Uncertainty and Vagueness Basics

Uncertainty & Vagueness: Basic Concepts

We recall that under:

- Uncertainty:
  - a statement is either true or false (all concepts have a precise definition)
  - due to lack of knowledge we can only estimate to which probability/possibility/necessity degree they are true or false
  - We will restrict our attention to Probability Theory
- Vagueness:
  - a statement may have a degree of truth in $[0, 1]$, as concepts without precise definition are involved
  - We will restrict our attention to Fuzzy Set Theory

Basic Concepts under Probability Theory

- Let W be a set of possible worlds $w \in W$
  - E.g., $W = \{1, 2, 3, 4, 5, 6\}$ is the set of possible outcomes in throwing a dice
- An event E is a subset $E \subseteq W$ of possible worlds
  - E.g., $E = \{2, 4, 6\}$ is the event "the outcome is even"
- If E, E' are events, so are $E \cap E', E \cup E', \overline{E} = W \setminus E$

Some properties on events

Commutative laws:

$$
\begin{aligned}
E_1 \cup E_2 &= E_2 \cup E_1 \\
E_1 \cap E_2 &= E_2 \cap E_1
\end{aligned}
$$

Associative laws:

$$
\begin{aligned}
E_1 \cup (E_2 \cup E_3) &= (E_1 \cup E_2) \cup E_3 \\
E_1 \cap (E_2 \cap E_3) &= (E_1 \cap E_2) \cap E_3
\end{aligned}
$$

Distributive laws:

$$
\begin{aligned}
E_1 \cap (E_2 \cup E_3) &= (E_1 \cap E_2) \cup (E_1 \cap E_3) \\
E_1 \cup (E_2 \cap E_3) &= (E_1 \cup E_2) \cap (E_1 \cup E_3)
\end{aligned}
$$

$$
\begin{aligned}
\overline{\overline{E}} &= E \\
E \cap W &= E \\
E \cup W &= W \\
E \cap \emptyset &= \emptyset \\
E \cup \emptyset &= E \\
E \cap \overline{E} &= \emptyset \\
E \cup \overline{E} &= W \\
E \cap E &= E \\
E \cup E &= E
\end{aligned}
$$

Some properties on events

De Morgan laws:

$$\begin{aligned}\overline{E_1 \cup E_2} &= \overline{E_1} \cap \overline{E_2}\\ \overline{E_1 \cap E_2} &= \overline{E_1} \cup \overline{E_2}\end{aligned}$$

De Morgan Theorem: For an index set (denumerable set) I

$$\begin{aligned}\overline{\bigcup_{i\in I} E_i} &= \bigcap_{i\in I} \overline{E_i}\\ \overline{\bigcap_{i\in I} E_i} &= \bigcup_{i\in I} \overline{E_i}\end{aligned}$$

Disjoint or Mutually Exclusive Events

- Events E_1, E_2 are disjoint or mutually exclusive iff $E_1 \cap E_2 = \emptyset$
- Events $E_1, E_2, \ldots$ are disjoint or mutually exclusive iff $E_i \cap E_j = \emptyset$ for every $i \neq j$

$$
\begin{aligned}
E &= (E \cap E') \cup (E \cap \overline{E'}) \\
\emptyset &= (E \cap E') \cap (E \cap \overline{E'}) \\
E &= E \cap E', \text{ if } E \subseteq E' \\
E' &= E \cup E', \text{ if } E \subseteq E'
\end{aligned}
$$

Event Space

- A set of events $\mathcal{E}$ is an event space iff
  1. $W \in \mathcal{E}$
  2. If $E \in \mathcal{E}$, then $\overline{E} \in \mathcal{E}$
  3. If $E_1 \in \mathcal{E}$ and $E_2 \in \mathcal{E}$, then $E_1 \cup E_2 \in \mathcal{E}$
- An event space $\mathcal{E}$ is a boolean algebra
  1. $\emptyset \in \mathcal{E}$
  2. If $E_1 \in \mathcal{E}$ and $E_2 \in \mathcal{E}$, then $E_1 \cap E_2 \in \mathcal{E}$
  3. If $E_1, E_2, \ldots, E_n \in \mathcal{E}$, then $\bigcup_{i=1}^{n} E_i \in \mathcal{E}$ and $\bigcap_{i=1}^{n} E_i \in \mathcal{E}$

Probability Function

- Probability Function: A probability function is a function $Pr : \mathcal{E} \rightarrow [0, 1]$ such that
  1. $Pr(E) \geq 0$ for every $E \in \mathcal{E}$
  2. $Pr(W) = 1$
  3. if $E_1, E_2, \ldots$ is an infinite, denumerable sequence of disjoint events in $\mathcal{E}$ then

$$Pr(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} Pr(E_i)$$

Some Properties

- $Pr(\emptyset) = 0$
- if $E_1, E_2, \ldots, E_n$ are disjoint events in $\mathcal{E}$ then

$$Pr(\bigcup_{i=1}^{n} E_i) = \sum_{i=1}^{n} Pr(E_i)$$

- $Pr(\overline{E}) = 1 - Pr(E)$
- $Pr(E) = Pr(E \cap E') + Pr(E \cap \overline{E'})$
- $Pr(E_1 \setminus E_2) = Pr(E_1 \cap \overline{E_2}) = Pr(E_1) - Pr(E_1 \cap E_2)$
- $Pr(E_1 \cup E_2) = Pr(E_1) + Pr(E_2) - Pr(E_1 \cap E_2)$
- For events $E_1, E_2, \ldots, E_n$,

$$Pr(\bigcup_{i=1}^{n} E_i) = \sum_{j=1} Pr(E_j) - \sum_{i<j} Pr(E_i \cap E_j) + \sum_{i<j<k} Pr(E_i \cap E_j \cap E_k) - \ldots + (-1)^{n+1} Pr(E_1 \cap E_2 \cap \ldots \cap E_n)$$

- If $E_1 \subseteq E_2$ then $Pr(E_1) \leq Pr(E_2)$
- (Boole's inequality) if $E_1, E_2, \ldots, E_n$ events in $\mathcal{E}$ then

$$Pr(\bigcup_{i=1}^{n} E_i) \leq \sum_{i=1}^{n} Pr(E_i)$$

Finite Possibility World with Equally Likely Worlds

- For many random experiments, there is a finite number of outcomes, i.e. $N = |W|$ (the cardinality of W) is finite
- Often it is realistic to assume that the probability of each outcome $w \in W$ is $1/N$
- An equally likely probability function Pr is such that
  1. $Pr(\{w\}) = 1/|W|$ for all $w \in W$
  2. $Pr(E) = |E|/|W|$
- E.g., in throwing two dices, the probability that the sum is seven is determined as follows:
  1. $W = \{(x, y) \mid x, y \in \{1, 2, 3, 4, 5, 6\}\}$
  2. For all $w \in W$, $Pr(w) = 1/|W| = 1/36$
  3. E is the event "the sum is seven", i.e, $E = \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$

$$Pr(E) = |E|/|W| = 6/36 = 1/6$$

Conditional probability

- The conditional probability of event E_1 given event E_2 is

$$Pr(E_1 \mid E_2) = \begin{cases} \frac{Pr(E_1 \cap E_2)}{Pr(E_2)} & \text{if} Pr(E_2) > 0 \\ 1 & \text{otherwise} \end{cases}$$

- Remark: if $Pr(E_1)$ and $Pr(E_2)$ are nonzero then

$$Pr(E_1 \cap E_2) = Pr(E_1 \mid E_2) \cdot Pr(E_2) = Pr(E_2 \mid E_1) \cdot Pr(E_1)$$

- For equally likely probability functions

$$Pr(E_1 \mid E_2) = \begin{cases} \frac{|E_1 \cap E_2|}{|E_2|} & \text{if} |E_2| > 0 \\ 1 & \text{otherwise} \end{cases}$$

- E.g., in tossing two coins, what is the probability of two heads given a head of the first coin?
  1. $W = \{(x, y) \mid x, y \in \{T, H\}\}$
  2. For all $w \in W$, $Pr(w) = 1/|W| = 1/4$
  3. E_1 is the event "head on first coin", $E_1 = \{(H, H), (H, T)\}$
  4. E_2 is the event "head on second coin", $E_2 = \{(H, H), (T, H)\}$
  5. E is the event "two heads", $E = E_1 \cap E_2 = \{(H, H)\}$

$$Pr(E|E_1) = \frac{Pr(E \cap E_1)}{Pr(E_1)} = \frac{|E_1 \cap E_2|}{|E_1|} = \frac{1/4}{1/2} = 1/2$$

Conditional probability: Properties

Assume $Pr(E) > 0$.

- $Pr(\emptyset \mid E) = 0$
- If $E_1, E_2, \ldots, E_n$ are disjoint events in $\mathcal{E}$ then

$$Pr(E_1 \cup \ldots \cup E_n \mid E) = \sum_{i=1}^{n} Pr(E_i \mid E)$$

- For event E'

$$Pr(\overline{E'} \mid E) = 1 - Pr(E' \mid E)$$

- For two events E_1, E_2

$$\begin{aligned}
Pr(E_1 \mid E) &= Pr(E_1 \cap E_2 \mid E) + Pr(E_1 \cap \overline{E_2} \mid E) \\
Pr(E_1 \cup E_2 \mid E) &= Pr(E_1 \mid E) + Pr(E_2 \mid E) - Pr(E_1 \cap E_2 \mid E) \\
Pr(E_1 \mid E) &\leq Pr(E_2 \mid E) \qquad \text{if } E_1 \subseteq E_2
\end{aligned}$$

- For events $E_1, \ldots, E_n$

$$Pr(E_1 \cup \ldots \cup E_n \mid E) \leq \sum_{i=1}^{n} Pr(E_i \mid E)$$

Theorem of Total Probabilities

- If $E_1, E_2, \ldots, E_n$ are disjoint events in $\mathcal{E}$ such that $Pr(E_i) > 0$ and $W = \bigcup_{i=1}^{n} E_i$ then
$$Pr(E) = \sum_{i=1}^{n} Pr(E \mid E_i) \cdot Pr(E_i)$$
- Remark. If $Pr(E_2) > 0$ then
$$Pr(E_1) = Pr(E_1 \mid E_2) \cdot Pr(E_2) + Pr(E_1 \mid \overline{E_2}) \cdot Pr(\overline{E_2})$$
- The theorem of total probabilities can be used to combine classifiers
  1. Assume we have n different classifiers CL_i for category C (e.g. C is "an image is about sportcars")
  2. What is the probability of classifying an image object o as being a sportcar?
  $$Pr(C \mid o) \approx \sum_{i=1}^{n} Pr(C \mid o, CL_i) \cdot Pr(CL_i)$$
  where
     - $Pr(C \mid o)$ is the probability of classifying o in category C
     - $Pr(C \mid o, CL_i)$ is the probability that classifier CL_i classifies o in category C
     - $Pr(CL_i)$ is the overall effectiveness of classifier CL_i

Bayes' Theorem

- Bayes' Theorem: there are several variants

$$Pr(E_1 \mid E_2) = \frac{Pr(E_2 \mid E_1) \cdot Pr(E_1)}{Pr(E_2)}$$

- Each term in Bayes' theorem has a conventional name:
  - $Pr(E_1)$ is the prior probability or marginal probability of E_1. It is "prior" in the sense that it does not take into account any information about E_2
  - $Pr(E_1 \mid E_2)$ is called the posterior probability because it is derived from or depends upon the specified value of E_2
  - $Pr(E_2)$ is the prior or marginal probability of E_2, and acts as a normalizing constant

Example: Students

- Students at school
  1. There are 60% boys and 40% girls
  2. Girl students wear trousers or skirts in equal numbers
  3. The boys all wear trousers
- An observer sees a (random) student from a distance wearing trousers
- What is the probability this student is a girl?
  1. The event A is that the student observed is a girl
  2. Event B is that the student observed is wearing trousers
  3. We want to compute $Pr(A \mid B)$

$$Pr(A \mid B) = \frac{Pr(B \mid A) \cdot Pr(A)}{Pr(B)} = \frac{0.5 \cdot 0.4}{0.8} = 0.25$$

3.1 $Pr(A)$ is the probability that the student is a girl, $Pr(A) = 0.4$

3.2 $Pr(\overline{A})$ is the probability that the student is a boy, $Pr(\overline{A}) = 0.6$

3.3 $Pr(B \mid A)$ is the the probability of the student wearing trousers given that the student is a girl, $Pr(B \mid A) = 0.5$

3.4 $Pr(B \mid \overline{A})$ is the the probability of the student wearing trousers given that the student is a boy, $Pr(B \mid \overline{A}) = 1.0$

3.5 $Pr(B)$ is the probability of a (randomly selected) student wearing trousers,

$$Pr(B) = Pr(B \mid A) \cdot Pr(A) + Pr(B \mid \overline{A}) \cdot Pr(\overline{A}) = 0.5 \cdot 0.4 + 1 \cdot 0.6 = 0.8$$

Example: Drug test

- Suppose a certain drug test is 99% sensitive and 99% specific, that is,
  - the test will correctly identify a drug user as testing positive 99% of the time (sensitivity)
  - will correctly identify a non-user as testing negative 99% of the time (specificity)
- This would seem to be a relatively accurate test, but Bayes' theorem will reveal a potential flaw
- A corporation decides to test its employees for opium use, and 0.5% of the employees use the drug
- We want to know the probability that, given a positive drug test, an employee is actually a drug user
- Let D be the event "being a drug user", let N be the event "not being a drug user", and let $+$ be the event "positive drug test"
- We want to compute $Pr(D \mid +)$

Example: Drug test (cont.)

$$
\begin{aligned}
Pr(D \mid +) &= \frac{Pr(+ \mid D) \cdot P(D)}{Pr(+)} \\
&= \frac{Pr(+ \mid D) \cdot P(D)}{Pr(+ \mid D) \cdot Pr(D) + Pr(+ \mid N) \cdot Pr(N)} \\
&= \frac{0.99 \cdot 0.005}{0.99 \cdot 0.005 + 0.01 \cdot 0.995} \\
&= 0.3322
\end{aligned}
$$

where

- $Pr(D)$ is the probability that a random employee is a drug user, $Pr(D) = 0.005$ (0.5% of the employees are drug users)
- $Pr(N)$ is the probability that a random employee is not a drug user, $Pr(N) = 1 - Pr(D) = 0.995$
- $Pr(+ \mid D)$ is the probability that the test is positive, given that the employee is a drug user, $Pr(+ \mid D) = 0.99$
- $Pr(+ \mid N)$ is the probability that the test is positive, given that the employee is not a drug user, $Pr(+ \mid N) = 0.01$ (since the test will produce a false positive for 1% of non-users)
- $Pr(+)$ is the probability of a positive test,

  $$Pr(+) = Pr(+ \mid D) \cdot Pr(D) + Pr(+ \mid N) \cdot Pr(N) = 0.99 \cdot 0.005 + 0.01 \cdot 0.995 = 0.495$$

Bayes' Theorem (cont.)

- Bayes' Theorem: there are several variants

$$Pr(E_1 \mid E_2) = \frac{Pr(E_2 \mid E_1) \cdot Pr(E_1)}{Pr(E_2 \mid E_1) \cdot Pr(E_1) + Pr(E_2 \mid \overline{E_1}) \cdot Pr(\overline{E_1})}$$

- General Bayes' Theorem If $E_1, \ldots E_n$ are disjoint events such that $W = \bigcup_{i=1}^{n} E_i$

$$Pr(E_k \mid E) = \frac{Pr(E \mid E_k) \cdot Pr(E_k)}{\sum_{i=1}^{n} Pr(E \mid E_i) \cdot Pr(E_i)}$$

- Multiplication Rule. If If $E_1, \ldots E_n$ are events such that $Pr(E_1 \cap \ldots \cap E_{n-1}) > 0$ then

$$Pr(E_1 \cap \ldots \cap E_n) = Pr(E_1) \cdot Pr(E_2 \mid E_1) \cdot Pr(E_3 \mid E_1 \cap E_2) \cdot Pr(E_n \mid E_1 \cap \ldots \cap E_{n-1})$$

- Useful for experiments defined in terms of stages: $Pr(E_j \mid E_1 \cap \ldots \cap E_{j-1})$ is the probability of an event described in terms of what happens on stage j conditioned on what happens on stages $1, 2, \ldots j-1$

Extensions of Bayes' Theorem

$$Pr(E \mid E_1 \cap E_2) = \frac{Pr(E) \cdot Pr(E_1 \mid E) \cdot Pr(E_2 \mid E_1 \cap E_2)}{Pr(E_1) \cdot Pr(E_2 \mid E_1)}$$

$$Pr(E \mid E_1 \cap E_2) = \frac{Pr(E_1 \mid E \cap E_2) \cdot Pr(E \mid E_1)}{Pr(E_1 \mid E_2)}$$

Independence of Events

- Events E_1, E_2 are independent iff one of the following conditions hold

$$\begin{aligned} Pr(E_1 \cap E_2) &= Pr(E_1) \cdot Pr(E_2) \\ Pr(E_1 \mid E_2) &= Pr(E_1), \qquad \text{if } Pr(E_2) > 0 \\ Pr(E_2 \mid E_1) &= Pr(E_2), \qquad \text{if } Pr(E_1) > 0 \end{aligned}$$

- Events $E_1, E_2, \ldots, E_n$ are independent iff

$$\begin{aligned} Pr(E_i \cap E_j) &= Pr(E_i) \cdot Pr(E_j), \qquad \text{for } i \neq j \\ Pr(E_i \cap E_j \cap E_k) &= Pr(E_i) \cdot Pr(E_j) \cdot Pr(E_k), \qquad \text{for } i \neq j, i \neq k, j \neq k \\ \vdots \quad &\vdots \quad \vdots \\ Pr(\bigcap_{i=1}^{n} E_i) &= \prod_{i=1}^{n} Pr(E_i) \end{aligned}$$

- If E_1 and E_2 are independent , then
  1. E_1 and $\overline{E_2}$ are independent, $\overline{E_1}$ and E_2 are independent
  2. $Pr(E_1 \cup E_2) = Pr(E_1) + Pr(E_2) - Pr(E_1) \cdot Pr(E_2)$

Discrete distributions

- Assume W is a countable set of possible worlds. We may assume that $W \subseteq \mathbb{N}$
- A discrete probability distribution over W is a function $\mu : W \to [0, 1]$ such that

$$\sum_{x \in W} \mu(x) = 1$$

- $\mu(x)$ indicates the probability that the world $x \in W$ is indeed the actual one

$$Pr(\{x\}) = \mu(x)$$

- Uniform distribution: W finite and all worlds are equal likely

$$\mu(x) = 1/|W|$$

- Probability of event E under distribution μ:

$$Pr(E) \quad = \quad \sum_{x \in E} \mu(x)$$

- Expectation of event E under distribution μ:

$$\mathcal{E}[E] \quad = \quad \sum_{x \in E} x \cdot \mu(x)$$

Example

- Throwing two dices and take the sum
- $W = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$
- Probability distribution:

	1	2	3	4	5	6	7	8	9	10	11	12
$\mu(x)$	0	$1/36$	$2/36$	$3/36$	$4/36$	$5/36$	$6/36$	$5/36$	$4/36$	$3/26$	$2/36$	$1/36$

- Let E be the event "the sum is at most 5", $E = \{1, 2, 3, 4, 5\}$

$$Pr(E) = \sum_{x \in E} \mu(x) = 0 + \frac{1}{36} + \frac{2}{36} + \frac{3}{36} + \frac{4}{36} = \frac{10}{36} = 0.2777$$

$$\mathcal{E}[E] = \sum_{x \in E} x \cdot \mu(x) = 1 \cdot 0 + 2 \cdot \frac{1}{36} + 3 \cdot \frac{2}{36} + 4 \cdot \frac{3}{36} + 5 \cdot \frac{4}{36} = \frac{40}{36} = 1.1111$$

- Remark

$$Pr(W) = \sum_{x \in W} \mu(x) = 1$$

$$\mathcal{E}[W] = \sum_{x \in W} x \cdot \mu(x) = 7$$

Probability & Logic

- Any statement φ is either true or false
- Due to lack of knowledge we can only estimate to which probability degree they are true or false
- Usually we have a possible world semantics with a distribution over possible worlds
- Possible world: any classical interpretation I, mapping any statement φ into $\{0, 1\}$

$$W = \{I \text{ classical interpretation}\}, \quad I(\varphi) \in \{0, 1\}$$

- Probability distribution: a mapping

$$\mu \colon W \to [0, 1], \quad \mu(I) \in [0, 1]$$

such that

$$\sum_{I \in W} \mu(I) = 1$$

- $\mu(I)$ indicates the probability that the world I is indeed the actual one

- A statement φ corresponds to the event M_ϕ "the set of models of φ", i.e.

$$M_\varphi = \{I \mid I \models \varphi\}$$

- The probability of a statement φ is determined as

$$Pr(\varphi) = Pr(M_\varphi) = \sum_{I \models \varphi} \mu(I)$$

Example

Probabilistic setting:

$$\varphi = sprinklerOn \vee wet$$

W	*sprinklerOn*	*wet*	μ
I_1	0	0	0.1
I_2	0	1	0.2
I_3	1	0	0.4
I_4	1	1	0.3

$$1 = \sum_{I \in W} \mu(I)$$

$$\begin{aligned} Pr(\varphi) &= Pr(\{I_2, I_3, I_4\}) \\ &= 0.2 + 0.4 + 0.3 = 0.9 \end{aligned}$$

Properties of probabilistic formulae

$$
\begin{array}{lcl}
Pr(\varphi \wedge \psi) & = & Pr(\varphi) + Pr(\psi) - Pr(\varphi \vee \psi) \\
Pr(\varphi \wedge \psi) & \leq & \min(Pr(\varphi), Pr(\psi)) \\
Pr(\varphi \wedge \psi) & \geq & \max(0, Pr(\varphi) + Pr(\psi) - 1) \\
Pr(\varphi \vee \psi) & = & Pr(\varphi) + Pr(\psi) - Pr(\varphi \wedge \psi) \\
Pr(\varphi \vee \psi) & \leq & \min(1, Pr(\varphi) + Pr(\psi)) \\
Pr(\varphi \vee \psi) & \geq & \max(Pr(\varphi), Pr(\psi)) \\
Pr(\neg\varphi) & = & 1 - Pr(\varphi) \\
Pr(\bot) & = & 0 \\
Pr(\top) & = & 1
\end{array}
$$

Probabilistic Knowledge Bases

- Finite nonempty set of basic events $\Phi = \{p_1, \ldots, p_n\}$.
- Event φ: Boolean combination of basic events
- Logical constraint $\psi \Leftarrow \varphi$: events ψ and φ: "φ implies ψ".
- Conditional constraint $(\psi|\varphi)[l, u]$: events ψ and φ, and $l, u \in [0, 1]$: "conditional probability of ψ given φ is in $[l, u]$".
- $\psi \geq l$ is a shortcut for $(\psi|\top)[l, 1]$, $\psi \leq u$ is a shortcut for $(\psi|\top)[0, u]$
- Probabilistic knowledge base $KB = (L, P)$:
  - finite set of logical constraints L,
  - finite set of conditional constraints P.

Example

Probabilistic knowledge base $KB = (L, P)$:

- $L = \{bird \Leftarrow eagle\}$:

  "Eagles are birds".

- $P = \{(have\_legs \mid bird)[1, 1], (fly \mid bird)[0.95, 1]\}$:

  "Birds have legs".

  "Birds fly with a probability of at least 0.95".

- World I: truth assignment to all basic events in Φ.
- $\mathcal{I}_\Phi$: all worlds for Φ.
- Probabilistic interpretation Pr: probability distribution on $\mathcal{I}_\Phi$.
- $Pr(\varphi)$: sum of all $Pr(I)$ such that $I \in \mathcal{I}_\Phi$ and $I \models \varphi$.
- $Pr(\psi|\varphi)$: if $Pr(\varphi) > 0$, then $Pr(\psi|\varphi) = Pr(\psi \wedge \varphi) \,/\, Pr(\varphi)$.
- Truth under Pr:
  - $Pr \models \psi \Leftarrow \varphi$ iff $Pr(\psi \wedge \varphi) = Pr(\varphi)$ (iff $Pr(\psi \Leftarrow \varphi) = 1$).
  - $Pr \models (\psi|\varphi)[l, u]$ iff $Pr(\psi \wedge \varphi) \in [l, u] \cdot Pr(\varphi)$ (iff either $Pr(\varphi) = 0$ or $Pr(\psi|\varphi) \in [l, u]$).

Example

- Set of basic propositions $\Phi = \{bird, fly\}$.
- $\mathcal{I}_\Phi$ contains exactly the worlds I_1, I_2, I_3, and I_4 over Φ:

	fly	$\neg$*fly*
bird	I_1	I_2
$\neg$*bird*	I_3	I_4

- Some probabilistic interpretations:

Pr_1	*fly*	$\neg$*fly*
bird	19/40	1/40
$\neg$*bird*	10/40	10/40

Pr_2	*fly*	$\neg$*fly*
bird	0	1/3
$\neg$*bird*	1/3	1/3

- $Pr_1(fly \wedge bird) = 19/40$ and $Pr_1(bird) = 20/40$.
- $Pr_2(fly \wedge bird) = 0$ and $Pr_2(bird) = 1/3$.
- $\neg fly \Leftarrow bird$ is false in Pr_1, but true in Pr_2.
- $(fly \mid bird)[.95, 1]$ is true in Pr_1, but false in Pr_2.

Satisfiability and Logical Entailment

- Pr is a model of $KB = (L, P)$ iff $Pr \models F$ for all $F \in L \cup P$.
- KB is satisfiable iff a model of KB exists.
- $KB \Vdash (\psi|\varphi)[l, u]$: $(\psi|\varphi)[l, u]$ is a logical consequence of KB iff every model of KB is also a model of $(\psi|\varphi)[l, u]$.
- $KB \Vdash_{tight} (\psi|\varphi)[l, u]$: $(\psi|\varphi)[l, u]$ is a tight logical consequence of KB iff l (resp., u) is the infimum (resp., supremum) of $Pr(\psi|\varphi)$ subject to all models Pr of KB with $Pr(\varphi) > 0$.

Example

- Probabilistic knowledge base:

$$KB = (\{bird \Leftarrow eagle\}, \{(have\_legs \mid bird)[1,1], (fly \mid bird)[0.95,1]\}).$$

- KB is satisfiable, since

  Pr with $Pr(bird \wedge eagle \wedge have\_legs \wedge fly) = 1$ is a model.

- Some conclusions under logical entailment:

$$KB \Vdash (have\_legs \mid bird)[0.3,1], \quad KB \Vdash (fly \mid bird)[0.6,1].$$

- Tight conclusions under logical entailment:

$$KB \Vdash_{tight} (have\_legs \mid bird)[1,1], \quad KB \Vdash_{tight} (fly \mid bird)[0.95,1],$$
$$KB \Vdash_{tight} (have\_legs \mid eagle)[1,1], \quad KB \Vdash_{tight} (fly \mid eagle)[0,1].$$

Exercise

Encode the Student Example

Deciding Model Existence / Satisfiability

Theorem: The probabilistic knowledge base $KB = (L, P)$ has a model Pr iff the following system of linear constraints over the variables y_r ($r \in R$), where $R = \{I \in \mathcal{I}_\Phi \mid I \models L\}$, is solvable:

$$
\begin{aligned}
\sum_{r \in R,\, r \models \neg\psi \wedge \varphi} -l\, y_r + \sum_{r \in R,\, r \models \psi \wedge \varphi} (1 - l)\, y_r &\geq 0 \quad (\forall (\psi|\varphi)[l, u] \in P) \\
\sum_{r \in R,\, r \models \neg\psi \wedge \varphi} u\, y_r + \sum_{r \in R,\, r \models \psi \wedge \varphi} (u - 1)\, y_r &\geq 0 \quad (\forall (\psi|\varphi)[l, u] \in P) \\
\sum_{r \in R,\, r \models \alpha} y_r &= 1 \\
y_r &\geq 0 \quad (\text{for all } r \in R)
\end{aligned}
$$

Explanation

- A probability distribution Pr is a model of $(\psi|\varphi)[l, u]$ iff

$$
\begin{array}{rll}
Pr(\psi \mid \varphi) \in [l, u] & \text{iff} & Pr(\psi \wedge \varphi)/Pr(\varphi) \in [l, u] \\
& \text{iff} & Pr(\psi \wedge \varphi) \in [l \cdot Pr(\varphi), u \cdot Pr(\varphi)] \\
& \text{iff} & Pr(\psi \wedge \varphi) \geq l \cdot Pr(\varphi) \text{ and , } Pr(\psi \wedge \varphi) \leq u \cdot Pr(\varphi) \\
Pr(\psi \wedge \varphi) \geq l \cdot Pr(\varphi) & \text{iff} & Pr(\psi \wedge \varphi) - l \cdot Pr(\varphi) \geq 0 \\
& \text{iff} & Pr(M_{\psi \wedge \varphi}) - l \cdot Pr(M_{\varphi}) \geq 0 \\
& \text{iff} & Pr(M_{\psi \wedge \varphi}) - l \cdot Pr(M_{\psi \wedge \varphi} \cup M_{\neg\psi \wedge \varphi}) \geq 0 \\
& \text{iff} & Pr(M_{\psi \wedge \varphi}) - l \cdot Pr(M_{\psi \wedge \varphi}) - l \cdot Pr(M_{\neg\psi \wedge \varphi}) \geq 0 \\
& \text{iff} & (1 - l) \cdot Pr(M_{\psi \wedge \varphi}) - l \cdot Pr(M_{\neg\psi \wedge \varphi}) \geq 0 \\
& \text{iff} & (1 - l) \sum_{r \models \psi \wedge \varphi} \mu(r) - l \sum_{lr \models \neg\psi \wedge \varphi} \mu(r) \geq 0 \\
& \text{iff} & \sum_{r \models \psi \wedge \varphi} (1 - l)\mu(r) + \sum_{l \models \neg\psi \wedge \varphi} (-l)\mu(r) \geq 0
\end{array}
$$

- As we are looking for the values of $\mu(r)$, by setting $y_r = \mu(r)$, any solution to the variables y_r under

$$
\begin{array}{rcl}
\sum_{r \models \psi \wedge \varphi} (1 - l)y_r + \sum_{l \models \neg\psi \wedge \varphi} (-l)y_r & \geq & 0 \\
\sum_{r \in W} y_r & = & 1 \\
y_r & \geq & 0 \text{ for all } r \in W
\end{array}
$$

is a probabilistic model of $(\psi|\varphi)[l, 1]$. The equations for the upper bound are derived similarly.

Computing Tight Logical Consequences

Theorem: Suppose $KB = (L, P)$ has a model Pr such that $Pr(\alpha) > 0$. Then, l (resp., u) such that $KB \,\|\!\!\models_{tight} (\beta|\alpha)[l, u]$ is given by the optimal value of the following linear program over the variables y_r ($r \in R$), where $R = \{I \in \mathcal{I}_\Phi \mid I \models L\}$:

$$
\text{minimize (resp., maximize)} \sum_{r \in R,\, r \models \beta \wedge \alpha} y_r \quad \text{subject to}
$$

$$
\begin{aligned}
\sum_{r\in R,\, r\models\neg\psi\wedge\varphi} -l\, y_r + \sum_{r\in R,\, r\models\psi\wedge\varphi} (1-l)\, y_r &\geq 0 \quad (\forall(\psi|\varphi)[l,u] \in P)\\
\sum_{r\in R,\, r\models\neg\psi\wedge\varphi} u\, y_r + \sum_{r\in R,\, r\models\psi\wedge\varphi} (u-1)\, y_r &\geq 0 \quad (\forall(\psi|\varphi)[l,u] \in P)\\
\sum_{r\in R,\, r\models\alpha} y_r &= 1\\
y_r &\geq 0 \quad (\text{for all } r \in R)
\end{aligned}
$$

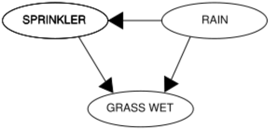
Bayesian Networks

Bayesian network (BN): compact specification of a joint distribution, based on a graphical notation for conditional independencies:

- a set of nodes; each node represents a random variable
- a directed, acyclic graph (link $\approx$ "directly influences")
- a conditional distribution for each node given its parents: $\mathbf{P}(X_i | Parents(X_i))$
- $$Pr(X_1, \ldots, X_n) = \Pi_{i=1}^{n} Pr(X_i \mid parents(X_i)) \ .$$

Any joint distribution can be represented as a BN.

	SPRINKLER	
RAIN	T	F
F	0.4	0.6
T	0.01	0.99



RAIN	
T	F
0.2	0.8

		GRASS WET	
SPRINKLER	RAIN	T	F
F	F	0.0	1.0
F	T	0.8	0.2
T	F	0.9	0.1
T	T	0.99	0.01

Joint probability function is

$$\begin{aligned} Pr(\texttt{GrassWet}, \texttt{Sprinkler}, \texttt{Rain}) &= Pr(\texttt{GrassWet} \mid \texttt{Sprinkler}, \texttt{Rain}) \\ &\quad \cdot Pr(\texttt{Sprinkler} \mid \texttt{Rain}) \cdot Pr(\texttt{Rain}) \,. \end{aligned} \tag{2}$$

The model can answer questions like “What is the probability that it is raining, given the grass is wet?”

$$\begin{aligned} Pr(\texttt{Rain} = T \mid \texttt{GrassWet} = T) &= \frac{Pr(\texttt{Rain} = T, \texttt{GrassWet} = T)}{Pr(\texttt{GrassWet} = T)} \\ &= \frac{\sum_{Y \in \{T,F\}} Pr(\texttt{Rain} = T, \texttt{GrassWet} = T, \texttt{Sprinkler} = Y)}{\sum_{Y_1, Y_2 \in \{T,F\}} Pr(\texttt{GrassWet} = T, (\texttt{Rain} = Y_1, \texttt{Sprinkler} = Y_2))} \\ &= \frac{0.99 \cdot 0.01 \cdot 0.2 + 0.8 \cdot 0.99 \cdot 0.2}{0.99 \cdot 0.01 \cdot 0.2 + 0.9 \cdot 0.4 \cdot 0.8 + 0.8 \cdot 0.99 \cdot 0.2 + 0 \cdot 0.6 \cdot 0.8} \\ &\approx 0.3577 \,. \end{aligned}$$

Encoding of Bayesian Network in Probabilistic Propositional Logic

- For every node a, we use a propositional letters $a(T)$ (a is true), $a(F)$ (a is false)
- We also need $(a(T) \leftrightarrow \neg a(F)) = 1)$
- If a node a has no parents: $a(T) = p$, where p is its associated probability
- If a node has parents, we encode its associated conditional probability table using conditional probability formulae

$$
\begin{aligned}
(\texttt{Sprinkler}(T) \mid \texttt{Rain}(F)) &= 0.4 \\
(\texttt{Sprinkler}(T) \mid \texttt{Rain}(T)) &= 0.01
\end{aligned}
$$

$$
\begin{aligned}
(\texttt{GrassWet}(T) \mid \texttt{Sprinkler}(F) \wedge \texttt{Rain}(F)) &= 0.0 \\
(\texttt{GrassWet}(T) \mid \texttt{Sprinkler}(F) \wedge \texttt{Rain}(T)) &= 0.8 \\
(\texttt{GrassWet}(T) \mid \texttt{Sprinkler}(T) \wedge \texttt{Rain}(F)) &= 0.9 \\
(\texttt{GrassWet}(T) \mid \texttt{Sprinkler}(T) \wedge \texttt{Rain}(T)) &= 0.99\ .
\end{aligned}
$$

Independent Choice Logic: Propositional Case

- A knowledge base $KB = \langle P, \mathcal{C} \rangle$ is a set of propositional formulae P together with a choice space $\mathcal{C}$
- A choice space $\mathcal{C}$ is a set $\mathcal{C}$ of choices of the form $\{(A_1 : \alpha_1), \ldots, (A_n : \alpha_n)\}$, where A_i is an atom and the α_i sum-up to 1
- A total choice T is a set of atoms such that from each choice $C_j \in \mathcal{C}$ there is exactly one atom $A_i^j \in C_j$ in T
- The probability of a total choice T is $Pr(T) = Pr(\bigwedge_{A_i^j \in T} A_i^j) = \prod_{A_i^j \in T} \alpha_i^j$
- A query is a propositional formula q. The probability of q w.r.t. KB is

$$Pr(q \mid KB) = \sum_{\{T \mid P \cup T \models q\}} Pr(T)$$

- Example:

$$P = \{a \rightarrow c, b \rightarrow c\}$$
$$\mathcal{C} = \{C_1 = \{a : 0.7, \neg a : 0.3\}, C_2 = \{b : 0.6, \neg b : 0.4\}\}$$

	Total Choice	$Pr(T)$
T_1	$\{a, b\}$	0.42
T_2	$\{a, \neg b\}$	0.28
T_3	$\{\neg a, b\}$	0.18
T_4	$\{\neg a, \neg b\}$	0.12

$$Pr(c \mid KB) = Pr(T_1) + Pr(T_2) + Pr(T_3) = 1 - Pr(T_4) = 0.88$$

Exercise

Show that Bayesian Networks may be simulated using ICL

Vagueness & Logic

- Statements involve concepts for which there is no exact definition, such as
  - tall, small, close, far, cheap, expensive, “is about”, “similar to”.
- A statements is true to some degree, which is taken from a truth space
- E.g., “Hotel Verdi is close to the train station to degree 0.83”
- E.g., “The image is about a sun set to degree 0.75”
- Truth space: set of truth values L and an partial order $\leq$
- Many-valued Interpretation: a function I mapping formulae into L, i.e. $I(\varphi) \in L$
- Mathematical Fuzzy Logic: $L = [0,1]$, but also $\{\frac{0}{n}, \frac{1}{n}, \ldots, \frac{n}{n}\}$ for an integer $n \geq 1$

- Problem: what is the interpretation of e.g. $\varphi \wedge \psi$?
  - E.g., if $I(\varphi) = 0.83$ and $I(\psi) = 0.2$, what is the result of $0.83 \wedge 0.2$?
- More generally, what is the result of $n \wedge m$, for $n, m \in [0, 1]$?
- The choice cannot be any arbitrary computable function, but has to reflect some basic properties that one expects to hold for a “conjunction”
- Norms: functions that are used to interpret connectives such as $\wedge, \vee, \neg, \rightarrow$
  - t-norm: interprets conjunction
  - s-norm: interprets disjunction
- Norms are compatible with classical two-valued logic

Axioms for t-norms and s-norms

Axiom Name	T-norm	S-norm
Tautology / Contradiction	$a \wedge 0 = 0$	$a \vee 1 = 1$
Identity	$a \wedge 1 = a$	$a \vee 0 = a$
Commutativity	$a \wedge b = b \wedge a$	$a \vee b = b \vee a$
Associativity	$(a \wedge b) \wedge c = a \wedge (b \wedge c)$	$(a \vee b) \vee c = a \vee (b \vee c)$
Monotonicity	if $b \leq c$, then $a \wedge b \leq a \wedge c$	if $b \leq c$, then $a \vee b \leq a \vee c$

Axioms for implication and negation functions

Axiom Name	Implication Function	Negation Function
Tautology / Contradiction	$0 \to b = 1$	$\neg 0 = 1, \ \neg 1 = 0$
	$a \to 1 = 1$	
Antitonicity	if $a \leq b$, then $a \to c \geq b \to c$	if $a \leq b$, then $\neg a \geq \neg b$
Monotonicity	if $b \leq c$, then $a \to b \leq a \to c$	

Usually,

$$a \to b = \sup\{c\colon a \wedge c \leq b\}$$

is used and is called r-implication and depends on the t-norm only

Typical norms

	Lukasiewicz Logic	Gödel Logic	Product Logic	Zadeh
$\neg x$	$1 - x$	if $x = 0$ then 1 else 0	if $x = 0$ then 1 else 0	$1 - x$
$x \wedge y$	$\max(x + y - 1, 0)$	$\min(x, y)$	$x \cdot y$	$\min(x, y)$
$x \vee y$	$\min(x + y, 1)$	$\max(x, y)$	$x + y - x \cdot y$	$\max(x, y)$
$x \Rightarrow y$	if $x \leq y$ then 1 else $1 - x + y$	if $x \leq y$ then 1 else y	if $x \leq y$ then 1 else y/x	$\max(1 - x, y)$

Note: for Lukasiewicz Logic and Zadeh, $x \Rightarrow y \equiv \neg x \vee y$

- Any other t-norm can be obtained as a combination of Lukasiewicz, Gödel and Product t-norm
- Zadeh: not interesting for mathematical fuzzy logicians: its a sub-logic of Łukasiewicz and, thus, rarely considered by fuzzy logicians

$$
\begin{aligned}
\neg_Z x &= \neg_Ł x \\
x \wedge_Z y &= x \wedge_Ł (x \rightarrow_Ł y) \\
x \rightarrow_Z y &= \neg_Ł x \vee_Ł y
\end{aligned}
$$

Some additional properties of t-norms, s-norms, implication functions, and negation functions of various fuzzy logics.

Property	Łukasiewicz Logic	Gödel Logic	Product Logic	Zadeh Logic
$x \wedge \neg x = 0$	•	•	•	
$x \vee \neg x = 1$	•			
$x \wedge x = x$		•		•
$x \vee x = x$		•		•
$\neg \neg x = x$	•			•
$x \Rightarrow y = \neg x \vee y$	•			•
$\neg (x \Rightarrow y) = x \wedge \neg y$	•			•
$\neg (x \wedge y) = \neg x \vee \neg y$	•	•	•	•
$\neg (x \vee y) = \neg x \wedge \neg y$	•	•	•	•
$x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z)$		•		•
$x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z)$		•		•

- Note: If all conditions in the upper part of a column have to be satisfied then we collapse to classical two-valued logic, i.e. $L = \{0, 1\}$

Propositional Fuzzy Logic

- Formulae: propositional formulae
- Truth space is $[0, 1]$
- Formulae have a a degree of truth in $[0, 1]$
- Interpretation: is a mapping $\mathcal{I} : Atoms \rightarrow [0, 1]$
- Interpretations are extended to formulae using norms to interpret connectives $\wedge, \vee, \neg, \rightarrow$

$$
\begin{array}{lcl}
\mathcal{I}(\varphi \wedge \psi) & = & \mathcal{I}(\varphi) \wedge \mathcal{I}(\psi) \\
\mathcal{I}(\varphi \vee \psi) & = & \mathcal{I}(\varphi) \vee \mathcal{I}(\psi) \\
\mathcal{I}(\varphi \rightarrow \psi) & = & \mathcal{I}(\varphi) \rightarrow \mathcal{I}(\psi) \\
\mathcal{I}(\neg \varphi) & = & \neg \mathcal{I}(\varphi)
\end{array}
$$

- Rational $r \in [0, 1]$ may appear as atom in formula, where $\mathcal{I}(r) = r$

Example

In Lukasiewicz logic:

$$\varphi = \mathit{Cold} \wedge \mathit{Cloudy}$$

$\mathcal{I}$	*Cold*	*Cloudy*	$\mathcal{I}(\varphi)$
$\mathcal{I}_1$	0	0.1	$\max(0, 0 + 0.1 - 1) = 0.0$
$\mathcal{I}_2$	0.3	0.4	$\max(0, 0.3 + 0.4 - 1) = 0.0$
$\mathcal{I}_3$	0.7	0.8	$\max(0, 0.7 + 0.8 - 1) = 0.5$
$\mathcal{I}_4$	1	1	$\max(0, 1 + 1 - 1) = 1.0$
$\vdots$	$\vdots$	$\vdots$	$\vdots$

- Note:

$$\mathcal{I}(r \rightarrow \varphi) = 1 \quad \textit{iff} \quad \mathcal{I}(\varphi) \geq r$$
$$\mathcal{I}(\varphi \rightarrow r) = 1 \quad \textit{iff} \quad \mathcal{I}(\varphi) \leq r$$

- We use $\varphi \geq r$ as an abbreviation of $r \rightarrow \varphi$ and $\varphi \leq r$ as an abbreviation of $\varphi \rightarrow r$
- Semantics:

$$\begin{array}{rll} I \models \varphi & \textit{iff} & \mathcal{I}(\varphi) = 1 \\ \mathcal{I} \models KB & \text{iff} & \mathcal{I} \models \varphi \text{ for all } \varphi \in KB \\ KB \models \varphi & \text{iff} & \text{for all } \mathcal{I}. \text{ if } \mathcal{I} \models KB \text{ then } \mathcal{I} \models \varphi \end{array}$$

- Deduction rule is valid: for $r, s \in [0, 1]$:

$$r \rightarrow \varphi, s \rightarrow (\varphi \rightarrow \psi) \models (r \wedge s) \rightarrow \psi$$

Informally,

From $\varphi \geq r$ and $(\varphi \rightarrow \psi) \geq s$ infer $\psi \geq r \wedge s$

Example

In Lukasiewicz logic:

$$\varphi = 0.4 \rightarrow (Cold \wedge Cloudy)$$

Read: $Cold \wedge Cloudy \geq 0.4$

$\mathcal{I}$	*Cold*	*Cloudy*	$\mathcal{I}(\varphi)$
$\mathcal{I}_1$	0	0.1	$0.4 \rightarrow 0.0 = \min(1, 1 - 0.4 + 0.0) = 0.6$
$\mathcal{I}_2$	0.3	0.4	$0.4 \rightarrow 0.0 = \min(1, 1 - 0.4 + 0.0) = 0.6$
$\mathcal{I}_3$	0.7	0.8	$0.4 \rightarrow 0.5 = \min(1, 1 - 0.4 + 0.5) = 1.0$
$\mathcal{I}_4$	1	1	$0.4 \rightarrow 1.0 = \min(1, 1 - 0.4 + 1.0) = 1.0$
$\vdots$	$\vdots$	$\vdots$	$\vdots$

$$\begin{array}{ccc}
\mathcal{I}_1 & \not\models & \varphi \\
\mathcal{I}_2 & \not\models & \varphi \\
\mathcal{I}_3 & \models & \varphi \\
\mathcal{I}_4 & \models & \varphi \\
\vdots & \vdots & \vdots
\end{array}$$

- Let

$$bsd(KB, \phi) = \sup\{\mathcal{I}(\phi) \mid \mathcal{I} \models KB\} \text{ (Best Satisfiability Degree (BSD))}$$

$$bed(KB, \phi) = \sup\{r \mid KB \models \varphi \geq r\} \text{ (Best Entailment Degree (BED))}$$

- Then

$$bed(KB, \phi) = \min x. \text{ such that } KB \cup \{\phi \leq x\} \text{ satisfiable.}$$

- Assume KB is a set of formulae $\phi \geq n$ or $\phi \leq n$
- For a formula ϕ consider a variable x_ϕ (that the degree of truth of ϕ is greater or equal to x_ϕ)
- E.g., for Łukasiewicz logic, use Mixed Integer Linear Programming

$$\begin{array}{rl} bed(KB, \phi) = \min x. \text{ such that } & x \in [0,1], x_\phi \leq x, \sigma(\phi), \\ & \text{for all } \phi' \geq n \in KB, x_{\phi'} \geq n, \sigma(\phi'), \\ & \text{for all } \phi' \leq n \in KB, x_{\phi'} \leq n, \sigma(\phi') \end{array}$$

$$\sigma(\phi) = \begin{cases} x_p \in [0,1] & \text{if} \quad \phi = p \\ x_r = r & \text{if} \quad \phi = r, r \in [0,1] \\ x_{\phi'} = \ominus x_\phi, x_\phi \in [0,1] & \text{if} \quad \phi = \neg\phi' \\ \begin{array}{l} x_{\phi_1} \otimes x_{\phi_2} = x_\phi, \\ \sigma(\phi_1), \sigma(\phi_2), x_\phi \in [0,1] \end{array} & \text{if} \quad \phi = \phi_1 \wedge \phi_2 \\ x_{\phi_1} \oplus x_{\phi_2} = x_\phi & \text{if} \quad \phi = \phi_1 \vee \phi_2 \\ \sigma(\neg\phi_1 \vee \phi_2) & \text{if} \quad \phi = \phi_1 \rightarrow \phi_2 . \end{cases}$$

where

$$\begin{array}{rcl} x_1 = \ominus x_2 & \mapsto & x_1 = 1 - x_2 \\ x_1 \oplus x_2 = z & \mapsto & \{y \leq z, x_1 + x_2 \geq y, z \leq x_1 + x_2 \leq z + y, y \in \{0,1\}\} \\ x_1 \otimes x_2 = z & \mapsto & \{z \leq y, x_1 + x_2 - 1 \geq y, z - y \leq x_1 + x_2 - 1 \leq z, y \in \{0,1\}\} \end{array}$$

- In a similar way, we may determine $bsd(KB, \phi)$ as

$$\begin{aligned} \min -x.\ \text{such that } & x \in [0,1], x_\phi \geq x, \sigma(\phi), \\ & \text{for all } \phi' \geq n \in KB, x_{\phi'} \geq n, \sigma(\phi'), \\ & \text{for all } \phi' \leq n \in KB, x_{\phi'} \leq n, \sigma(\phi') \end{aligned}$$

Example

- Consider $KB = \{p \geq 0.6, p \rightarrow q \geq 0.7\}$
- Let us show that $bed(q, KB) = 0.3$
- Recall that $bed(q, KB)$ is

$$\begin{aligned} &\min x.\ \text{such that } x \in [0,1], x_q \leq x, \sigma(q), \\ &\quad \text{for all } \phi' \geq n \in KB, x_{\phi'} \geq n, \sigma(\phi'), \\ &\quad \text{for all } \phi' \leq n \in KB, x_{\phi'} \leq n, \sigma(\phi') \end{aligned}$$

$$\begin{aligned} p \geq 0.6 &\quad\mapsto\quad x_p \geq 0.6, x_p \in [0,1] \\ p \rightarrow q \geq 0.7 &\quad\mapsto\quad x_{p\rightarrow q} \geq 0.7, x_{p\rightarrow q} \in [0,1], \sigma(p \rightarrow q) \\ \sigma(q) &\quad\mapsto\quad x_q \in [0,1] \\ \sigma(p \rightarrow q) &\quad\mapsto\quad x_{\neg p \vee q} = x_{p\rightarrow q}, \sigma(\neg p \vee q) \\ \sigma(\neg p \vee q) &\quad\mapsto\quad x_{\neg p} + x_q = x_{\neg p \vee q}, \sigma(\neg p), \sigma(q), x_{\neg p \vee q} \in [0,1] \\ \sigma(\neg p) &\quad\mapsto\quad x_p = 1 - x_{\neg p}, x_p \in [0,1] \end{aligned}$$

It follows that $0.3 = \min x. \ldots$

Fuzzy Concrete Domains

- Allows us to deal with concepts such as young, cheap, cold, etc.
- We allow also crisp constraints such as AlarmSystem $\wedge$ (price $>$ 26,000), AlarmSystem $\rightarrow$ (deliverytime $\geq$ 30)
- Fuzzy membership functions: usually of the form

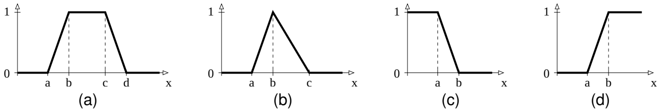


Figure: (a) Trapezoidal function $trz(a, b, c, d)$, (b) triangular function $tri(a, b, c)$, (c) left shoulder function $ls(a, b)$, and (d) right shoulder function $rs(a, b)$.

- For instance, AlarmSystem $\wedge$ (price $ls(18000, 22000)$)

Fuzzy Concrete Domains (cont.)

Definition (The language $\mathcal{P}(\mathcal{N})$)

Let $\mathcal{A}$ be a set of propositional atoms, and $\mathcal{F}$ a set of pairs $\langle f, D_f \rangle$ each made of a feature name and an associated concrete domain D_f, and let k be a value in D_f. Then the following formulae are in $\mathcal{P}(\mathcal{N})$:

1. every atom $A \in \mathcal{A}$ is a formula
2. if $\langle f, D_f \rangle \in \mathcal{F}$, $k \in D_f$, and $c \in \{\geq, \leq, =\}$ then $(f\ c\ k)$ is a formula
3. if $\langle f, D_f \rangle \in \mathcal{F}$ and c is of the form $ls(a, b)$, $rs(a, b)$, $tri(a, b, c)$, $trz(a, b, c, d)$ then $(f\ c)$ is a formula
4. if ψ and φ are formulae and $n \in [0, 1]$ then so are $\neg\psi$, $\psi \wedge \varphi$, $\psi \vee \varphi$, $\psi \rightarrow \varphi$. We use $\psi \leftrightarrow \varphi$ in place of $(\psi \rightarrow \varphi) \wedge (\varphi \rightarrow \psi)$,
5. if $\psi_1, \ldots, \psi_n$ are formulae, then $w_1 \cdot \psi_1 + \ldots + w_n \cdot \psi_n$ is a formula, where $w_i \in [0, 1]$ and $\sum_i w_i \leq 1$
6. if ψ is a formula and $n \in [0, 1]$ then $\langle \psi, n \rangle$ is a formula in $\mathcal{P}(\mathcal{N})$. If n is omitted, then $\langle \psi, 1 \rangle$ is assumed

Definition (Interpretation and models)

An interpretation $\mathcal{I}$ for $\mathcal{P}(\mathcal{N})$ is a function (denoted as a superscript $\cdot^{\mathcal{I}}$ on its argument) that maps each atom in $\mathcal{A}$ into a truth value $A^{\mathcal{I}} \in [0, 1]$, each feature name f into a value $f^{\mathcal{I}} \in D_f$, and assigns truth values in $[0, 1]$ to formulas as follows:

- for hard constraints, $(f\ c\ k)^{\mathcal{I}} = 1$ iff the relation $f^{\mathcal{I}}\ c\ k$ is true in D_f, $(f\ c\ k)^{\mathcal{I}} = 0$ otherwise
- for soft constraints, $(f\ c)^{\mathcal{I}} = c(f^{\mathcal{I}})$, i.e., the result of evaluating the fuzzy membership function c on the value $f^{\mathcal{I}}$
- $(\neg\psi)^{\mathcal{I}} = \neg\psi^{\mathcal{I}}$, $(\psi \wedge \varphi)^{\mathcal{I}} = \psi^{\mathcal{I}} \wedge \varphi^{\mathcal{I}}$, $(\psi \vee \varphi)^{\mathcal{I}} = \psi^{\mathcal{I}} \vee \varphi^{\mathcal{I}}$, $(\psi \rightarrow \varphi)^{\mathcal{I}} = \psi^{\mathcal{I}} \Rightarrow \varphi^{\mathcal{I}}$ and $(w_1 \cdot \psi_1 + \ldots + w_n \cdot \psi_n)^{\mathcal{I}} = \sum_i w_i \cdot \psi_i^{\mathcal{I}}$
- $\mathcal{I} \models \langle \psi, n \rangle$ iff $\psi^{\mathcal{I}} \geq n$.

Example: Matchmaking

- Suppose we have a buyer and a seller (agents)
  - A car seller sells a sedan car
  - A buyer is looking for a second hand passenger car
  - Both the buyer as well as the seller have preferences (restrictions)
  - There is some background knowledge
- The objective is determine “an optimal” (Pareto optimal) agreement among the two

Matchmaking Example: the Background Knowledge

1. A sedan is a passenger car
2. A satellite alarm system is an alarm system
3. The navigator pack is a satellite alarm system with a GPS system
4. The Insurance Plus package is a driver insurance together with a theft insurance
5. The car colours are black or grey

Matchmaking Example: Buyer's preferences

1. He does not want to pay more than 26000 euro (buyer reservation value)
2. He wants an alarm system in the car and he is completely satisfied with paying no more than 23000 euro, but he can go up to 26000 euro to a lesser degree of satisfaction
3. He wants a driver insurance and either a theft insurance or a fire insurance
4. He wants air conditioning and the external colour should be either black or grey
5. Preferably the price is no more than 22000 euro, but he can go up to 24000 euro to a lesser degree of satisfaction
6. The kilometer warranty is preferrably at least 140000, but he may go down to 160000 to a lesser degree of satisfaction
7. The weights of the preferences 2-6 are, (0.1, 0.2, 0.1, 0.2, 0.4). The higher the value the more important is the preference

Matchmaking Example: Seller's preferences

1. He wants to sell no less than 24000 euro (seller reservation value)
2. If there is an navigator pack system in the car then he is completely satisfied with selling no less than 26000 euro, but he can go down to 24000 euro to a lesser degree of satisfaction
3. Preferably the seller sells the Insurance Plus package
4. The kilometer warranty is preferrably at most 150000, but he may go up to 170000 to a lesser degree of satisfaction
5. If the color is black then the car has air conditioning
6. The weights of the preferences 2-5 are, (0.3, 0.1, 0.4, 0.2). The higher the value the more important is the preference

Matchmaking Example: Encoding

$$\mathcal{T} = \begin{cases} \text{Sedan} \longrightarrow \text{PassengerCar} \\ \text{ExternalColorBlack} \longrightarrow \neg\text{ExternalColorGray} \\ \text{SatelliteAlarm} \longrightarrow \text{AlarmSystem} \\ \text{InsurancePlus} \leftrightarrow \text{DriverInsurance} \wedge \text{TheftInsurance} \\ \text{NavigatorPack} \leftrightarrow \text{SatelliteAlarm} \wedge \text{GPS\_system} \end{cases}$$

Buyer's request:

β = PassengerCar $\wedge$ price $\leq$ 26000
β_1 = AlarmSystem $\Rightarrow$ (price , ls(23000, 26000))
β_2 = DriverInsurance $\wedge$ (TheftInsurance $\vee$ FireInsurance)
β_3 = AirConditioning $\wedge$ (ExternalColorBlack $\vee$ ExternalColorGray)
β_4 = (price , ls(22000, 24000))
β_5 = (km_warranty , rs(140000, 160000))
$\mathcal{B} = 0.1 \cdot \beta_1 + 0.2 \cdot \beta_2 + 0.1 \cdot \beta_3 + 0.2 \cdot \beta_4 + 0.2 \cdot \beta_5$

Seller's request:

σ = Sedan $\wedge$ price $\geq$ 24000
σ_1 = NavigatorPack $\wedge$ (price , rs(24000, 26000))
σ_2 = InsurancePlus
σ_3 = (km_warranty , ls(150000, 170000))
σ_4 = ExternalColorBlack $\wedge$ AirConditioning
$\mathcal{S} = 0.3 \cdot \sigma_1 + 0.1 \cdot \sigma_2 + 0.4 \cdot \sigma_3 + 0.2 \cdot \sigma_4$

Let

$$KB = \mathcal{T} \cup \{\beta, \sigma\} \cup \{\text{buy} \leftrightarrow \mathcal{B}, \text{sell} \leftrightarrow \mathcal{S}\}$$

Pareto optimal solution:

$$bsd(KB, \text{buy} \wedge_{\Pi} \text{sell}) = 0.651$$

In particular, the final agreement is:

Sedan$^{\bar{\mathcal{I}}} = 1.0$, PassengerCar$^{\bar{\mathcal{I}}} = 1.0$, InsurancePlus$^{\bar{\mathcal{I}}} = 1.0$, AlarmSystem$^{\bar{\mathcal{I}}} = 1.0$,
DriverInsurance$^{\bar{\mathcal{I}}} = 1.0$, AirConditioning$^{\bar{\mathcal{I}}} = 1.0$, NavigatorPack$^{\bar{\mathcal{I}}} = 1.0$,
(km_warranty ls(150000, 170000))$^{\bar{\mathcal{I}}} = 0.5$, i.e. km_warranty$^{\bar{\mathcal{I}}} = 160000$,
(price, ls(23000, 26000))$^{\bar{\mathcal{I}}} = 0.33$, i.e. price$^{\bar{\mathcal{I}}} = 24000$,
TheftInsurance$^{\bar{\mathcal{I}}} = 1.0$, FireInsurance$^{\bar{\mathcal{I}}} = 1.0$, ExternalColorBlack$^{\bar{\mathcal{I}}} = 1.0$, ExternalColorGray$^{\bar{\mathcal{I}}} = 0.0$.

Example: (Fuzzy) Multi-Criteria Decision Making

- We have to decide which offer to choose for the development of a Public School
- There are 3 offers (Alternatives), which have been evaluated by an expert according to 3 Criteria
  - Cost, DeliveryTime, Quality

Preliminaries: MCDM Basics

- Alternatives A_i: different choices of action available to the decision maker to be ranked
- Decision criteria C_j: different dimensions from which the alternatives can be viewed and evaluated
- Decision weights w_j: importance of a criteria
- Performance weights a_{ij}: performance of alternative w.r.t. a decision criteria

		Criteria				
		w_1	w_2	.	.	w_m
Alternatives		C_1	C_2	.	.	C_m
x_1	A_1	a_{11}	a_{12}	.	.	a_{1m}
x_2	A_2	a_{21}	a_{22}	.	.	a_{2m}
.	.	.	.	.	.	.
.	.	.	.	.	.	.
x_n	A_n	a_{n1}	a_{n2}	.	.	a_{nm}

(3)

- Final ranking value x_i:

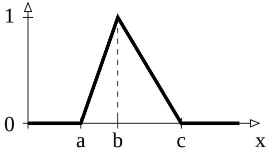
$$x_i = \sum_{j=1}^{m} a_{ij} w_j$$

- Optimal alternative A^*:

$$A^* = \arg\max_{A_i} x_i$$

Preliminaries: Fuzzy MCDM Basics

- Principal difference: weights w_i and performance a_{ij} are fuzzy numbers
- Fuzzy number $\tilde{n}$: fuzzy set over relas with triangular membership function $tri(a, b, c)$. Intended being an approximation of the number b



- Any real value n is seen as the fuzzy number $tri(n, n, n)$
- Arithmetic operators $+, -, \cdot$ and $\div$ are extended to fuzzy numbers
  - For $* \in \{+, \cdot\}$, $\tilde{n}_1 * \tilde{n}_2 = tri(a_1 * a_2, b_1 * b_2, c_1 * c_2)$
  - For $* \in \{-, \div\}$, $\tilde{n}_1 * \tilde{n}_2 = tri(a_1 * c_2, b_1 * b_2, c_1 * a_2)$
- Final ranking value x_i: fuzzy number

$$\tilde{x}_i = \sum_{j=1}^{m} \tilde{a}_{ij} \cdot \tilde{w}_j$$

- Optimal alternative A^*:

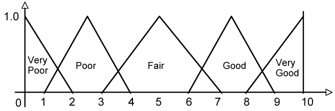
$$A^* = \arg\max_{A_i} x_i$$

  using some fuzzy number ranking method. E.g., Best Non-Fuzzy Performance (BNP): $(a + b + c)/3$

Example: (Fuzzy) Multi-Criteria Decision Making

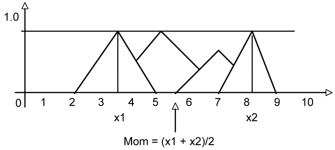
- We have to decide which offer to choose for the development of a Public School
- There are 3 offers (Alternatives), which have been evaluated by an expert according to 3 Criteria
- The importance of alternative A_i against criteria C_j is $a_{ij} \in \{\text{VeryPoor}, \text{Poor}, \text{Fair}, \text{Good}, \text{VeryGood}\}$
- The importance of the criteria is weighted $w_{ij} \in [0, 1]$, $\sum_i w_{ij} = 1$ ($w_1 = 0.3$, $w_2 = 0.2$, $w_3 = 0.5$)

Offer	Cost 0.3	DeliveryTime 0.2	Quality 0.5
A_1	VeryPoor	Fair	Good
A_2	Good	VeryGood	Poor
A_3	Fair	Fair	Poor



$$
\begin{array}{rcl}
KB & = & \{A_1, A_2, A_3\} \text{ where} \\
A_i & \leftrightarrow & w_1 \cdot (\text{hasScore } a_{i1}) + w_2 \cdot (\text{hasScore } a_{i2}) + w_3 \cdot (\text{hasScore } a_{i3})
\end{array}
$$

- The Final Rank Value, $rn(KB, A_i)$, of alternative A_i is defined as the Middle of Maxima (MOM) de-fuzzification method



$$rn(KB, A_1) = 0.75,\ rn(KB, A_2) = 0.25,\ rn(KB, A_3) = 0.375$$

- So, we may choose offer A_1

Note: Computing Middle of Maxima (MOM)

- Middle of Maxima (MOM) = (Largest of Maxima (LOM) + Smallest of Maxima (SOM))/2
- LOM is implemented in the following steps
  1. Compute $n = bsd(A_i, KB)$
  2. Maximise the value of the (internal) variable representing the value of hasScore, i.e. the variable x_{hasScore}, given $KB \cup \{A_i \geq n\}$
- SOM is implemented in the following steps
  1. Compute $n = bsd(A_i, KB)$
  2. Minimise the variable x_{hasScore}, given $KB \cup \{A_i \geq n\}$
- MOM is implemented in the following steps
  1. Compute $n = bsd(A_i, KB)$
  2. Maximise the variable x_{hasScore}, given $KB \cup \{A_i \geq n\}$
  3. Minimise the variable x_{hasScore}, given $KB \cup \{A_i \geq n\}$
  4. Take the average of the two values obtained from the two maximisation and minimisation problems

Predicate Fuzzy Logics Basics

- Formulae: First-Order Logic formulae, *terms* are either variables or constants
  - we may introduce functions symbols as well, with crisp semantics (but uninteresting), or we need to discuss also fuzzy equality (which we leave out here)
- Truth space is $[0, 1]$
- Formulae have a a degree of truth in $[0, 1]$
- Interpretation: is a mapping $\mathcal{I} : Atoms \to [0, 1]$
- Interpretations are extended to formulae as follows:

$$
\begin{array}{rcl}
\mathcal{I}(\neg\phi) & = & \mathcal{I}(\phi) \to 0 \\
\mathcal{I}(\phi \wedge \psi) & = & \mathcal{I}(\phi) \wedge \mathcal{I}(\psi) \\
\mathcal{I}(\phi \to \psi) & = & \mathcal{I}(\phi) \to \mathcal{I}(\psi) \\
\mathcal{I}(\exists x \phi) & = & \sup_{c \in \Delta^{\mathcal{I}}} \mathcal{I}_x^c(\phi) \\
\mathcal{I}(\forall x \phi) & = & \inf_{c \in \Delta^{\mathcal{I}}} \mathcal{I}_x^c(\phi)
\end{array}
$$

  where $\mathcal{I}_x^c$ is as $\mathcal{I}$, except that variable x is mapped into individual c
- Definitions of $\mathcal{I} \models \langle \phi, n \rangle$, $\mathcal{I} \models \mathcal{T}$, $\mathcal{T} \models \langle \phi, n \rangle$, $bed(KB, \phi)$ and $bsd(KB, \phi)$ are as for the propositional case

- $\neg\forall \mathbf{x}\, \varphi(\mathbf{x}) \equiv \exists \mathbf{x}\, \neg\varphi(\mathbf{x})$ true in Ł, but does not hold for logic G and Π
- $(\neg\forall x\, p(x)) \wedge (\neg\exists x\, \neg p(x))$ has no classical model. In Gödel logic it has no finite model, but has an infinite model: for integer $n \geq 1$, let $\mathcal{I}$ such that $\mathcal{I}(p(n)) = 1/n$

$$\begin{aligned}\mathcal{I}(\forall x\, p(x)) &= \inf_n 1/n = 0 \\ \mathcal{I}(\exists x\, \neg p(x)) &= \sup_n \neg 1/n = \sup 0 = 0\end{aligned}$$

- Note: If $\mathcal{I} \models \exists x\, \phi(x)$ then not necessarily there is $c \in \Delta^{\mathcal{I}}$ such that $\mathcal{I} \models \phi(c)$.

$$\begin{aligned}\Delta_{\mathcal{I}} &= \{n \mid \text{ integer } n \geq 1\} \\ \mathcal{I}(p(n)) &= 1 - 1/n < 1, \text{ for all } n \\ \mathcal{I}(\exists x\, p(x)) &= \sup_n 1 - 1/n = 1\end{aligned}$$

- Witnessed formula: $\exists x\, \phi(x)$ is witnessed in $\mathcal{I}$ iff there is $c \in \Delta_{\mathcal{I}}$ such that $\mathcal{I}(\exists x\, \phi(x)) = \mathcal{I}(\phi(c))$ (similarly for $\forall x\, \phi(x)$)
- Witnessed interpretation: $\mathcal{I}$ witnessed if all quantified formulae are witnessed in $\mathcal{I}$

Proposition

In Ł, ϕ is satisfiable iff there is a witnessed model of ϕ.

The proposition does not hold for logic G and Π

Fuzzy Concrete Domains

- Allows us to deal with concepts such as young, cheap, cold, etc.
- Fuzzy membership functions: usually of the form

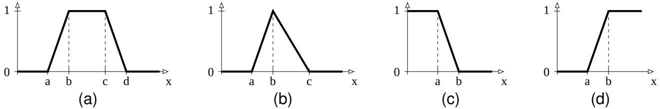


Figure: (a) Trapezoidal function $trz(a, b, c, d)$, (b) triangular function $tri(a, b, c)$, (c) left shoulder function $ls(a, b)$, and (d) right shoulder function $rs(a, b)$.

- Works similarly as for propositional case:
  - We consider a concrete domain over rational numbers with concrete predicates:

$$\geq (x, y), \leq (x, y), = (x, y), ls(a, b)(x), rs(a, b)(x), tri(a, b, c)(x), trz(a, b, c, d)(x)$$

  - Formulae may contain concrete predicates as atom
  - There are variables and constants for rational numbers
  - Formula example

$$\langle \exists r.\mathrm{AlarmSystem}(avs) \land \mathrm{price}(avs, r) \land ls(350, 500)(r), n\rangle$$

- The semantics is an obvious extension of the fuzzy FOL case