Limitations and potential of lifted probabilistic inference

David Poole

Department of Computer Science,
University of British Columbia

IJCAI 2011 Tutorial

Outline

Assumptions behind current lifted algorithms

- Known population size
- Conditioning and querying on conjunctions of ground atoms
- Unique names assumption: different references to individuals denote different individuals. There is no identity uncertainty.
- No querying about equality
- No existence uncertainty

Modelling Assumptions

A priori, individuals are indistinguishable and so share the same probabilities (exchangeability)

- unique names, known # individuals
- unique names, unknown # individuals
- identity uncertainty
- identity uncertainty, existence/type uncertainty

What is Observed / Queried?

- conjunction of ground assignments for few individuals
- conjunction of ground assignments for all individuals
- arbitrary propositions of ground assignments
- quantified (first-order) query, without equality
- quantified (first-order) query, with equality

Example Observation

Suppose we observe exactly one person asking a question:

$$\exists x \; asks\_question(x)$$

Example Observation

Suppose we observe exactly one person asking a question:

$$\exists x \; asks\_question(x) \\ \quad \wedge \forall y \; y \neq x \rightarrow \neg asks\_question(y)$$

[What is a reasonable language for observations?
What is an agent physically able to observe?]

Limitations to Lifted Inference

Jaeger [AIJ 2000] show that

- If the query/observation language includes first-order logic with equality
- then there are queries that are not polynomial in population size.

Limitations to Lifted Inference

Jaeger [AIJ 2000] show that

- If the query/observation language includes first-order logic with equality
- then there are queries that are not polynomial in population size.

Are there weaker languages with equality that avoid this proof?
Are there stronger results that are possible?

Outline

Problems with Lifted VE

- Intermediate representations (parfactors, counting formulae...) are not closed under lifted operations.
  - We need to ground sometimes
  - We need better intermediate representations
  - We need alternatives to lifted VE.

Lifted Search

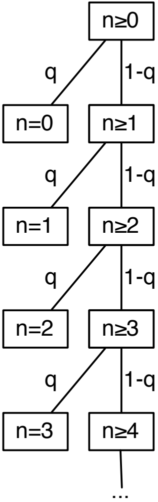
- Variable elimination is the dynamic programming variant of recursive conditioning (and related search methods).
- VE creates intermediate representations.
- Search just evaluates factors when fully instantiated.
- In search, everything is evaluated in a particular context — perhaps we don't need complex intermediate representations that need to anticipate all eventualities.

Outline

What if we don't know the population size?

- The previous methods assumed that we know how many individuals there are.
- What if we don't know the population size, but only have a distribution over population size?

Geometric population size



If n is distributed according to a geometric distribution with $\forall k \; q = P(n = k | n \geq k)$, the expected value of p^n is

$$\frac{q}{1 - p(1-q)}$$

Proof:

$$\begin{aligned}
\mathcal{E}_n(p^n) &= \sum_{n=0}^{\infty} q(1-q)^n p^n \\
&= q \sum_{n=0}^{\infty} (p(1-q))^n \\
&= \frac{q}{1-p(1-q)}
\end{aligned}$$

Poisson population size

If n is distributed according to a Poisson distribution $f(n, \lambda) = \frac{\lambda^n e^{-\lambda}}{n!}$ (λ is the expected number of individuals) the expected value of p^n is

$$e^{-\lambda(1-p)}.$$

Proof:

$$\begin{aligned}\mathcal{E}_n(p^n) &= \sum_{n=0}^{\infty} \frac{\lambda^n e^{-\lambda}}{n!} p^n \\ &= e^{-\lambda} e^{\lambda p} \sum_{n=0}^{\infty} \frac{(\lambda p)^n e^{-\lambda p}}{n!} \\ &= e^{-\lambda(1-p)}\end{aligned}$$

Counting for unknown population size

- Open problem: Is there an analog of counting formulae for unknown population?
- Can we do lifted inference after finding some evidence about the number of objects? (Is there a conjugate distribution for counting with an unknown population?)

What is the population size is infinite?

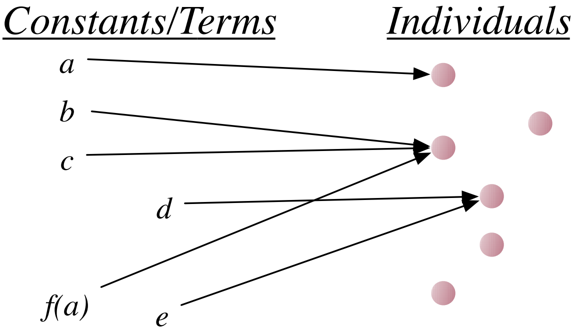
- If $n = \infty$ then

$$p^n = \begin{cases} 1 \text{ if } p = 1 \\ 0 \text{ if } p < 1 \end{cases}$$

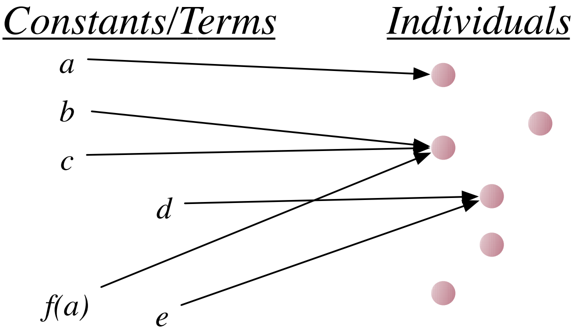
- Is there a (finite) counting formula for $n = \infty$?

Outline

Symbol Denotations

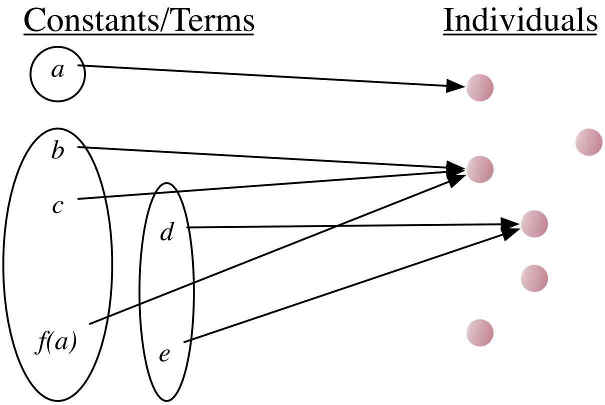


Symbol Denotations



In logic, $x = y$ is true if x and y refer to the same individual.
$a \neq b$, $b = c$, $b = f(a)$, $d = e$, $d \neq b$, ...

Symbol Partitioning



Equality

Equality can be axiomatized with:

- $x = x$
- $x = y \Rightarrow y = x$
- $x = y \wedge y = z \Rightarrow x = z$
- $y = z \Rightarrow f(x_1, \ldots, y, \ldots, x_n) = f(x_1, \ldots, z, \ldots, x_n)$
- $y = z \wedge p(x_1, \ldots, y, \ldots, x_n) \Rightarrow p(x_1, \ldots, z, \ldots, x_n)$

The most common theorem-proving method is paramodulation: map each equivalence class of equal terms to a canonical element.

Probability and Identity

- Have a probability distribution over partitions of the terms
- The number of partitions grows faster than any exponential (Bell number)
- The most common method is to use MCMC: one step is to move a term to a new or different partition.
- Can we do this in a analytic / lifted manner?

Existence Uncertainty

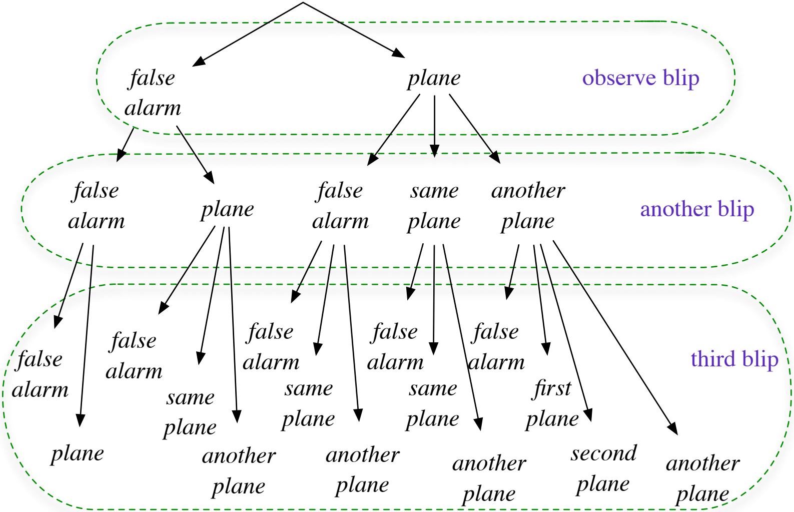
- What is the probability there is a plane in this area?
- What is the probability there is a large gold reserve in some region?
- What is the probability that there is a third bathroom given there are two bedrooms?
- What is the probability that there are (exactly) three bathrooms given there are two bedrooms?

Existence Uncertainty

Two approaches:

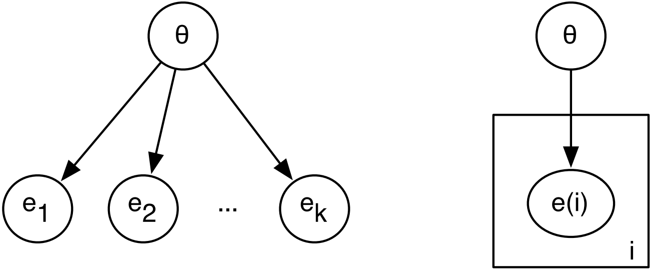
- BLOG: you have a distribution over the number of objects, then for each number you can reason about the correspondence.
- NP-BLOG: keep asking: is there one more?
  e.g., if you observe a radar blip, there are three hypotheses:
  - the blip was produced by plane you already hypothesized
  - the blip was produced by another plane
  - the blip wasn't produced by a plane

Existence Example



Lifted Learning

For a Bayesian there is only inference.



Model: $\forall i \; P(e(i) = 1|\theta) = \theta$

Infer: $P(\theta|e(1) \ldots e(k)) \propto \theta^{|e(i)=1|}(1-\theta)^{|e(i)=0|}P(\theta)$

Conclusion

- Big gap between what we know how to do and the potential of lifted inference.
- Limited knowledge of the limitations of lifted inference.
- Exact inference forms the foundations of approximate inference (e.g, Rao-Blackwellization, variational methods)
- How to do lifted inference for richer languages?
- Lifted planning, lifted MDPs, lifted POMDPs, lifted RL....
- What do we need for real applications?