Uncertain KR&R

Chapter 9

Outline

- Probability
- Bayesian networks
- Fuzzy logic

Probability

FOL fails for a domain due to:

Laziness: too much to list the complete set of rules, too hard to use the enormous rules that result

Theoretical ignorance: there is no complete theory for the domain

Practical ignorance: have not or cannot run all necessary tests

Probability

- Probability = a degree of belief
- Probability comes from:

  Frequentist: experiments and statistical assessment

  Objectivist: real aspects of the universe

  Subjectivist: a way of characterizing an agent's beliefs

- Decision theory = probability theory + utility theory

Probability

Prior probability: probability in the absence of any other information

$$P(Dice = 2) = 1/6$$

random variable: Dice

domain = <1, 2, 3, 4, 5, 6>

probability distribution: P(Dice) = <1/6, 1/6, 1/6, 1/6, 1/6, 1/6>

Probability

Conditional probability: probability in the presence of some evidence

$P(\text{Dice} = 2 \mid \text{Dice is even}) = 1/3$

$P(\text{Dice} = 2 \mid \text{Dice is odd}) = 0$

$P(A \mid B) = P(A \wedge B)/P(B)$

$P(A \wedge B) = P(A \mid B).P(B)$

Probability

Example:

S = stiff neck

M = meningitis

$P(S \mid M) = 0.5$

$P(M) = 1/50000$

$P(S) = 1/20$

$P(M \mid S) = P(S \mid M).P(M)/P(S) = 1/5000$

Probability

Joint probability distributions:

$X: \langle x_1, \ldots, x_m \rangle \quad Y: \langle y_1, \ldots, y_n \rangle$

$P(X = x_i, Y = y_j)$

Probability

Axioms:

- $0 \leq P(A) \leq 1$
- $P(true) = 1$ and $P(false) = 0$
- $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$

Probability

Derived properties:

- $P(\neg A) = 1 - P(A)$
- $P(U) = P(A_1) + P(A_2) + \ldots + P(A_n)$

  $U = A_1 \vee A_2 \vee \ldots \vee A_n$ collectively exhaustive

  $A_i \wedge A_j = \text{false}$ mutually exclusive

Probability

Bayes' theorem:

$$P(H_i \mid E) = P(E \mid H_i).P(H_i)/\sum_i P(E \mid H_i).P(H_i)$$

H_i's are collectively exhaustive & mutually exclusive

Probability

Problem: a full joint probability distribution $P(X_1, X_2, ..., X_n)$ is sufficient for computing any (conditional) probability on X_i's, but the number of joint probabilities is exponential.

Probability

- Independence:

  $P(A \wedge B) = P(A).P(B)$

  $P(A) = P(A \mid B)$

- Conditional independence:

  $P(A \wedge B \mid E) = P(A \mid E).P(B \mid E)$

  $P(A \mid E) = P(A \mid E \wedge B)$

Example:

$P(\text{Toothache} \mid \text{Cavity} \wedge \text{Catch}) = P(\text{Toothache} \mid \text{Cavity})$

$P(\text{Catch} \mid \text{Cavity} \wedge \text{Toothache}) = P(\text{Catch} \mid \text{Cavity})$

Probability

"In John's and Mary's house, an alarm is installed to sound in case of burglary or earthquake. When the alarm sounds, John and Mary may make a call for help or rescue."

Probability

"In John's and Mary's house, an alarm is installed to sound in case of burglary or earthquake. When the alarm sounds, John and Mary may make a call for help or rescue."

Q1: If earthquake happens, how likely will John make a call?

Probability

"In John's and Mary's house, an alarm is installed to sound in case of burglary or earthquake. When the alarm sounds, John and Mary may make a call for help or rescue."

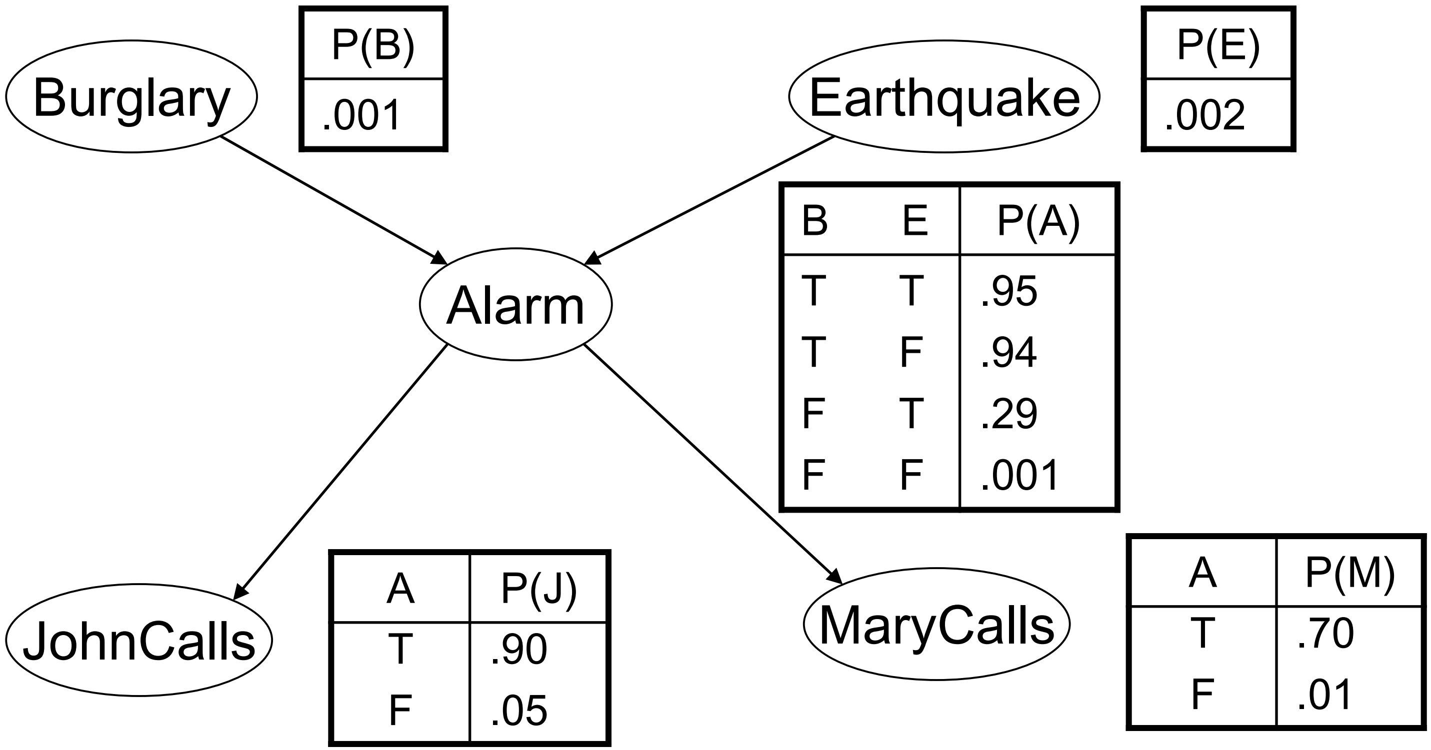
Q1: If earthquake happens, how likely will John make a call?

Q2: If the alarm sounds, how likely is the house burglarized?

Bayesian Networks

- Pearl, J. (1982). Reverend Bayes on Inference Engines: A Distributed Hierarchical Approach, presented at the Second National Conference on Artificial Intelligence (AAAI-82), Pittsburgh, Pennsylvania

Bayesian Networks



P(B)
.001

P(E)
.002

B	E	P(A)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

A	P(J)
T	.90
F	.05

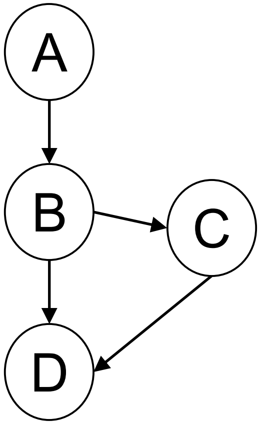
A	P(M)
T	.70
F	.01

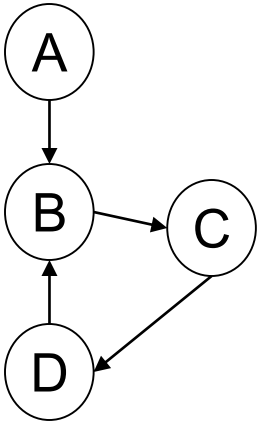
Bayesian Networks

Syntax:

- A set of random variables makes up the nodes
- A set of directed links connects pairs of nodes
- Each node has a conditional probability table that quantifies the effects of its parent nodes
- The graph has no directed cycles

yes





no

Bayesian Networks

Semantics:

- An ordering on the nodes: X_i is a predecessor of $X_j \Rightarrow i < j$
- $P(X_1, X_2, \ldots, X_n)$

  $= P(X_n \mid X_{n-1}, \ldots, X_1).P(X_{n-1} \mid X_{n-2}, \ldots, X_1). \ldots .P(X_2 \mid X_1).P(X_1)$

  $= \Pi_i P(X_i \mid X_{i-1}, \ldots, X_1) = \Pi_i P(X_i \mid Parents(X_i))$

$P(X_i \mid X_{i-1}, \ldots, X_1) = P(X_i \mid Parents(X_i)) \quad Parents(X_i) \subseteq \{X_{i-1}, \ldots, X_1\}$

Each node is conditionally independent of its predecessors given its parents

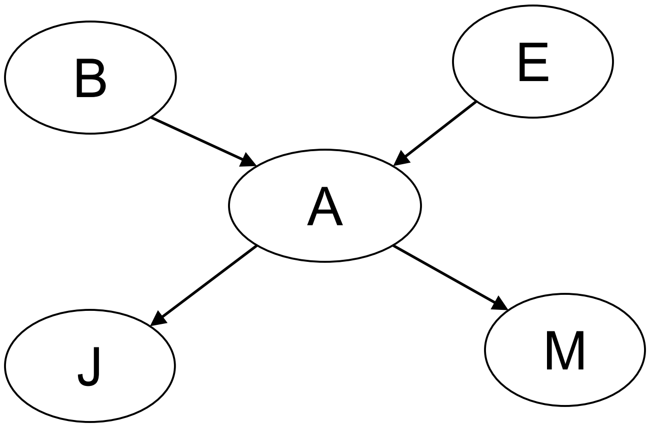
Bayesian Networks

Example:

$P(J \wedge M \wedge A \wedge \neg B \wedge \neg E)$

$= P(J \mid A).P(M \mid A).P(A \mid \neg B \wedge \neg E).P(\neg B).P(\neg E)$

$= 0.00062$



Bayesian Networks

- Why Bayesian Networks?

Uncertain Question Answering

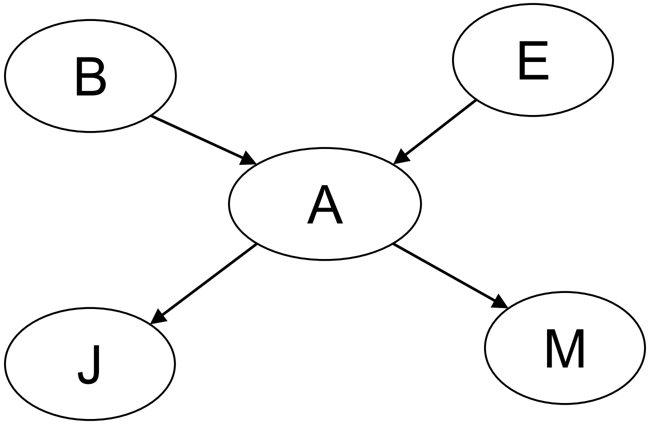
P(Query | Evidence) = ?

Diagnostic (from effects to causes): $P(B \mid J)$

Causal (from causes to effects): $P(J \mid B)$

Intercausal (between causes of a common effect): $P(B \mid A, E)$

Mixed: $P(A \mid J, \neg E)$, $P(B \mid J, \neg E)$



Uncertain Question Answering

- The independence assumptions in a Bayesian Network simplify computation of conditional probabilities on its variables

Uncertain Question Answering

Q1: If earthquake happens, how likely will John make a call?

Q2: If the alarm sounds, how likely is the house burglarized?

Q3: If the alarm sounds, how likely both John and Mary make calls?

Uncertain Question Answering

$P(B \mid A)$

$= P(B \wedge A)/P(A)$

$= \alpha P(B \wedge A)$

$P(\neg B \mid A)$

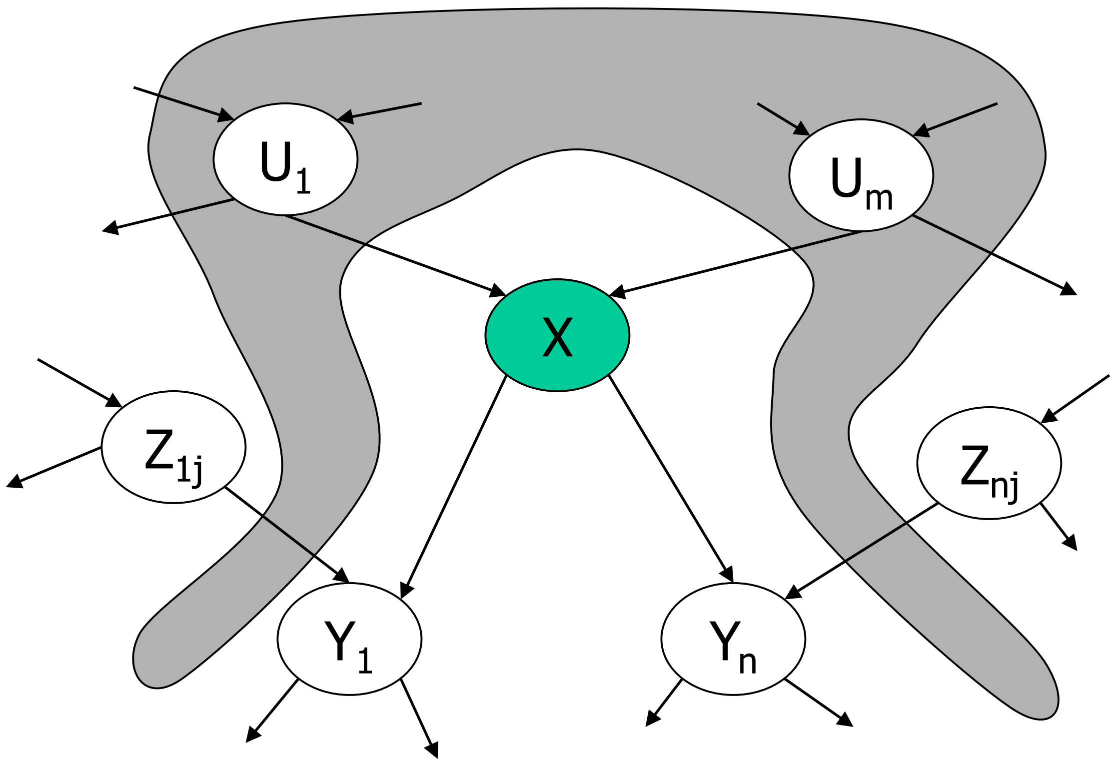
$= \alpha P(\neg B \wedge A)$

$\Rightarrow \alpha = 1/(P(B \wedge A) + P(\neg B \wedge A))$

General Conditional Independence

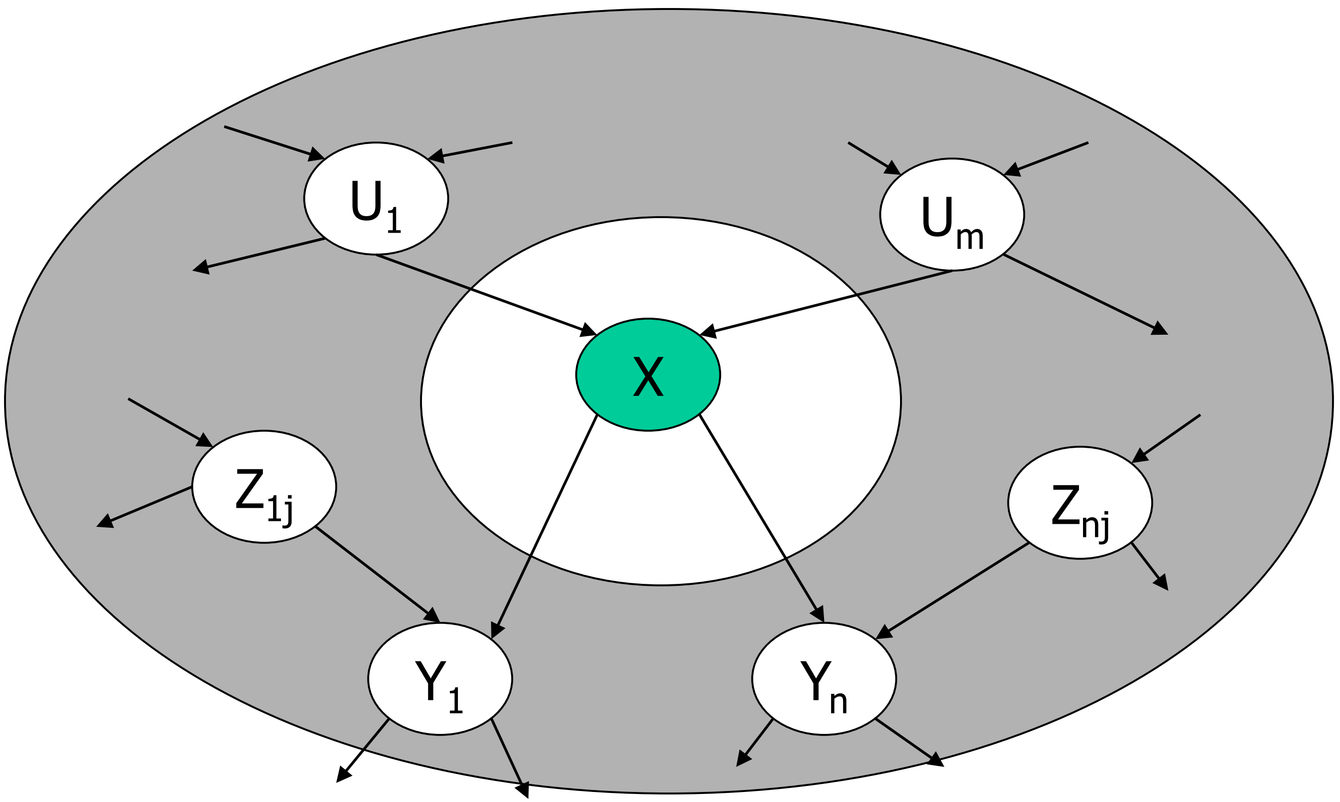
- A Bayesian Network implies all conditional independence among its variables

General Conditional Independence



A node (X) is conditionally independent of its **non-descendents** (Z_{ij}'s), given its **parents** (U_i's)

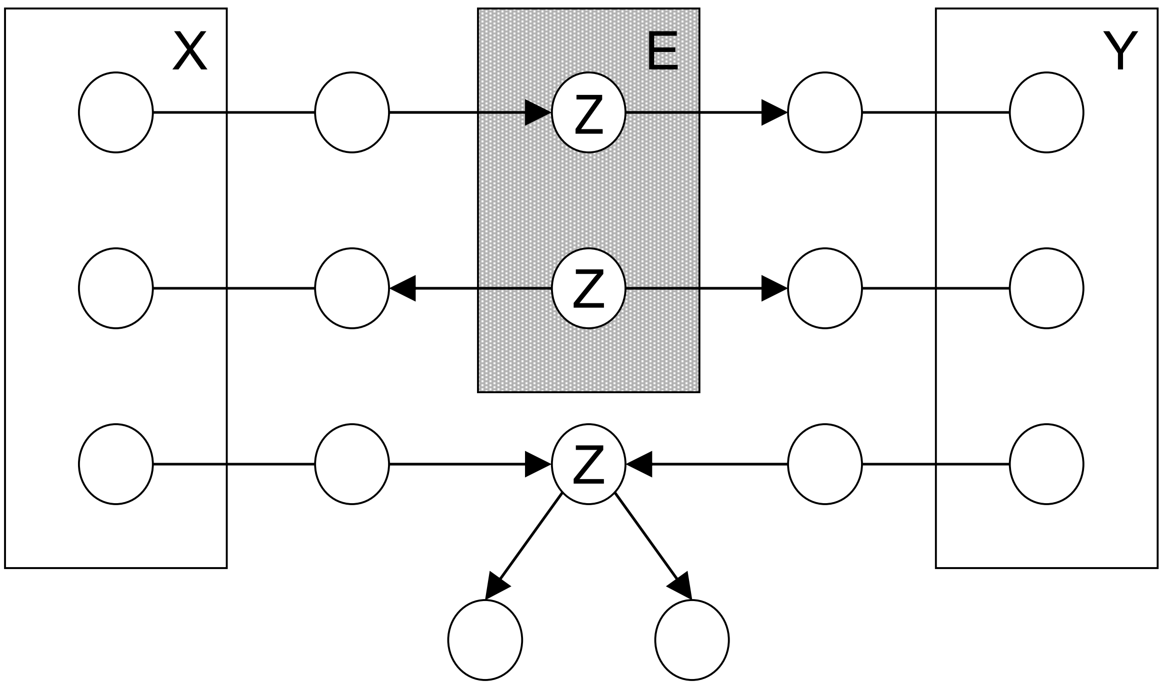
General Conditional Independence



A node (X) is conditionally independent of **all other nodes**, given its **parents** (U_i's), **children** (Y_i's), and **children's parents** (Z_{ij}'s)

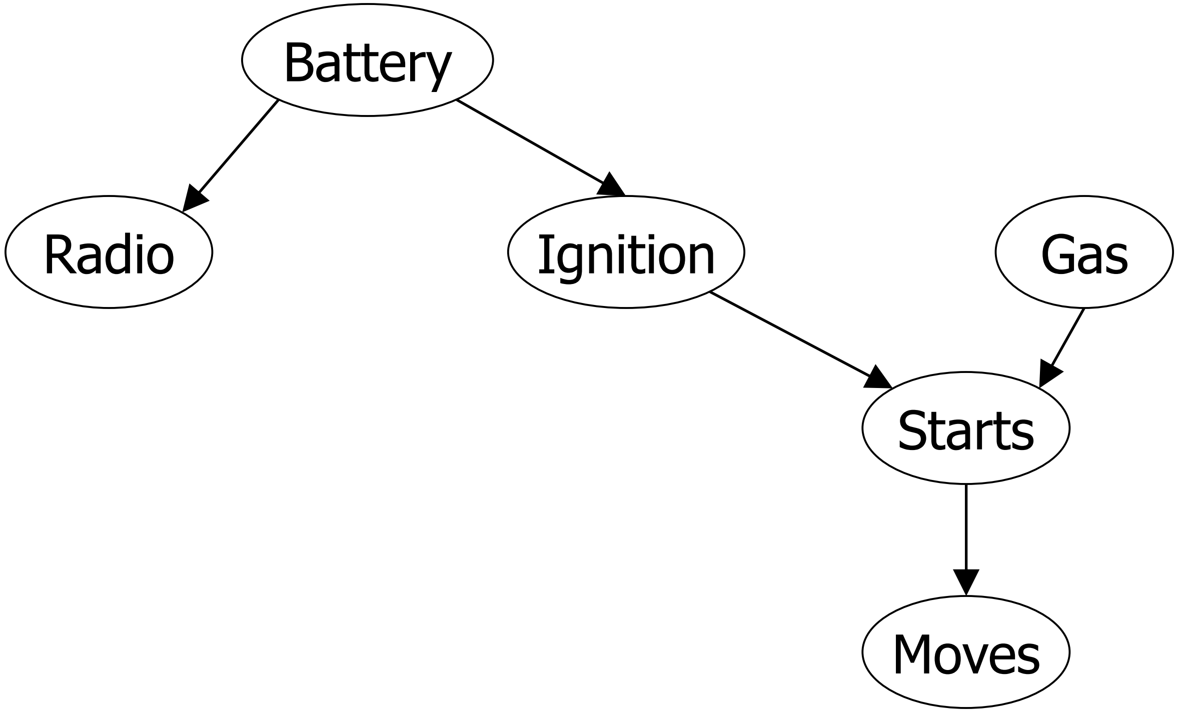
General Conditional Independence

X and Y are conditionally independent given E



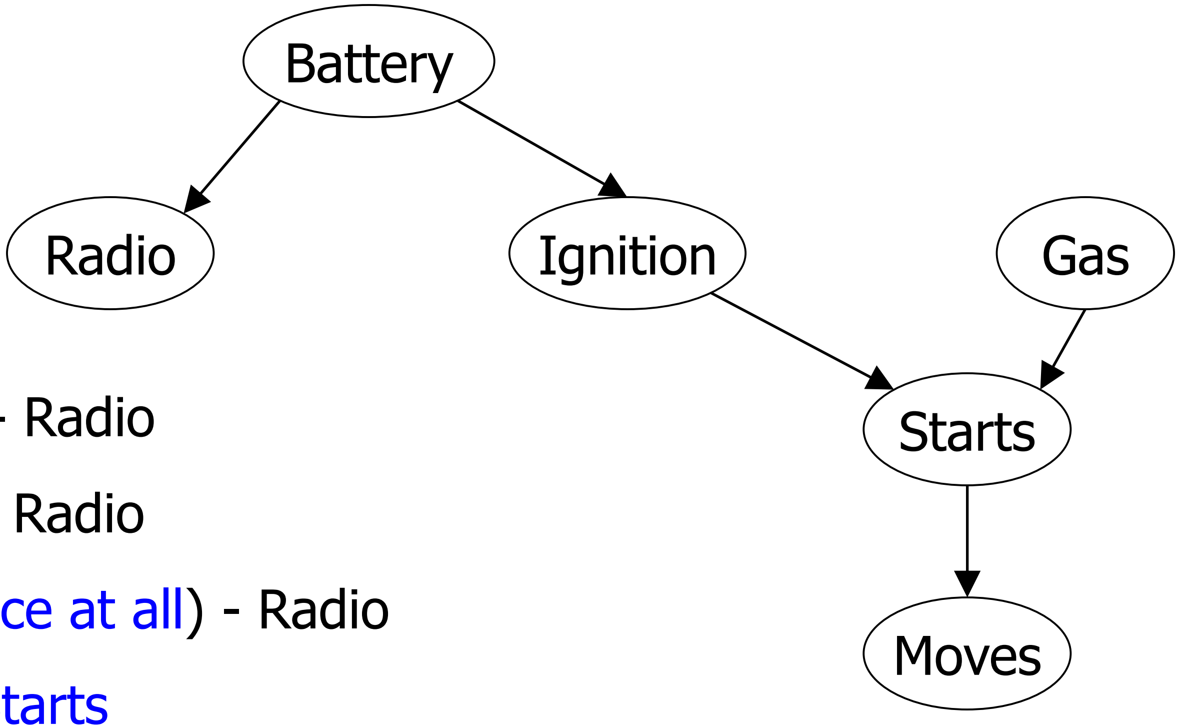
General Conditional Independence

Example:



General Conditional Independence

Example:



Gas - (Ignition) - Radio

Gas - (Battery) - Radio

Gas - (no evidence at all) - Radio

Gas $\leftrightarrow$ Radio | Starts

Gas $\leftrightarrow$ Radio | Moves

Vagueness

- The Oxford Companion to Philosophy (1995):

  "Words like smart, tall, and fat are vague since in most contexts of use there is no bright line separating them from not smart, not tall, and not fat respectively …"

Vagueness

- Imprecision vs. Uncertainty:

  The bottle is about half-full.

  vs.

  It is likely to a degree of 0.5 that the bottle is full.

Fuzzy Sets

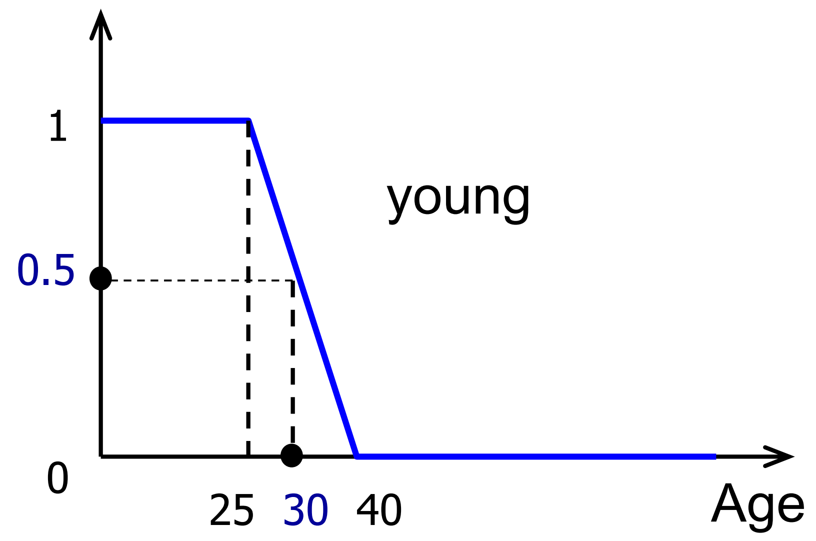
- Zadeh, L.A. (1965). Fuzzy Sets
  Journal of Information and Control

Fuzzy Set Definition

A fuzzy set is defined by a membership function that maps elements of a given domain (a crisp set) into values in [0, 1].

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

$\mu_A \leftrightarrow A$



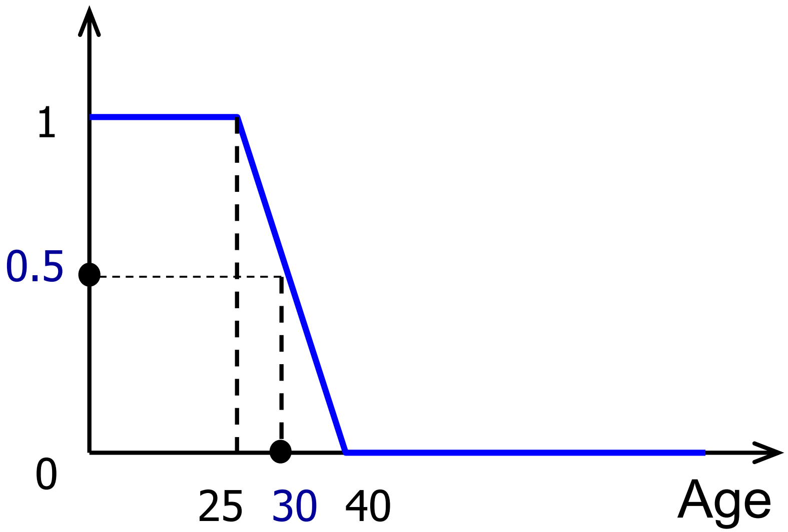
Fuzzy Set Representation

- Discrete domain:

  high-dice score: {1:0, 2:0, 3:0.2, 4:0.5, 5:0.9, 6:1}

- Continuous domain:

  A(u) = 1 for $u \in [0, 25]$
  A(u) = (40 - u)/15 for $u \in [25, 40]$
  A(u) = 0 for $u \in [40, 150]$



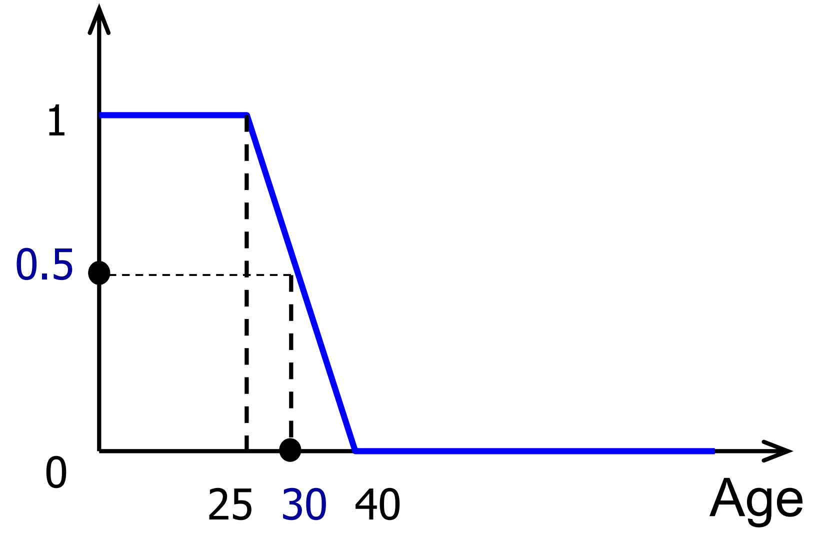
Fuzzy Set Representation

- α-cuts:

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

$A^{\alpha+} = \{u \mid A(u) > \alpha\}$ strong α-cut

$A^{0.5} = [0, 30]$



Fuzzy Set Representation

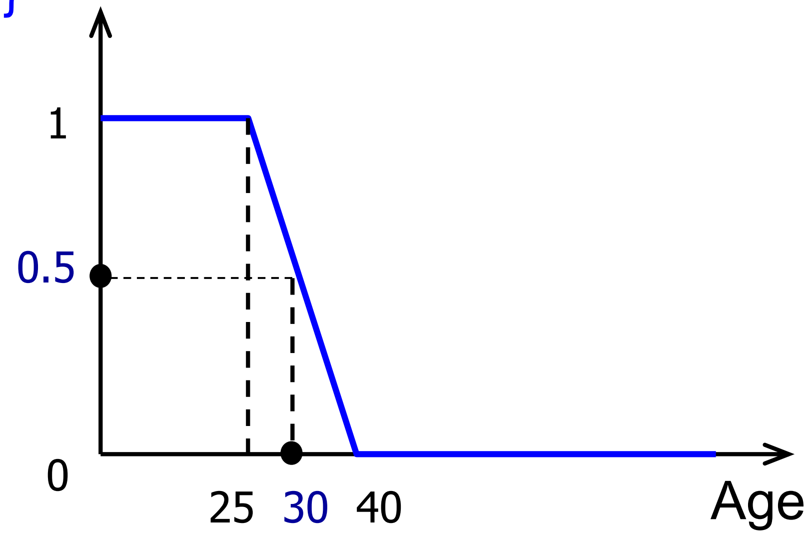
- α-cuts:

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

  $A^{\alpha+} = \{u \mid A(u) > \alpha\}$ strong α-cut

  $A(u) = \sup \{\alpha \mid u \in A^{\alpha}\}$

  $A^{0.5} = [0, 30]$



Fuzzy Set Representation

- Support:

  $\text{supp}(A) = \{u \mid A(u) > 0\} = A^{0+}$

- Core:

  $\text{core}(A) = \{u \mid A(u) = 1\} = A^{1}$

- Height:

  $h(A) = \sup_U A(u)$

Fuzzy Set Representation

- Normal fuzzy set: $h(A) = 1$
- Sub-normal fuzzy set: $h(A) < 1$

Membership Degrees

- Subjective definition

Membership Degrees

- Subjective definition

- Voting model:

  Each voter has a subset of U as his/her own crisp definition of the concept that A represents.

  A(u) is the proportion of voters whose crisp definitions include u.

Membership Degrees

- Voting model:

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}
1										
2										
3	x	x								
4	x	x	x	x	x					
5	x	x	x	x	x	x	x	x	x	
6	x	x	x	x	x	x	x	x	x	x

Fuzzy Subset Relations

$A \subseteq B$ iff $A(u) \leq B(u)$ for every $u \in U$

A is more "specific" than B

"X is A" entails "X is B"

Fuzzy Set Operations

- Standard definitions:

Complement: $\overline{A}(u) = 1 - A(u)$

Intersection: $(A \cap B)(u) = \min[A(u), B(u)]$

Union: $(A \cup B)(u) = \max[A(u), B(u)]$

Fuzzy Set Operations

- Example:

not young = $\overline{\text{young}}$

not old = $\overline{\text{old}}$

middle-age = not young $\cap$ not old

old = $\neg$young

Fuzzy Relations

- Crisp relation:

$$R(U_1, ..., U_n) \subseteq U_1 \times ... \times U_n$$

$$R(u_1, ..., u_n) = 1 \text{ iff } (u_1, ..., u_n) \in R \text{ or } = 0 \text{ otherwise}$$

Fuzzy Relations

- Crisp relation:

  $R(U_1, \ldots, U_n) \subseteq U_1 \times \ldots \times U_n$

  $R(u_1, \ldots, u_n) = 1$ iff $(u_1, \ldots, u_n) \in R$ or $= 0$ otherwise

- Fuzzy relation: a fuzzy set on $U_1 \times \ldots \times U_n$

Fuzzy Relations

- Fuzzy relation:

U_1 = {New York, Paris}, U_2 = {Beijing, New York, London}

R = “very far”

	NY	Paris
Beijing	1	.9
NY	0	.7
London	.6	.3

R = {(NY, Beijing): 1, ...}

Fuzzy Numbers

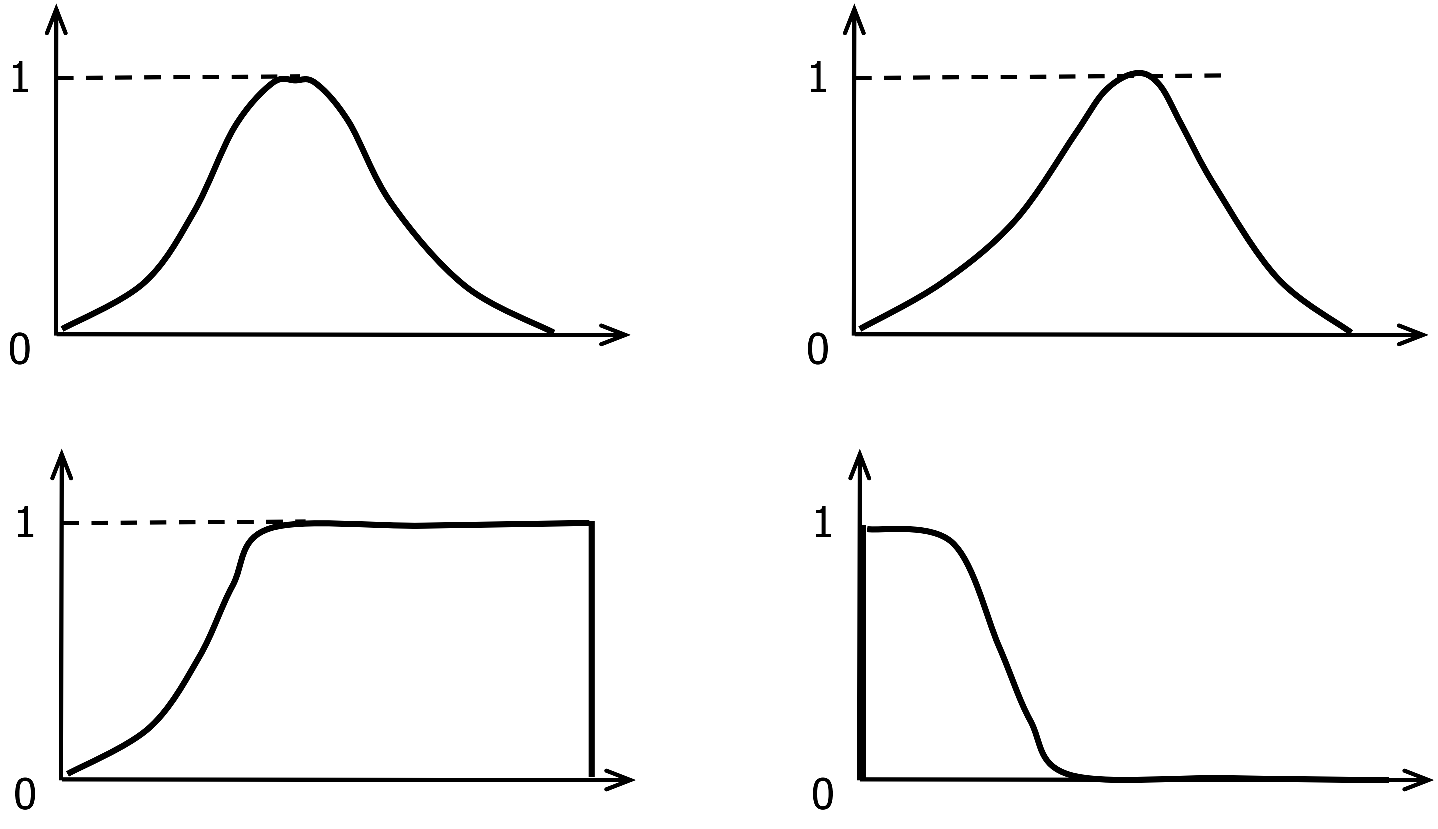
- A fuzzy number A is a fuzzy set on R:

  A must be a normal fuzzy set

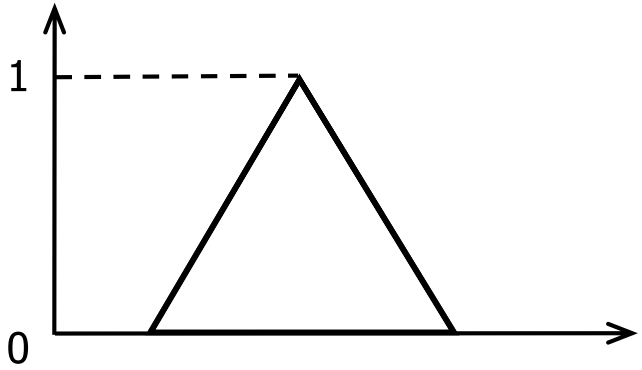
  A^{α} must be a closed interval for every $\alpha \in (0, 1]$

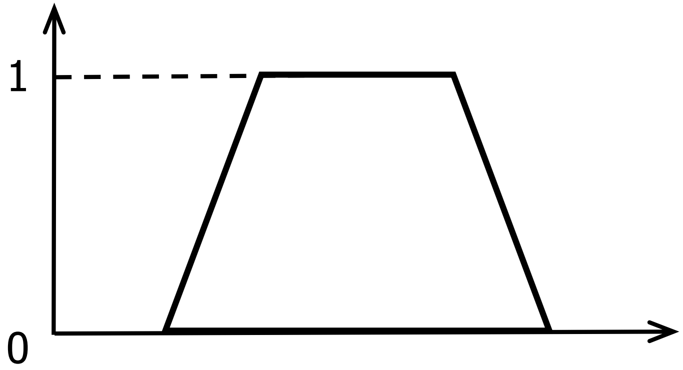
  supp(A) = A^{0+} must be bounded

Basic Types of Fuzzy Numbers



Basic Types of Fuzzy Numbers





Operations of Fuzzy Numbers

- Interval-based operations:

$$(A \circ B)^{\alpha} = A^{\alpha} \circ B^{\alpha}$$

Operations of Fuzzy Numbers

- Arithmetic operations on intervals:

$$[a, b] \circ [d, e] = \{f \circ g \mid a \leq f \leq b, d \leq g \leq e\}$$

Operations of Fuzzy Numbers

- Arithmetic operations on intervals:

$$[a, b] \circ [d, e] = \{f \circ g \mid a \le f \le b,\ d \le g \le e\}$$

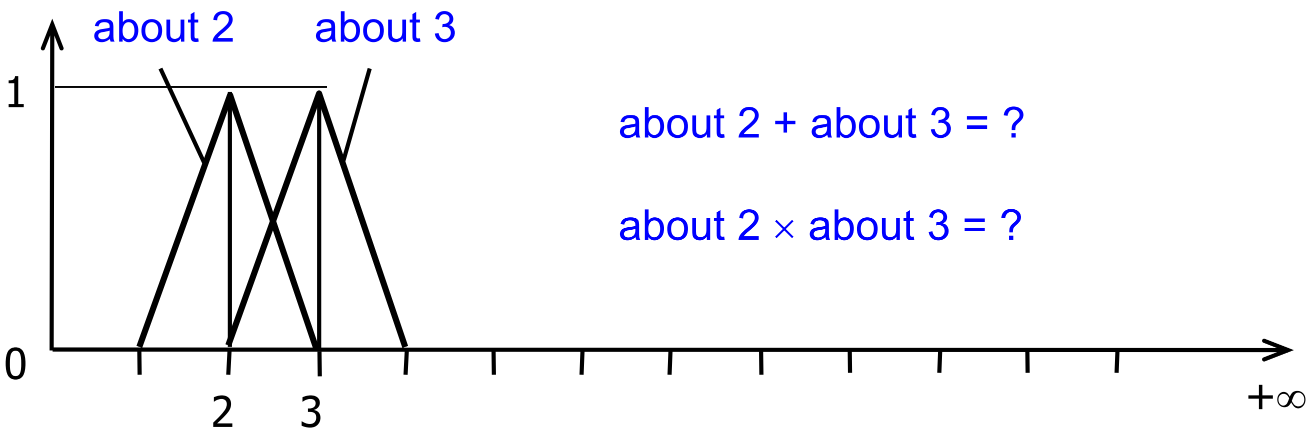
$$[a, b] + [d, e] = [a + d, b + e]$$

$$[a, b] - [d, e] = [a - e, b - d]$$

$$[a, b] * [d, e] = [\min(ad, ae, bd, be), \max(ad, ae, bd, be)]$$

$$[a, b] / [d, e] = [a, b] * [1/e, 1/d] \quad 0 \notin [d, e]$$

Operations of Fuzzy Numbers



about 2 + about 3 = ?

about 2 × about 3 = ?

Operations of Fuzzy Numbers

- Discrete domains:

$$A = \{x_i: A(x_i)\} \qquad B = \{y_i: B(y_i)\}$$

$$A \circ B = ?$$

Operations of Fuzzy Numbers

- Extension principle:

$f: U_1 \times U_2 \rightarrow V$

induces

$g: \tilde{U}_1 \times \tilde{U}_2 \rightarrow \tilde{V}$

$[g(A, B)](v) = \sup_{\{(u1,\ u2)\ |\ v\ =\ f(u1,\ u2)\}} \min\{A(u_1), B(u_2)\}$

Operations of Fuzzy Numbers

- Discrete domains:

$$A = \{x_i: A(x_i)\} \qquad B = \{y_i: B(y_i)\}$$

$$(A \circ B)(v) = \sup_{\{(xi,\ yj)\ |\ v\ =\ xi^\circ yj)\}} \min\{A(x_i), B(y_j)\}$$

Fuzzy Logic

if x is A then y is B

x is A*

----

y is B*

Fuzzy Logic

- View a fuzzy rule as a fuzzy relation
- Measure similarity of A and A*

Fuzzy Controller

- As special expert systems
- When difficult to construct mathematical models
- When acquired models are expensive to use

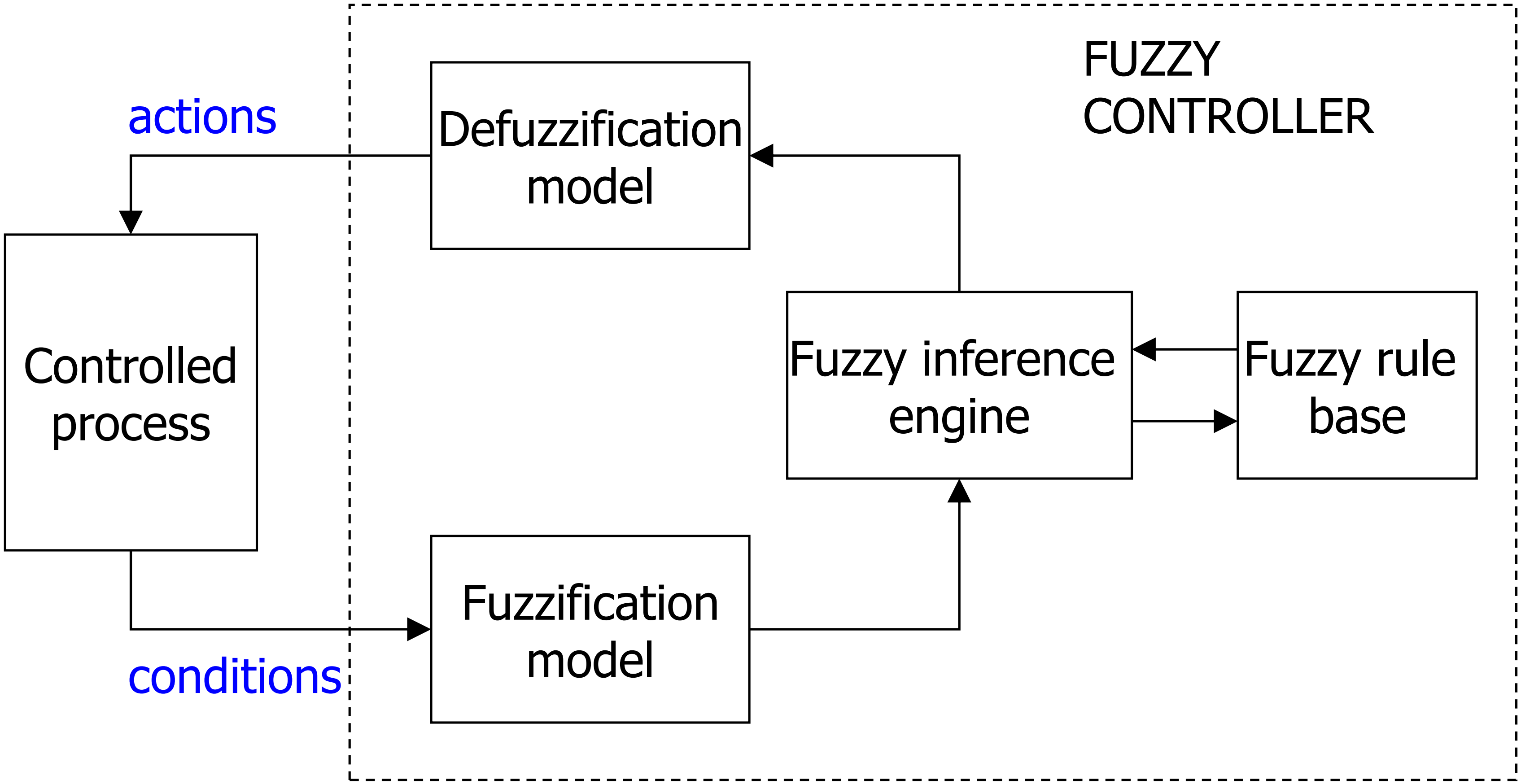
Fuzzy Controller

IF the temperature is very high

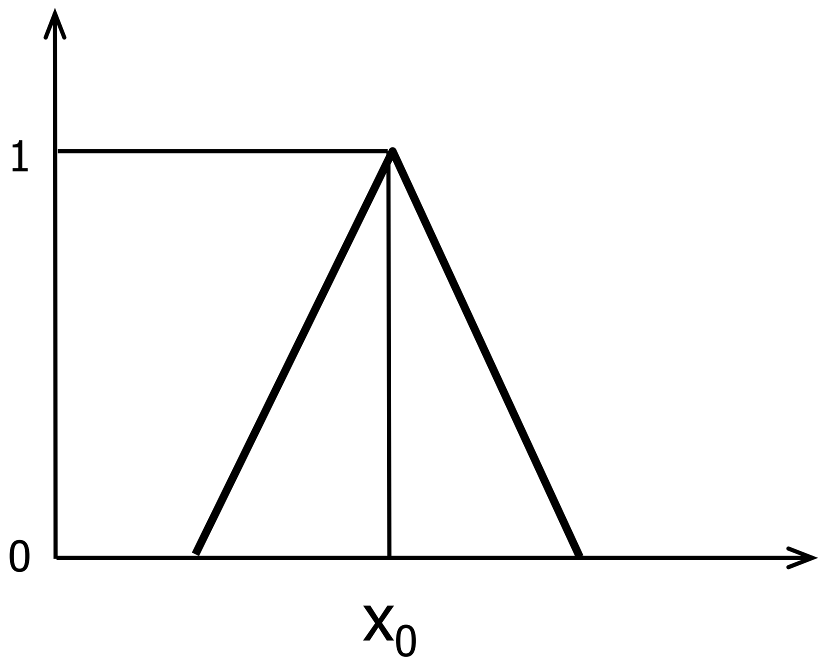
AND the pressure is slightly low

THEN the heat change should be slightly negative

Fuzzy Controller



Fuzzification



Defuzzification

- Center of Area:

$$x = (\sum A(z).z)/\sum A(z)$$

Defuzzification

- Center of Maxima:

$$M = \{z \mid A(z) = h(A)\}$$

$$x = (\min M + \max M)/2$$

Defuzzification

- Mean of Maxima:

$$M = \{z \mid A(z) = h(A)\}$$

$$x = \sum z / |M|$$

Exercises

- In Klir-Yuan’s textbook: 1.9, 1.10, 2.11, 2.12, 4.5