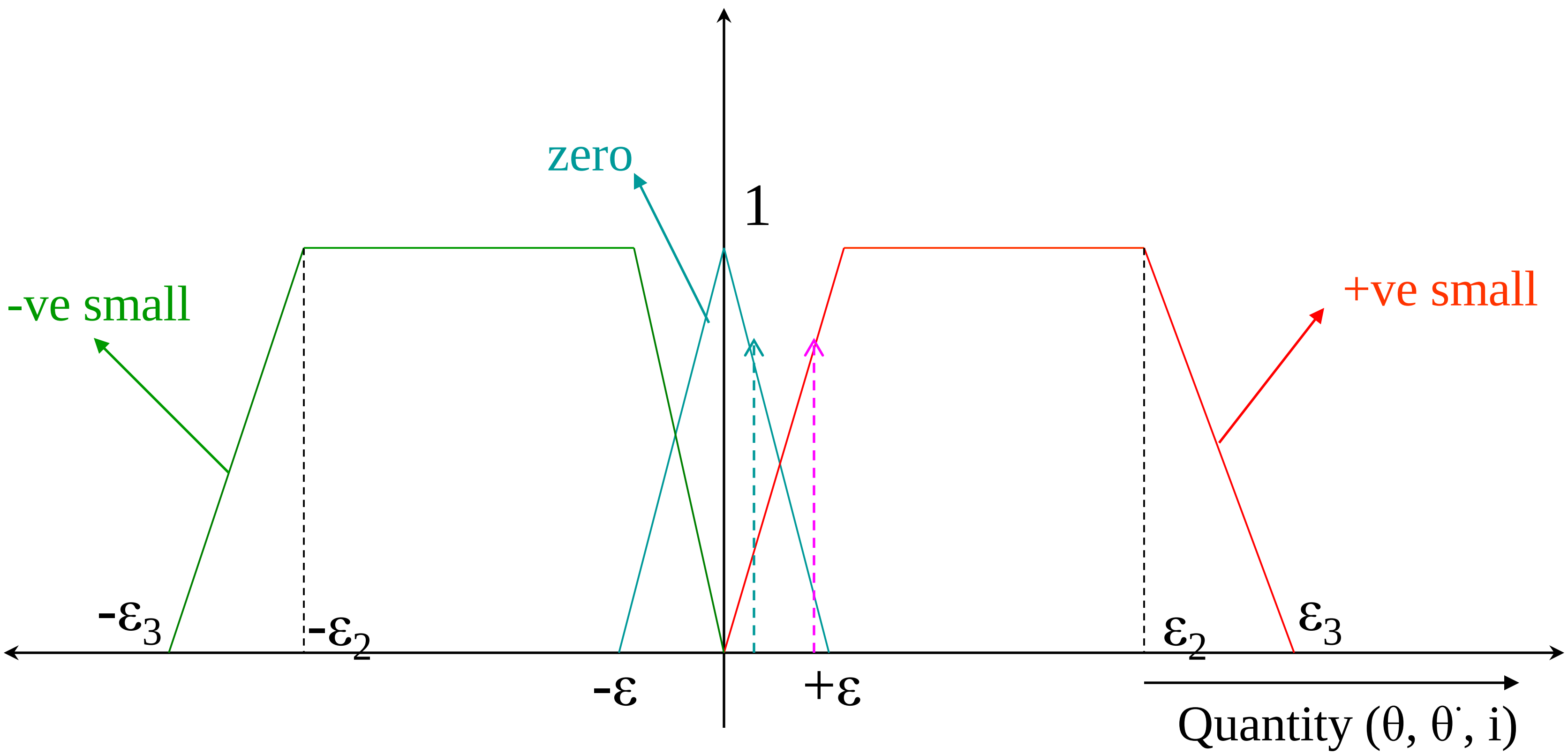
Region of interest

Inference procedure

1. Read actual numerical values of θ and $\dot{\theta}$
2. Get the corresponding μ values μ_{Zero}, $\mu_{(+ve\ small)}$, $\mu_{(-ve\ small)}$. This is called FUZZIFICATION
3. For different rules, get the fuzzy I-values from the R.H.S of the rules.
4. "Collate" by some method and get ONE current value. This is called DEFUZZIFICATION
5. Result is one numerical value of 'i'.

Rules Involved

if θ is Zero and dθ/dt is Zero then i is Zero
if θ is Zero and dθ/dt is +ve small then i is –ve small
if θ is +ve small and dθ/dt is Zero then i is –ve small
if θ +ve small and dθ/dt is +ve small then i is -ve medium



Fuzzification

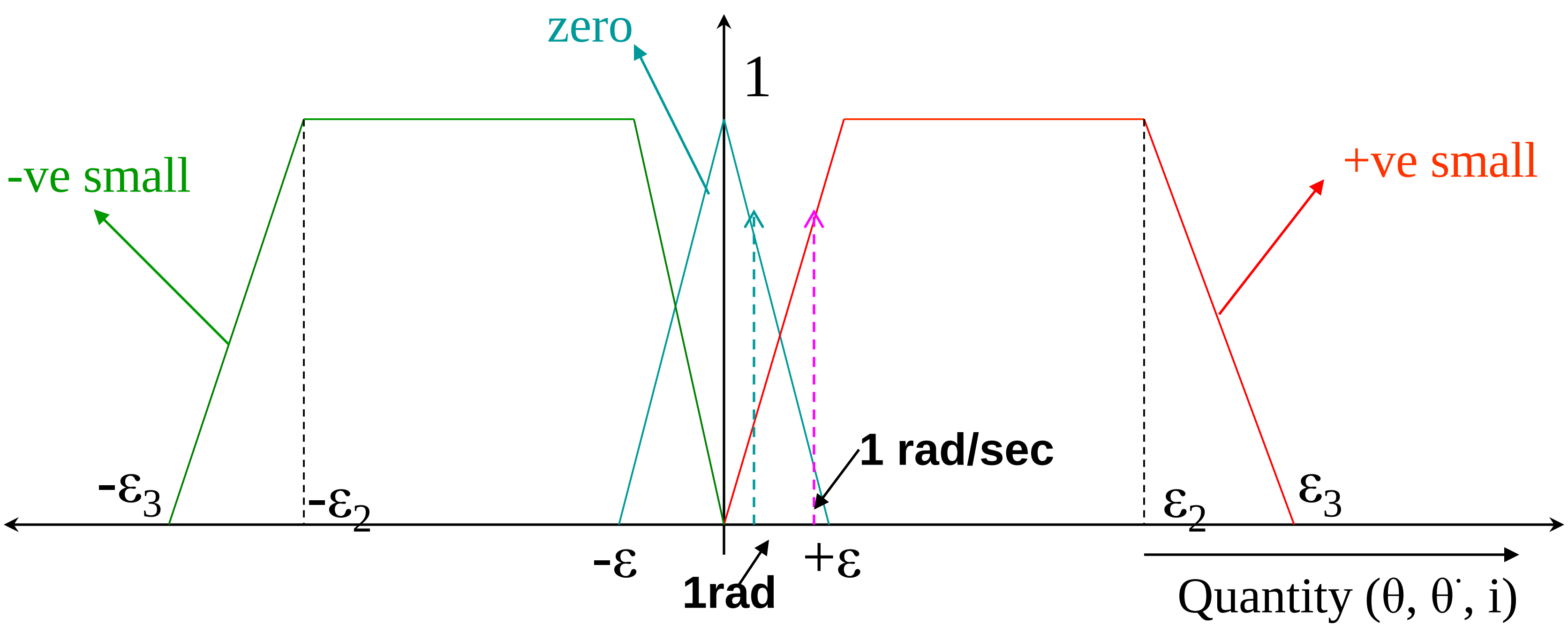
Suppose θ is 1 radian and dθ/dt is 1 rad/sec

μ_{zero}(θ =1)=0.8 (say)

$M_{+ve\text{-}small}$(θ =1)=0.4 (say)

μ_{zero}(dθ/dt =1)=0.3 (say)

$\mu_{+ve\text{-}small}$(dθ/dt =1)=0.7 (say)

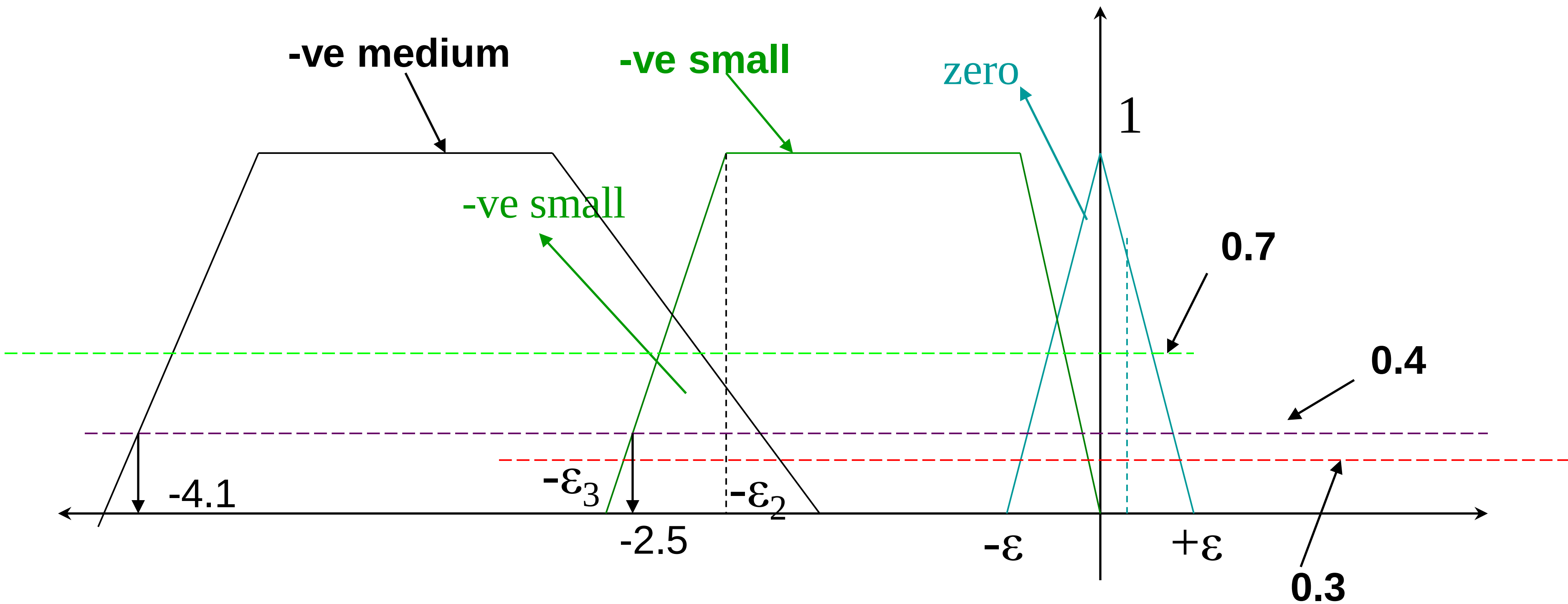


Fuzzification

Suppose θ is 1 radian and dθ/dt is 1 rad/sec
$\mu_{zero}(\theta=1)=0.8$ (say)
$\mu_{+ve\text{-}small}(\theta=1)=0.4$ (say)
$\mu_{zero}(d\theta/dt=1)=0.3$ (say)
$\mu_{+ve\text{-}small}(d\theta/dt=1)=0.7$ (say)

if θ is Zero and dθ/dt is Zero then i is Zero
min(0.8, 0.3)=0.3
hence $\mu_{zero}(i)=0.3$
if θ is Zero and dθ/dt is +ve small then i is –ve small
min(0.8, 0.7)=0.7
hence $\mu_{-ve\text{-}small}(i)=0.7$
if θ is +ve small and dθ/dt is Zero then i is –ve small
min(0.4, 0.3)=0.3
hence μ-ve-small(i)=0.3
if θ +ve small and dθ/dt is +ve small then i is -ve medium
min(0.4, 0.7)=0.4
hence $\mu_{-ve\text{-}medium}(i)=0.4$

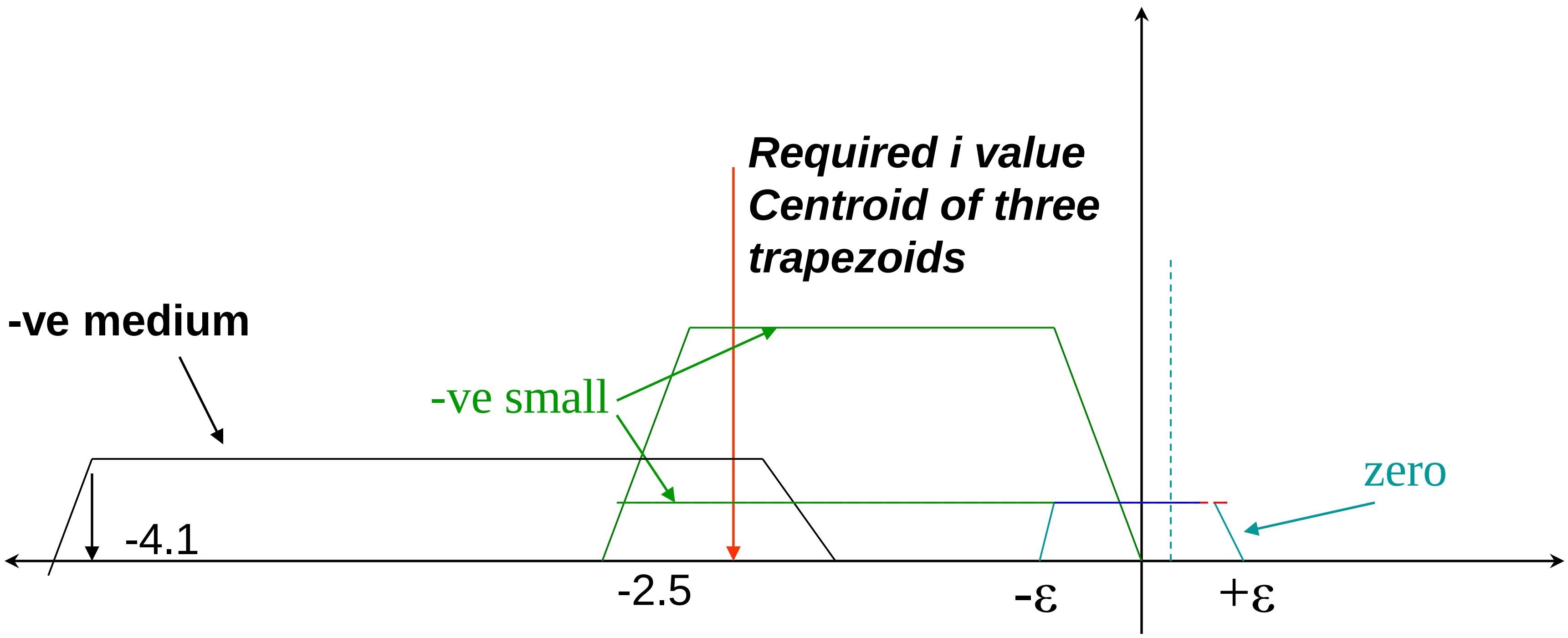
Finding *i*



Possible candidates:

i=0.5 and -0.5 from the "zero" profile and μ=0.3
i=-0.1 and -2.5 from the "-ve-small" profile and μ=0.3
i=-1.7 and -4.1 from the "-ve-small" profile and μ=0.3

Defuzzification: Finding *i* by the *centroid* method



Possible candidates:

i is the x-coord of the centroid of the areas given by the blue trapezium, the green trapeziums and the black trapezium

End Recap

A look at reasoning

- Deduction: $p, p \rightarrow q \vdash q$
- Induction: $p_1, p_2, p_3, \ldots \vdash for\_all\ p$
- Abduction: $q, p \rightarrow q \vdash p$
- Default reasoning: Non-monotonic reasoning: Negation by failure
  - If something cannot be proven, its negation is asserted to be true
  - E.g., in Prolog

How to define subset hood?

Meaning of fuzzy subset

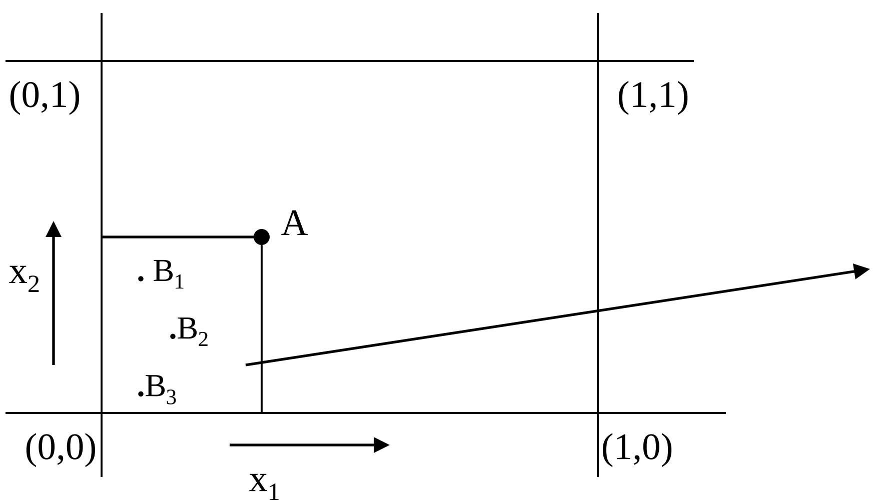
Suppose, following classical set theory we say

$$B \subset A$$

if

$$\mu_B(x) \leq \mu_A(x) \forall x$$

Consider the n-hyperspace representation of A and B



Region where $\mu_B(x) \leq \mu_A(x)$

This effectively means

$B \in P(A)$ CRISPLY

P(A) = Power set of *A*

Eg: Suppose

A = {0,1,0,1,0,1,……………….,0,1} – 10^4 elements

B = {0,0,0,1,0,1,………………….,0,1} – 10^4 elements

Isn't $B \subset A$ with a degree? (only differs in the 2nd element)

Subset operator is the “odd man” out

- AUB, $A \cap B$, A^c are all “Set Constructors” while $A \subseteq B$ is a Boolean Expression or predicate.
- According to classical logic
  - In Crisp Set theory $A \subseteq B$ is defined as

$$\forall x \quad x \in A \Rightarrow x \in B$$

  - So, in fuzzy set theory $A \subseteq B$ can be defined as

$$\forall x \quad \mu_A(x) \Rightarrow \mu_B(x)$$

Zadeh's definition of subsethood goes against the grain of fuzziness theory

- Another way of defining $A \subseteq B$ is as follows:

$$\forall x \quad \mu_A(x) \leq \mu_B(x)$$

But, these two definitions imply that $\mu_{P(B)}(A)=1$
where P(B) is the power set of B

Thus, these two definitions violate the fuzzy principle that every belongingness except Universe is fuzzy

Fuzzy definition of subset

Measured in terms of "fit violation", i.e. violating the condition $\mu_B(x) \leq \mu_A(x)$

Degree of subset hood S(A,B)= 1- degree of superset

$$= \quad 1 - \frac{\sum_x \max(0, \mu_B(x) - \mu_A(x))}{m(B)}$$

m(B) = cardinality of B

$$= \quad \sum_x \mu_B(x)$$

We can show that $E(A) = S(A \cup A^c, A \cap A^c)$

Exercise 1:

Show the relationship between entropy and subset hood

Exercise 2:

Prove that

$$S(B, A) = m(A \cap B) / m(B)$$

Subset hood of B in A

Fuzzy sets to fuzzy logic

Forms the foundation of fuzzy rule based system or fuzzy expert system

Expert System

Rules are of the form

If

$$C_1 \wedge C_2 \wedge \ldots\ldots\ldots\ldots C_n$$

then

A_i

Where C_is are conditions

Eg: C_1=Colour of the eye yellow

C_2= has fever

C_3=high bilurubin

A = hepatitis

In fuzzy logic we have fuzzy predicates

<u>Classical logic</u>

$P(x_1,x_2,x_3.....x_n) = 0/1$

<u>Fuzzy Logic</u>

$P(x_1,x_2,x_3.....x_n) = [0,1]$

<u>Fuzzy OR</u>

$$P(x) \vee Q(y) = \max(P(x), Q(y))$$

<u>Fuzzy AND</u>

$$P(x) \wedge Q(y) = \min(P(x), Q(y))$$

<u>Fuzzy NOT</u>

$$\sim P(x) = 1 - P(x)$$

Fuzzy Implication

- Many theories have been advanced and many expressions exist
- The most used is Lukasiewitz formula
- t(P) = truth value of a proposition/predicate. In fuzzy logic t(P) = [0,1]
- $\underbrace{t(P \rightarrow Q) = \min[1, 1 - t(P) + t(Q)]}_{\text{Lukasiewitz definition of implication}}$

Fuzzy Inferencing

- Two methods of inferencing in classical logic
  - Modus Ponens
    - Given p and $p \rightarrow q$, infer q
  - Modus Tolens
    - Given $\sim q$ and $p \rightarrow q$, infer $\sim p$
- How is fuzzy inferencing done?

Fuzzy Modus Ponens in terms of truth values

- Given $t(p)=1$ and $t(p \rightarrow q)=1$, infer $t(q)=1$
- In fuzzy logic,
  - given $t(p)>=a$, $0<=a<=1$
  - *and* $t(p \rightarrow >q)=c$, $0<=c<=1$
  - What is $t(q)$
- How much of truth is transferred over the channel

$$p \Rightarrow q$$

Lukasiewitz formula for Fuzzy Implication

- t(P) = truth value of a proposition/predicate. In fuzzy logic t(P) = [0,1]
- t($P \rightarrow Q$) = min[1,1 -t(P)+t(Q)]

Lukasiewitz definition of implication

Use Lukasiewitz definition

- $t(p \rightarrow q) = min[1, 1 - t(p) + t(q)]$
- We have $t(p\text{-}>q)=c$, *i.e.*, $min[1,1 - t(p)+t(q)]=c$
- Case 1:
- $c=1$ gives $1 - t(p)+t(q)>=1$, *i.e.*, $t(q)>=a$
- Otherwise, $1 - t(p)+t(q)=c$, *i.e.*, $t(q)>=c+a-1$
- Combining, $t(q)=max(0,a+c-1)$
- This is the amount of truth transferred over the channel $p \rightarrow q$

Two equations consistent

$$Sub(B,A) = 1 - Sup(B,A)$$

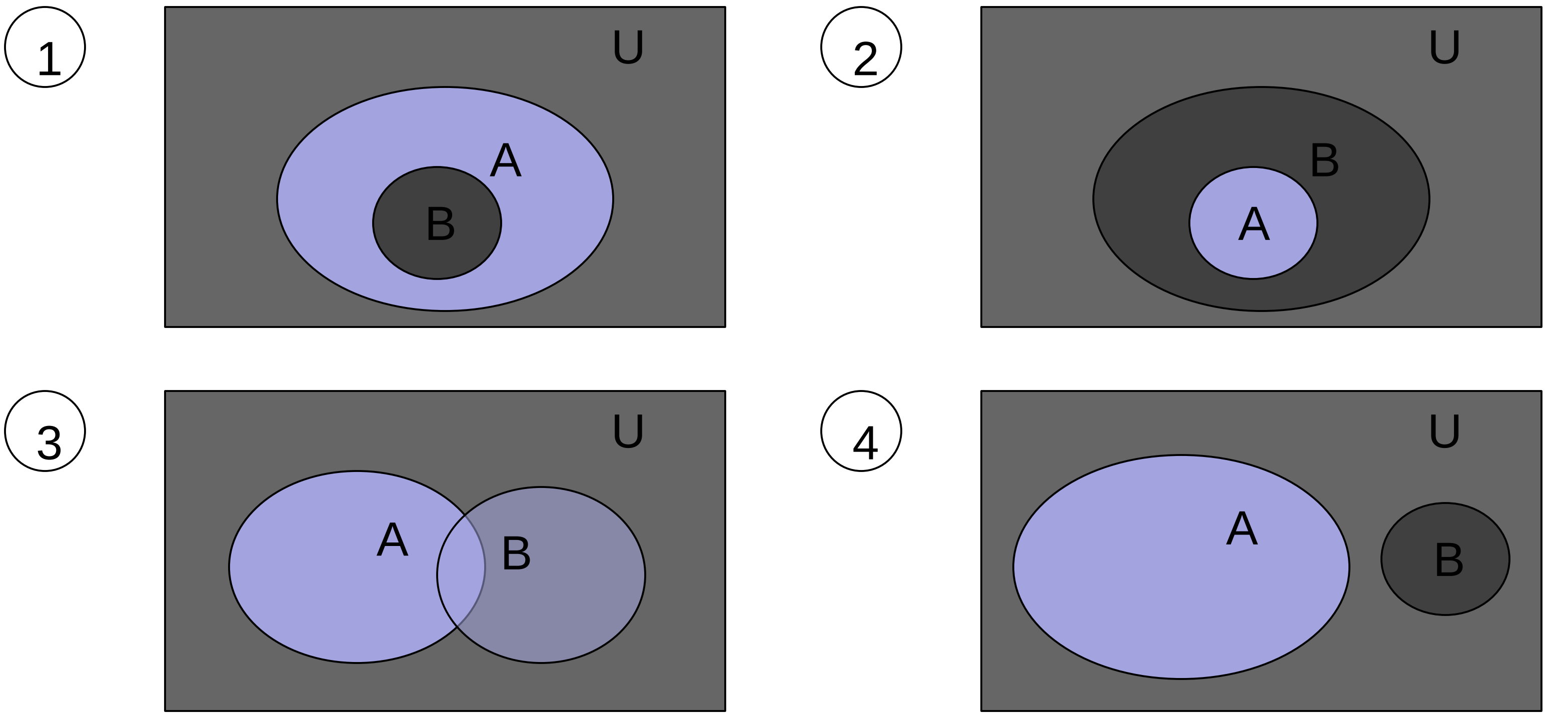
$$= 1 - \frac{\sum_{x_i \in U} \max(0, \mu_B(x_i) - \mu_A(x_i))}{\sum_{x_i \in U} \mu_B(x_i)} \qquad \text{where } U = \{x_1, x_2, \ldots, x_n\}$$

$$t(\mu_B(x_i) \rightarrow \mu_A(x_i)) = \min(1, 1 - t(\mu_B(x_i)) + t(\mu_A(x_i)))$$

- These two equations are consistent with each other

Proof

- Let us consider two crisp sets A and B



Proof (contd…)

- Case I:
  - So, $\mu_A(x_i) = 1$ only when $\mu_B(x_i) = 1$ $\quad So, \mu_B(x_i) - \mu_A(x_i) <= 0$

$$Sub(B, A) = 1 - \frac{\sum_{x_i \in U} \max(0, \mu_B(x_i) - \mu_A(x_i))}{\sum_{x_i \in U} \mu_B(x_i)}$$

$$= 1 - \frac{0}{\sum_{x_i \in U} \mu_B(x_i)} = 1$$

Proof (contd…)

$$Since \qquad \mu_B(x_i) \rightarrow \mu_A(x_i) <= 0$$

$$L = t(\mu_B(x_i) \rightarrow \mu_A(x_i)) = \min(1, 1 - (t(\mu_B(x_i)) - t(\mu_A(x_i))))$$

$$= \min(1, 1 - (-ve)) = 1$$

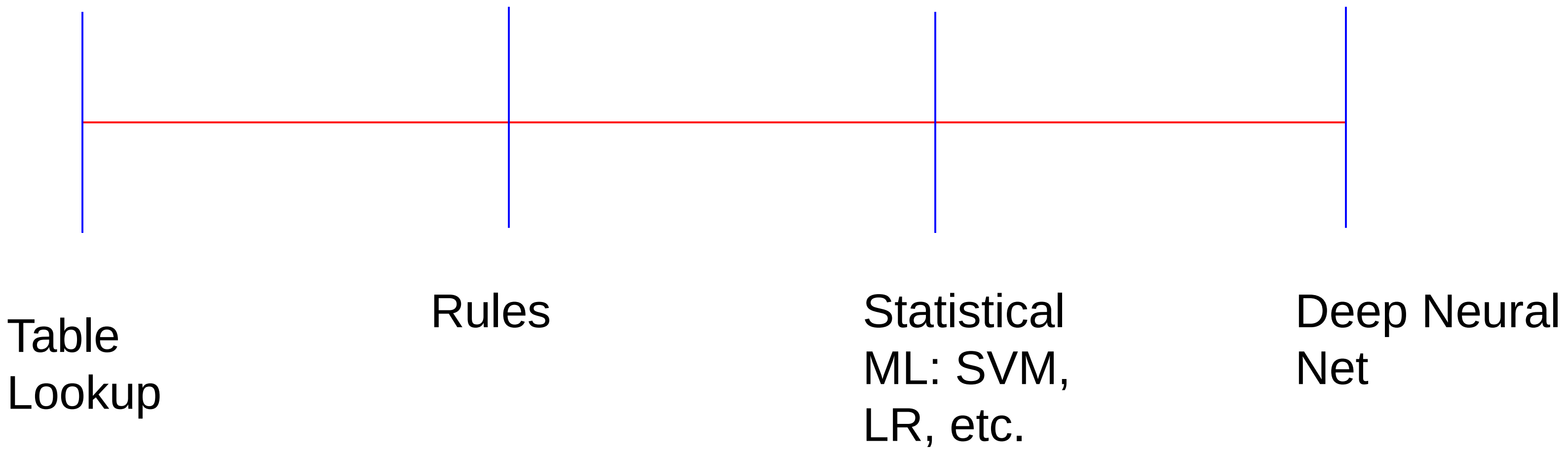
- Thus, in case I these two equations are consistent with each other
- Prove them for other three cases

Closure

Lecture 1

Why probability

A Perspective on Machine Learning



A Practical Problem

- A bridge is being built. The weight it can tolerate has a normal distribution with $\mu=400$ and $\sigma=40$. A car that goes on the bridge has weight distribution (again normal) given by $\mu=3$ and $\sigma=0.3$. We want the probability that the bridge is damaged to be less than *0.1*. How many cars can we allow to go on the bridge?

When does the bridge break?

$$W_{total} > W_{tolerance}$$

We want this event…

$$(W_{total} - W_{tolerance}) > 0$$

$$\Rightarrow \frac{(W_{total} - W_{tolerance}) - (3N - 400)}{\sqrt{0.09N + 1600}} > \frac{-(3N - 400)}{\sqrt{0.09N + 1600}}$$

$$\Rightarrow z > \frac{-(3N - 400)}{\sqrt{0.09N + 1600}}$$

When will this Probability exceed 0.1

$$P\left(z > \frac{-(3N - 400)}{\sqrt{0.09N + 1600}} \right) > 0.1$$

Solving this gives N <= 117

How?

V=1.28

Standard Normal Probabilities

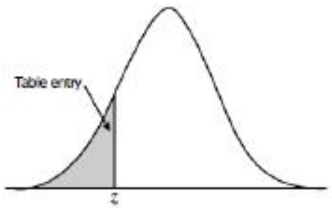


Table entry for z is the area under the standard normal curve to the left of z.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

Standard Normal Probabilities

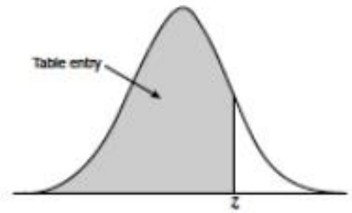


Table entry for z is the area under the standard normal curve to the left of z.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Get N from…

$$1.28 = \frac{-(3N - 400)}{\sqrt{1600 + 0.09N}}$$

N= ~117

End Lecture 1

Lecture 2

Why probability cntd; using the Z-score datasheet

Another Problem

- We have to estimate the percentage of sand grains in a pile of sand resulting from the fragmentation of a mineral compound which fall in a particular range.

Z-score table

Standard Normal Probabilities

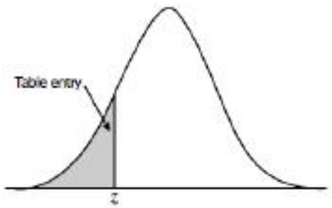


Table entry for z is the area under the standard normal curve to the left of z.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

Standard Normal Probabilities

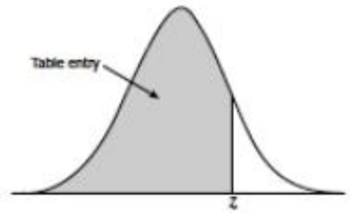
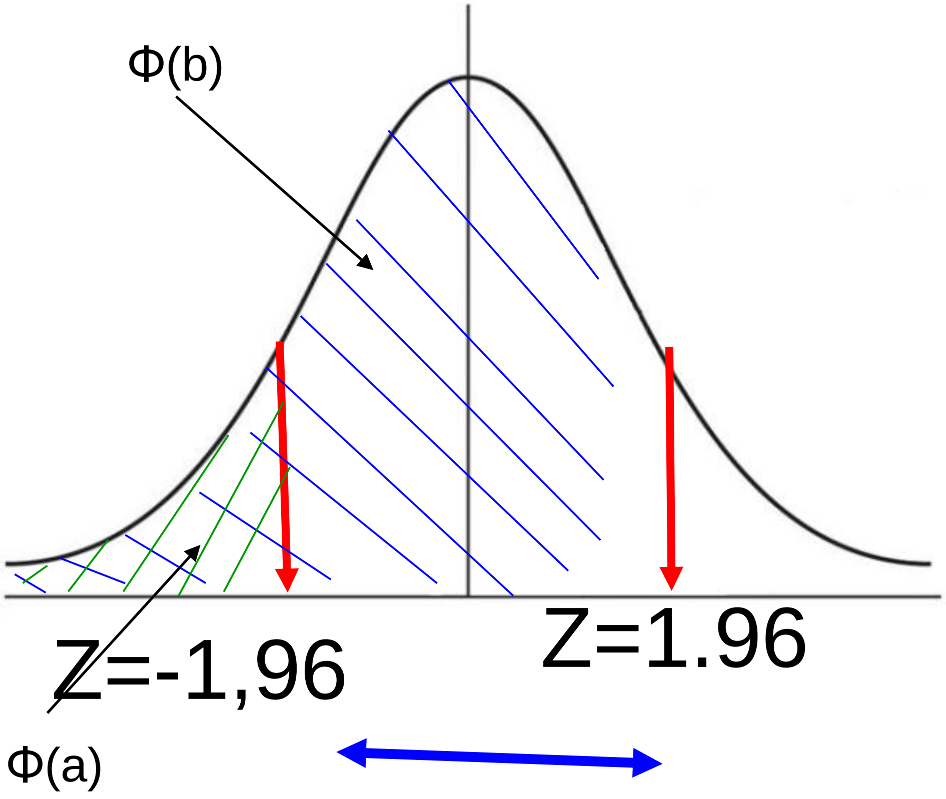


Table entry for z is the area under the standard normal curve to the left of z.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

The 95% confidence interval



-1.96 <= Z <= +1.96

- Φ(1.96)=0.9750
- By symmetry

Φ(-1.96)= 1-Φ(1.96)

➔Φ(1.96)-Φ(-1.96)= 2.Φ(1.96)-1= 2 X 0.975-1.0=1.95-1.0=0.95

Interval Estimate

95% confidence interval; bounds on μ

$$P\left[-1.96 < \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < 1.96\right] = 0.95$$

$$\Rightarrow P\left[-1.96\frac{\sigma}{\sqrt{n}} < \bar{X}-\mu < 1.96\frac{\sigma}{\sqrt{n}}\right] = 0.95$$

$$\Rightarrow P\left[\bar{X}-1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{X}+1.96\frac{\sigma}{\sqrt{n}}\right] = 0.95$$

End Lecture 2

Lecture 3

Central Limit Theorem: the foundation of Hypothesis Testing; MGF and proof of CLT

Statement of Central Limit Theorem

- Let X_1, X_2, X_3, …, X_n be n independent random variables, each with mean μ and variance σ^2
- Also let

$$S_n = X_1 + X_2 + X_3, \ldots + X_n$$

- Then,

the following is **standard normal**

$$S_n^* = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

Mathematical adjustment

$$S_n^{\ *} = \frac{S_n - n\mu}{\sigma\sqrt{n}},\ gives$$

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{\dfrac{S_n}{n} - \mu}{\dfrac{\sigma\sqrt{n}}{n}} = \frac{\overline{X} - \mu}{\dfrac{\sigma}{\sqrt{n}}}$$

Eqv Statement of CLT

Let $X_1, X_2, X_3, \ldots, X_n$ be n independent random variables forming a sample from a population with mean μ and variance σ^2.

Then the sample mean is normally distributed with mean μ and variance σ^2/n.

$$\frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

MGF

Moment Generating Function

$M_X(t)=E(e^{tX})$,

X is a random Variable and

$f(x_j)=P(X=x_j)$

$$M_X(t)=\sum_{j=1}^{n} e^{tx_j} f(x_j)$$

for discrete distribution

$$M_X(t)=\int_{-\infty}^{+\infty} e^{tx} f(x)dx$$

for continuous distribution

Uniqueness Theorem

- Suppose *X* and *Y* are random variables having moment generating functions $M_X(t)$ and $M_Y(t)$ respectively.

- Then *X* and *Y* have the same probability distribution if and only if $M_X(t)=M_Y(t)$ identically.

MGF of *N(0,1)*

$$MGF = \int_{-\infty}^{+\infty} e^{ty} \frac{1}{\sqrt{2\pi}} e^{\left(-y^2/2\right)} dy$$

$$= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{\left(-y^2/2+ty\right)} dy$$

$$= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y^2-2yt+t^2)} . e^{\frac{t^2}{2}} dy$$

$$= e^{\frac{t^2}{2}} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-t)^2} dy$$

$$= e^{\frac{t^2}{2}}$$

Proof of CLT

- To prove that

$$S_n^* = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

- Is standard normal, we will show that

$$M_{S_n^*}(t) = M_Z(t)$$

- i.e., the moment generating function of S_n* is equal to the moment generating function of standard normal r.v.

As n tends to infinity…

$$E(e^{tS_n^*}) = (1 + \frac{t^2}{2n} + \ldots)^n$$

$$Study \;\; L_n = (1 + \frac{t^2}{2n} + \ldots)^n, \; as \; n -> \infty$$

$$\log L_n = n \log(1 + \frac{t^2}{2n} + \ldots)$$

$$= \frac{\log(1 + \frac{t^2}{2n} + \ldots)}{1/n}$$

$$Both \;\; num \; and \;\; denom \rightarrow 0, \quad as \; n -> \infty$$

As n tends to infinity…

$$take\ derivative\ of\ numerator\ and$$

$$numerator\ as\ per\ L'Hospital\ rule$$

$$= \frac{\dfrac{(-\dfrac{t^2}{2n^2})}{(1+\dfrac{t^2}{2n}+\ldots)}}{-1/n^2} = \frac{\dfrac{t^2}{2}}{(1+\dfrac{t^2}{2n}+\ldots)}$$

$$= \frac{t^2}{2},\ as\ n \to \infty$$

$$same\ as\ the\ mgf\ of\ Z$$

End Lecture 3

Lecture 4

H0, HA, Hypothesis testing in logic

Null and Alternative Hypothesis

- H_0: Null Hypothesis → the hypothesis we want to reject
- H_A or H_1: Alternative Hypothesis → opposite of H_0
- We use the sample statistics, trying to reject H_0

Type I and Type II error

- Type I: incorrectly reject H_0, when it should have been accepted.

- Type II: incorrectly accept H_0 when it should have been rejected.

Digression: Hypothesis Testing in Logic

Using Predicate Calculus

Himalayan Club example

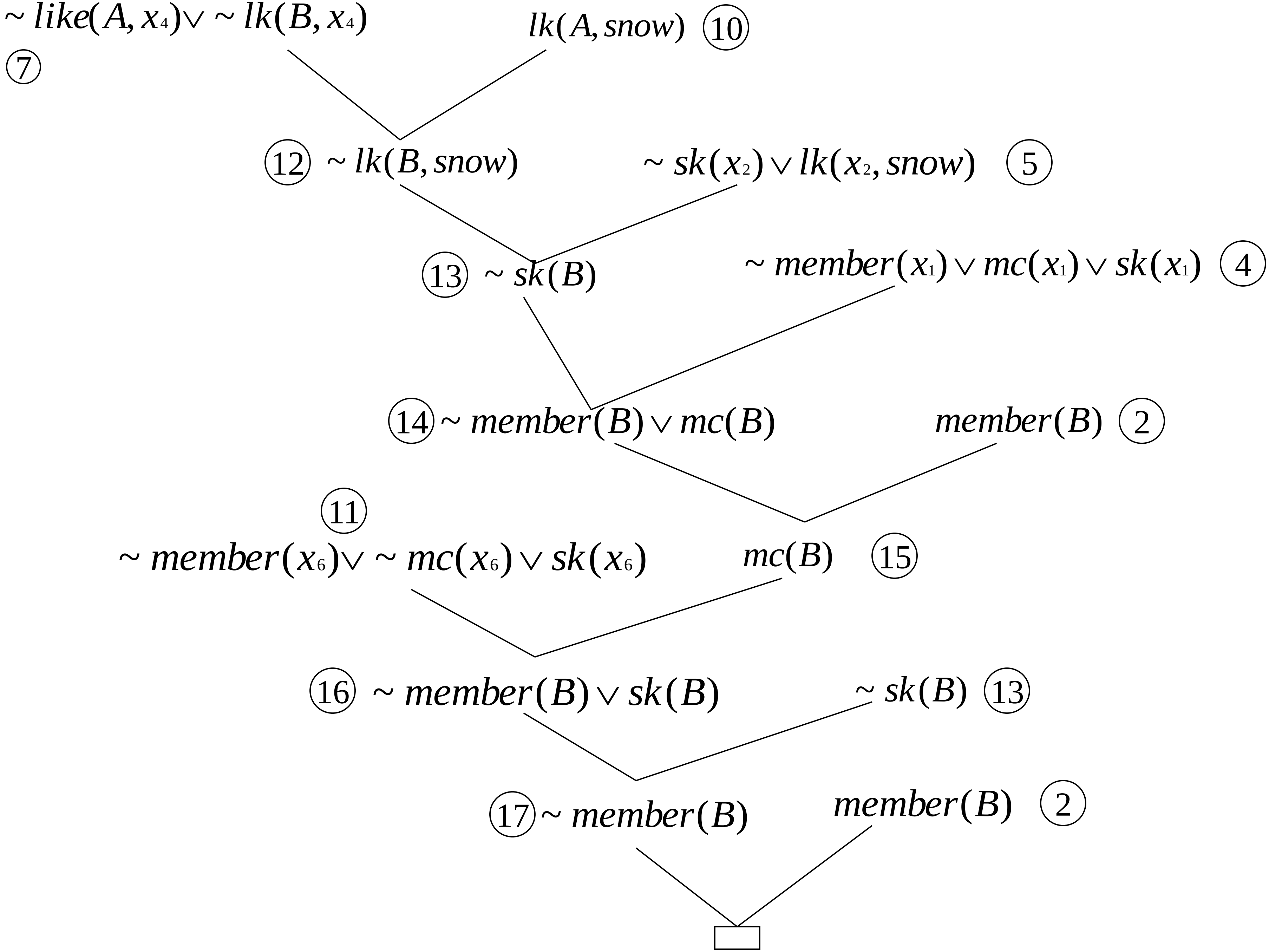
- Introduction through an example *(Zohar Manna, 1974):*
  - Problem: A, B and C belong to the Himalayan club. Every member in the club is either a mountain climber or a skier or both. A likes whatever B dislikes and dislikes whatever B likes. A likes rain and snow. No mountain climber likes rain. Every skier likes snow. *Is there a member who is a mountain climber and not a skier?*
- Given knowledge has:
  - Facts
  - Rules

Null Hypothesis: H_0

- H_0: The club does NOT have any member that is a mountain climber (MC) and not a skier (SK)

- Key question: Under H_0, is the observation valid?

- In other words: is the hypothesis consistent with the data?



~ like(A, x4) ∨ ~ lk(B, x4) 7
lk(A, snow) 10
12 ~ lk(B, snow)
~ sk(x2) ∨ lk(x2, snow) 5
13 ~ sk(B)
~ member(x1) ∨ mc(x1) ∨ sk(x1) 4
14 ~ member(B) ∨ mc(B)
member(B) 2
11
~ member(x6) ∨ ~ mc(x6) ∨ sk(x6)
mc(B) 15
16 ~ member(B) ∨ sk(B)
~ sk(B) 13
17 ~ member(B)
member(B) 2

Principle

- If Hypothesis not consistent with data, hypothesis must be rejected
- Data cannot be rejected
- Data is GOLD!

End Lecture 4

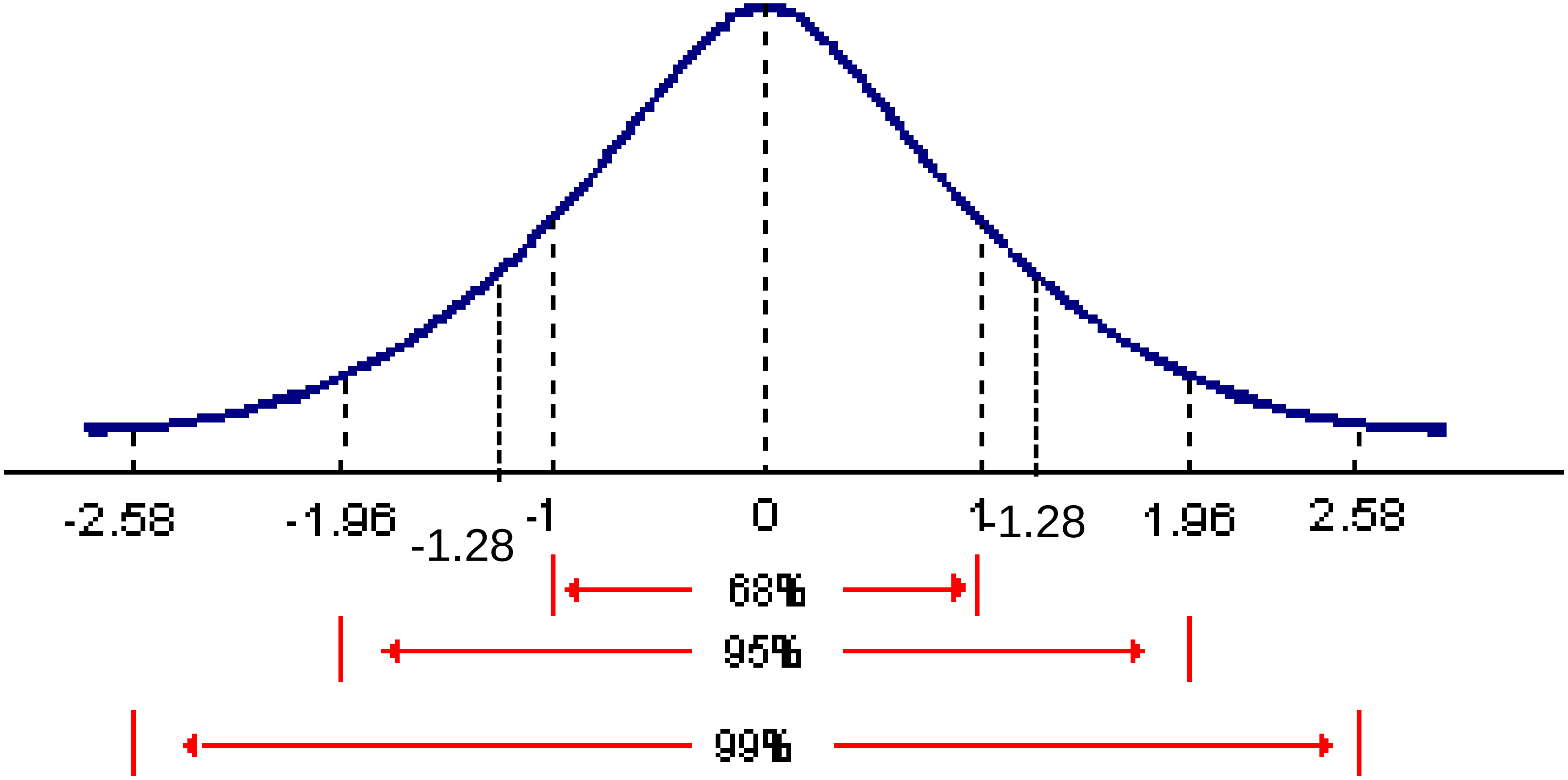
Lecture 5

Instruments of HT

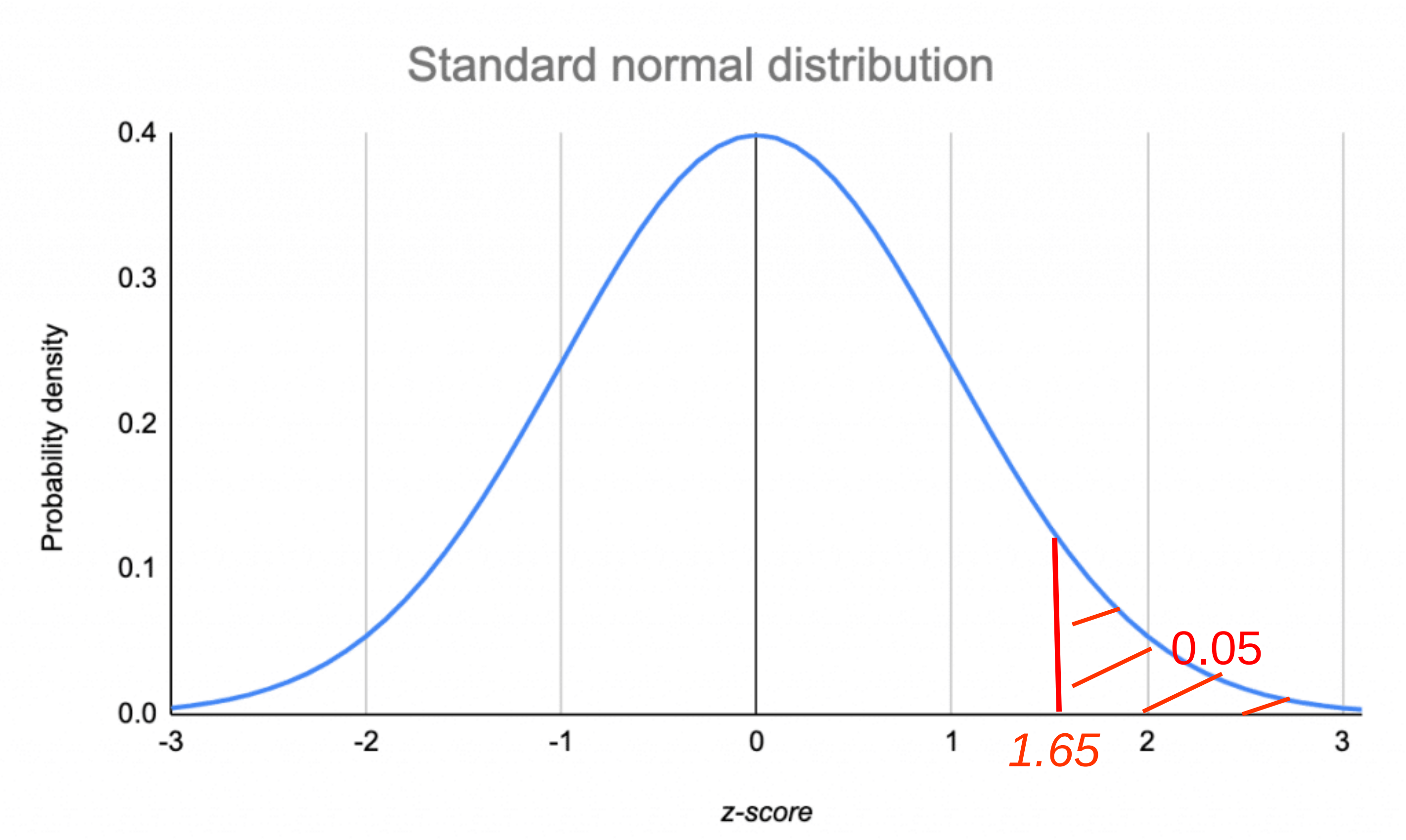
A useful table

test-type (col)			
vs.	Two-Tail	1 sided to +inf	1 sided from -inf
Confidence Interval (significance level)			
90% (0.10)	(- and +) 1.65	-1.28 to +inf	-inf to +1.28
95% (0.05)	(- and +) 1.96	-1.65 to +inf	-inf to +1.65
99% (0.01)	(- and +) 2.58	-2.33 to +inf	-inf to 2.33

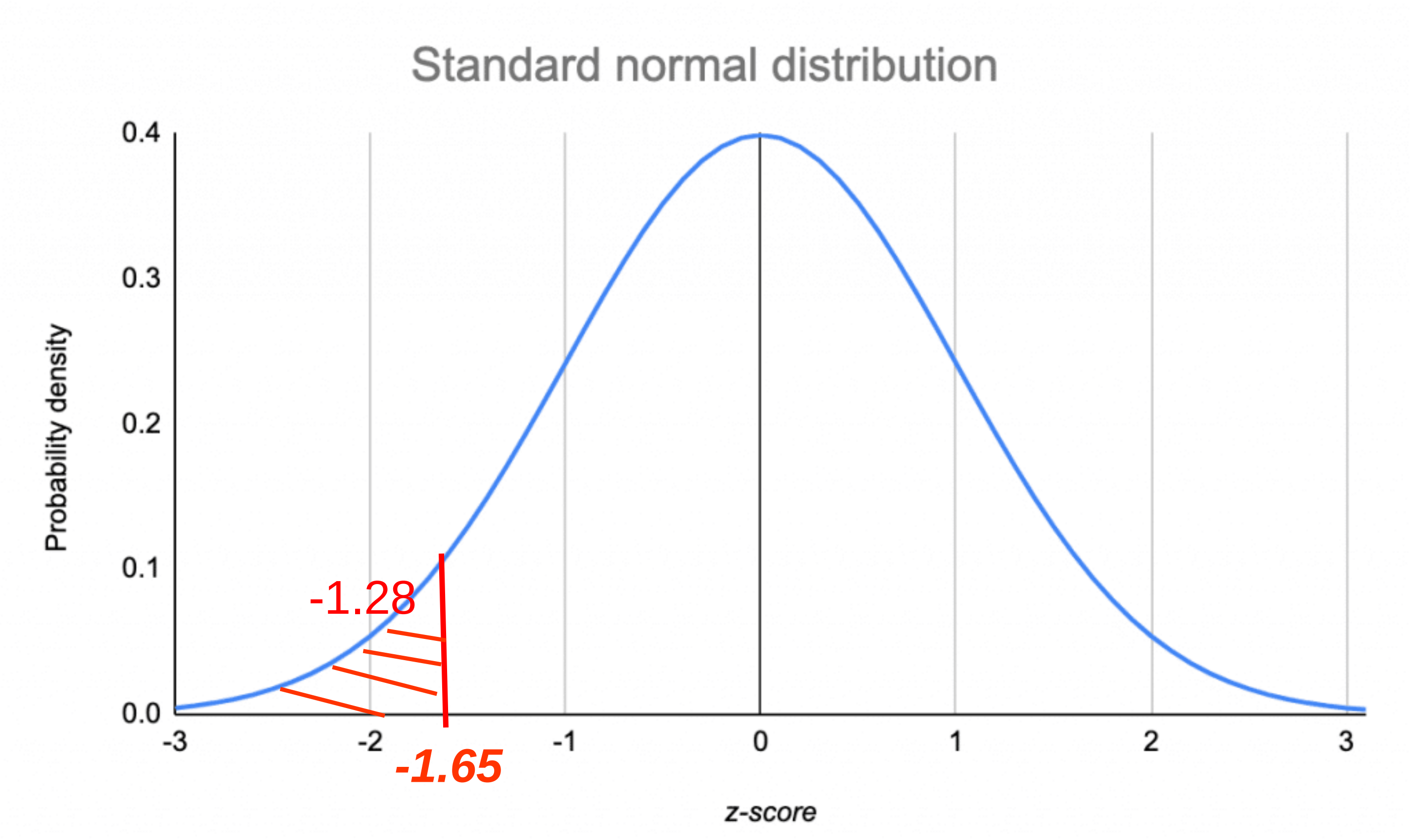
2 sided 95% confidence interval



95% 1-sided confidence interval (from –inf)



95% 1-sided confidence interval (to +inf)



Problem Statement: bottling of fluid

- A factory has a machine that- the factory claims- dispenses 80mL of fluid in a bottle. This needs to be tested. A sample of 40 bottles is taken. The average amount of fluid is 78mL with standard deviation of 2.5. Verify the factory's claim.

https://www.youtube.com/watch?v=zJ8e_wAWUzE

Solution

- Claimed population mean, $\mu=80$
- $n=40$, sample mean, $\mu_{obs}=78$, sample standard deviation, $\sigma_{obs}=2.5$
- H_0: $\mu=80$
- H_A:
  - $\mu \neq 80$ (2-sided test)

2-tailed analysis

- Z_c= +- 1.96
- Z_{obs}=

$$\frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{78 - 80}{\frac{2.5}{\sqrt{40}}}$$

$$= -5\,(approx.)$$

- Falls in rejection region

Z-test based observation (2-tailed)

- $-5 < -1.96$
- We reject the null hypothesis
- The claim that **the machine fills bottles with 80mL** fluid is rejected based on the evidence

99% confidence interval, Z_c=+-2.58

- $-5.0 < -2.58$
- So for 99% confidence interval also the hypothesis is rejected

90% confidence interval, Z_c=+-1.28

- -5.0 < -1.28
- So for 90% confidence interval also the hypothesis is rejected

End Lecture 5

Lecture 6

Type-I and Type-II errors: **Always wrt Null Hypothesis H_0**

as per data / actual	ACCEPT	REJECT
TRUE	***No Error***	***Type- I error***
FALSE	***Type-II Error***	***No Error***

Significance of level of significance α

- 90% confidence interval, $\alpha=0.10$
  - ➔Prepared to tolerate 10% Type-I error
  - ➔Probability of wrong rejection of H_0 is 10%
- 95% confidence interval, $\alpha=0.05$
  - ➔Prepared to tolerate 5% Type-I error
  - ➔Probability of wrong rejection of H_0 is 5%
- 99% confidence interval, $\alpha=0.01$
  - ➔Prepared to tolerate 1% Type-I error
  - ➔Probability of wrong rejection of H_0 is 1%

Nicotine problem

Problem statement *(Sheldon M. Ross, PSES, 2004)*

All cigarettes presently on the market have an average nicotine content of at least 1.6mg per cigarette. A firm that produces cigarettes claims that it has discovered a new way to cure tobacco leaves that will result in the average nicotine content of a cigarette being less than 1.6 mg. To test this claim, a sample of 20 of the firms cigarettes were analysed. If it is known that the standard deviation of a cigarette’s nicotine content is 0.8 mg., what conclusions can be drawn at the 5% level of significance if the average nicotine content of the 20 cigarettes is 1.54?

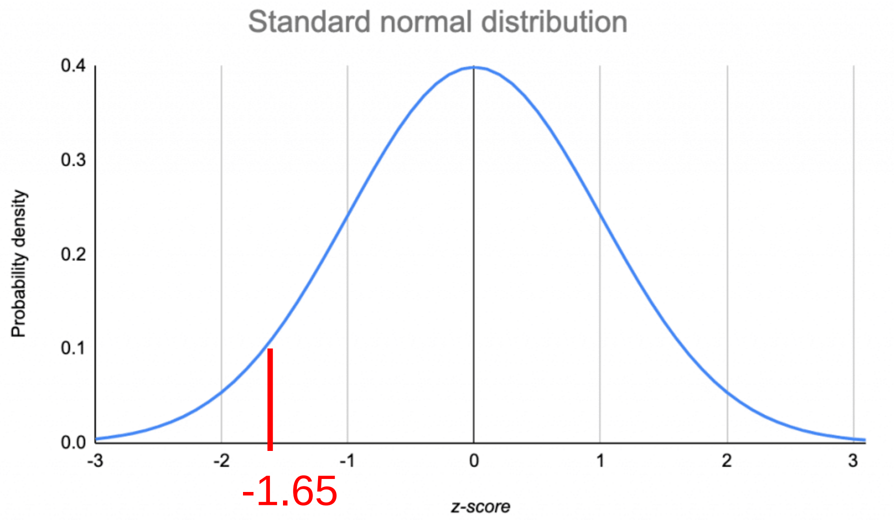
Solution to the Nicotine problem

H_0: μ>=1.6 versus H_1: μ<1.6

With μ=1.6,

$$Z_o = \frac{\sqrt{n}(\overline{X} - \mu\ )}{\sigma} = \frac{\sqrt{20}(1.54 - 1.6)}{0.8} = -0.33$$

$Z_o > Z_c$, so cannot reject H_0



Conclusion from the nicotine problem

- H_0 cannot be rejected
- Data suggests nicotine content is >=1.6 mg
- Company’s claim that the new method of cigarette making ensures < 1.6mg nicotine content is not consistent with the data

End Lecture 6

Lecture 7

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 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

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)

Get ChiSquare Observed

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

$$= 0.675$$

and Compare with ChiSq Critical

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

$$= 9.48$$

$\text{ChiSq}_{\text{obseved}} < \text{ChiSq}_{\text{critical}}$

No reason to reject null hypothesis

H_0= Data follows Poisson distribution

Fitting Normal Distribution

Cricket Score problem

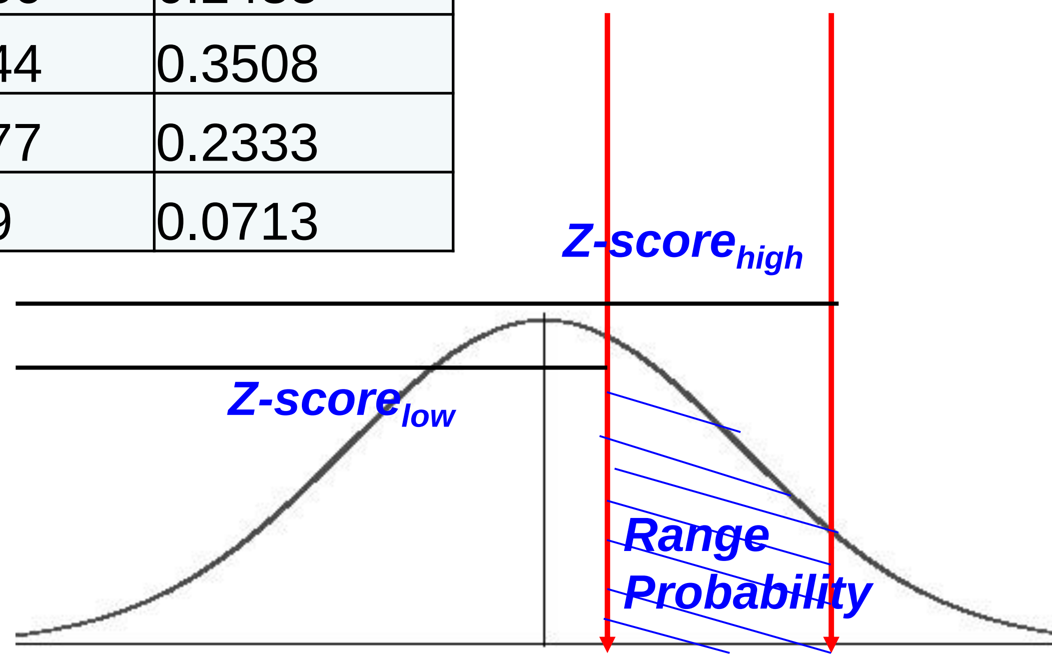
Range	Midpoint (MP)	#innings (I)
0-20	10	10
21-40	30	20
41-60	50	40
61-80	70	20
81-100	90	10
		100

Z_{low} $[=(X_{low}-\mu)/\sigma]$ and Z_{high} $[=(X_{high}-\mu)/\sigma)$

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



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 Lecture 7

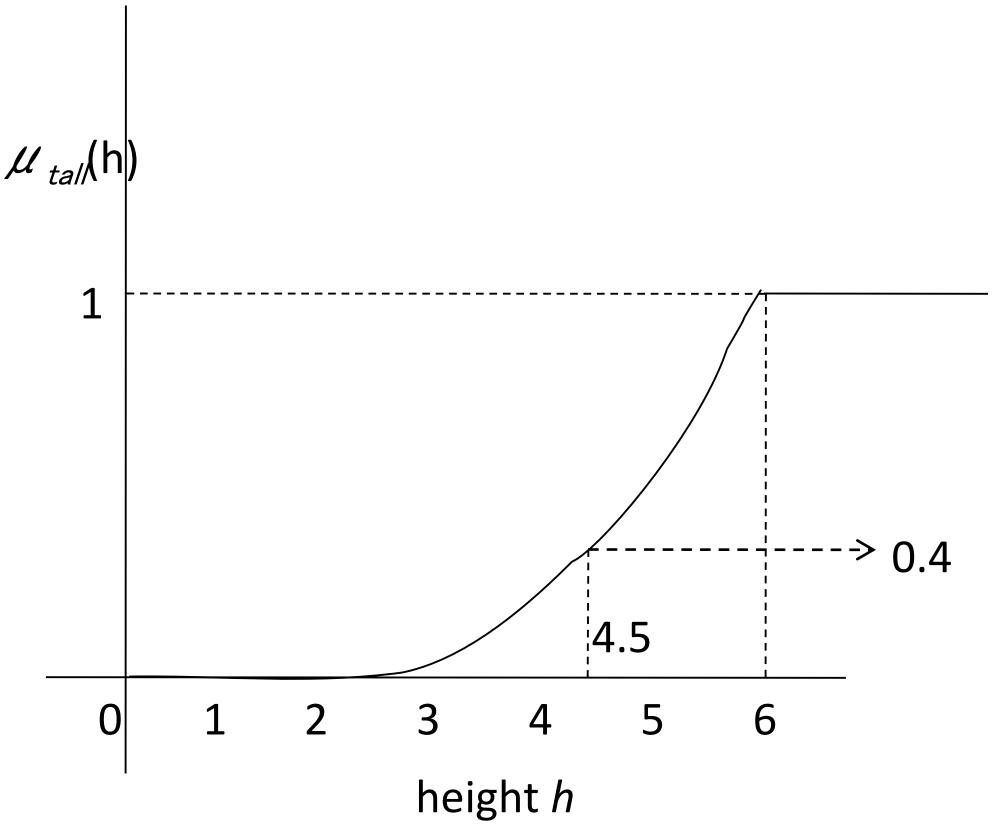
Lecture 8

Fuzzy sets and logic

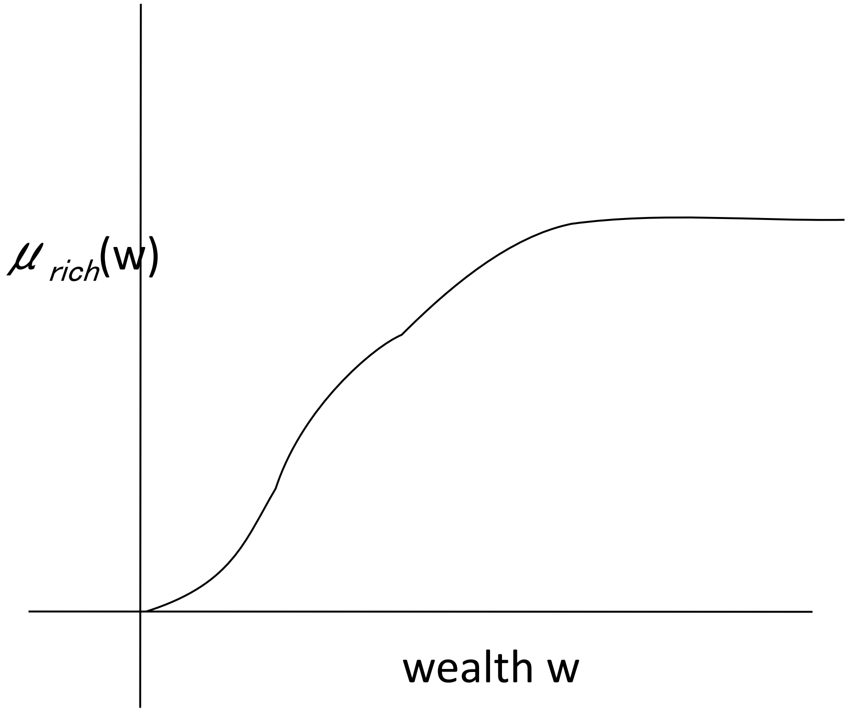
Linguistic Variables

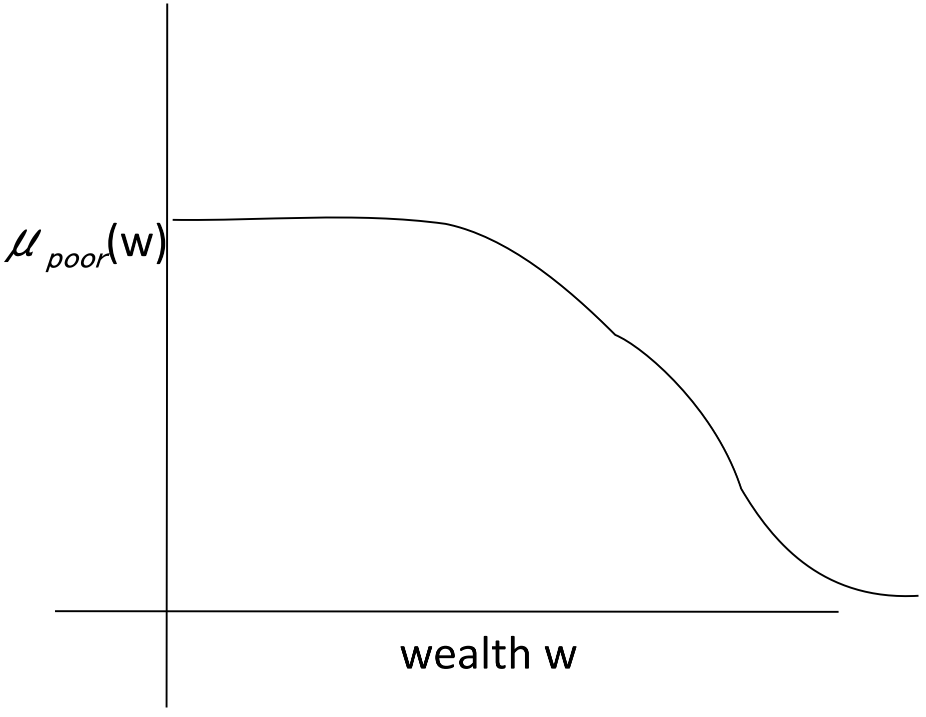
- Fuzzy sets are named by Linguistic Variables (typically adjectives).
- Underlying the LV is a numerical quantity

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

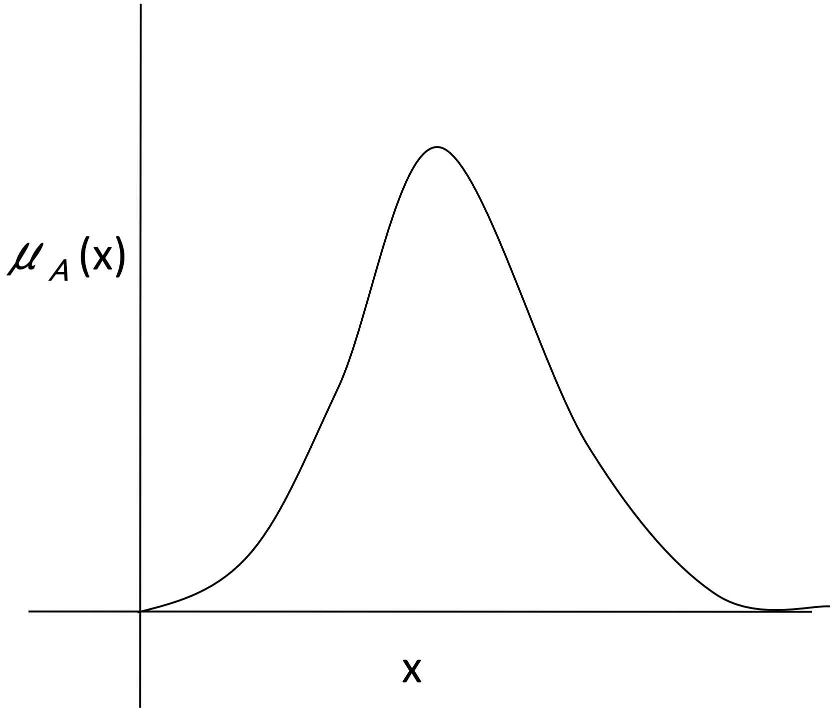


Example Profiles

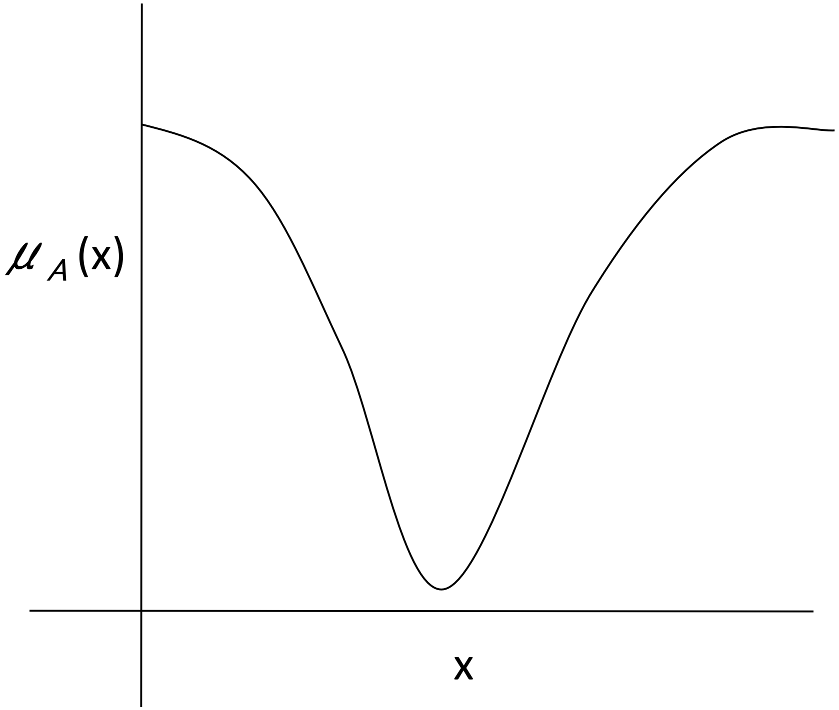




Example Profiles



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



Profile representing
extreme

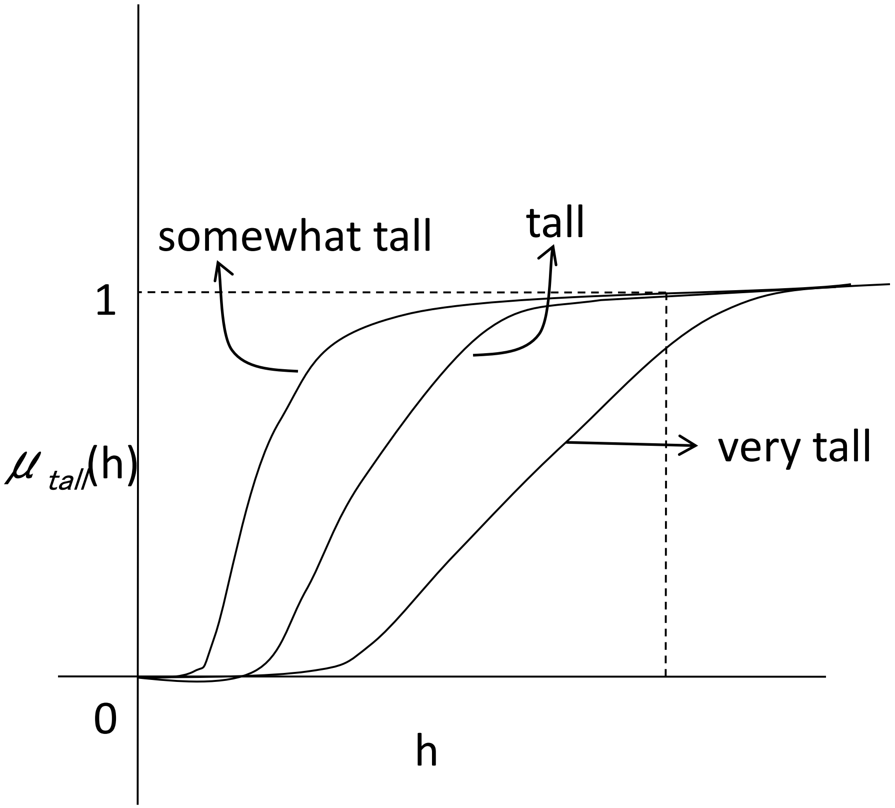
Concept of Hedge

- 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 Set Theory

- 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"

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

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

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

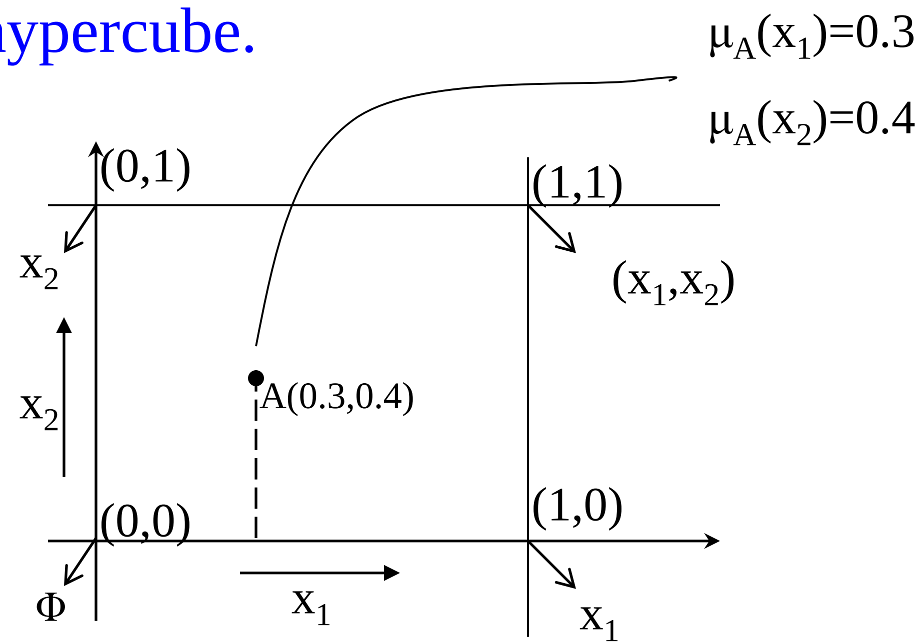
Representation of Fuzzy sets

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 $\{\mu_A(x_1), \mu_A(x_2), \ldots\ldots \mu_A(x_n)\}$

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.

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.

End Lecture 8

Lecture 9

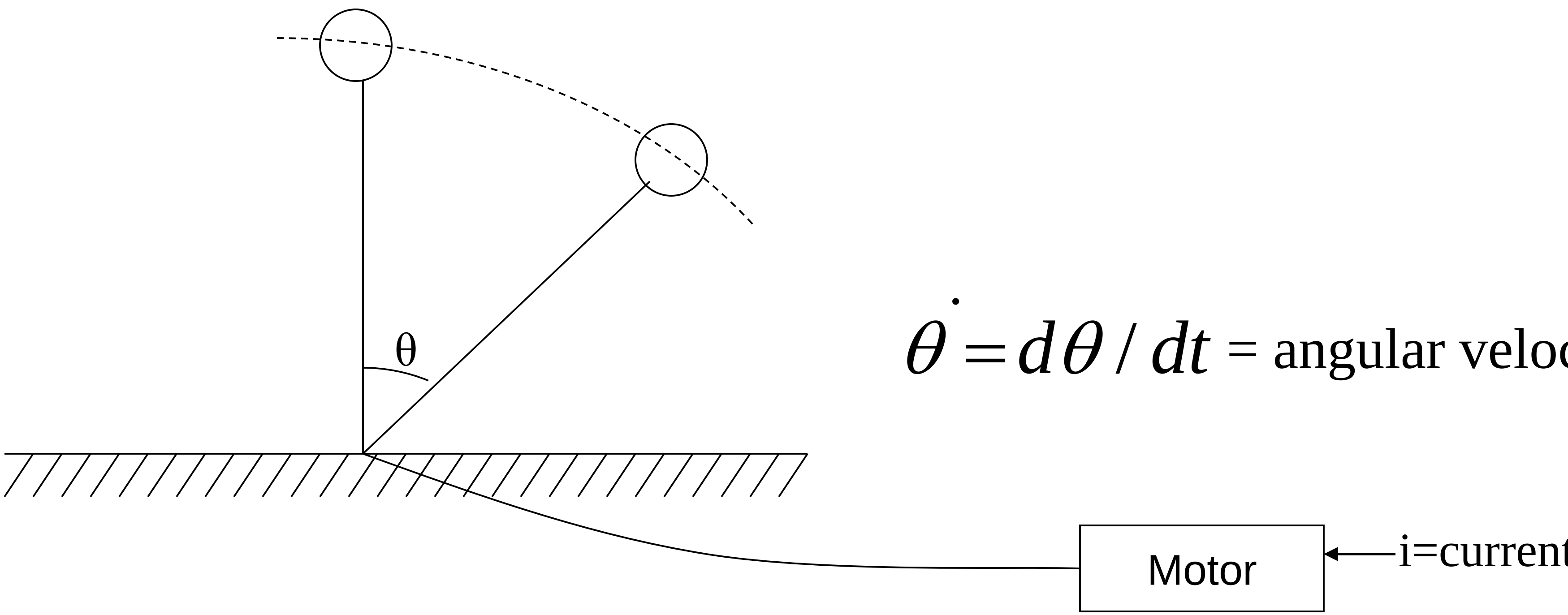
Fuzzy Logic Application

Application of Fuzzy Logic

An example

An Example

Controlling an inverted pendulum:



$$\dot{\theta} = d\theta / dt \text{ = angular velocity}$$

The goal: To keep the pendulum in vertical position ($\theta=0$) in dynamic equilibrium. Whenever the pendulum departs from vertical, a torque is produced by sending a current ‘i’

Controlling factors for appropriate current

Angle θ, Angular velocity $\dot{\theta}$

<u>Some intuitive rules</u>

If θ is +ve small and $\dot{\theta}$ is –ve small

then current is zero

If θ is +ve small and $\dot{\theta}$ is +ve small

then current is –ve medium

Control Matrix

$\dot{\theta}$ \ θ	-ve med	-ve small	Zero	+ve small	+ve med	
-ve med						
-ve small		+ve med	+ve small	Zero		
Zero		+ve small	Zero	-ve small		
+ve small		Zero	-ve small	-ve med		
+ve med						

Region of interest

Each cell is a rule of the form

If θ is <> and $\dot{\theta}$ is <>

then i is <>

<u>4 “Centre rules”</u>

1. if θ = = Zero and $\dot{\theta}$ = = Zero then i = Zero
2. if θ is +ve small and $\dot{\theta}$ = = Zero then i is –ve small
3. if θ is –ve small and $\dot{\theta}$ = = Zero then i is +ve small
4. if θ = = Zero and $\dot{\theta}$ is +ve small then i is –ve small
5. if θ = = Zero and $\dot{\theta}$ is –ve small then i is +ve small

End Lecture 9

Lecture 10

ML and HT, Fuzzification-Defuzziifcation

ML and Hypothesis Testing

Bias detection

	English				
	Muril	**XLMR**	**Mbert**	**Bernice**	**IndicBERT**
Gender	47.08	**50.26**	47.61	56.61	52.38
Socioeconomic	58.77	64.03	**53.50**	54.38	63.15
Age	**49.15**	44.06	47.45	45.76	54.23
Physical-appearance	58.13	58.13	**53.48**	62.79	62.79
Disability	65.51	68.96	58.62	72.41	**51.72**
Crows Pair	52.87	55.17	**50.75**	56.09	56.55

From our AAAI submitted paper on Bias detection

Hypothesis testing on bias

- H_0: $\mu_1 = \mu_2$ (there is no diff in mean biasedness)
- H_A: $\mu_1 <> \mu_2$
- Under H_0, $\mu_{\theta 1-\theta 2}=0$, θ_1 and θ_2 are sample means
- $\sigma_{\theta 1-\theta 2}= [37.84/6+ 21.56/6]^{1/2}=3.15$
- Z

$=[\{(\theta_1-\theta_2)-0\}/ \sigma_{\theta 1-\theta 2}]$

$=[(54.83-56.33)/3.15]=-0.48$

Examination of H_0

test-type (col)			
vs.	Two-Tail	1 sided to +inf	1 sided from -inf
Confidence Interval (significance level)			
90% (0.10)	(- and +) 1.65	-1.28 to +inf	-inf to +1.28
95% (0.05)	(- and +) 1.96	-1.65 to +inf	-inf to +1.65
99% (0.01)	(- and +) 2.58	-2.33 to +inf	-inf to 2.33

- Cannot reject H_0: for 95% CI, since -1.96<-0.48<+1.96
- Nor for 99%, nor for 90%

Conclusion for relative biasedness of MURIL and INDICBERT

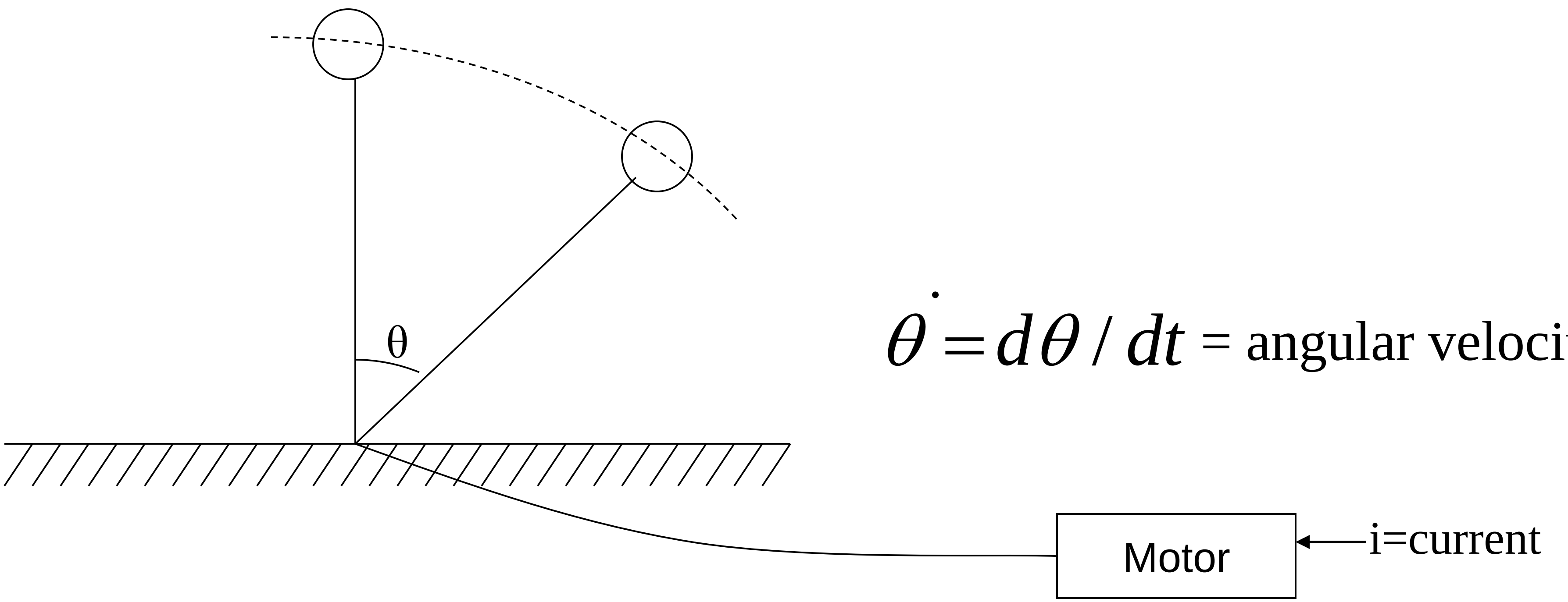
- There is 95% probability that MURIL and INDICBERT are equally biased

	English				
	Muril	**XLMR**	**Mbert**	**Bernice**	**IndicBERT**
Gender	47.08	**50.26**	47.61	56.61	52.38
Socioeconomic	58.77	64.03	**53.50**	54.38	63.15
Age	**49.15**	44.06	47.45	45.76	54.23
Physical-appearance	58.13	58.13	**53.48**	62.79	62.79
Disability	65.51	68.96	58.62	72.41	**51.72**
Crows Pair	52.87	55.17	**50.75**	56.09	56.55

Fuzzification and Defuzzification- Control of inverted pendulum

An Example

Controlling an inverted pendulum:



$$\dot{\theta} = d\theta / dt$$ = angular velocity

Control Matrix

$\dot{\theta}$ \ θ	-ve med	-ve small	Zero	+ve small	+ve med	
-ve med						
-ve small		+ve med	+ve small	Zero		
Zero		+ve small	Zero	-ve small		
+ve small		Zero	-ve small	-ve med		
+ve med						

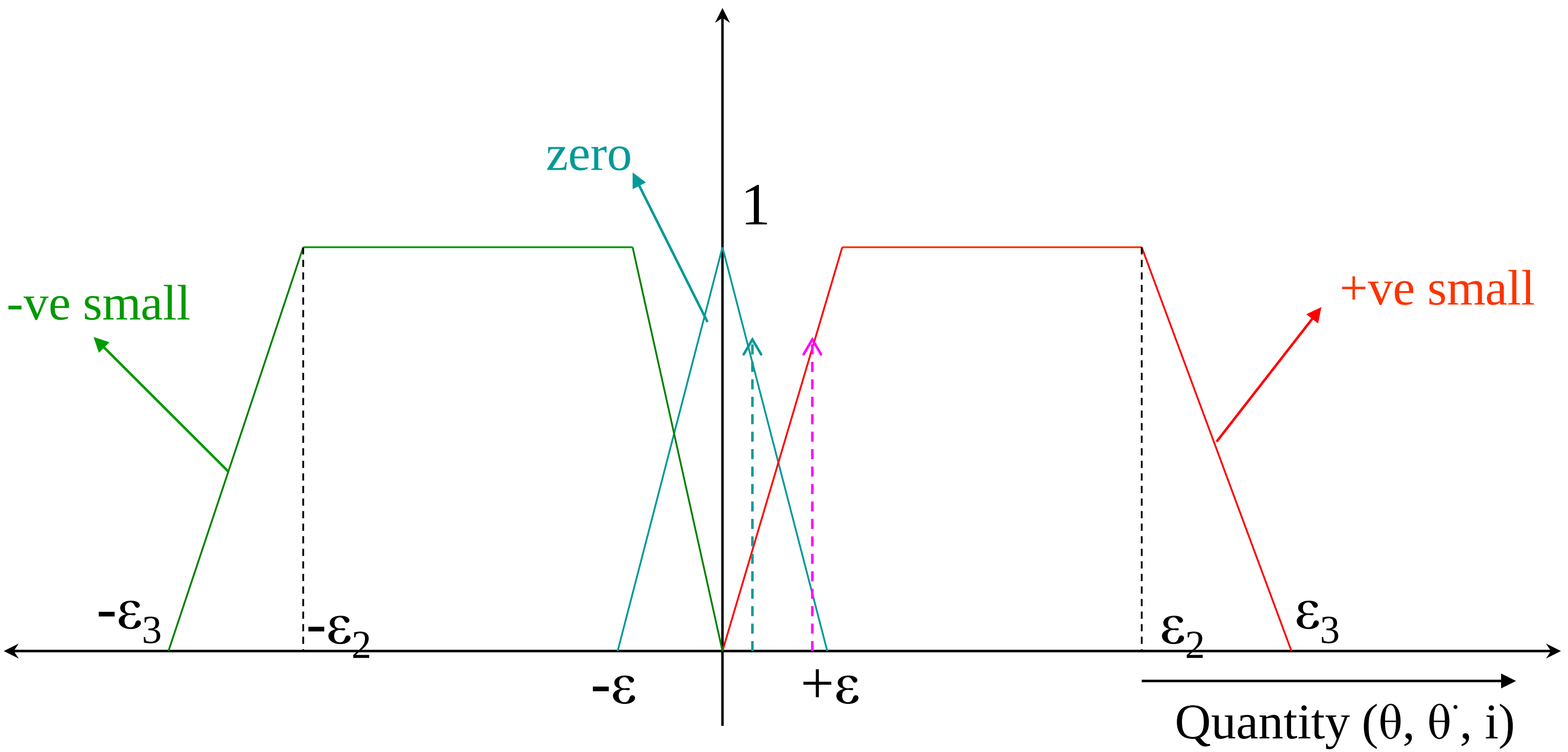
Region of interest

Inference procedure

1. Read actual numerical values of θ and $\dot{\theta}$
2. Get the corresponding μ values μ_{Zero}, $\mu_{(+ve\ small)}$, $\mu_{(-ve\ small)}$. This is called FUZZIFICATION
3. For different rules, get the fuzzy I-values from the R.H.S of the rules.
4. "Collate" by some method and get ONE current value. This is called DEFUZZIFICATION
5. Result is one numerical value of 'i'.

Rules Involved

if θ is Zero and dθ/dt is Zero then i is Zero
if θ is Zero and dθ/dt is +ve small then i is –ve small
if θ is +ve small and dθ/dt is Zero then i is –ve small
if θ +ve small and dθ/dt is +ve small then i is -ve medium



Fuzzification

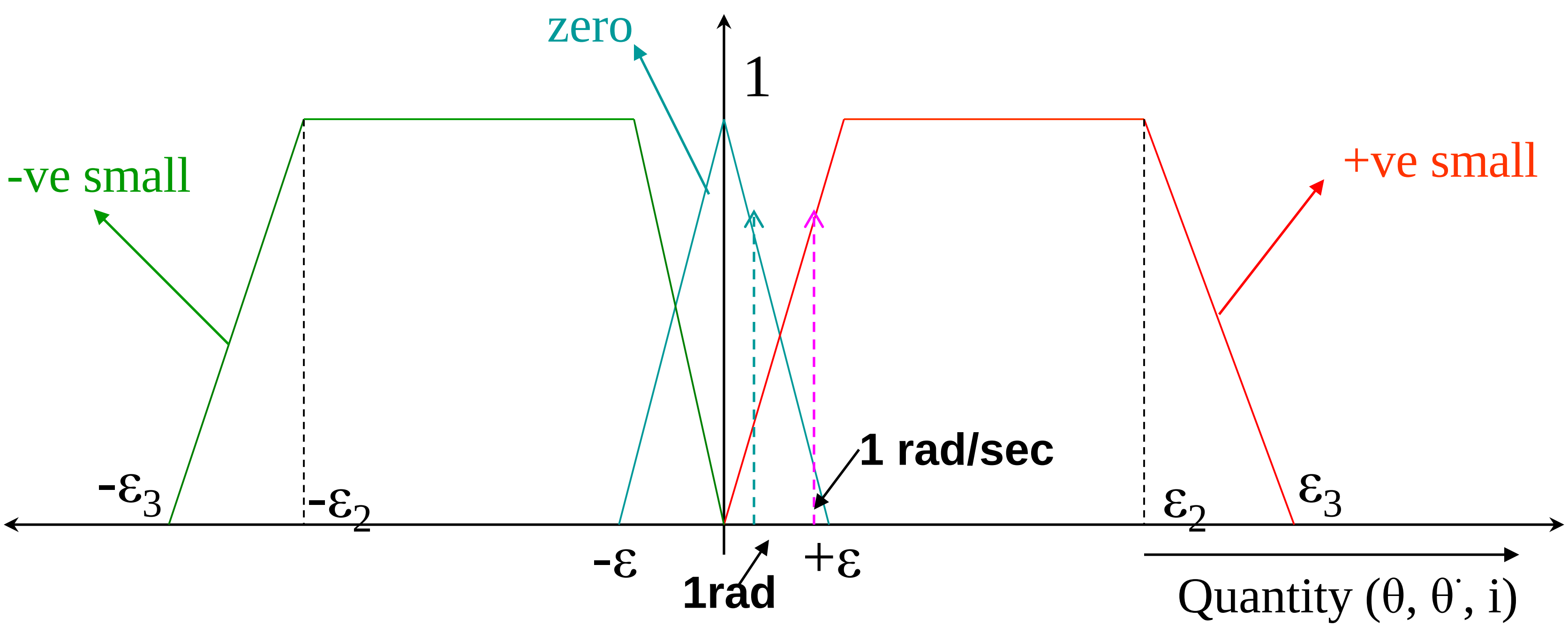
Suppose θ is 1 radian and dθ/dt is 1 rad/sec

μ_{zero}(θ =1)=0.8 (say)

$M_{+ve\text{-}small}$(θ =1)=0.4 (say)

μ_{zero}(dθ/dt =1)=0.3 (say)

$\mu_{+ve\text{-}small}$(dθ/dt =1)=0.7 (say)



Fuzzification

Suppose θ is 1 radian and dθ/dt is 1 rad/sec

$\mu_{zero}(\theta = 1) = 0.8$ (say)

$\mu_{+ve\text{-}small}(\theta = 1) = 0.4$ (say)

$\mu_{zero}(d\theta/dt = 1) = 0.3$ (say)

$\mu_{+ve\text{-}small}(d\theta/dt = 1) = 0.7$ (say)

if θ is Zero and dθ/dt is Zero then i is Zero

min(0.8, 0.3)=0.3

hence $\mu_{zero}(i) = 0.3$

if θ is Zero and dθ/dt is +ve small then i is –ve small

min(0.8, 0.7)=0.7

hence $\mu_{-ve\text{-}small}(i) = 0.7$

if θ is +ve small and dθ/dt is Zero then i is –ve small

min(0.4, 0.3)=0.3

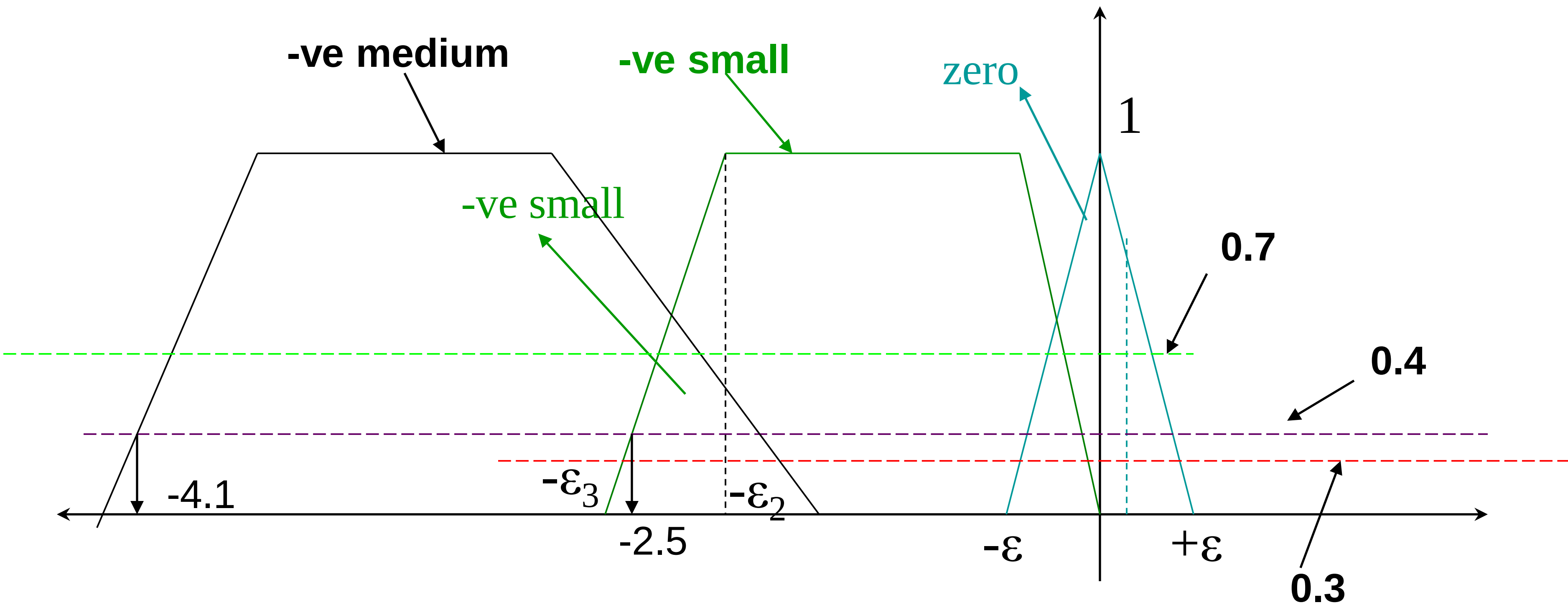
hence *μ-ve-small(i)=0.3*

if θ +ve small and dθ/dt is +ve small then i is -ve medium

min(0.4, 0.7)=0.4

hence $\mu_{-ve\text{-}medium}(i) = 0.4$

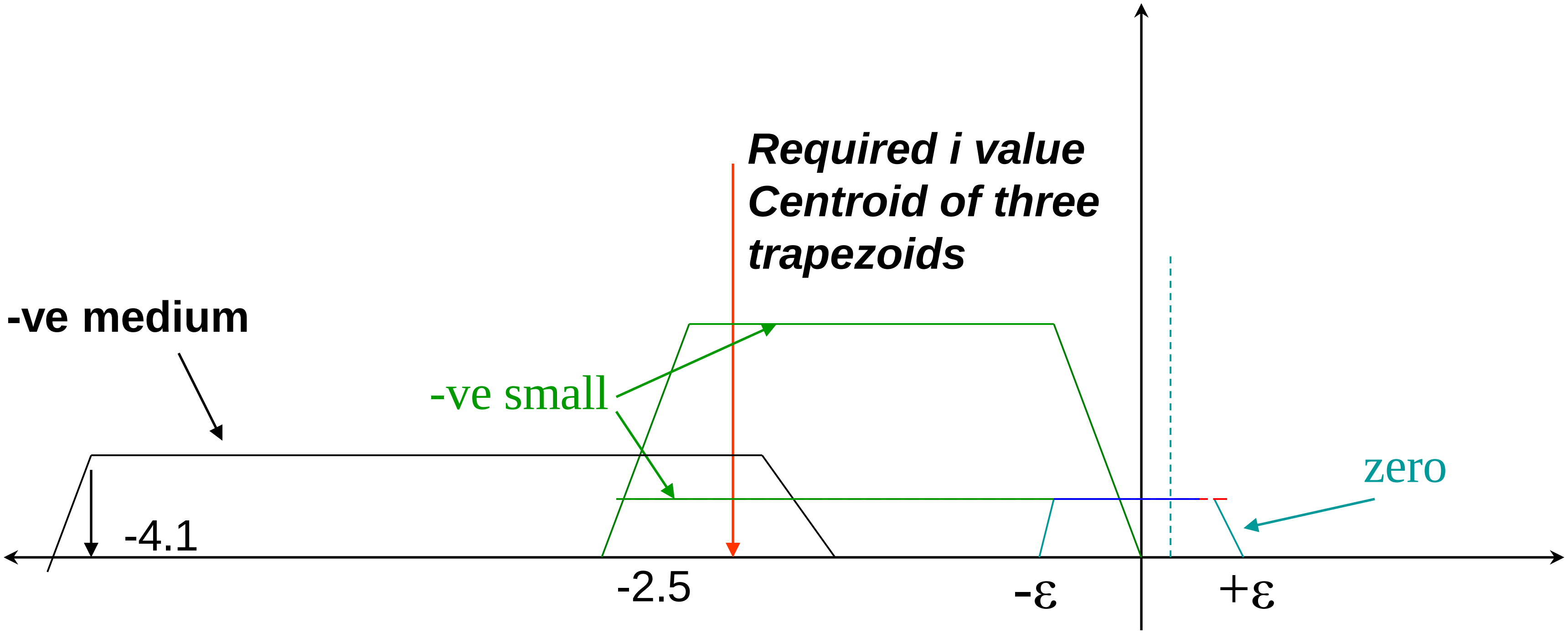
Finding *i*



Possible candidates:

i=0.5 and -0.5 from the "zero" profile and μ=0.3
i=-0.1 and -2.5 from the "-ve-small" profile and μ=0.3
i=-1.7 and -4.1 from the "-ve-small" profile and μ=0.3

Defuzzification: Finding *i* by the *centroid* method



Possible candidates:

i is the x-coord of the centroid of the areas given by the blue trapezium, the green trapeziums and the black trapezium

End Lecture 10

Lecture 11

Fuzzy Implication

Fuzzy definition of subset

Measured in terms of "fit violation", i.e. violating the condition $\mu_B(x) \leq \mu_A(x)$

Degree of subset hood S(A,B)= 1- degree of superset

$$= \quad 1 - \frac{\sum_x \max(0, \mu_B(x) - \mu_A(x))}{m(B)}$$

m(B) = cardinality of B

$$= \quad \sum_x \mu_B(x)$$

Fuzzy Implication

- Many theories have been advanced and many expressions exist
- The most used is Lukasiewitz formula
- t(P) = truth value of a proposition/predicate. In fuzzy logic t(P) = [0,1]
- $t(P \rightarrow Q) = \min[1, 1 - t(P) + t(Q)]$

Lukasiewitz definition of implication

Fuzzy Inferencing

- Two methods of inferencing in classical logic
  - Modus Ponens
    - Given p and $p \rightarrow q$, infer q
  - Modus Tolens
    - Given $\sim q$ and $p \rightarrow q$, infer $\sim p$
- How is fuzzy inferencing done?

Fuzzy Modus Ponens in terms of truth values

- Given $t(p)=1$ and $t(p \rightarrow q)=1$, infer $t(q)=1$
- In fuzzy logic,
  - given $t(p)>=a$, $0<=a<=1$
  - *and* $t(p \rightarrow >q)=c$, $0<=c<=1$
  - What is $t(q)$
- How much of truth is transferred over the channel

$$p \Rightarrow q$$

Lukasiewitz formula for Fuzzy Implication

- t(P) = truth value of a proposition/predicate. In fuzzy logic t(P) = [0,1]
- t($P \rightarrow Q$) = min[1,1 -t(P)+t(Q)]

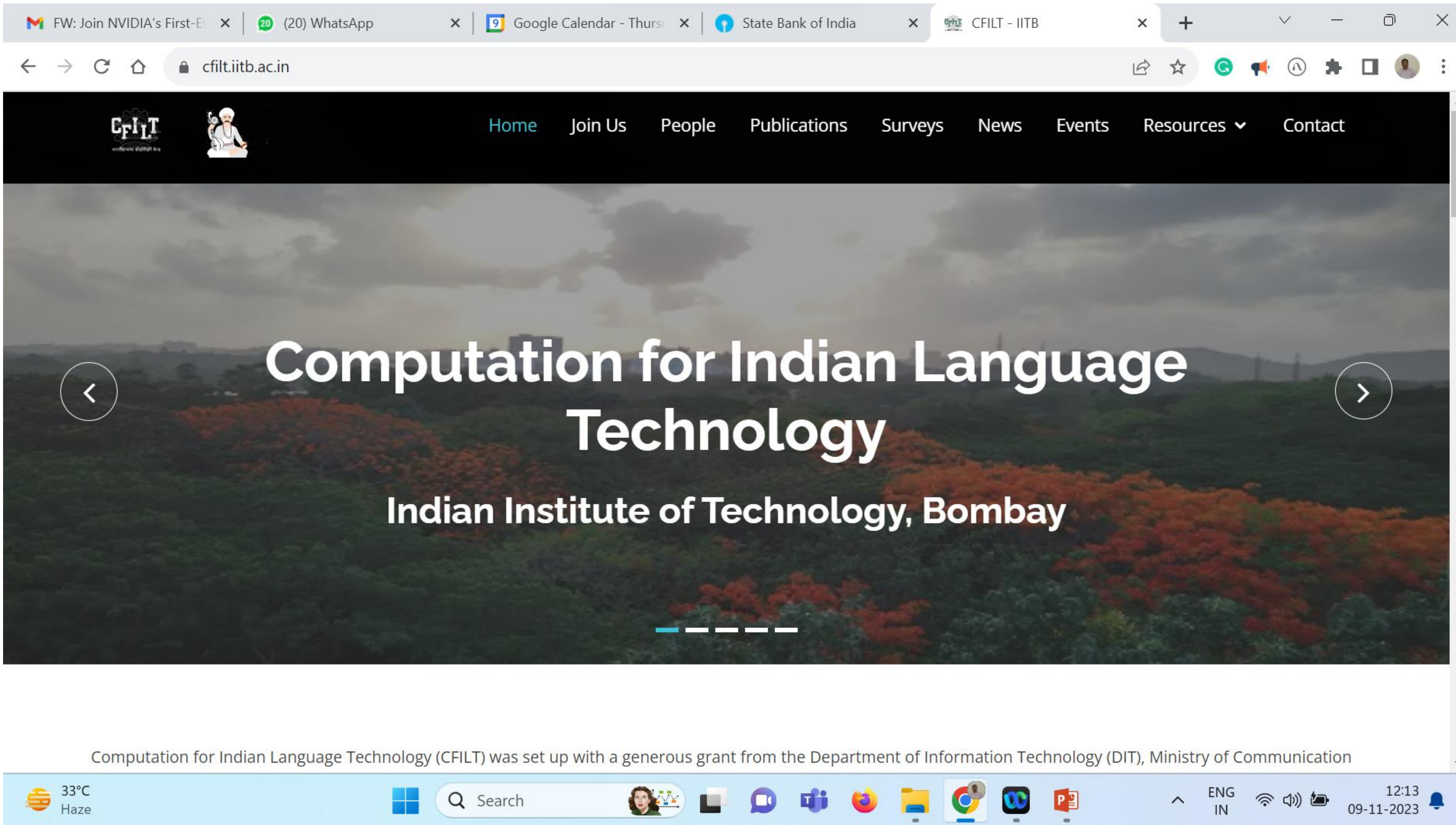
Lukasiewitz definition of implication

Use Lukasiewitz definition

- $t(p \rightarrow q) = min[1, 1 - t(p) + t(q)]$
- We have $t(p\text{-}>q)=c$, *i.e.*, $min[1,1 - t(p)+t(q)]=c$
- Case 1:
- $c=1$ gives $1 - t(p)+t(q)>=1$, *i.e.*, $t(q)>=a$
- Otherwise, $1 - t(p)+t(q)=c$, *i.e.*, $t(q)>=c+a-1$
- Combining, $t(q)=max(0,a+c-1)$
- This is the amount of truth transferred over the channel $p \rightarrow q$

End Lecture 11

Our NLP Lab, CFILT (www.cfilt.iitb.ac.in_



Research areas

- **Lexical Resources**
- **Lexical and Structural Disambiguation**
- **Shallow Parsing**
- **Cross Lingual Information Retrieval**
- **Machine Translation**
- **Text Entailment**
- **Sentiment Analysis**
- **Cognitive NLP**

All the best