Logic and Probability

Lecture 2: Probability and Nonmonotonicity

Wesley Holliday & Thomas Icard
UC Berkeley & Stanford

August 12, 2014
ESSLLI, Tübingen

- End of last time: Jaynes on Probability as Extended Logic.
- Motivated by patterns of plausible reasoning (Pólya).
- Another important inference pattern: nonmonotonicity.
- Today:
  - Nonmonotonic Logic, Belief Revision, and Probabilistic Semantics
  - Graphical Models
  - Markov Logic

Part I:
Nonmonotonic Logic, Belief Revision, and Probabilistic Semantics

Example

Recall the example of the jewelry thief. We had:

$$\frac{\text{The masked man across the street is carrying a bag of jewelry}}{\text{That man is dishonest}}$$

But now suppose the police officer learns this man was at a masquerade, that he was the owners of the store, and that the broken window was due to a rock from a passing truck. In fact, he is only protecting his store. This additional information renders the pattern much less reasonable:

$$\frac{\begin{array}{c}\text{The masked man across the street is carrying a bag of jewelry}\\ \text{The man owns this store and tonight was a masquerade party}\end{array}}{\text{That man is dishonest}}$$

Nonmonotonicity of Inference

Classical logic:

$$\Gamma \vdash \varphi \quad \Longrightarrow \quad \Gamma \cup \{\psi\} \vdash \varphi \,.$$

Note that this pattern fails for conditional probability. It may well be that

$$P(\varphi|\chi) > P(\varphi|\psi \wedge \chi) \,.$$

Some people think it is possible to try to save monotonicity by saying that what was in your mind was not a general rule about birds flying but a probabilistic rule. So far these people have not worked out any detailed epistemology for this approach, i.e. exactly what probabilistic sentences should be used. Instead AI has moved to directly formalizing nonmonotonic logical reasoning.

—John McCarthy, 1990

Many systems developed for nonmonotonic reasoning:

- Circumscription (McCarthy, ...)
- Default Logic (Reiter, ...)
- Autoepistemic Logic (Moore, ...)
- Inheritance Networks (Touretzky et al., ...)

$\vdots$

Gabbay (1985) first suggested studying nonmonotonic from an abstract proof theoretic point of view, investigating consequences relations

$$\varphi \mathrel{|\!\sim} \psi$$

between individual propositional formulas φ and ψ.

All such consequence relations will allow for nonmonotonicity—there may be φ, ψ, and χ for which

$$\varphi \mathrel{|\!\sim} \psi \qquad \text{but} \qquad \varphi \wedge \chi \mathrel{|\!\not\sim} \psi \, .$$

Preferential Consequence Relations: Logic **P**

$$\frac{}{\varphi \mathrel{|\!\sim} \varphi} \quad (\textsc{Refl.})$$

$$\frac{\vdash \varphi_1 \leftrightarrow \varphi_2 \qquad \varphi_1 \mathrel{|\!\sim} \psi}{\varphi_2 \mathrel{|\!\sim} \psi} \ (\text{L-E}) \qquad \frac{\vdash \psi_1 \rightarrow \psi_2 \qquad \varphi \mathrel{|\!\sim} \psi_1}{\varphi \mathrel{|\!\sim} \psi_2} \ (\text{R-W})$$

$$\frac{\varphi \mathrel{|\!\sim} \psi_1 \qquad \varphi \mathrel{|\!\sim} \psi_2}{\varphi \mathrel{|\!\sim} \psi_1 \wedge \psi_2} \quad (\textsc{And}) \qquad \frac{\varphi_1 \mathrel{|\!\sim} \psi \qquad \varphi_2 \mathrel{|\!\sim} \psi}{\varphi_1 \vee \varphi_2 \mathrel{|\!\sim} \psi} \quad (\textsc{Or})$$

$$\frac{\varphi \mathrel{|\!\sim} \psi_1 \qquad \varphi \mathrel{|\!\sim} \psi_2}{\varphi \wedge \psi_1 \mathrel{|\!\sim} \psi_2} \quad (\textsc{Cautious Monotonicity})$$

Semantics for Logic **P**: Preferential Models

A preferential model is a triple $(W, \preceq, V)$ of a set W of 'worlds', a preorder $\preceq$ on W, and a propositional valuation $V : \text{Prop} \to \wp(W)$.

Given a model $\mathcal{M} = (W, \preceq, V)$, define:

$$[\![\varphi]\!]_{\mathcal{M}} = \{w \in W \mid w \in \hat{V}(\varphi)\}$$

$$Best_{\mathcal{M}}(A) = \{w \in A \mid \text{there is no } w' \in A, \text{ such that } w' \prec w\}$$

We can then extend our propositional language $\mathcal{L}$ to include a conditional belief operator $B^{\varphi}(\psi)$:

$$\mathcal{M} \models B^{\varphi}(\psi) \qquad \text{iff} \qquad Best_{\mathcal{M}}([\![\varphi]\!]_{\mathcal{M}}) \subseteq [\![\psi]\!]_{\mathcal{M}} \; .$$

Theorem (Kraus et al. 1990; *inter alia*)

For any consequence relation $\mid\!\sim$ satisfying the axioms of **P**, there is a model $\mathcal{M}$, such that:

$$\varphi \mid\!\sim \psi \qquad \text{iff} \qquad \mathcal{M} \models B^{\varphi}\psi \, .$$

Conversely, given any model $\mathcal{M}$, there is a **P**-consequence relation $\mid\!\sim$ satisfying the above equivalence.

- How does system **P** relate to probability?
- Recall Adams' semantics for probabilistic entailment:

$$\Gamma \vDash_a \varphi \quad \text{iff for all } \epsilon > 0, \text{ there is } \delta > 0 \text{ s.t. for all } P:$$

$$\text{if } P(\neg\gamma) < \delta \text{ for all } \gamma \in \Gamma, P(\neg\varphi) < \epsilon \; ,$$

- Similar idea here. Main theme: relate a 'default' or 'conditional' $\varphi \mathrel{|\!\sim} \psi$ to conditional probability $P(\psi|\varphi)$.

Adams Conditionals

Let Δ be a finite set of default statements $\{\alpha_1 \mid\!\sim \beta_1, \ldots, \alpha_n \mid\!\sim \beta_n\}$.

We write

$$\Delta \vDash_A \varphi \mid\!\sim \psi \, ,$$

if for all $\epsilon > 0$, there is a $\delta > 0$, such that for all $P : \mathcal{L} \to [0, 1]$:

$$\text{if } P(\beta_i|\alpha_i) > 1 - \delta \text{ for all } i \leq n, \text{ then } P(\psi|\varphi) > 1 - \epsilon.$$

Theorem (Adams 1966)

A default statement $\varphi \mid\!\sim \psi$ is derivable from a set Δ of default statements using the rules of **P**, if and only if $\Delta \vDash_A \varphi \mid\!\sim \psi$.

Example (Soundness of Cautious Monotonicity)

$$\frac{\varphi \mid\sim \psi_1 \qquad \varphi \mid\sim \psi_2}{\varphi \wedge \psi_1 \mid\sim \psi_2}$$

We need to show that for any ϵ there is a δ such that,

if $P(\psi_1|\varphi), P(\psi_2|\varphi) > 1 - \delta$, then $P(\psi_2|\varphi \wedge \psi_1) > 1 - \epsilon$.

Given ϵ, let $\delta = \frac{1}{2}\epsilon$. Then:

$$\begin{aligned}
P(\psi_2|\varphi \wedge \psi_1) &= \frac{P(\psi_1 \wedge \psi_2|\varphi)}{P(\psi_1|\varphi)} \\
&\geq P(\psi_1 \wedge \psi_2|\varphi) \\
&= P(\psi_1|\varphi) + P(\psi_2|\varphi) - P(\psi_1 \vee \psi_2|\varphi) \\
&\geq P(\psi_1|\varphi) + P(\psi_2|\varphi) - 1 \\
&> (1 - \frac{1}{2}\epsilon) + (1 - \frac{1}{2}\epsilon) - 1 \\
&= 1 - \epsilon\ .
\end{aligned}$$

Example (Soundness of CAUTIOUS MONOTONICITY)

$$\frac{\varphi \mathrel{|\!\sim} \psi_1 \qquad \varphi \mathrel{|\!\sim} \psi_2}{\varphi \wedge \psi_1 \mathrel{|\!\sim} \psi_2}$$

We need to show that for any ϵ there is a δ such that,

$$\text{if } P(\psi_1|\varphi), P(\psi_2|\varphi) > 1 - \delta, \text{ then } P(\psi_2|\varphi \wedge \psi_1) > 1 - \epsilon.$$

Given ϵ, let $\delta = \frac{1}{2}\epsilon$. Then:

$$\begin{aligned}
P(\psi_2|\varphi \wedge \psi_1) &= \frac{P(\psi_1 \wedge \psi_2|\varphi)}{P(\psi_1|\varphi)} \\
&\geq P(\psi_1 \wedge \psi_2|\varphi) \\
&= P(\psi_1|\varphi) + P(\psi_2|\varphi) - P(\psi_1 \vee \psi_2|\varphi) \\
&\geq P(\psi_1|\varphi) + P(\psi_2|\varphi) - 1 \\
&> (1 - \frac{1}{2}\epsilon) + (1 - \frac{1}{2}\epsilon) - 1 \\
&= 1 - \epsilon\ .
\end{aligned}$$

Rational Consequence Relations: Logic **R**

Some have proposed a stronger rule than CAUTIOUS MONOTONICITY:

$$\frac{\varphi \mid\sim \psi_1 \qquad \varphi \not\mid\sim \neg\psi_2}{\varphi \wedge \psi_2 \mid\sim \psi_1} \qquad \text{(RATIONAL MONOTONICITY)}$$

The system that results from adding this rule to **P** is called **R**.

Example

$$\frac{\begin{array}{c}\text{If I don't see a car, it's okay to cross the street} \\ \text{Not seeing a car is no reason to think it's not August}\end{array}}{\text{If I don't see a car and it's August, it's okay to cross the street}}$$

Semantics for Logic **R**

The models for **R** are just like those for **P**, namely triples $(W, \preceq, V)$, except that $\preceq$ is now assumed to be a total preorder:

$$\text{for all } w, w' \in W, \text{ either } w \preceq w' \text{ or } w' \preceq w \text{ (or both).}$$

Theorem (Lehmann & Magidor 1992; *inter alia*)

R consequence relations correspond to conditional belief in total models.

Alchourrón-Gärdenfors-Makinson (AGM) Belief Revision

1. $\mathcal{K} * \varphi$ is consistent and logically closed ;
2. $\varphi \in \mathcal{K} * \varphi$;
3. $\mathcal{K} * \varphi \subseteq Cl(\mathcal{K} \cup \{\varphi\})$;
4. If $\neg\varphi \notin \mathcal{K}$, then $Cl(\mathcal{K} \cup \{\varphi\}) \subseteq \mathcal{K} * \varphi$;
5. $\mathcal{K} * \varphi = Cl(\{\bot\})$, if and only if $\vdash \neg\varphi$;
6. If $\vdash \varphi_1 \leftrightarrow \varphi_2$, then $\mathcal{K} * \varphi_1 = \mathcal{K} * \varphi_2$;
7. $\mathcal{K} * (\varphi \wedge \psi) \subseteq Cl(\mathcal{K} * \varphi \cup \{\psi\})$;
8. If $\neg\psi \notin \mathcal{K} * \varphi$, then $Cl(\mathcal{K} * \varphi \cup \{\psi\}) \subseteq \mathcal{K} * (\varphi \wedge \psi)$.

Theorem (Folklore)

1. Suppose we are given a knowledge base $\mathcal{K}$ and a revision operation $*$. Define a consequence relation $\mid\!\sim$ so that

$$\varphi \mid\!\sim \psi \qquad \text{iff} \qquad \psi \in \mathcal{K} * \varphi \,.$$

   Then $\mid\!\sim$ satisfies all the rules of **R**.

2. Suppose $\mid\!\sim$ is a consequence relation satisfying **R**, and such that $\varphi \mid\!\sim \bot$ only if $\vdash \neg\varphi$. If we define

$$\mathcal{K} := \{\psi \mid \top \mid\!\sim \psi\} \qquad \text{and} \qquad \mathcal{K} * \varphi := \{\psi : \varphi \mid\!\sim \psi\} \,,$$

   then this gives us an AGM belief revision operation.

- What is the relation between logic **R** (and hence AGM belief revision) and probability?
- It is easy to see that Rational Monotonicity is not sound with respect to Adams' semantics.
- But we can obtain an adequate probabilistic semantics by moving to so called Popper functions, taking conditional probability as primitive and relating defaults to statements of conditional probability 1.

Popper Functions

1. For some φ, ψ, we have $P(\varphi|\psi) \neq 1$
2. $P(\varphi|\varphi) = 1$
3. $P(\neg\varphi|\psi) = 1 - P(\varphi|\psi)$
4. $P(\varphi \wedge \psi|\chi) = P(\varphi|\psi \wedge \chi)P(\psi|\chi)$

(Cf. Jaynes; also Leitgeb 2012 on semantics of counterfactuals.)

Theorem (Harper 1975; Hawthorne 1998)

1. Given Popper function P, define a consequence relation $\varphi \mid\sim \psi$, which holds just in case $P(\psi \mid \varphi) = 1$. Then $\mid\sim$ is an **R** relation.

2. For any **R** relation $\mid\sim$, there is a Popper function P such that:

   - $P(\psi \mid \varphi) = 1$, iff $\varphi \mid\sim \psi$;
   - $P(\psi \mid \varphi) = 0$, iff $\varphi \mid\sim \neg\psi$;
   - $0 < P(\psi \mid \varphi) < 1$, iff $\varphi \not\mid\sim \psi$ and $\varphi \not\mid\sim \neg\psi$.

- The probabilistic semantics of these logics all involve what one might call 'extreme probabilities'.
- If we use 'intermediate probabilities' more straightforwardly, and e.g., define $\varphi \mid\!\sim \psi$ to hold when $P(\psi|\varphi) > \theta$, for some fixed threshold θ, the resulting logic merely validates $\textsc{Refl}$, L-E, and R-W.
- Pearl (1988) argued that both styles of reasoning may be appropriate in different circumstances. For instance, the $\textsc{Cautious Monotonicity}$ rule is analogous to a commonsense—but strictly speaking, invalid—pattern for natural language *Most*:

$$\frac{\text{Most students will get an `A'} \qquad \text{Most students are male}}{\text{Most male students will get an `A'}}$$

- Nonetheless, for many other intuitive reasoning patterns—including some we saw in Lecture 1—we seem to need genuine *degrees*.
- In the late 80s / early 90s, there was a sense that a theory of defeasible reasoning needs a theory of causality.
- The *locus classics* for probabilistic graphical models with a causal interpretation is Pearl (1988).

Part II:
Graphical Models

More Patterns of Plausible Reasoning

1. Nonmonotonicity

2. Explaining Away: If $X \Rightarrow Z$ and $Y \Rightarrow Z$ independently, and Z is true, then finding out Y makes X less credible. (Recall jewel thief.)

3. Screening Off: If both $X \Rightarrow Y$ and $X \Rightarrow Z$, then *a priori*, learning about Y can tell you something about Z, and *vice versa*. However, after learning that X is true, Y and Z become independent.

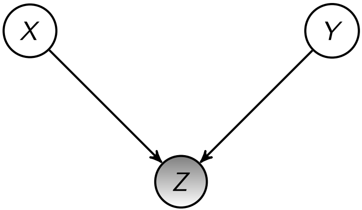
Several others. See Pearl (1988).

We take it for granted that probability calculus is unique in the way it handles context-dependent information and that no competing calculus exists that closely covers so many aspects of plausible reasoning.

—Judea Pearl, 1988

(Though see Darwiche & Ginsburg 1992; Halpern 2003.)

Explaining Away (and Nonmonotonicity)

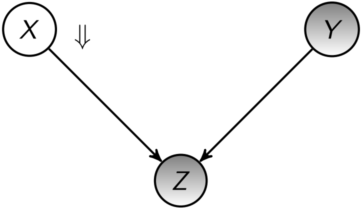


X: The man across the street is stealing jewelry.

Y: The man across the street owns the store.

Z: The man across the street is running with a bag of jewelry.

Explaining Away (and Nonmonotonicity)

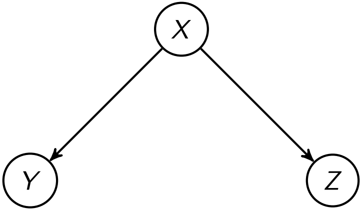


X: The man across the street is stealing jewelry.

Y: The man across the street owns the store.

Z: The man across the street is running with a bag of jewelry.

Screening Off

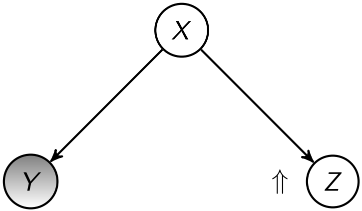


X: Hannes Leitgeb is giving a talk tonight

Y: Konrad is here today

Z: The lecture tonight will be interesting

Screening Off

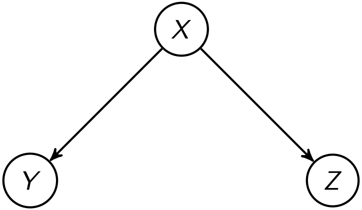


X: Hannes Leitgeb is giving a talk tonight

Y: Konrad is here today

Z: The lecture tonight will be interesting

Screening Off

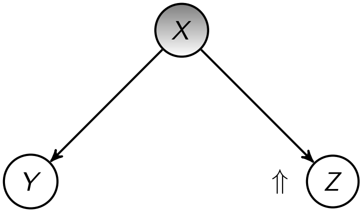


X: Hannes Leitgeb is giving a talk tonight

Y: Konrad is here today

Z: The lecture tonight will be interesting

Screening Off

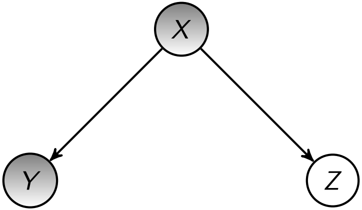


X: Hannes Leitgeb is giving a talk tonight

Y: Konrad is here today

Z: The lecture tonight will be interesting

Screening Off



X: Hannes Leitgeb is giving a talk tonight

Y: Konrad is here today

Z: The lecture tonight will be interesting

Bayesian Networks

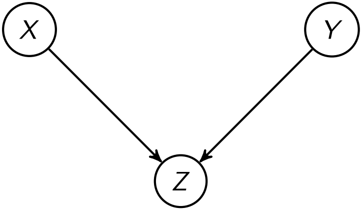
Suppose we have a number of random variables $\mathcal{X} = X_1, \ldots, X_n$, with possible values $Val(X_1), \ldots, Val(X_n)$.

A Bayesian network over $\mathcal{X}$ is a DAG (V, E), in which the nodes V correspond to the variables $\mathcal{X}$, and the distribution on $\mathcal{X}$ 'factorizes' as:

$$P(X_1, \ldots, X_n) \quad = \quad \prod_{i=1}^{n} P(X_i | Pa_{X_i})$$

where Pa_{X_i} denotes the set of parents of X_i in the graph.

$P(X) = 0.1$



$P(Y) = 0.2$

$$P(Z|X,Y) = 0.9$$
$$P(Z|X,\overline{Y}) = 0.8$$
$$P(Z|\overline{X},Y) = 0.6$$
$$P(Z|\overline{X},\overline{Y}) = 0.2$$

Typical computational tasks:

1. Determining independence: d-separation.
2. Finding the right DAG.
3. Estimating the parameters of the graph.
4. Computing marginal probabilities.
5. Computing conditional probabilities.

For tasks 4 and 5 there are many approaches. One of the most interesting, and arguably most powerful, uses logic in a central way.

Weighted Model Counting

Given a Bayes net over random variables $X_1, \ldots, X_n$, define a new propositional logical language with a new set $\textsc{Atom}$ of atoms:

- A propositional atom B_v for each value $v \in Val(X_i)$;
- A propositional atom $C_{v,\vec{u}}$ for each value $v \in Val(X_i)$ and vector of values $\vec{u} \in Val(Pa_{X_i})$.

Weighted Model Counting

Define the weight of an atom $W(A)$ as follows

$$W(A) = \begin{cases} P(X_i = v \mid Pa_{X_i} = \vec{u}) & \text{if } A = C_{v,\vec{u}} \\ 1 & \text{otherwise} \end{cases}$$

Define the weight of an atomic valuation $\nu : \textsc{Atom} \to \{0, 1\}$ by:

$$W(\nu) = \prod_{A:\nu(A)=1} W(A) \; .$$

Finally, define the weight of a set of formulas Γ to be:

$$W(\Gamma) = \sum_{\nu:\nu(\Gamma)=1} W(\nu) \; .$$

Weighted Model Counting

We finally build a particular theory Γ as follows:

1. For each $i \leq n$, where $Val(X_i) = \{v_1, \ldots, v_k\}$, we add to Γ

$$B_{v_1} \oplus \cdots \oplus B_{v_k} \ ,$$

   where $\oplus$ is shorthand for 'exclusive or'.

2. For each $v \in Val(X_i)$ and $\vec{u} \in Val(Pa_{X_i})$, we add

$$C_{v,\vec{u}} \leftrightarrow (B_v \wedge B_{u_1} \wedge \cdots \wedge B_{u_m}) \ .$$

Then, in general we have

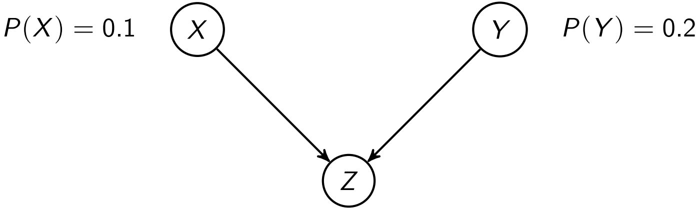
$$W(\Gamma) = 1, \text{ while, for example, } W(\Gamma \cup \{B_v\}) = P(X_i = v) \ .$$

Weighted Model Counting

- To determine $P(X_i|\vec{Y})$, we can simply ask for the weight ratio:

$$\frac{W(\Gamma \cup \{B_v, B_{u_1}, \ldots, B_{u_m}\})}{W(\Gamma \cup \{B_{u_1}, \ldots, B_{u_m}\})} \, .$$

- Boolean model counting is well studied, a canonical example of a #P-complete problem. Inference in graphical models is also #P.
- Weighted model counting is a powerful inference technique for inference in Bayes nets. See, e.g., Chavira & Darwiche (2008).



$$P(Z|X, Y) = 0.9$$
$$P(Z|X, \overline{Y}) = 0.8$$
$$P(Z|\overline{X}, Y) = 0.6$$
$$P(Z|\overline{X}, \overline{Y}) = 0.2$$

Variables: $B_{X0}, B_{X1}, B_{Y0}, B_{Y1}, B_{Z0}, B_{Z1}, C_{X0}, C_{X1}, C_{Y0}, C_{Y1}$
$C_{Z0,X0,Y0}, C_{Z1,X0,Y0}, C_{Z0,X1,Y0}, C_{Z0,X0,Y1}$
$C_{Z1,X1,Y0}, C_{Z1,X0,Y1}, C_{Z0,X1,Y1}, C_{Z1,X1,Y1}$

B formulas: weight 1; C formulas: according to conditional probabilities.

Variables: $B_{X0}, B_{X1}, B_{Y0}, B_{Y1}, B_{Z0}, B_{Z1}, C_{X0}, C_{X1}, C_{Y0}, C_{Y1}$
$C_{Z0,X0,Y0}, C_{Z1,X0,Y0}, C_{Z0,X1,Y0}, C_{Z0,X0,Y1}$
$C_{Z1,X1,Y0}, C_{Z1,X0,Y1}, C_{Z0,X1,Y1}, C_{Z1,X1,Y1}$

B formulas: weight 1; C formulas: according to conditional probabilities.

$$B_{X0} \oplus B_{X1} \qquad B_{Y0} \oplus B_{Y1} \qquad B_{Z0} \oplus B_{Z1}$$

$$C_{Z0,X0,Y0} \leftrightarrow (B_{Z0} \wedge B_{X0} \wedge B_{Y0}) \qquad C_{Z1,X1,Y0} \leftrightarrow (B_{Z1} \wedge B_{X1} \wedge B_{Y0})$$

$$\vdots$$

Then, e.g.,

$$P(X, Y, \overline{Z}) = W(\Gamma \cup \{B_{X1}, B_{Y1}, B_{Z0}\}) = 0.1 * 0.2 * 0.1 = 0.002\ .$$

- Bayes nets limit what kinds of independence structure among variables can be represented: no cycles, asymmetric.

- A more general class of models can be defined using potential functions, i.e., functions $\phi_i : Val(\mathcal{X}') \rightarrow \mathbb{R}^+$ from sets of values for some subset $\mathcal{X}' \subseteq \mathcal{X}$ to real-valued weights.

- A set of potential functions $\phi_1, \ldots, \phi_t$ determines a probability distribution over $\mathcal{X}$:

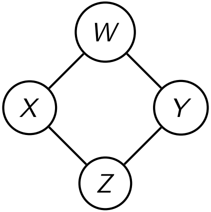
$$P(X_1, \ldots, X_n) \quad \propto \quad \prod_{i=1}^{t} \phi(\mathcal{X}') \, .$$

- Bayes nets are a special case, with a potential function for each variable node together with its parents.

- But so are, e.g., log-linear models:

$$P(X_1, \ldots, X_n) \quad \propto \quad exp\Big(\sum_i w_i f_i(\mathcal{X}')\Big) \, .$$

Example (Markov Network)



$$\phi_1 : Val(\{W, X\}) \to \mathbb{R}$$
$$\phi_2 : Val(\{W, Y\}) \to \mathbb{R}$$
$$\phi_3 : Val(\{X, Z\}) \to \mathbb{R}$$
$$\phi_4 : Val(\{Y, Z\}) \to \mathbb{R}$$

Part III:
Markov Logic

- Basic idea (see, e.g., Lowd & Domingos 2009): use (first-order) logic to build (potentially quite large) graphical models.

- Serve as an interface layer between users working in concrete domains, and engineers working on general inference techniques.

Begin with a pure first-order language without constants, and a finite set of weighted formulas:

$$(\varphi_1, w_1), \ldots, (\varphi_n, w_n)$$

Given a set of constants $\mathcal{C}$, we build a Markov network as follows:

- For each grounding of each atomic formula from $\varphi_1, \ldots, \varphi_n$, we include a node in the network, i.e., a new binary random variable.
- For each grounding of each φ_i, we include a feature that receives value 1 if this grounding is true, 0 otherwise. The weight of the feature is w_i.
- The probability of a given state, i.e., valuation of atomic formulas is:

$$P(v) \;\; \propto \;\; exp\Big(\sum_{i \leq n} w_i n_i(v)\Big) \, ,$$

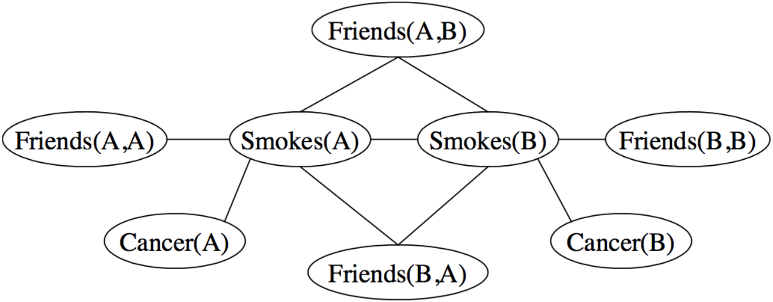
where $n_i(v)$ is the number of true groundings of φ_i under v.

Example

Suppose we have predicates $\mathit{Friend}(x, y)$, $\mathit{Smoke}(x)$, and $\mathit{Cancer}(x)$:

- $\forall x \forall y \forall z\ \mathit{Friend}(x, y) \land \mathit{Friend}(y, z) \supset \mathit{Friend}(x, z)$ 0.7 ;
- $\forall x \big(\neg \exists y\ \mathit{Friend}(x, y)\big) \supset \mathit{Smoke}(x)$ 2.3 ;
- $\forall x\ \mathit{Smoke}(x) \supset \mathit{Cancer}(x)$ 1.5 ;
- $\forall x \forall y\ \mathit{Friend}(x, y) \supset \big(\mathit{Smoke}(x) \equiv \mathit{Smoke}(y)\big)$ 1.1 ;

Given constants A and B, we generate the following Markov network:



We can then make queries as in general graphical models, e.g.:

$$P\Big(Cancer(B) \mid Smokes(A)\Big)$$

- General inference techniques are available, but one can also combine probabilistic inference techniques with logical inference techniques, e.g., resolution-based methods, to devise lifted inference algorithms.
- Many applications: named entity recognition, information extraction, textual entailment, etc. See especially work by P. Domingos & co.
- Note: The role of first-order logic is in defining Markov networks, given a set of constants. We do not, e.g., obtain probabilities over first-order expressions as such. That is the topic for tomorrow.

Summary and Preview

- Popular nonmonotonic logics capture what many view as intuitive reasoning patterns.
- From the perspective of probability, these can be seen as reasoning with 'extreme probabilities', witness Adams' semantics and Popper function semantics.
- For other types of reasoning graphical models have become a useful tool, indeed ubiquitous in AI, machine learning, NLP, etc.
- Such models nonetheless can be seen as a generalization of logic, viz. weighted model counting.
- Tomorrow: probabilistic programs (stochastic λ-calculus), and first-order logics of probability.