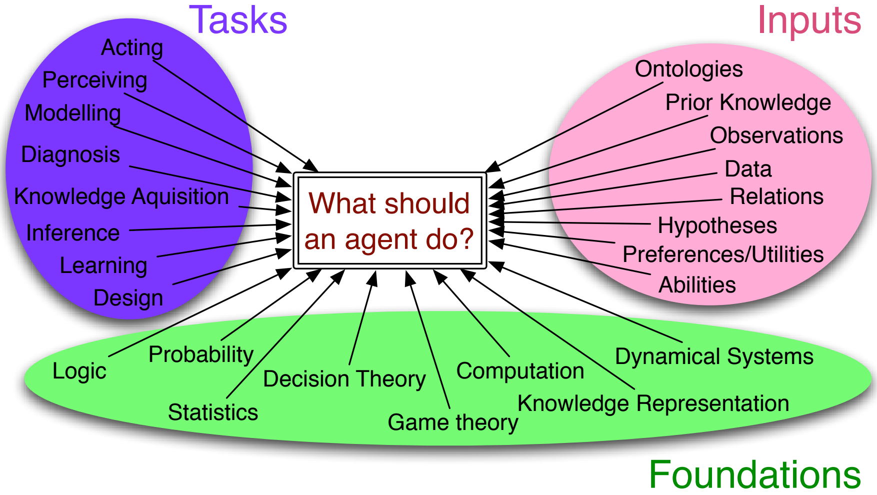
Logic, Probability and Computation: Statistical Relational AI and Beyond

David Poole

Department of Computer Science,
University of British Columbia

October 2014

AI: computational agents that act intelligently



Outline

1. Logic and Probability
   - Relational Probabilistic Models
   - Probabilistic Logic Programs
2. Lifted Inference
3. Undirected models, Directed models, and Weighted Formulae
4. Existence and Identity Uncertainty

Why Logic?

Logic provides a **semantics** linking

- the symbols in our language
- the (real or imaginary) world we are trying to characterise

Suppose K represents our knowledge of the world

- If
  $$K \models g$$
  then g must be true of the world.
- If
  $$K \not\models g$$
  there is a model of K in which g is false.

Thus logical consequence seems like the correct notion for prediction.

First-order Predicate Calculus

The world (we want to represent) is made up of individuals (things) and relationships between things.

Classical (first order) logic lets us represent:

- individuals in the world
- relations amongst those individuals
- conjunctions, disjunctions, negations of relations
- quantification over individuals

Why Probability?

- There is lots of uncertainty about the world, but agents still need to act.
- Predictions are needed to decide what to do:
  - definitive predictions: you will be run over tomorrow
  - point probabilities: probability you will be run over tomorrow is 0.002
  - probability ranges: you will be run over with probability in range [0.001,0.34]
- Acting is gambling: agents who don't use probabilities will lose to those who do — Dutch books.
- Probabilities can be learned from data.
  Bayes' rule specifies how to combine data and prior knowledge.

Bayes' Rule

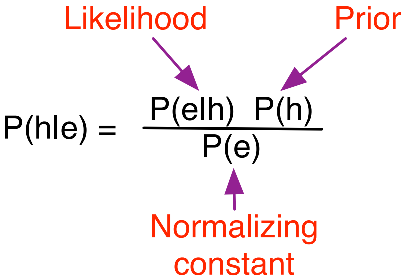
Likelihood Prior

$$P(h|e) = \frac{P(e|h)\ P(h)}{P(e)}$$

Normalizing constant

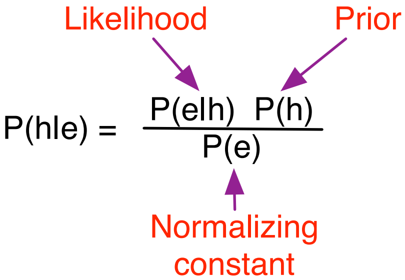
What if e is an electronic health record?

Bayes' Rule



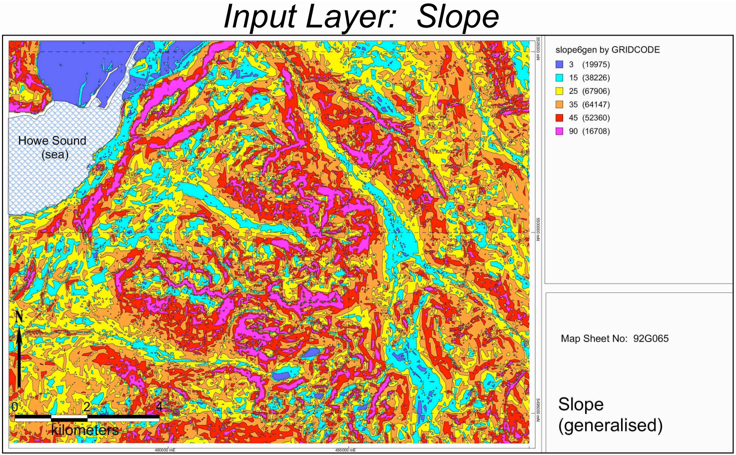
What if e is an electronic health record?
What if e is all the electronic health records?

Bayes' Rule

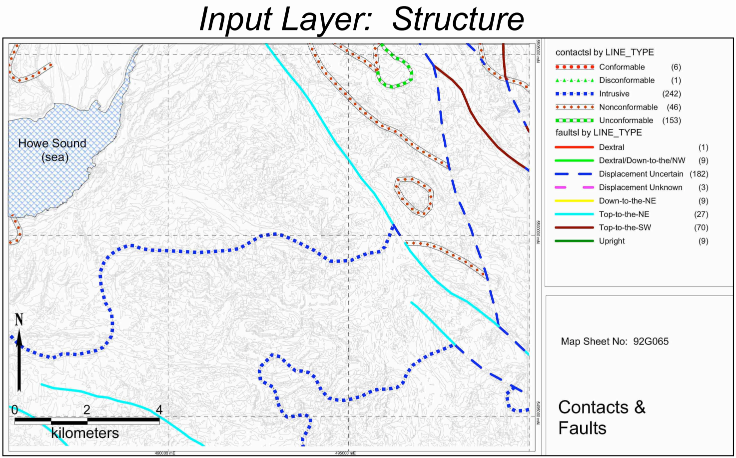


What if e is an electronic health record?
What if e is all the electronic health records?
What if e is a description of everything known about the geology of Earth?

Example Observation, Geology



Example Observation, Geology



Outline

1. Logic and Probability
   - Relational Probabilistic Models
   - Probabilistic Logic Programs
2. Lifted Inference
3. Undirected models, Directed models, and Weighted Formulae
4. Existence and Identity Uncertainty

Relational Learning

- Often the values of properties are not meaningful values but names of individuals.
- It is the properties of these individuals and their relationship to other individuals that needs to be learned.
- Relational learning has been studied under the umbrella of "Inductive Logic Programming" as the representations are often logic programs.

Example: trading agent

What does Joe like?

Individual	Property	Value
joe	*likes*	*resort_14*
joe	*dislikes*	*resort_35*
...	...	...
resort_14	*type*	*resort*
resort_14	*near*	*beach_18*
beach_18	*type*	*beach*
beach_18	*covered_in*	*ws*
ws	*type*	*sand*
ws	*color*	*white*
...	...	...

Example: trading agent

Possible theory that could be learned:

$$
\begin{aligned}
& prop(joe, likes, R) \leftarrow \\
& \quad prop(R, type, resort) \wedge \\
& \quad prop(R, near, B) \wedge \\
& \quad prop(B, type, beach) \wedge \\
& \quad prop(B, covered\_in, S) \wedge \\
& \quad prop(S, type, sand).
\end{aligned}
$$

Joe likes resorts that are near sandy beaches.

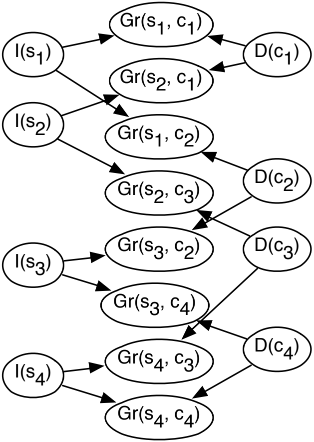
- But we want probabilistic predictions.

Example: Predicting Relations

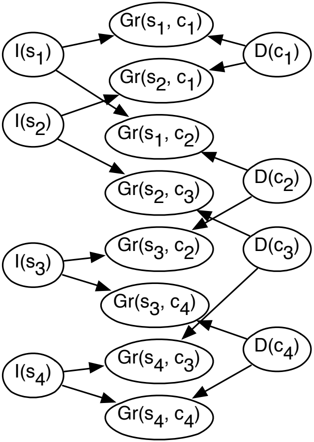
Student	*Course*	*Grade*
s_1	c_1	A
s_2	c_1	C
s_1	c_2	B
s_2	c_3	B
s_3	c_2	B
s_4	c_3	B
s_3	c_4	?
s_4	c_4	?

- Students s_3 and s_4 have the same averages, on courses with the same averages.
- Which student would you expect to better?

From Relations to Belief Networks



From Relations to Belief Networks



$I(S)$	$D(C)$	$Gr(S, C)$		
		A	B	C
true	*true*	0.5	0.4	0.1
true	*false*	0.9	0.09	0.01
false	*true*	0.01	0.09	0.9
false	*false*	0.1	0.4	0.5

$$P(I(S)) = 0.5$$
$$P(D(C)) = 0.5$$

"parameter sharing"

Example: Predicting Relations

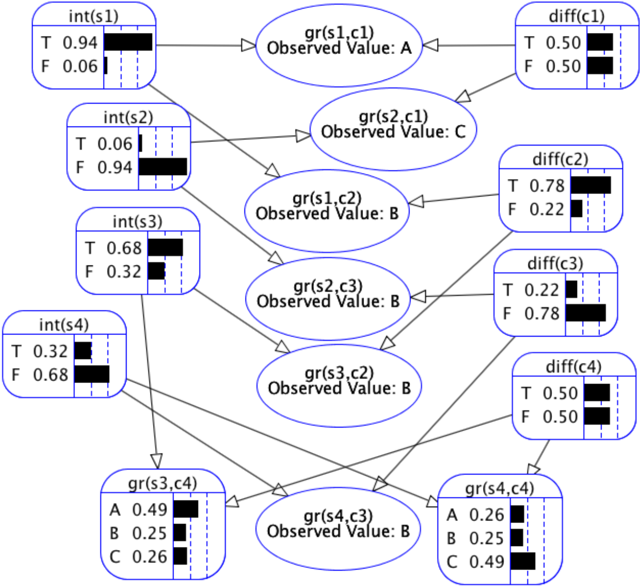
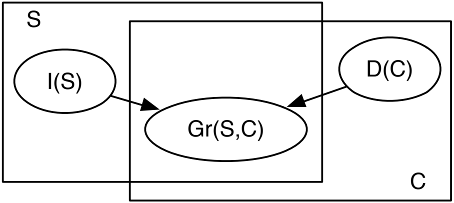
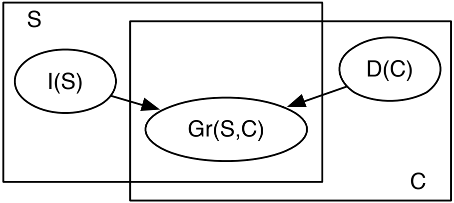


Plate Notation



- S, C logical variable representing students, courses
- the set of individuals of a type is called a population
- $I(S)$, $Gr(S,C)$, $D(C)$ are parametrized random variables

Plate Notation

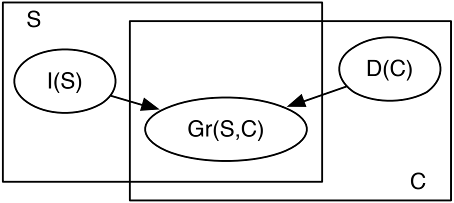


- S, C logical variable representing students, courses
- the set of individuals of a type is called a population
- $I(S)$, $Gr(S,C)$, $D(C)$ are parametrized random variables

Grounding:

- for every student s, there is a random variable $I(s)$
- for every course c, there is a random variable $D(c)$
- for every s, c pair there is a random variable $Gr(s,c)$
- all instances share the same structure and parameters

Plate Notation

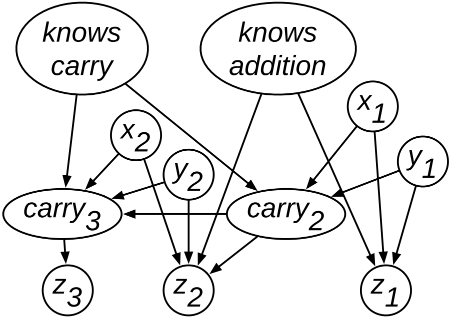


- If there were 1000 students and 100 courses:
  Grounding contains
  - 1000 $I(s)$ variables
  - 100 $D(C)$ variables
  - 100000 $Gr(s,c)$ variables

  total: 101100 variables
- Numbers to be specified to define the probabilities:
  1 for $I(s)$, 1 for $D(C)$, 8 for $Gr(S,C)$ = 10 parameters.

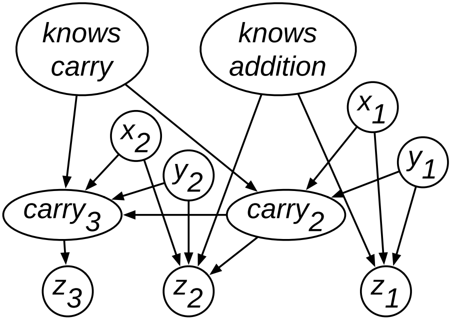
Bayesian Networks

$$\begin{array}{rrrr} & & x_2 & x_1 \\ + & & y_2 & y_1 \\ \hline & z_3 & z_2 & z_1 \end{array}$$



Bayesian Networks

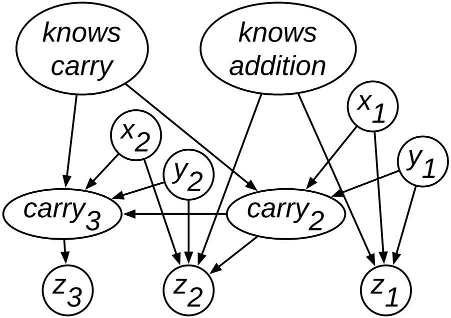
$$\begin{array}{cccc} & & x_2 & x_1 \\ + & & y_2 & y_1 \\ \hline & z_3 & z_2 & z_1 \end{array}$$



What if there were multiple digits

Bayesian Networks

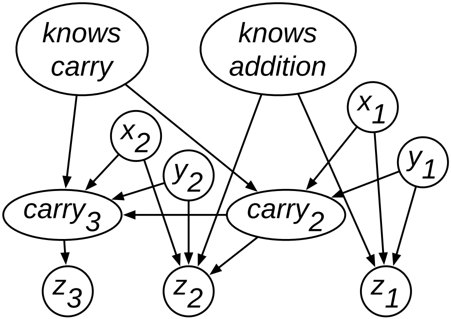
$$\begin{array}{cccc} & & x_2 & x_1 \\ + & & y_2 & y_1 \\ \hline & z_3 & z_2 & z_1 \end{array}$$



What if there were multiple digits, problems

Bayesian Networks

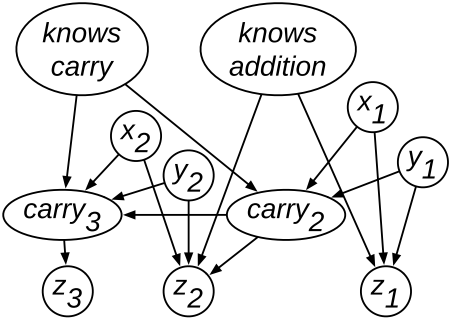
$$
\begin{array}{rrrr}
 & & x_2 & x_1 \\
+ & & y_2 & y_1 \\
\hline
 & z_3 & z_2 & z_1
\end{array}
$$



What if there were multiple digits, problems, students

Bayesian Networks

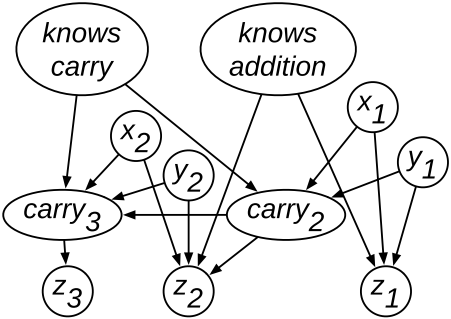
		x_2	x_1
$+$		y_2	y_1
	z_3	z_2	z_1



What if there were multiple digits, problems, students, times?

Bayesian Networks

$$\begin{array}{cccc} & & x_2 & x_1 \\ + & & y_2 & y_1 \\ \hline & z_3 & z_2 & z_1 \end{array}$$

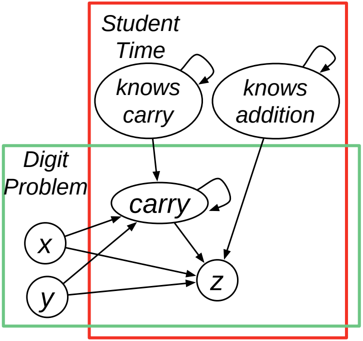


What if there were multiple digits, problems, students, times?

How can we build a model before we know the individuals?

Multi-digit addition with parametrized BNs / plates

$$\begin{array}{rcccc} & x_{j_x} & \cdots & x_2 & x_1 \\ + & y_{j_z} & \cdots & y_2 & y_1 \\ \hline & z_{j_z} & \cdots & z_2 & z_1 \end{array}$$



Random Variables: $x(D, P)$, $y(D, P)$, $knowsCarry(S, T)$, $knowsAddition(S, T)$, $carry(D, P, S, T)$, $z(D, P, S, T)$
for each: digit D, problem P, student S, time T

☛ parametrized random variables

Relational Probabilistic Models

Often we want random variables for combinations of individual in populations

- build a probabilistic model before knowing the individuals
- learn the model for one set of individuals
- apply the model to new individuals
- allow complex relationships between individuals

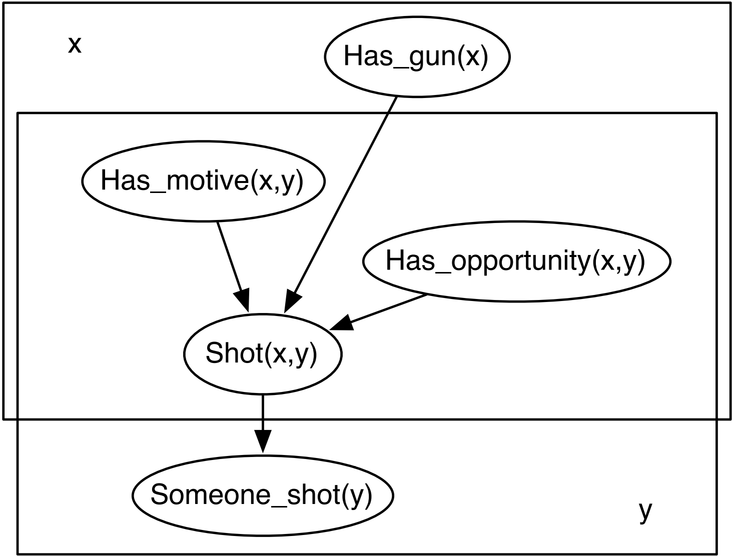
Exchangeability

- Before we know anything about individuals, they are indistinguishable, and so should be treated identically.

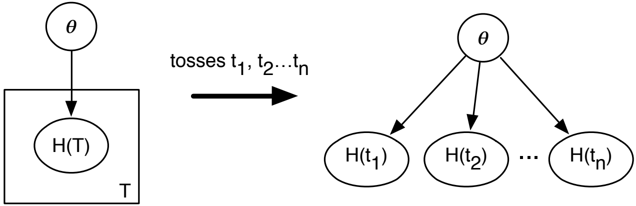
Representing Conditional Probabilities

- $P(grade(S, C) \mid intelligent(S), difficult(C))$ — parameter sharing — individuals share probability parameters.
- $P(happy(X) \mid friend(X, Y), mean(Y))$ — needs aggregation — $happy(a)$ depends on an unbounded number of parents.
- There can be more structure about the individuals
  - the carry of one digit depends on carry of the previous digit
  - probability that two authors collaborate depends on whether they have a paper authored together

Example: Aggregation



Example Plate Notation for Learning Parameters



- T is a logical variable representing tosses of a thumb tack
- $H(t)$ is a Boolean variable that is true if toss t is heads.
- θ is a random variable representing the probability of heads.
- Range of θ is $\{0.0, 0.01, 0.02, \ldots, 0.99, 1.0\}$ or interval $[0, 1]$.
- $P(H(t_i){=}true \mid \theta{=}p) = p$
- $H(t_i)$ is independent of $H(t_j)$ (for $i \neq j$) given θ: i.i.d. or independent and identically distributed.

Outline

1. Logic and Probability
   - Relational Probabilistic Models
   - Probabilistic Logic Programs
2. Lifted Inference
3. Undirected models, Directed models, and Weighted Formulae
4. Existence and Identity Uncertainty

Independent Choice Logic (ICL)

- A language for relational probabilistic models.
- Idea: combine logic and probability, where all uncertainty in handled in terms of Bayesian decision theory, and logic specifies consequences of choices.
- An ICL theory consists of a choice space with probabilities over choices and a logic program that gives consequences of choices.
- History: parametrized Bayesian networks, abduction and default reasoning $\longrightarrow$ probabilistic Horn abduction (IJCAI-91); richer language (negation as failure + choices by other agents $\longrightarrow$ independent choice logic (AIJ 1997).

Independent Choice Logic

- An atomic hypothesis is an atomic formula.
  An alternative is a set of atomic hypotheses.
  $\mathcal{C}$, the choice space is a set of disjoint alternatives.
- $\mathcal{F}$, the facts is an acyclic logic program that gives consequences of choices (can contain negation as failure).
  No atomic hypothesis is the head of a rule.
- P_0 a probability distribution over alternatives:

$$\forall A \in \mathcal{C} \sum_{a \in A} P_0(a) = 1.$$

Meaningless Example

$$\mathcal{C} = \{\{c_1, c_2, c_3\}, \{b_1, b_2\}\}$$

$$\begin{aligned}\mathcal{F} = \{\ & f \leftarrow c_1 \land b_1, && f \leftarrow c_3 \land b_2, \\ & d \leftarrow c_1, && d \leftarrow\ \sim c_2 \land b_1, \\ & e \leftarrow f, && e \leftarrow\ \sim d\}\end{aligned}$$

$$\begin{array}{lll} P_0(c_1) = 0.5 & P_0(c_2) = 0.3 & P_0(c_3) = 0.2 \\ P_0(b_1) = 0.9 & P_0(b_2) = 0.1 & \end{array}$$

Semantics of ICL

- There is a possible world for each selection of one element from each alternative.
- The logic program together with the selected atoms specifies what is true in each possible world.
- The elements of different alternatives are independent.

Meaningless Example: Semantics

$$\mathcal{F} = \{\ f \leftarrow c_1 \wedge b_1,\quad f \leftarrow c_3 \wedge b_2,$$
$$d \leftarrow c_1,\quad d \leftarrow \sim c_2 \wedge b_1,$$
$$e \leftarrow f,\quad e \leftarrow \sim d\}$$

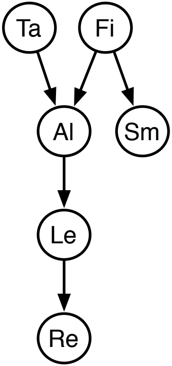
$P_0(c_1) = 0.5$ $P_0(c_2) = 0.3$ $P_0(c_3) = 0.2$
$P_0(b_1) = 0.9$ $P_0(b_2) = 0.1$

		selection		logic program			
w_1	$\models$	c_1	b_1	f	d	e	$P(w_1) = 0.45$
w_2	$\models$	c_2	b_1	$\sim f$	$\sim d$	e	$P(w_2) = 0.27$
w_3	$\models$	c_3	b_1	$\sim f$	d	$\sim e$	$P(w_3) = 0.18$
w_4	$\models$	c_1	b_2	$\sim f$	d	$\sim e$	$P(w_4) = 0.05$
w_5	$\models$	c_2	b_2	$\sim f$	$\sim d$	e	$P(w_5) = 0.03$
w_6	$\models$	c_3	b_2	f	$\sim d$	e	$P(w_6) = 0.02$

$P(e) = 0.45 + 0.27 + 0.03 + 0.02 = 0.77$

Belief Networks, Decision trees and ICL rules

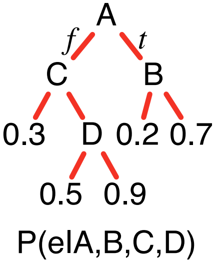
- There is a local mapping from belief networks into ICL.



prob ta : 0.02.
prob $fire$: 0.01.
$alarm \leftarrow ta \wedge fire \wedge atf$.
$alarm \leftarrow \sim ta \wedge fire \wedge antf$.
$alarm \leftarrow ta \wedge \sim fire \wedge atnf$.
$alarm \leftarrow \sim ta \wedge \sim fire \wedge antnf$.
prob atf : 0.5.
prob $antf$: 0.99.
prob $atnf$: 0.85.
prob $antnf$: 0.0001.
$smoke \leftarrow fire \wedge sf$.
prob sf : 0.9.
$smoke \leftarrow \sim fire \wedge snf$.
prob snf : 0.01.

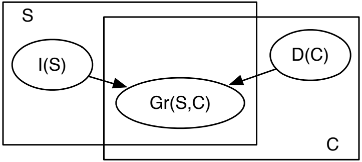
Belief Networks, Decision trees and ICL rules

- Rules can represent decision tree with probabilities:



$e \leftarrow a \wedge b \wedge h_1.$	$P_0(h_1) = 0.7$
$e \leftarrow a \wedge \sim b \wedge h_2.$	$P_0(h_2) = 0.2$
$e \leftarrow \sim a \wedge c \wedge d \wedge h_3.$	$P_0(h_3) = 0.9$
$e \leftarrow \sim a \wedge c \wedge \sim d \wedge h_4.$	$P_0(h_4) = 0.5$
$e \leftarrow \sim a \wedge \sim c \wedge h_5.$	$P_0(h_5) = 0.3$

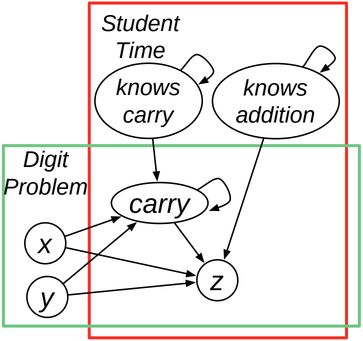
Predicting Grades



prob $int(S) : 0.5.$
prob $diff(C) : 0.5.$
$gr(S,C,G) \leftarrow int(S) \wedge diff(C) \wedge idg(S,C,G).$
prob $idg(S,C,a) : 0.5,\ idg(S,C,b) : 0.4,\ idg(S,C,c) : 0.1.$
$gr(S,C,G) \leftarrow int(S) \wedge \sim diff(C) \wedge indg(S,C,G).$
prob $indg(S,C,a) : 0.9,\ indg(S,C,b) : 0.09,\ indg(S,C,c) : 0.01.$
$gr(S,C,G) \leftarrow \sim int(S) \wedge diff(C) \wedge nidg(S,C,G).$
prob $nidg(S,C,a) : 0.01,\ nidg(S,C,b) : 0.09,\ nidg(S,C,c) : 0.9.$
$gr(S,C,G) \leftarrow \sim int(S) \wedge \sim diff(C) \wedge nindg(S,C,G).$
prob $nindg(S,C,a) : 0.1,\ nindg(S,C,b) : 0.4,\ nindg(S,C,c) : 0.5.$

Multi-digit addition with parametrized BNs / plates

$$\begin{array}{cccc} & x_{j_x} & \cdots & x_2 & x_1 \\ + & y_{j_z} & \cdots & y_2 & y_1 \\ \hline & z_{j_z} & \cdots & z_2 & z_1 \end{array}$$



Random Variables: $x(D,P)$, $y(D,P)$, $knowsCarry(S,T)$, $knowsAddition(S,T)$, $carry(D,P,S,T)$, $z(D,P,S,T)$
for each: digit D, problem P, student S, time T

☛ parametrized random variables

ICL rules for multi-digit addition

$$
\begin{aligned}
z(D,P,S,T) = V \leftarrow\ & \\
& x(D,P) = Vx \wedge \\
& y(D,P) = Vy \wedge \\
& carry(D,P,S,T) = Vc \wedge \\
& knowsAddition(S,T) \wedge \\
& \neg mistake(D,P,S,T) \wedge \\
& V \text{ is } (Vx + Vy + Vc) \text{ div } 10.
\end{aligned}
$$

$$
\begin{aligned}
z(D,P,S,T) = V \leftarrow\ & \\
& knowsAddition(S,T) \wedge \\
& mistake(D,P,S,T) \wedge \\
& selectDig(D,P,S,T) = V.
\end{aligned}
$$

$$
\begin{aligned}
z(D,P,S,T) = V \leftarrow\ & \\
& \neg knowsAddition(S,T) \wedge \\
& selectDig(D,P,S,T) = V.
\end{aligned}
$$

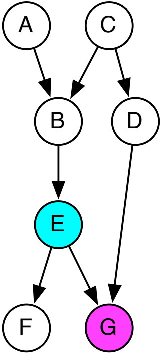
Alternatives:

$\forall DPST \{noMistake(D,P,S,T), mistake(D,P,S,T)\}$

$\forall DPST \{selectDig(D,P,S,T) = V \mid V \in \{0..9\}\}$

Outline

Bayesian Network Inference



$$P(E \mid g) = \frac{P(E \wedge g)}{p(g)}$$

$$P(E \wedge g) = \sum_F \sum_B \sum_C \sum_A \sum_D P(A)P(B \mid AC)$$

$$P(C)P(D \mid C)P(E \mid B)P(F \mid E)P(g \mid ED)$$

$$= \left(\sum_F P(F \mid E) \right)$$

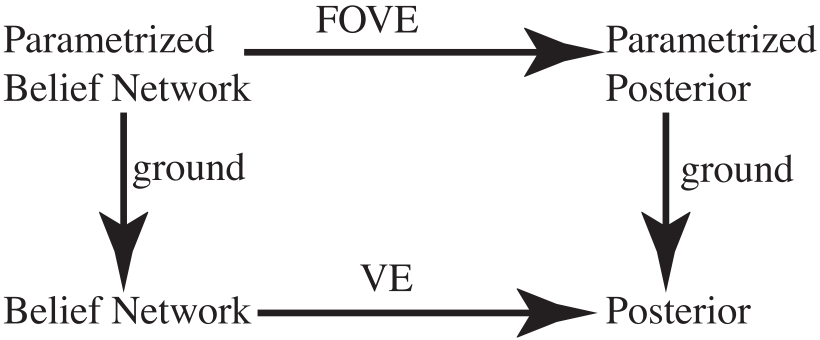
$$\sum_B P(e \mid B) \sum_C P(C) \left(\sum_A P(A)P(B \mid AC) \right)$$

$$\left(\sum_D P(D \mid C)P(g \mid ED) \right)$$

Lifted Inference

- Idea: treat those individuals about which you have the same information as a block; just count them.
- Use the ideas from lifted theorem proving - no need to ground.
- Potential to be exponentially faster in the number of non-differentialed individuals.
- Relies on knowing the number of individuals (the population size).

First-order probabilistic inference



Theorem Proving and Unification

In 1965, Robinson showed how unification allows many ground steps with one step:

$$\underbrace{f(X,Z) \vee p(X,a) \qquad \sim p(b,Y) \vee g(Y,W)}_{f(b,Z) \vee g(a,W)}$$

Substitution $\{X/b, Y/a\}$ is the most general unifier of $p(X,a)$ and $p(b,Y)$.

Variable Elimination and Unification

- Multiplying parametrized factors:

$$\underbrace{[f(X,Z),p(X,a)] \quad \times \quad [p(b,Y),g(Y,W)]}_{[f(b,Z),p(b,a),g(a,W)]}$$

  Doesn't work because the first parametrized factor can't subsequently be used for $X = b$ but can be used for other instances of X.

- We split $[f(X,Z),p(X,a)]$ into

  $[f(b,Z),p(b,a)]$
  $[f(X,Z),p(X,a)]$ with constraint $X \neq b$,

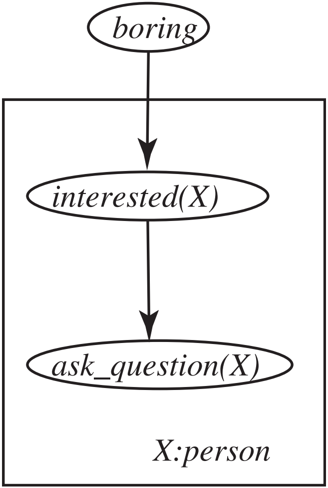
Parametric Factors

A parametric factor is a triple $\langle C, V, t \rangle$ where

- C is a set of inequality constraints on parameters,
- V is a set of parametrized random variables
- t is a table representing a factor from the random variables to the non-negative reals.

$$\left\langle \{X \neq sue\}, \{interested(X), boring\}, \begin{array}{|ll|l|} \hline interested & boring & Val \\ \hline yes & yes & 0.001 \\ yes & no & 0.01 \\ & \ldots & \\ \hline \end{array} \right\rangle$$

Removing a parameter when summing



n people

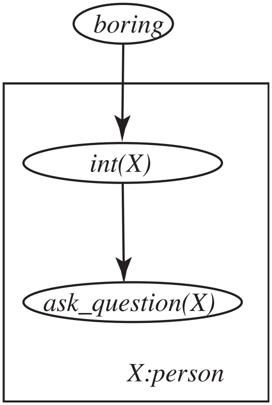
we observe no questions

Eliminate *interested*:

$$\langle \{\}, \{boring, interested(X)\}, t_1 \rangle$$
$$\langle \{\}, \{interested(X)\}, t_2 \rangle$$
$$\downarrow$$
$$\langle \{\}, \{boring\}, (t_1 \times t_2)^n \rangle$$

$(t_1 \times t_2)^n$ is computed pointwise; we can compute it in time $O(\log n)$.

Counting Elimination



$|people| = n$

[de Salvo Braz et al. 2007] and [Milch et al. 08]

Eliminate *boring*:

VE: factor on $\{int(p_1), \ldots, int(p_n)\}$

Size is $O(d^n)$ where d is size of range of interested.

Exchangeable: only the number of interested individuals matters.

Counting Formula:

#interested	Value
0	v_0
1	v_1
...	...
n	v_n

Complexity: $O(n^{d-1})$.

Potential of Lifted Inference

- Reduce complexity:

  polynomial $\longrightarrow$ *logarithmic*

  exponential $\longrightarrow$ *polynomial*

- We can now do lifting for unary relations, but we know we can't do all binary relations [Guy Van den Broeck, 2013]
- An active research area.

Outline

1. Logic and Probability
   - Relational Probabilistic Models
   - Probabilistic Logic Programs
2. Lifted Inference
3. Undirected models, Directed models, and Weighted Formulae
4. Existence and Identity Uncertainty

Logistic Regression

Logistic Regression, write $R(a_i)$ as R_i:

$$P(Q|R_1,\ldots,R_n) = \frac{1}{1+e^{w_0+w_1R_1+\cdots+w_nR_n}}$$

If all of the R_i are exchangeable $w_1,\ldots,w_n$ must all be the same:

$$P(Q|R_1,\ldots,R_n) = \frac{1}{1+e^{w_0+w_1(R_1+\cdots+R_n)}}$$

Logistic Regression

Logistic Regression, write $R(a_i)$ as R_i:

$$P(Q|R_1,\dots,R_n) = \frac{1}{1+e^{w_0+w_1R_1+\dots+w_nR_n}}$$

If all of the R_i are exchangeable $w_1,\dots,w_n$ must all be the same:

$$P(Q|R_1,\dots,R_n) = \frac{1}{1+e^{w_0+w_1(R_1+\dots+R_n)}}$$

If we learn the parameters for $n=10$ the prediction for $n=20$ depends on how values R_i are represented numerically:

- If *True* $=1$ and *False* $=0$ then $P(Q|R_1,\dots,R_n)$ depends on the number of R_i that are true.
- If *True* $=1$ and *False* $=-1$ then $P(Q|R_1,\dots,R_n)$ depends on how many more of R_i are true than false.
- If *True* $=0$ and *False* $=-1$ then $P(Q|R_1,\dots,R_n)$ depends on the number of R_i that are false.

Directed and Undirected models

- Weighted formula (WF): $\langle L, F, w \rangle$
  - L is a set of logical variables,
  - F is a logical formula: $\{$free logical variables in $F\} \subseteq L$
  - w is a real-valued weight.
- Instances of weighted formule obtained by assigning individuals to variables in L.

Directed and Undirected models

- Weighted formula (WF): $\langle L, F, w \rangle$
  - L is a set of logical variables,
  - F is a logical formula: $\{$free logical variables in $F\} \subseteq L$
  - w is a real-valued weight.
- Instances of weighted formule obtained by assigning individuals to variables in L.
- A world is an assignment of a value to each ground instance of each atom.
- Markov logic network (MLN): "undirected model" weighted formulae define measures on worlds.

Directed and Undirected models

- Weighted formula (WF): $\langle L, F, w \rangle$
  - L is a set of logical variables,
  - F is a logical formula: $\{$free logical variables in $F\} \subseteq L$
  - w is a real-valued weight.
- Instances of weighted formule obtained by assigning individuals to variables in L.
- A world is an assignment of a value to each ground instance of each atom.
- Markov logic network (MLN): "undirected model" weighted formulae define measures on worlds.
- Relational logistic regression (RLR): "directed model" weighted formulae define conditional probabilities.

Example

Weighted formulae:

$$\langle \{x\}, \mathit{funFor}(x), -5 \rangle$$
$$\langle \{x, y\}, \mathit{funFor}(x) \land \mathit{knows}(x, y) \land \mathit{social}(y), 10 \rangle$$

If *obs* includes observations for all $\mathit{knows}(x, y)$ and $\mathit{social}(y)$:

$$P(\mathit{funFor}(x) \mid \mathit{obs}) = \mathit{sigmoid}(-5 + 10 n_T)$$

n_T is the number of individuals y for which $\mathit{knows}(x, y) \land \mathit{social}(y)$ is *True* in *obs*.

$$\mathit{sigmoid}(x) = \frac{1}{1 + e^{-x}}$$

Abstract Example

$$\langle \{\}, q, \alpha_0 \rangle$$
$$\langle \{x\}, q \wedge \neg r(x), \alpha_1 \rangle$$
$$\langle \{x\}, q \wedge r(x), \alpha_2 \rangle$$
$$\langle \{x\}, r(x), \alpha_3 \rangle$$

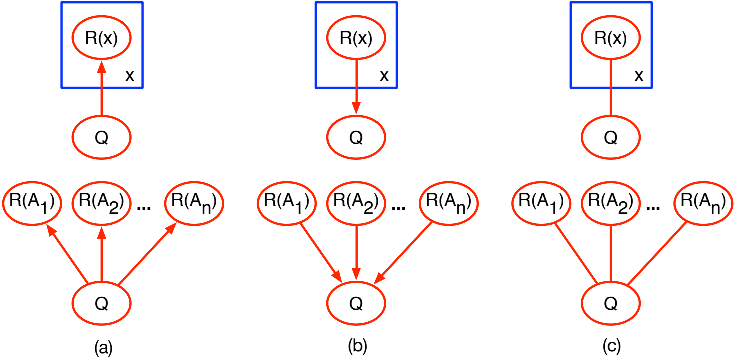
If $r(x)$ for every individual x is observed:

$$P(q \mid obs) = sigmoid(\alpha_0 + n_F \alpha_1 + n_T \alpha_2)$$

n_T is number of individuals for which $r(x)$ is true
n_F is number of individuals for which $r(x)$ is false

$$sigmoid(x) = \frac{1}{1 + e^{-x}}$$

Three Elementary Models

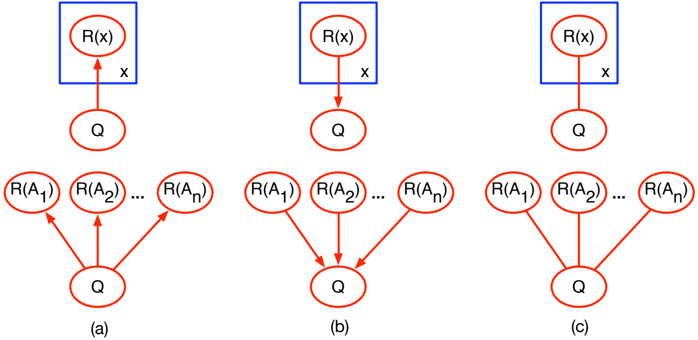


(a) Naïve Bayes

(b) (Relational) Logistic Regression

(c) Markov network

Independence Assumptions



- Naïve Bayes and Markov network: $R(x)$ and $R(y)$ (for $x \neq y$)
  - are independent given Q
  - are dependent not given Q.
- Directed model with aggregation: $R(x)$ and $R(y)$ (for $x \neq y$)
  - are dependent given Q,
  - are independent not given Q.

What happens as Population size n Changes: Simplest case

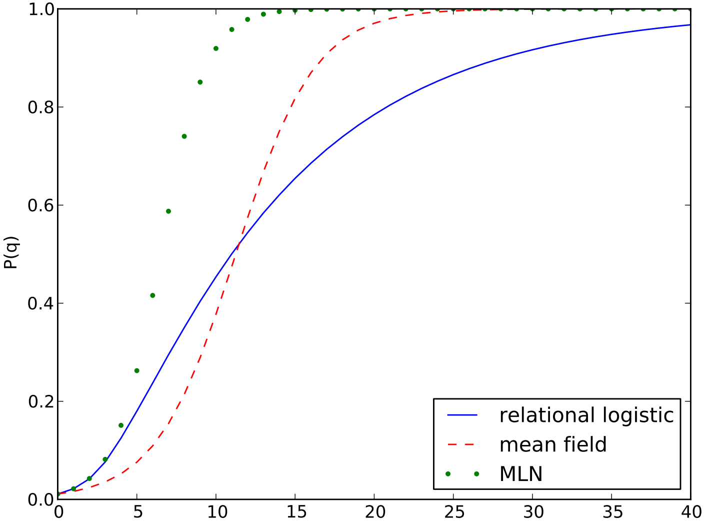
$\langle \{\}, q, \alpha_0 \rangle$
$\langle \{x\}, q \wedge \neg r(x), \alpha_1 \rangle$
$\langle \{x\}, q \wedge r(x), \alpha_2 \rangle$
$\langle \{x\}, r(x), \alpha_3 \rangle$

$$P_{MLN}(q \mid n) = sigmoid(\ \alpha_0 + n \log(e^{\alpha_2} + e^{\alpha_1 - \alpha_3})\ )$$

$$P_{RLR}(q \mid n) = \sum_{i=0}^{n} \binom{n}{i} sigmoid(\alpha_0 + i\alpha_1 + (n-i)\alpha_2)(1-p_r)^i p_r^{n-i}$$

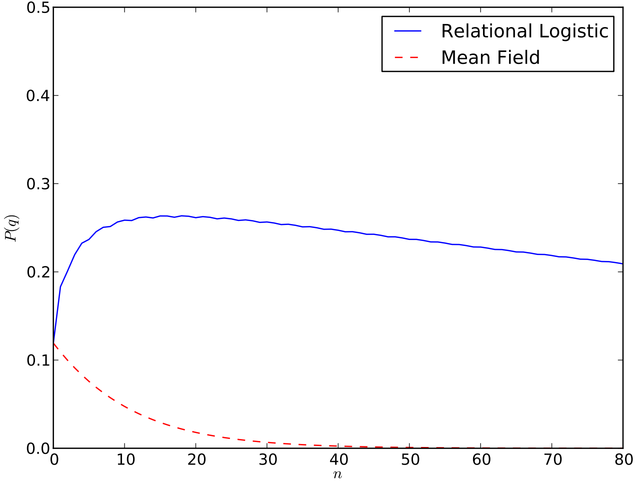
$$P_{MF}(q \mid n) = sigmoid(\alpha_0 + np_r\alpha_1 + n(1-p_r)\alpha_2)$$

Population Growth: $P(q \mid n)$



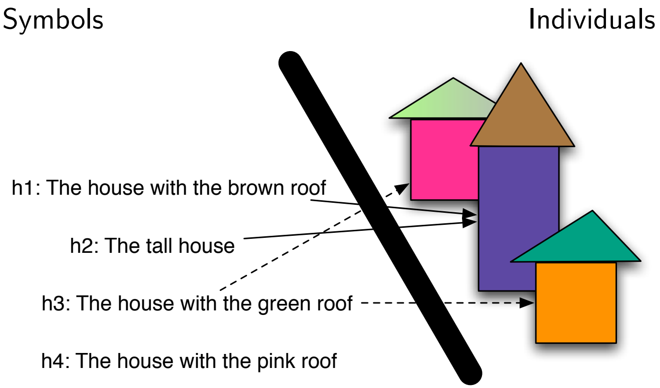
Population Growths: $P_{RLR}(q \mid n)$

Whereas this MLN is a sigmoid of n, RLR needn't be monotonic:



Outline

Correspondence Problem



c symbols and i individuals $\longrightarrow c^{i+1}$ correspondences

Clarity Principle

Clarity principle: probabilities must be over well-defined propositions.

- What if an individual doesn't exist?
  - $house(h4) \wedge roof\_colour(h4, pink) \wedge \neg exists(h4)$

Clarity Principle

Clarity principle: probabilities must be over well-defined propositions.

- What if an individual doesn't exist?
  - $house(h4) \land roof\_colour(h4, pink) \land \neg exists(h4)$
- What if more than one individual exists? Which one are we referring to?
  —In a house with three bedrooms, which is the second bedroom?

Clarity Principle

Clarity principle: probabilities must be over well-defined propositions.

- What if an individual doesn't exist?
  - $house(h4) \wedge roof\_colour(h4, pink) \wedge \neg exists(h4)$
- What if more than one individual exists? Which one are we referring to?
  —In a house with three bedrooms, which is the second bedroom?
- Reified individuals are special:
  — Non-existence means the relation is false.
  — Well defined what doesn't exist when existence is false.
  — Reified individuals with the same description are the same individual.

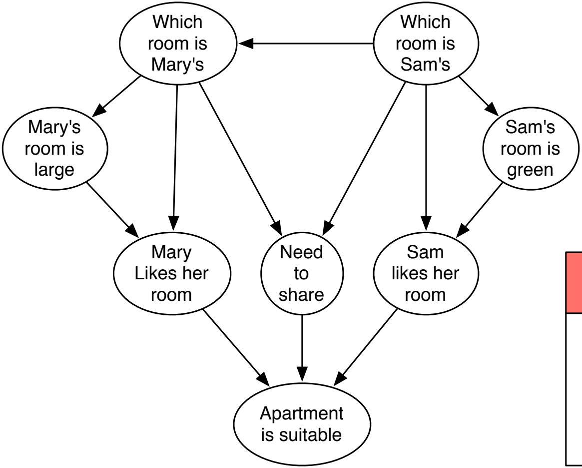
Role assignments

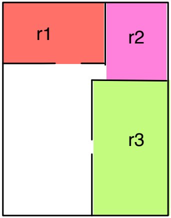
Hypothesis about what apartment Mary would like.

Whether Mary likes an apartment depends on:

- Whether there is a bedroom for daughter Sam
- Whether Sam's room is green
- Whether there is a bedroom for Mary
- Whether Mary's room is large
- Whether they share

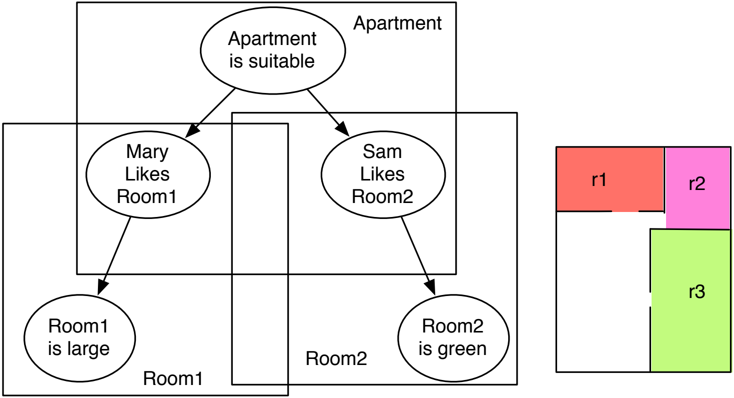
Bayesian Network Representation





How can we condition on the observation of the apartment?

Naive Bayes representation



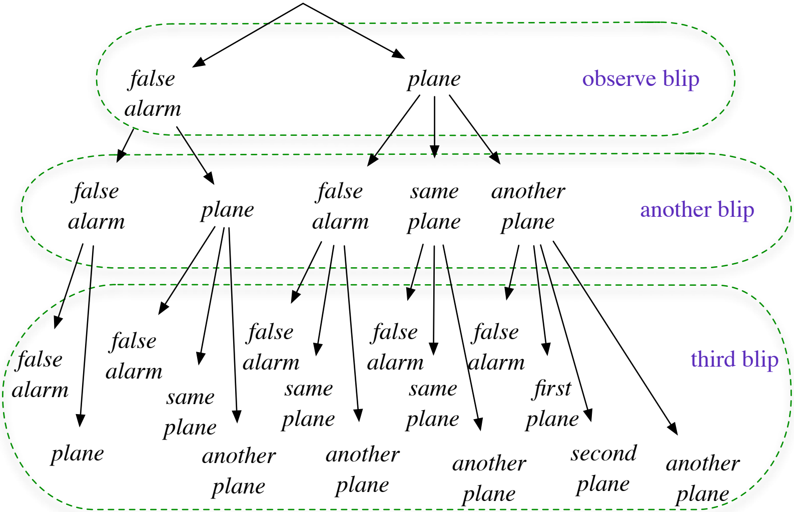
How do we specify that Mary chooses a room?
What about the case where they (have to) share?

Number and Existence Uncertainty

- PRMs (Pfeffer et al.), BLOG (Milch et al.): distribution over the number of individuals. For each number, reason about the correspondence.
- NP-BLOG (Carbonetto et al.): keep asking: is there one more?

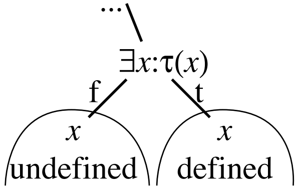
  e.g., if you observe a radar blip, there are three hypotheses:
  - the blip was produced by plane you already hypothesized
  - the blip was produced by another plane
  - the blip wasn't produced by a plane

Existence Example



First-order Semantic Trees

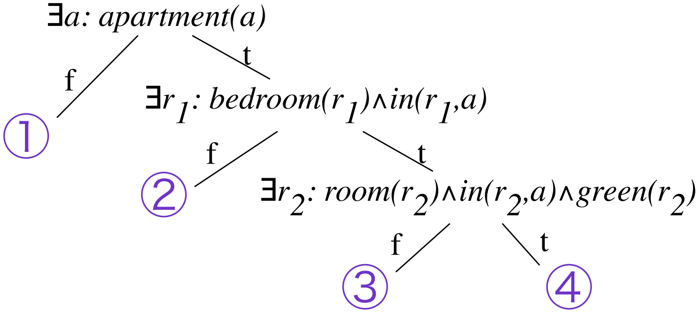
Split on quantified first-order formulae:



- The "true" sub-tree is in the scope of x
- The "false" sub-tree is not in the scope of x

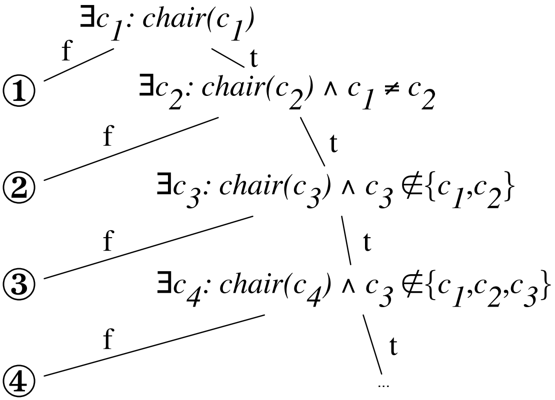
A logical generative model generates a first-order semantic tree.

First-order Semantic Tree (cont)

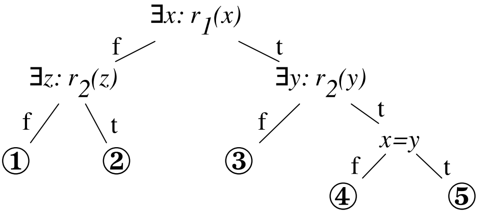


1. there is no apartment
2. there is no bedroom in the apartment
3. there is a bedroom but no green room
4. there is a bedroom and a green room

Distributions over number



Roles and Identity (1)



① there no individual filling either role

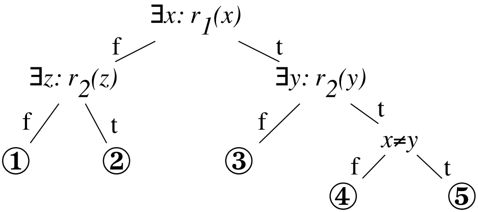
② there is an individual filling role r_2 but none filling r_1

③ there is an individual filling role r_1 but none filling r_2

④ only different individuals fill roles r_1 and r_2

⑤ some individual fills both roles r_1 and r_2

Roles and Identity (2)



① there no individual filling either role

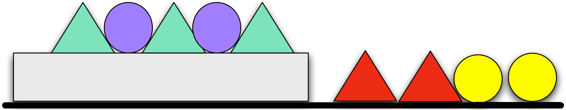
② there is an individual filling role r_2 but none filling r_1

③ there is an individual filling role r_1 but none filling r_2

④ only the same individual fill roles r_1 and r_2

⑤ there are different individuals that fill roles r_1 and r_2

Observation Protocols

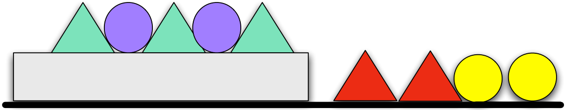


Observe a triangle and a circle touching. What is the probability the triangle is green?

$$P(green(x) \mid triangle(x) \land \exists y\ circle(y) \land touching(x, y))$$

The answer depends on how the x and y were chosen!

Protocol for Observing



$$P(green(x) \mid triangle(x) \wedge \exists y\ circle(y) \wedge touching(x, y))$$

$select(x)$	$select(y)$	$select(x, y)$
$select(y)$	$select(x)$	
$3/4$	$2/3$	$4/5$

Conclusion

- To decide what to do an agent should take into account its uncertainty and it preferences (utility).
- The field of "statistical relational AI" looks at how to combine first-order logic and probabilistic reasoning.
- We need models that can condition on observations that follow some protocol

Challenges

- Representation: heuristically and epistemologically adequate representations for probabilistic models + observations (+ actions + utilities + ontologies)
- Inference: compute posterior probabilities (or optimal actions) quickly enough to be useful
- Learning: get best hypotheses conditioned on all observations possible

AI: computational agents that act intelligently