Population Size Extrapolation in Relational Probabilistic Modelling

David Poole, David Buchman, Seyed Mehran Kazemi, Kristian Kersting and Sriraam Natarajan

University of British Columbia,
Technical University of Dortmund,
Indiana University

September 2014

Outline

Outline

Relational Probabilistic Models

Markov Logic Networks and Relational Logistic Regression

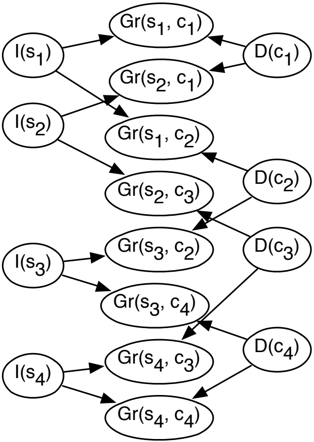
Varying Populations

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. Why should we make different predictions?
- How can we make predictions when the values of properties *Student* and *Course* are individuals?

From Relations to Belief Networks

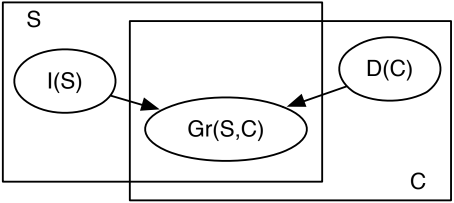


$I(S)$	$D(C)$	$Gr(S,C)$		
		A	B	C
true	*true*	0.5	0.4	0.1
true	*false*	0.9	0.09	0.01
false	*true*	0.01	0.1	0.9
false	*false*	0.1	0.4	0.5

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

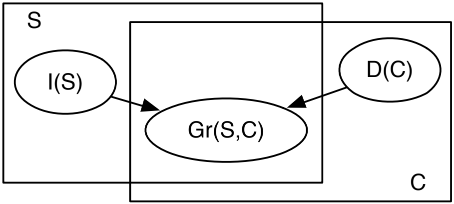
"parameter sharing"

Plate Notation



- S is a logical variable representing students
- C is a logical variable representing 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 is a logical variable representing students
- C is a logical variable representing courses
- the set of individuals of a type is called a population
- $I(S)$, $Gr(S,C)$, $D(C)$ are parametrized random variables
- for every student s, there is a random variable $I(s)$
- for every course c, there is a random variable $D(c)$
- for every student s and course c pair there is a random variable $Gr(s,c)$
- all instances share the same structure and parameters

Outline

Relational Probabilistic Models

Markov Logic Networks and Relational Logistic Regression

Varying Populations

Directed and Undirected models

- Weighted formula (WF): $\langle L, F, w \rangle$
  - L is a set of logical variables,
  - F is a logical formula: $\{\text{free logical variables in } F\} \subseteq L$
  - 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:
  Probability of a world is proportional to the exponent of the sum of the instances of the formulae true in the world.

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: Probability of a world is proportional to the exponent of the sum of the instances of the formulae true in the world.
- Relational logistic regression (RLR): "directed model" weighted formulae define conditional probabilities: Probability of a variable assignment given a parent assignment is proportional to the exponent of the sum of the weights the instances of the formulae true in the assignment.

Example

Weighted formulae:

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

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

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

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

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

Abstract Example

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

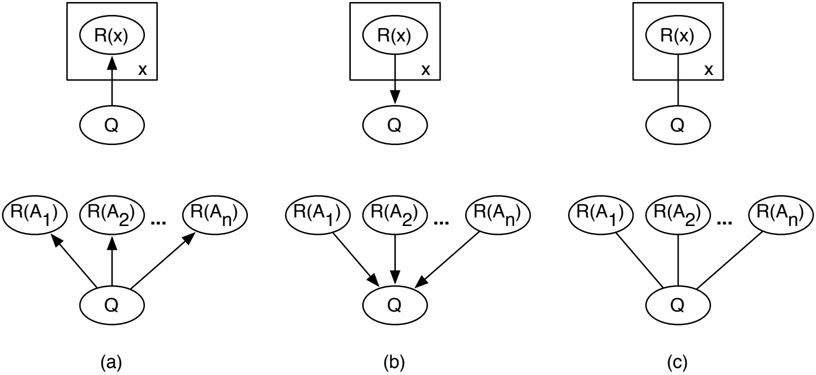
If $r(x)$ for every individual x is observed:

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

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

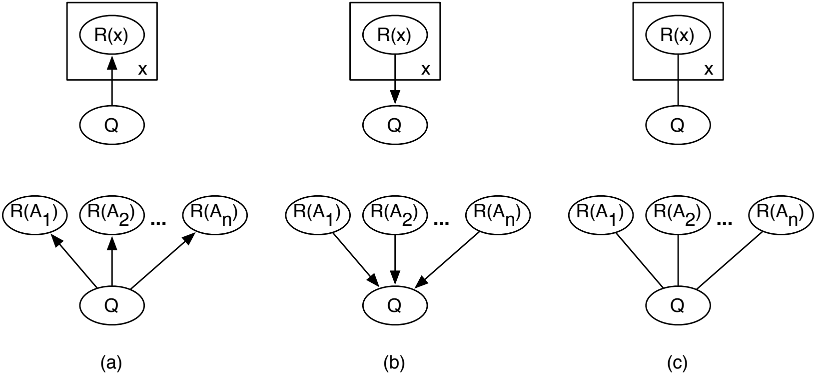


(a) Naïve Bayes

(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.

Outline

Relational Probabilistic Models

Markov Logic Networks and Relational Logistic Regression

Varying Populations

What happens as Population size n Changes: Simplest case

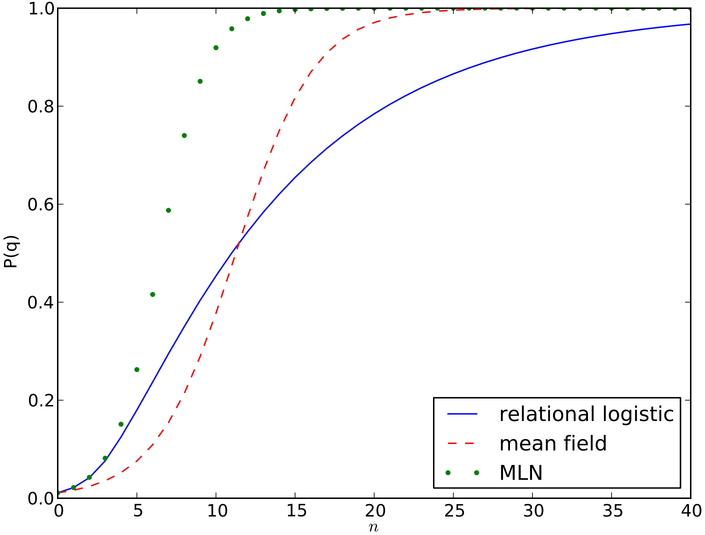
$\langle \{\}, q, \alpha_0 \rangle$
$\langle \{x\}, q \wedge \neg r(x), \alpha_1 \rangle$
$\langle \{x\}, q \wedge r(x), \alpha_2 \rangle$
$\langle \{x\}, r(x), \alpha_3 \rangle$

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

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

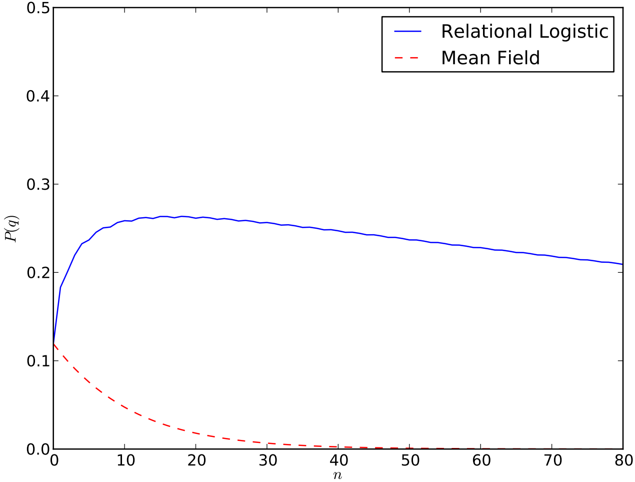
$$P_{MF}(q \mid n) = sigmoid(\alpha_0 + np_r\alpha_1 + n(1-p_r)\alpha_2)$$

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

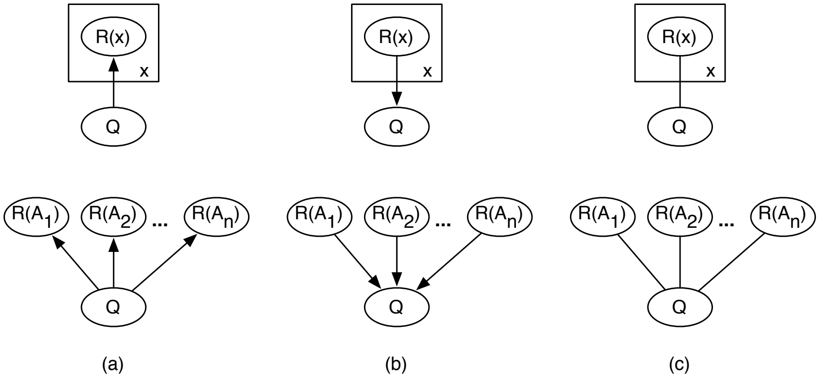


Population Growths: $P_{RLR}(q \mid n)$

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

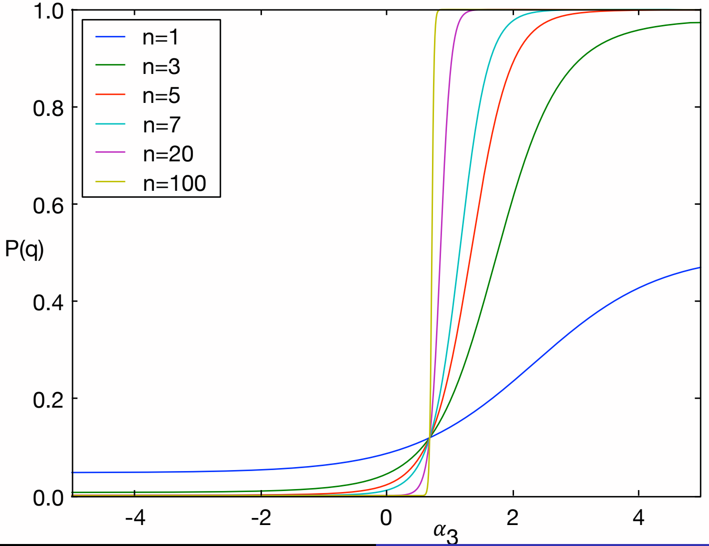


Dependence of $R(x)$ on population size

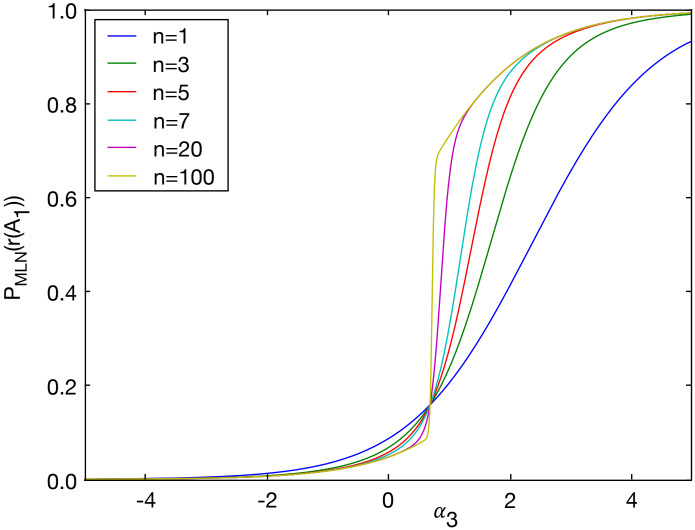


- In (b), the directed model with aggregation, $P(R(x))$ is not affected by the population size.
- In (c), $P_{MLN}(R(x))$ is unaffected by population size if and only if the MLN is equivalent to a Naïve Bayes model (a).
- For other MLNs...

$P_{MLN}(q \mid \alpha_3)$ for various n



$P_{MLN}(r(A_1) \mid \alpha_3)$ for various n



Results on population growth

- For RLR the probability of child given the parents is aways the sigmoid of a polynomial of the counts of the parents.
  All polynomials can be represented.

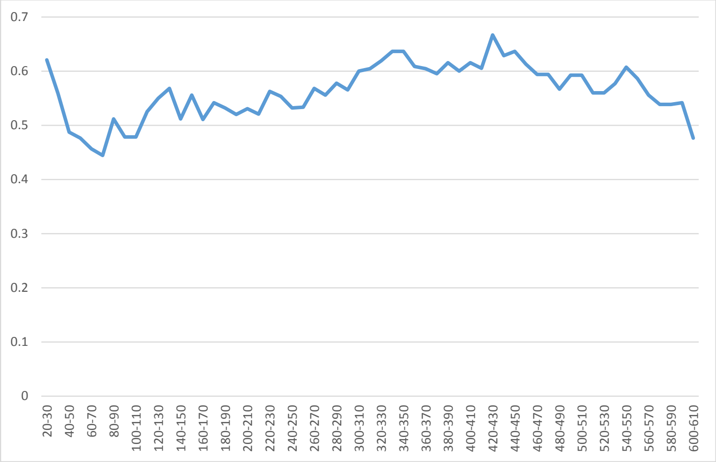
Results on population growth

- For RLR the probability of child given the parents is aways the sigmoid of a polynomial of the counts of the parents. All polynomials can be represented.
- In an MLN without infinite weights, if V is not in a formula with a logical variable of a population, then $P(V \mid n)$ is bounded away from 0 and 1 as population $n \to \infty$.

Results on population growth

- For RLR the probability of child given the parents is aways the sigmoid of a polynomial of the counts of the parents.
  All polynomials can be represented.
- In an MLN without infinite weights, if V is not in a formula with a logical variable of a population, then $P(V \mid n)$ is bounded away from 0 and 1 as population $n \to \infty$.
- In an MLN without infinite weights, if V is in a formula with some $R(X)$, where X does not appear in V and $R(X)$ doesn't unify with other formulae:
  then either $P(r)$ is independent of the population size n or $\lim_{n\to\infty} P_{MLN}(r)$ is either 1 or 0.

Real Data



Observed $P(25 < Age(p) < 45 \mid n)$, where n is number of movies watched from the Movielens dataset.

Example of polynomial dependence of population

$$\langle \{\}, q, \alpha_0 \rangle$$
$$\langle \{x\}, q \wedge true(x), \alpha_1 \rangle$$
$$\langle \{x\}, q \wedge r(x), \alpha_2 \rangle$$
$$\langle \{x\}, true(x), \alpha_3 \rangle$$
$$\langle \{x\}, r(x), \alpha_4 \rangle$$
$$\langle \{x, y\}, q \wedge true(x) \wedge true(y), \alpha_5 \rangle$$
$$\langle \{x, y\}, q \wedge r(x) \wedge true(y), \alpha_6 \rangle$$
$$\langle \{x, y\}, q \wedge r(x) \wedge r(y), \alpha_7 \rangle$$

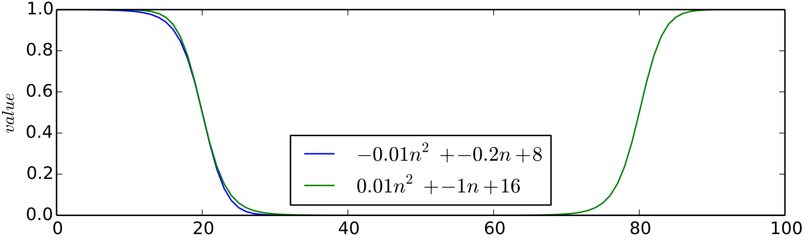
In RLR and in MLN, if all $R(A_i)$ are observed:

$$P(q \mid obs) = sigmoid(\alpha_0 + n\alpha_1 + n_T\alpha_2 + n^2\alpha_5 + n_T n\alpha_6 + n_T^2\alpha_7)$$

$R(x)$ is true for n_T individuals out of a population of n.

Danger of fitting to data without understanding the model

- RLR can fit sigmoid of any polynomial.
- Consider a polynomial of degree 2:



Conclusions

- The form of the formulae used gives prior information about the dependence on population.
- The model should fit with our prior knowledge.
- We are beginning to understand this dependence, but there is a lot we don't know.
- MLNs and RLR provide different modelling assumptions, which are applicable in different circumstances.