Methods of Artificial Intelligence: Lecture

9. Session: Vagueness and Uncertainty II

Kai-Uwe Kühnberger, Nohayr Muhammad

Winter Term 2022/2023

January 13th, 2023

Overview

- Probabilistic Logic
  - Properties of Probabilities
  - Propositional Probabilistic Logic
  - Conditional Probability
- Bayesian Networks
  - Joint Distributions
  - Conditional Independence
  - Example
  - Constructing Bayesian Networks

Probabilistic Logic: Properties of Probabilities

Modeling Uncertainty

Probability Space

- A (finite) probability space is a triple <Ω,Σ,*P*>
  - Ω is a finite non-empty set
  - Σ is the set of all subsets of Ω
  - P: $\Sigma \rightarrow$ **R** is a '*probability measure*'

- Ω is the set of elementary events
  (in other fields also called "possible worlds" or "states")

- Σ is the set of events
  - Each event *E* in Σ is a subset of Ω
  - Intuitively, each event *E* corresponds to a set of possible worlds that satisfy some statement

Probability Measures

- Probability measures assign probabilities to events
- They are characterized by Kolmogorov's axioms (here: finite version)

1. $P(E) \geq 0$ (Non-negativity)
2. $P(\Omega) = 1$ (Normalization)
3. If the events $E_1, E_2, \ldots, E_n$ are disjoint, then (Finite Additivity)

$$P(E_1 \cup E_2 \cup \ldots \cup E_n) = P(E_1) + P(E_2) + \ldots + P(E_n).$$

- The infinite version of Kolmogorov's axioms replaces the last axiom with:

The countable sequence of disjoint events $E_1, E_2, \ldots$ satisfies:

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$$

Further Properties

- We can derive many other properties from Kolmogorov’s axioms
- *For example, we have*

$P(A) \leq 1$ (Numeric bound)

$P(\Omega \setminus E) = 1 - P(E)$ (Complement)

$A \subseteq B$ implies $P(A) \leq P(B)$ (Monotonicity)

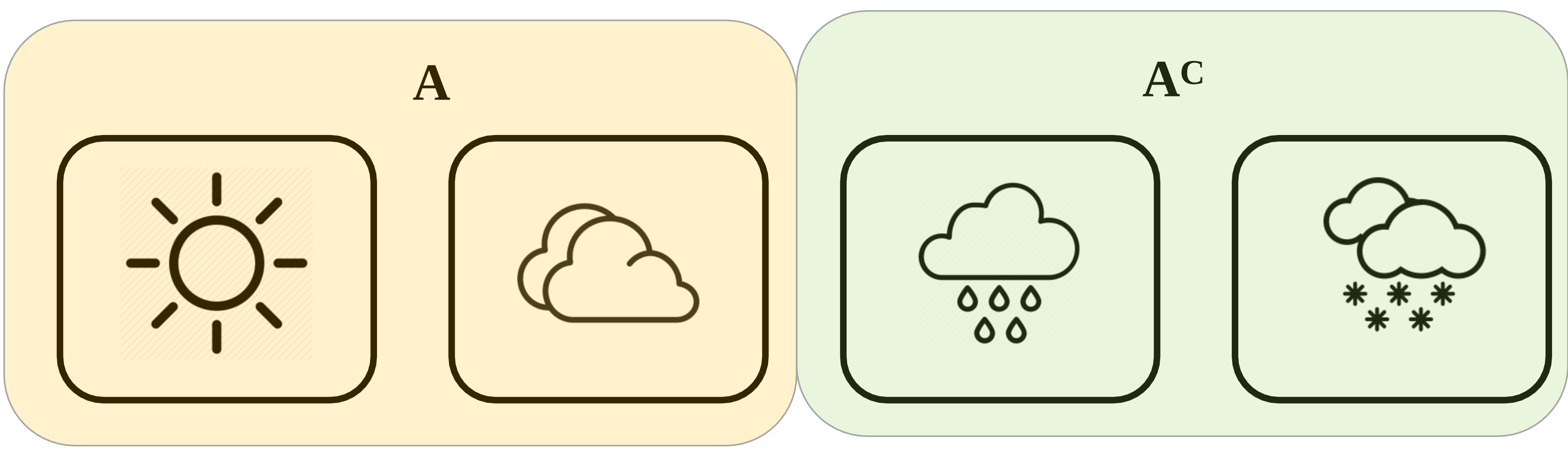
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$ (Sum Rule)

- Deriving some of these properties is good practice to get a better intuition for the axioms and for probabilities

Proof of P(A) ≤ 1

- To derive $P(A) \leq 1$, note that

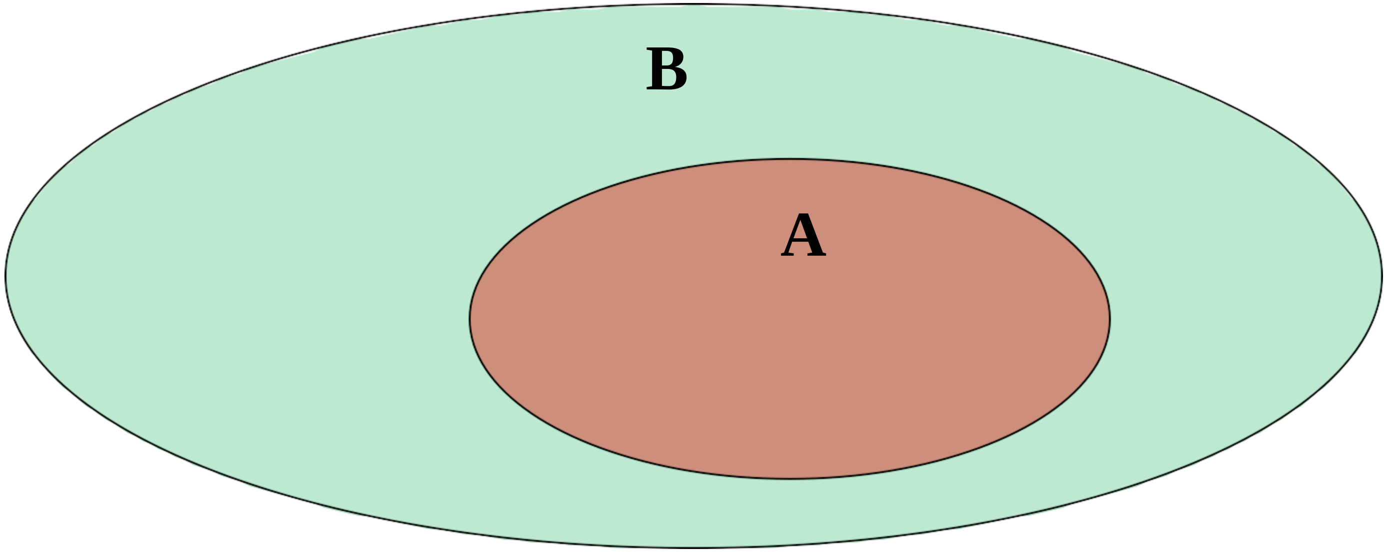
$$
\begin{aligned}
P(A) &= P(A) + 0 & \\
&\leq P(A) + P(A^C) & \text{(Non-negativity)} \\
&= P(A \cup A^C) & \text{(Additivity)} \\
&= P(\Omega) & \\
&= 1 & \text{(Normalization)}
\end{aligned}
$$



Proof of $A \subseteq B$ implies $P(A) \leq P(B)$

- To derive $A \subseteq B$ implies $P(A) \leq P(B)$, assume that $A \subseteq B$. Then

$P(B)$	$= P(A \cup (B \setminus A))$	$(A \subseteq B)$
	$= P(A) + P(B \setminus A)$	(Additivity)
	$\geq P(A)$	(Non-negativity)



Probabilistic Logic: Propositional Probabilistic Logic

Probabilistic Reasoning and
Probabilistic Knowledge Bases

Propositional Probabilistic Logic

- Similar to Fuzzy Logic, we want to extend classical logic
- We do so by regarding propositional classical interpretations as possible worlds
- Given a set of propositional atoms $A_1, \ldots, A_n$, let Ω be the set of classical interpretations $I: \{A_1, \ldots, A_n\} \rightarrow \{0,1\}$
- Furthermore, we identify formulas F with the event E_F that contains all interpretations that satisfy F
- A probability measure P can then assign probabilities to formulas via

$$P(F) = P(E_F)$$

Example

- Consider propositional atoms A_1, A_2
- We let Ω be the set of classical interpretations and consider a probability measure P (as we saw before, it suffices to define P for elementary events)

A_1	A_2	$P(A_1,A_2)$
0	0	0.2
0	1	0.3
1	0	0.15
1	1	0.35

$P(A_1) = P(\{(1,0), (1,1)\}) = P(\{(1,0)\}) + P(\{(1,1)\}) = 0.15 + 0.35 = 0.5$

$P(A_2) = P(\{(0,1), (1,1)\}) = P(\{(0,1)\} + P(\{(1,1)\}) = 0.3 + 0.35 = 0.65$

$P(A_1 \vee A_2) = 0.3 + 0.15 + 0.35 = 0.8$

Fuzzy vs. Probabilistic Logic

- Probabilistic logic maintains many logical properties
- If T is a tautological formula, we have

  $$P(T) = P(E_T) = P(\Omega) = 1$$

  - If C is a contradictory formula, we have $P(C) = P(E_C) =$

    $P(\emptyset) = 1 - P(\Omega) = 0$

  - Hence, each probability measure assigns probability 1 to tautological and probability 0 to contradictory formulas

- We do not have such a property for Fuzzy logics

(this is reasonable because $A \vee \neg A$ is not necessarily true for vague statements)

Simple Reasoning Example

- Probabilistic knowledge bases contain probabilistic formulas F: p
- A probability measure P satisfies the expression F:p iff P(F) = p

- Consider the knowledge base
- $A \land B \; : 0.3$
- $A \qquad : 0.8$

- We can derive $P(B) \leq 0.5$

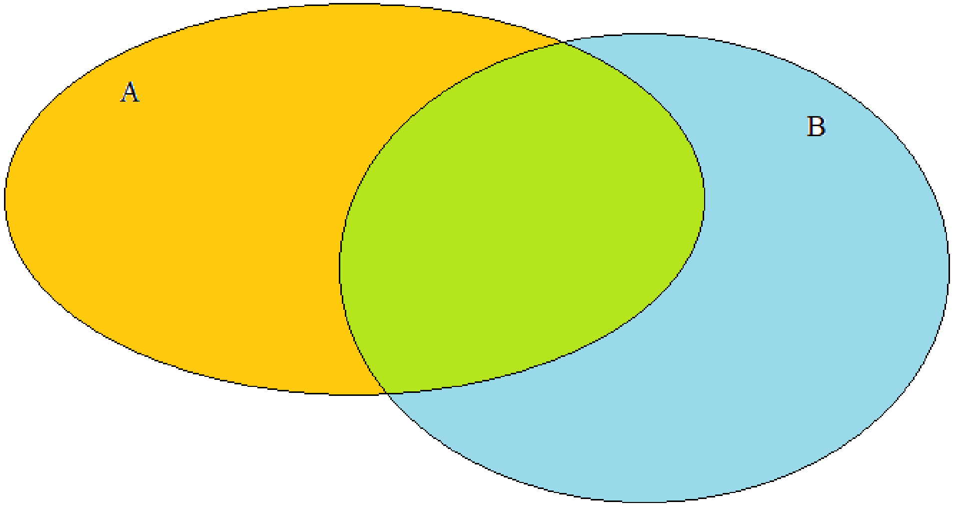
$P(B)$	$= P((A \lor \neg A) \land B)$	Tautology
	$= P((A \land B) \lor (\neg A \land B))$	Distributivity
	$= P(A \land B) + P(\neg A \land B)$	Additivity
	$= 0.3 + (1 - P(A \lor \neg B))$	Complementary Event
	$\leq 0.3 + 1 - P(A)$	$(E(A) \subseteq E(A \lor \neg B))$
	$= 0.3 + 1 - 0.8$	
	$= 0.5$	

Probabilistic Logic: Conditional Probability

Dependent Events

Conditional Probability

- We often want to express probability under particular conditions
  - Probability of disease given that patient shows certain symptoms
  - Probability of spam given certain keywords in email
  - Probability of car given certain features of image
- P(B | A) is the probability that event B is occurring given that event A has occurred
- Formally, conditional probability is defined as $P(B \mid A) = P(B \cap A) / P(A)$



Conditional Probability

- The definition can easily be transferred to probabilistic logics
  $P(G \mid F) = P(G \wedge F) / P(F)$
- Probabilistic knowledge bases can also contain conditional formulas
  $(G \mid F)[p]$
- P satisfies $(G \mid F)[p]$ if and only if $P(G \mid F) = p$

Conditional Reasoning Example

- Consider the knowledge base
  - Bird: 0.8
  - (Penguin | Bird)[0.2]
  - (Flies | Bird)[0.3]
  - (Bird | Penguin)[1]
  - (Flies | Penguin)[0]

- We can derive P(Penguin) ≥ 0.16 for all P that satisfy knowledge base

$$
\begin{aligned}
P(\text{Penguin}) &= P((\text{Bird} \lor \neg\text{Bird}) \land \text{Penguin}) && \text{(Tautology)}\\
&= P((\text{Bird} \land \text{Penguin}) \lor (\neg\text{Bird} \land \text{Penguin})) && \text{(Distributivity)}\\
&= P(\text{Bird} \land \text{Penguin}) + P(\neg\text{Bird} \land \text{Penguin}) && \text{(Additivity)}\\
&= P(\text{Penguin} \mid \text{Bird}) * P(\text{Bird}) + && \text{(Conditional)}\\
&\quad\ P(\text{Penguin} \mid \neg\text{Bird}) * P(\neg\text{Bird}) &&\\
&= 0.2 * 0.8 + P(\text{Penguin} \mid \neg\text{Bird}) * P(\neg\text{Bird}) && \text{(Non-negativity)}\\
&\geq 0.2 * 0.8 = 0.16 &&
\end{aligned}
$$

Summary

- In Fuzzy Theory, vague statements can be true to a certain degree.
- In Probability Theory, precise statements are either true or false, probabilities express our degree of belief

- Probabilistic Logic maintains more classical properties than Fuzzy Logic (e.g. Tautologies and Contradictions)

- Reasoning in Probabilistic Logic is often harder than in Fuzzy Logic

	Classical	Fuzzy	Probabilistic
Truth values	{0,1}	[0,1]	[0,1]
Propositions are	True or false	True to a certain degree	True or false
Truth values represent	True or false	Degree of "membership"	"Probability" that statement is true

Summary

- Some more features

	Probability Theory	Fuzzy Logic
Possible "values"	$0 \leq P(A) \leq 1$	Truth values are elements of the real interval [0,1]
"Complement operation"	$P(\Omega \setminus E) = 1 – P(E)$	$A^C : \mu_{A^C}(x) = 1 - \mu_A(x)$ [fuzzy set]
Involution of "negation"	Holds	*Holds for Fuzzy Logics we introduced (involutive monodial t-norm based algebras)*
"Conjunction"	$P(A \cap B) = P(A) \cdot P(B)$ [if A and B are independent]	$alg_t(x,y) = x \cdot y$ [example of a t-norm]
"Disjunction"	$P(A \cup B) = P(A) + P(B) – P(A \cap B)$	$alg_s(x,y) = x + y – (x \cdot y)$ [example of an s-norm]
"Neutral elements"	$P(\Omega) = 1$ $P(\emptyset) = 0$	1 Defined as $0 \Rightarrow 0$ 0 Defined as $\neg 1$

Bayesian Networks

Joint Distribution

Motivation

- Probabilistic Reasoning is computationally difficult
- Bayesian networks can improve performance by making independency assumptions
  - Flu does not depend on size of patient
  - Spam does not depend on font type of email
  - Rain does not depend on daylight
  - Etc.

Probability Distribution

- Bayesian networks express beliefs about a finite set of random variables $\{X_1, X_2, \ldots, X_n\}$

- Each random variable has a domain of values
  - Boolean $\{0,1\}$
  - Finite $\{a_1, a_2, \ldots, a_k\}$
  - Infinite $[0,1]$, $\mathbf{Q}$, $\mathbf{R}$

- A joint probability distribution over $\{X_1, X_2, \ldots, X_n\}$ is a function $P(X_1, X_2, \ldots, X_n)$ that maps variable assignments to $\mathbf{R}$ such that:

1. $P(x_1, x_2, \ldots, x_n) \geq 0$ for all assignments $(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$

2. $\sum_{x_1, x_2, \ldots, x_n} P(x_1, x_2, \ldots, x_n) = 1$

Inference by Enumeration

- Consider three Boolean varibles *toothache, cavity, catch*

Start with the joint distribution:

	toothache		¬ *toothache*	
	catch	¬ *catch*	*catch*	¬ *catch*
cavity	**.108**	**.012**	**.072**	**.008**
¬ *cavity*	**.016**	**.064**	**.144**	**.576**

For any proposition ϕ, sum the atomic events where it is true:

$$P(\phi) = \Sigma_{\omega:\omega \models \phi} P(\omega)$$

Inference by Enumeration

Start with the joint distribution:

	toothache		¬ *toothache*	
	catch	¬ *catch*	*catch*	¬ *catch*
cavity	**.108**	**.012**	**.072**	**.008**
¬ *cavity*	**.016**	**.064**	**.144**	**.576**

For any proposition ϕ, sum the atomic events where it is true:

$$P(\phi) = \Sigma_{\omega:\omega\models\phi} P(\omega)$$

$$P(toothache) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2$$

Inference by Enumeration

Start with the joint distribution:

	toothache		¬ *toothache*	
	catch	¬ *catch*	*catch*	¬ *catch*
cavity	**.108**	**.012**	**.072**	**.008**
¬ *cavity*	**.016**	**.064**	**.144**	**.576**

For any proposition ϕ, sum the atomic events where it is true:

$P(\phi) = \Sigma_{\omega:\omega\models\phi} P(\omega)$

$P(cavity \lor toothache) = 0.108 + 0.012 + 0.072 + 0.008 + 0.016 + 0.064 = 0.28$

Inference by Enumeration

Start with the joint distribution:

	toothache		$\neg$ *toothache*	
	catch	$\neg$ *catch*	*catch*	$\neg$ *catch*
cavity	**.108**	**.012**	**.072**	**.008**
$\neg$ *cavity*	**.016**	**.064**	**.144**	**.576**

Can also compute conditional probabilities:

$$P(\neg cavity | toothache) = \frac{P(\neg cavity \land toothache)}{P(toothache)}$$

$$= \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4$$

Bayesian Networks

Conditional Independence

Independence

- Formally, events A and B are called independent iff
  $P(A \cap B) = P(A) * P(B)$

- Analogously, formulas F and G are called independent iff
  $P(F \wedge G) = P(F) * P(G)$

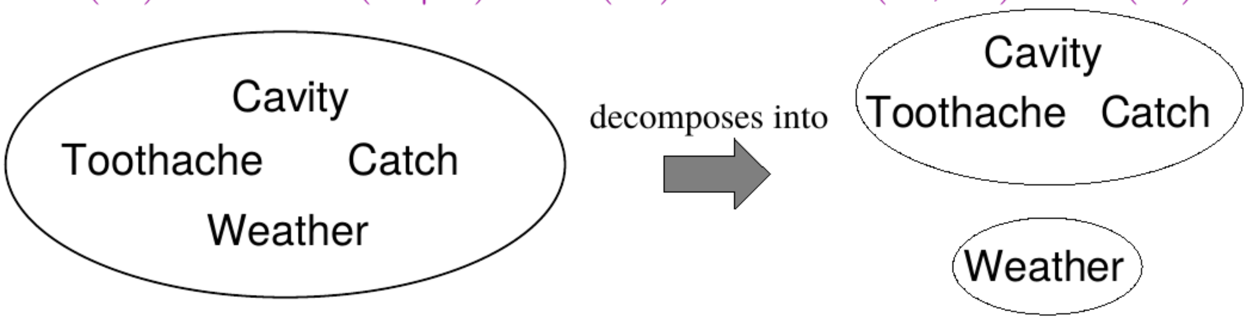
- If A is independent of B, then $P(B \mid A) = P(B)$

$$
\begin{aligned}
P(B \mid A) &= P(B \cap A) / P(A) \\
&= P(B) * P(A) / P(A) && \text{(Independence)} \\
&= P(B) * 1 \\
&= P(B)
\end{aligned}
$$

(A does not add any information about B – symmetrically $P(A \mid B) = P(A)$)

Independence

A and B are independent iff
$\mathbf{P}(A|B) = \mathbf{P}(A)$ or $\mathbf{P}(B|A) = \mathbf{P}(B)$ or $\mathbf{P}(A, B) = \mathbf{P}(A)\mathbf{P}(B)$



$$\mathbf{P}(Toothache, Catch, Cavity, Weather)$$
$$= \mathbf{P}(Toothache, Catch, Cavity)\mathbf{P}(Weather)$$

32 entries reduced to 12; for n independent biased coins, $2^n \rightarrow n$

Absolute independence powerful but rare

Dentistry is a large field with hundreds of variables, none of which are independent. What to do?

Conditional Independence

- Events A and B are called conditionally independent given C iff
  $P(A \cap B \mid C) = P(A \mid C) * P(B \mid C)$

- Analogously, formulas F and G are called independent given H iff
  $P(F \wedge G \mid H) = P(F \mid H) * P(G \mid H)$

- If A is independent of B given C, then $P(B \mid A \cap C) = P(B \mid C)$

$P(B \mid A \cap C) = P(B \cap A \cap C) / P(A \cap C)$	(Conditional Prob.)
$= P(B \cap A \cap C) / P(C)) * (P(C) / P(A \cap C)$	
$= P(A \cap B \mid C) * 1/P(A \mid C)$	(Conditional Prob.)
$= P(A \mid C) * P(B \mid C) / P(A \mid C)$	(Independence)
$= P(B \mid C)$	

(if C is known, A does not add any information about B)

Conditional Independence

If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:

(1) $P(catch|toothache, cavity) = P(catch|cavity)$

The same independence holds if I haven't got a cavity:

(2) $P(catch|toothache, \neg cavity) = P(catch|\neg cavity)$

$Catch$ is conditionally independent of $Toothache$ given $Cavity$:

$$\mathbf{P}(Catch|Toothache, Cavity) = \mathbf{P}(Catch|Cavity)$$

Equivalent statements:

$$\mathbf{P}(Toothache|Catch, Cavity) = \mathbf{P}(Toothache|Cavity)$$
$$\mathbf{P}(Toothache, Catch|Cavity) = \mathbf{P}(Toothache|Cavity)\mathbf{P}(Catch|Cavity)$$

Conditional independence can reduce complexity of calculations significantly.

Further Readings

Most topics can be found in:

- S. Russell, P. Norvig.
  Artificial Intelligence - A modern approach.
  Pearson Education. 2010. 3rd Edition.
    Sections 13.1: Introduction to acting under uncertainty
    Sections 13.2–3: Basic probability and probabilistic logic notions
    Sections 14.7.3: Fuzzy sets and fuzzy logics

- T.J. Ross.
  Fuzzy logic with engineering applications.
  John Wiley & Sons. 2009.

- D. Koller, N. Friedman.
  Probabilistic graphicals models: principles and techniques.
  MIT Press. 2009.