Bucket Elimination in Credal Networks

Exact and Approximation Algorithms

Denis Mauá, Cassio de Campos, Marco Zaffalon
IDSIA

Credal Networks

- Extend Bayesian nets to the imprecise setting.
- Knowledge is represented qualitatively by an interaction graph, and quantitatively by a collection of credal sets.
- Applications: e.g. expert systems, classifiers.

Credal Networks

- Use of imprecise probabilities leads to computational difficulties.
- Inferences in CNs are NP-hard even in cases where their Bayesian counterpart is polynomial.
- Current approximate algorithms for computing with CNs do not provide any bounds on the error.

Extensive Specification

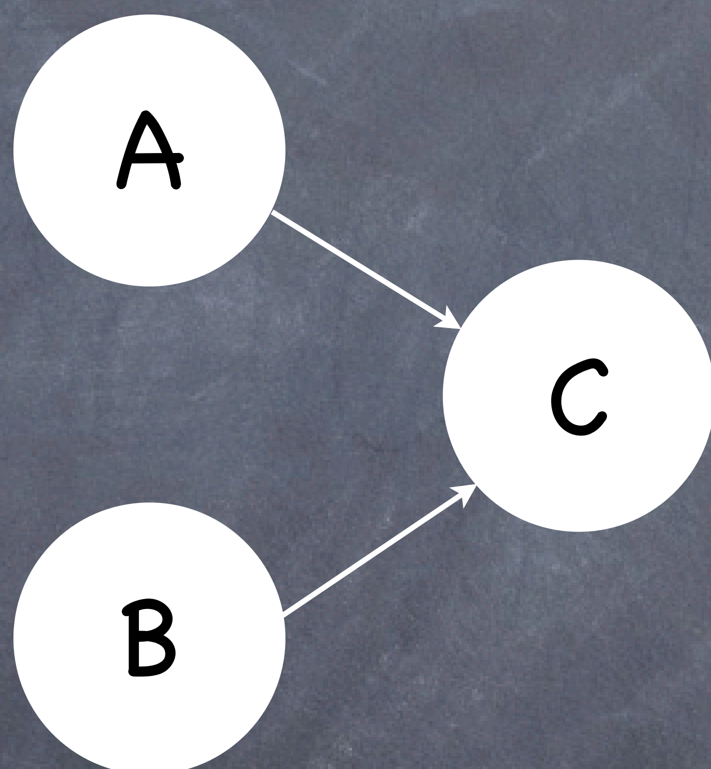
- We focus on extensively specified credal sets/networks.
- Credal sets are specified as sets of conditional probability tables.
- Results can be generalized to locally specified nets.

Example

K(A)

a	0.3
~a	0.7

P1(A)



K(B)

b	0.1
~b	0.9

P1(B)

b	0.7
~b	0.3

P2(B)

K(C|A,B)

c	a	b	0.2
~c	a	b	0.8
c	~a	b	0.6
~c	~a	b	0.4
c	a	~b	0.8
~c	a	~b	0.2
c	~a	~b	0.5
~c	~a	~b	0.5

P1(C|A)

c	a	b	0.4
~c	a	b	0.6
c	~a	b	0.5
~c	~a	b	0.5
c	a	~b	0.3
~c	a	~b	0.7
c	~a	~b	0.5
~c	~a	~b	0.5

P2(C|A)

Potentials

- A potential P(U|V) is mapping from the possibility space of U and V to the set of nonnegative rational numbers.
- It “may” represent (conditional) probability functions p(U|v) for each v.
- They inherit algebraic properties from probabilities (c.f. valuation algebras).

Potentials

$P(C|A)$

c	a	0.2
~c	a	0.8
c	~a	0.6
~c	~a	0.4

$$\frac{p(C|a)}{p(C|\sim a)}$$

$P(d,C|A)$

c	a	0.2
~c	a	0.4
c	~a	0.3
~c	~a	0.1

$$\frac{p(d,C|a)}{p(d,C|\sim a)}$$

Credal Sets

- We consider extensively specified credal sets.
- By abuse of terminology, we define a credal set K(U|V) as a finite set of potentials P(U|V).
- Equivalent to the extreme mass functions of a closed convex set of (precise) probability functions, where the choice of a cpf in a potential implies the choice of the others cpfs.

Credal Sets

- Let u be an element of the sample space of U, and v be an element of the sample space of V.
- $K(u|v)$ denotes a set of probabilities $p(u|v)$.

Example

K(A)

a	0.3
~a	0.7

P1(A)

K(B)

b	0.1
~b	0.9

P1(B)

b	0.7
~b	0.3

P2(B)

K(~a) = {0.7}

K(b) = {0.1, 0.7}

Credal Networks

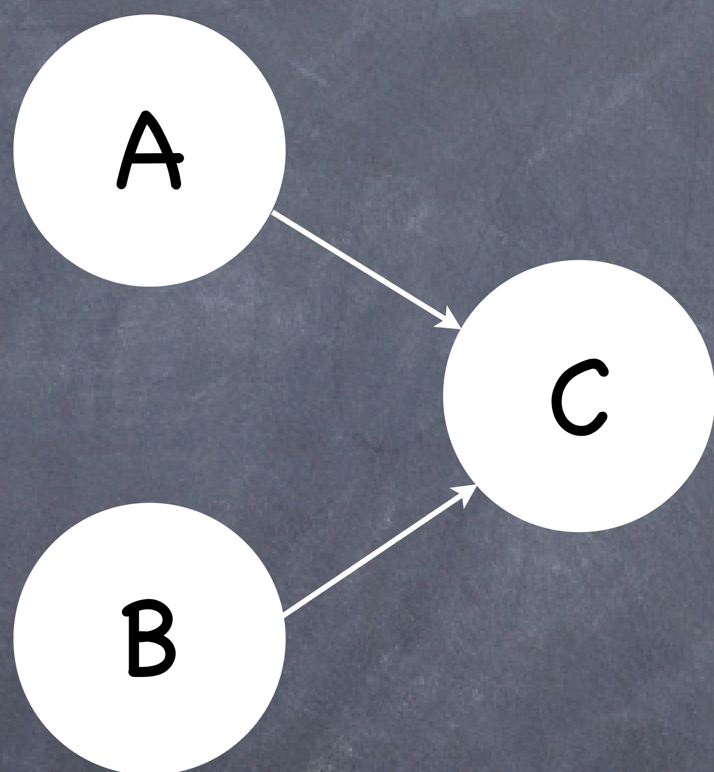
- Let X={X1,...,Xn} be a set of categorical random variables.
- Let G=(V,E) be a DAG where there is a node Vi for each Xi in X.
- Let K be a collection of credal sets K(Xi|pa(Xi)) for each Xi in X.
- A credal network is a pair CN=(G,K).

Example

K(A)

a	0.3
~a	0.7

P1(A)



K(B)

b	0.1	b	0.7
~b	0.9	~b	0.3

P1(B) P2(B)

K(C|A,B)

c	a	b	0.2
~c	a	b	0.8
c	~a	b	0.6
~c	~a	b	0.4
c	a	~b	0.8
~c	a	~b	0.2
c	~a	~b	0.5
~c	~a	~b	0.5

P1(C|A)

c	a	b	0.4
~c	a	b	0.6
c	~a	b	0.5
~c	~a	b	0.5
c	a	~b	0.3
~c	a	~b	0.7
c	~a	~b	0.5
~c	~a	~b	0.5

P2(C|A)

Algebra of Credal Sets

- Marginalization and multiplication operations generalizes marginalization and multiplication in probabilities.
- Operations in Credal sets are performed element-wise.

Algebra of Credal Sets

- $K1 = \{P1,P2\}$, $K2 = \{P3,P4\}$,
  $K3 = K1*K2 = \{P1P3,P1P4,P2P3,P2P4\}$
- $K(A) = \sum_B K(A,B)$
  $= \{ \sum_B P(A,B) : P(A,B) \text{ in } K(A,B) \}$
- $\sum_B K(A|B)K(C|D) = K(C|D) \sum_B K(A|B)$

Inference in CNs

- Given a CN=(G,K) the strict (strong) extension is given by $K(X) = \prod K(Xi|pa(Xi))$
- Belief updating consists in computing marginals conditional on evidence
  - $K(Q|E) = \sum_{X\backslash Q,E} K(X) / \sum_{X\backslash E} K(X)$

Inference in CNs

- In particular, if $E=\emptyset$
  - $K(Q) = \sum_{X \setminus Q} K(X)$
- It is possible to map a belief updating problem with evidence into an equivalent belief updating problem with no evidence.

Inference in CNs

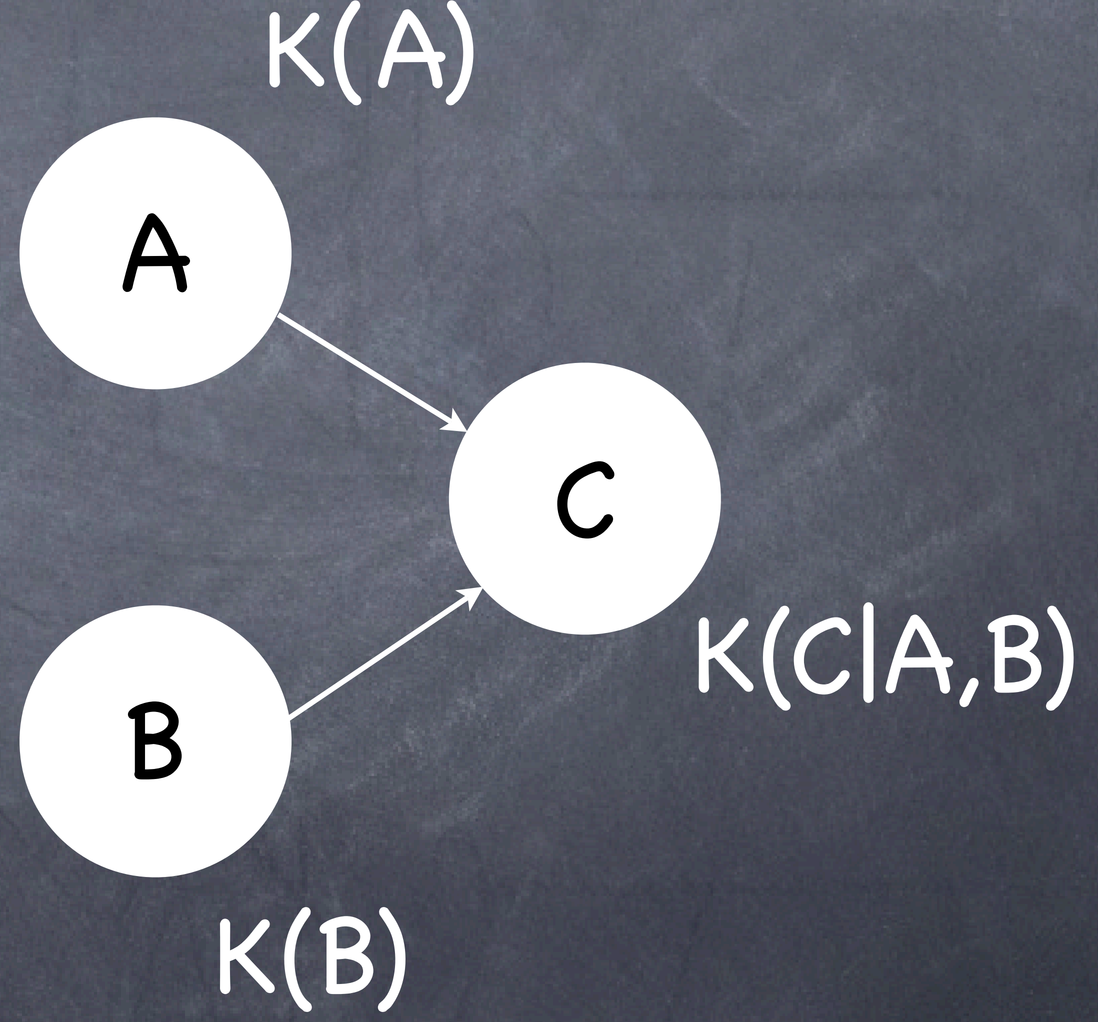
- For a given ordering $X1,\ldots,Xn$
  - $K(Q) = \sum_{X1,\ldots Xn \setminus Q} K(X)$
  - $K(Q) = \sum_{X1,\ldots Xn \setminus Q} \prod K(Xi|pa(Xi))$
- Chosen ordering matters (a lot).

Bucket Elimination

- Assume an ordering X1,...,Xn of the variables in X.
- Start with an ordered partition bucket 1,...,bucket n of K.
  - bucket i contains all credal sets whose highest variable is Xi.
- Recursion: for i := n to 1, do:
  - Compute $K(U|V) = \sum_{Xi \setminus Q} \prod_{\text{bucket } i} Kj$ and add it to the largest-index variable bucket.

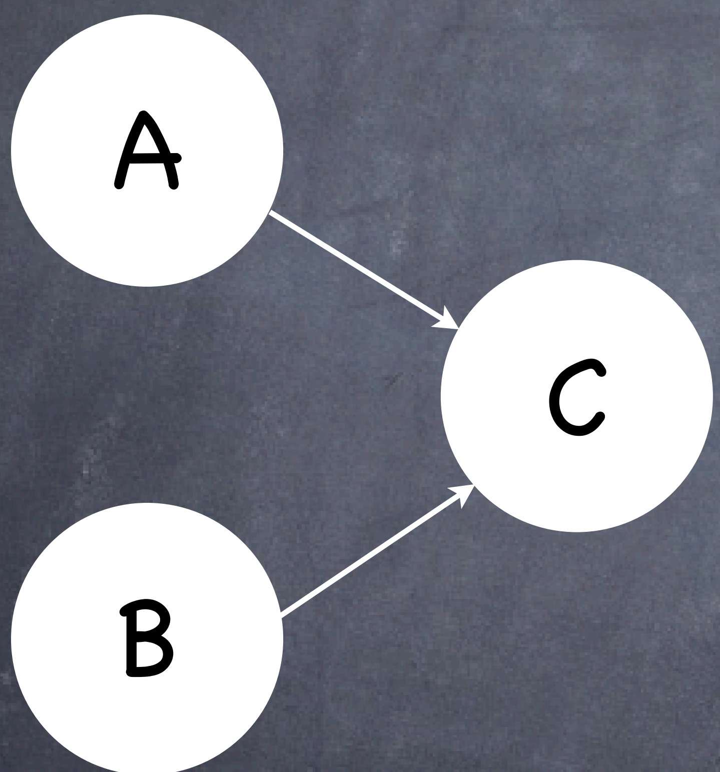
Example

- $G = (\{A,B,C\},\{(A,C),(B,C)\})$
- $K = \{K(A),K(B),K(C|A,B)\}$
- $K(C)$?



Example

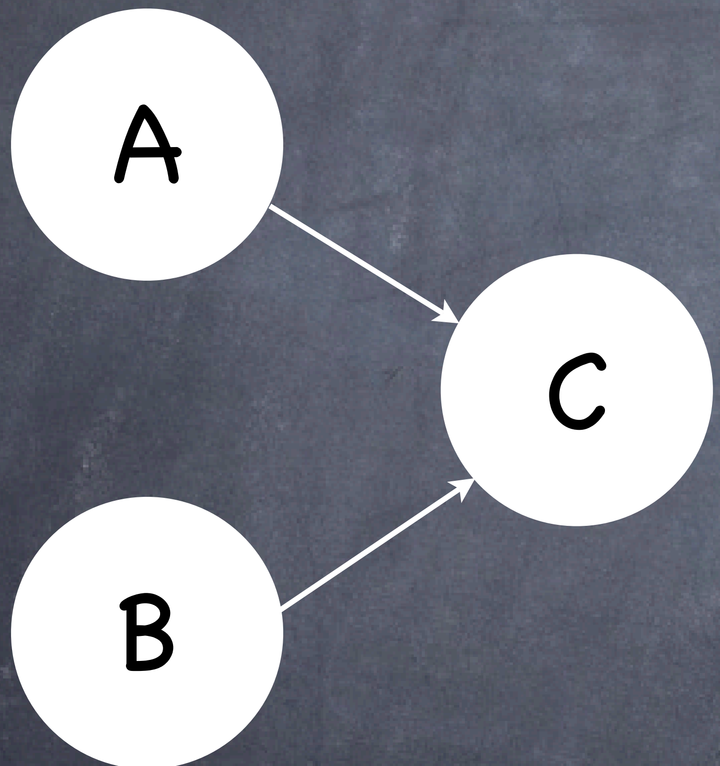
ordering: B, C, A



initialization:

bucket B	bucket C	bucket A
K(B)		K(C|A,B) K(A)

Example



ordering: B, C, A

recursion 1:

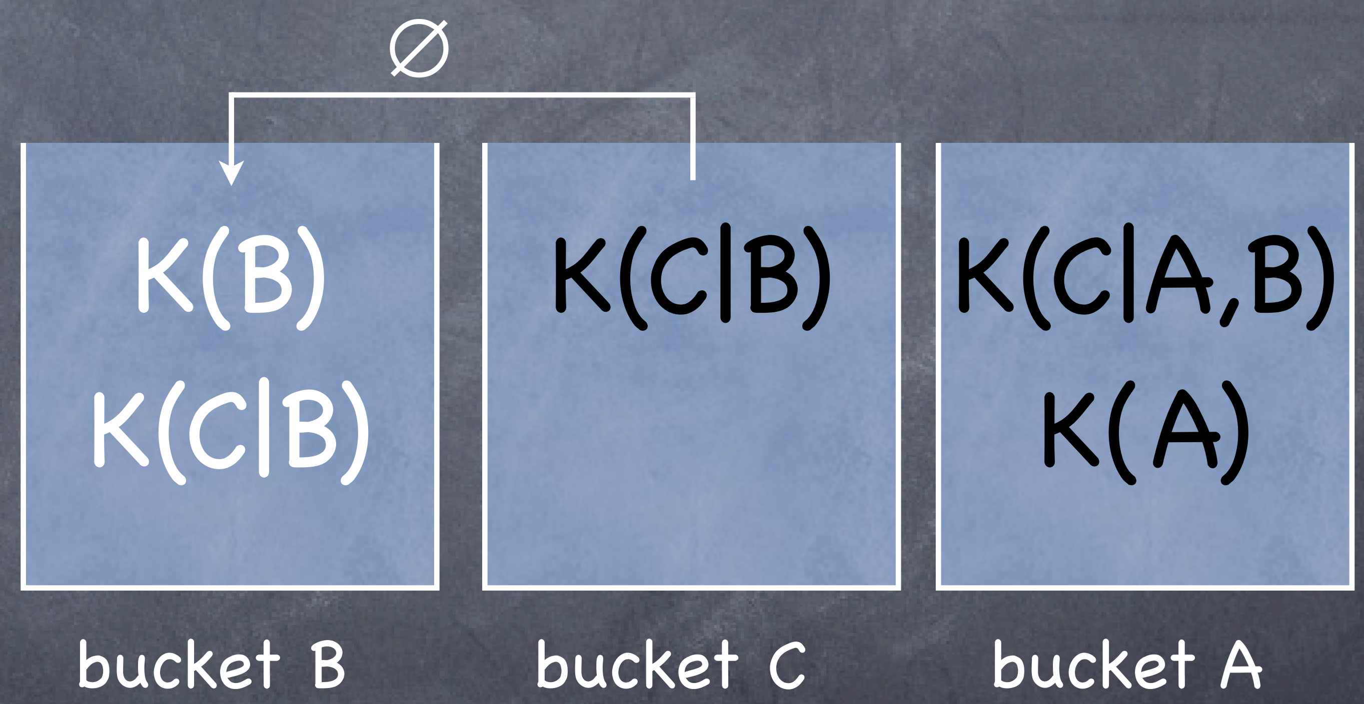
$K(C|B)=\sum_A K(A)K(C|A,B)$

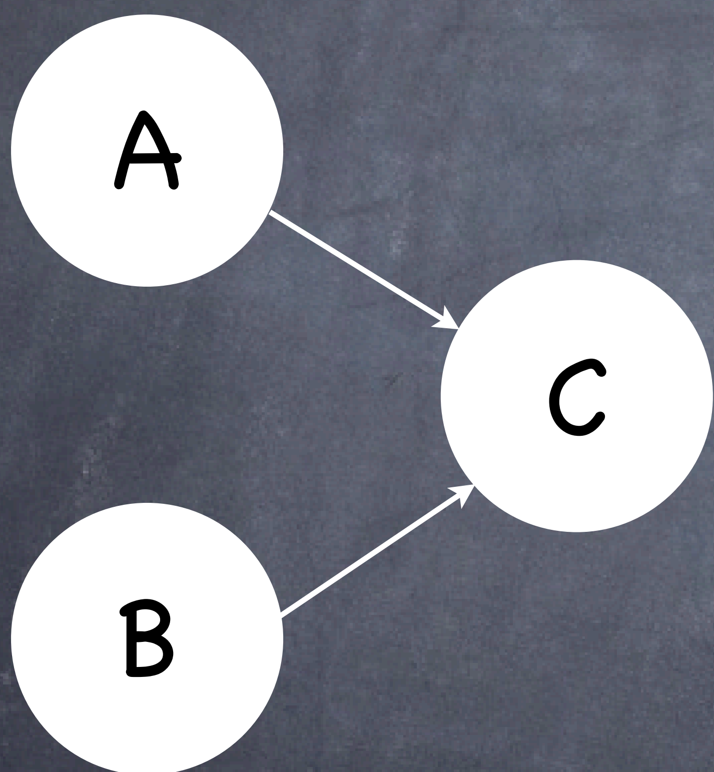
K(B)	K(C|B)	K(C|A,B) K(A)
bucket B	bucket C	bucket A

Example

ordering: A, B, C

recursion 2:

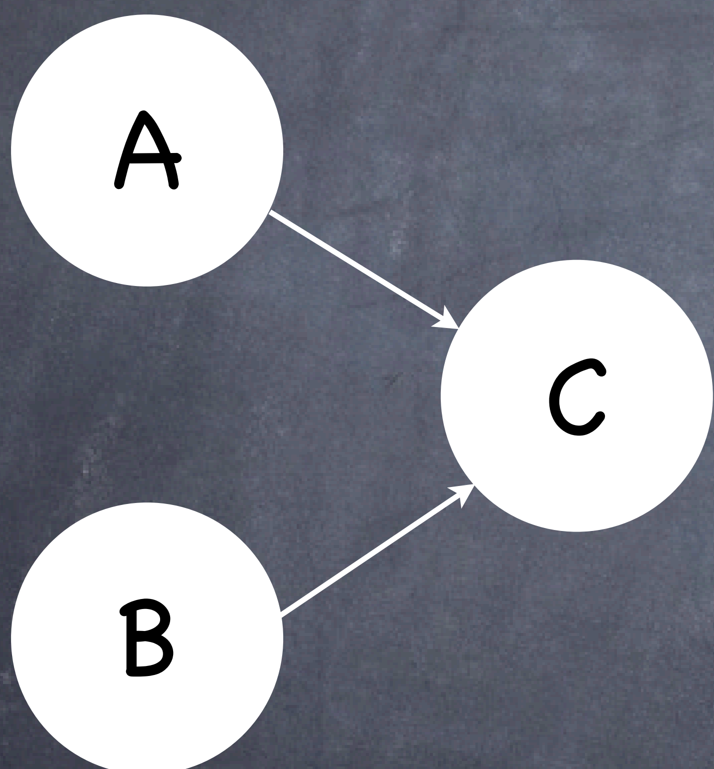




Example

ordering: A, B, C

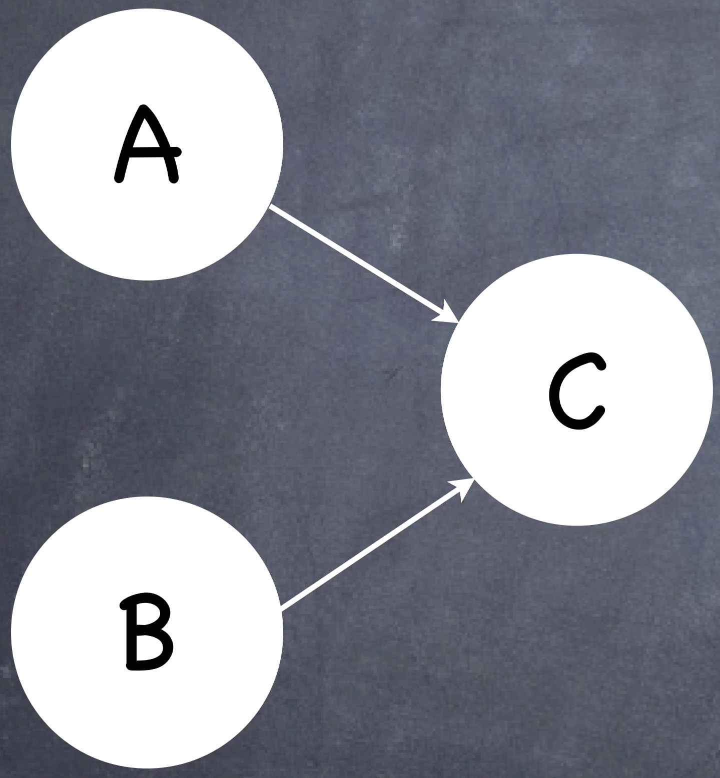
recursion 3:



bucket B	bucket C	bucket A
$K(B)$ $K(C\|B)$	$K(C\|B)$	$K(C\|A,B)$ $K(A)$

$$K(C)=\sum_B K(B)K(C|B)$$

Example



- Final computation:
  - $K(C)=\sum_B K(B)\sum_A K(A)K(C|A,B)$
  - $|K(C)| = |K(B)|\times|K(A)|\times|K(C|A,B)|$!!!

Inference

- Computing the marginal credal set is exponential in the number of potentials in K.
- Usually we are interested only in the extremes, not in the full credal set:
  - min p(q) = min K(q)
  - max p(q) = max K(q)

Pareto Dominance

- We say that a potential P1(X) Pareto dominates a potential P2(X) iff
  - $P1(x) \geq P2(x)$, for all x, and
  - $P1(x) > P2(x)$, for some x.
- Notation: $P1(X) > P2(X)$.

Pareto Set

- Given a credal set K(X)
- The Pareto set PS(X) is the set of non-dominated potentials in K(X)
  - PS(X) := { P(X) in K(X): there is no P'(X) in K(X) such that P'(X) > P(X) }.
  - $|PS(X)| \leq |K(X)|$.

Upper Probability

- The maxima can only be obtained at non dominated potentials:
  - max p(q) = max K(q) = max PS(q)
- Equivalently for min p(q) with a small modification in Pareto dominance.

Results

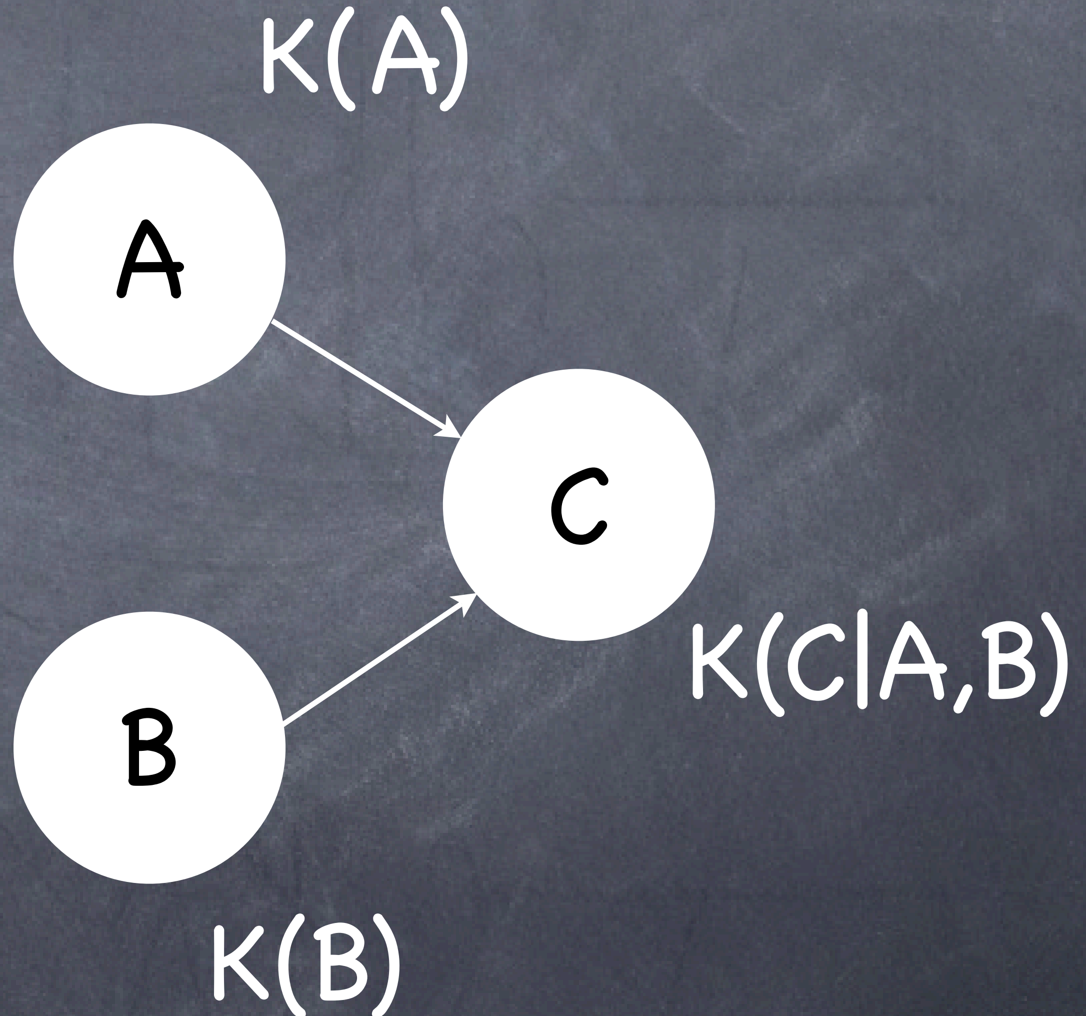
- Let $K(X) = \sum_Z K(XZ)K(YZ)$.
- **Theorem**. $PS(X) = PS(\sum_Z PS(XZ)PS(YZ))$
  - The "Pareto operation" is distributive.
- We can use this to reduce the size of the credal set propagated during bucket elimination.

Bucket Elimination

- Assume an ordering X1,...,Xn of the variables in X.
- Start with an ordered partition bucket 1,...,bucket n of K.
  - bucket i contains all credal sets whose highest variable is Xi **instantiated** at q.
- Recursion: for i := n to 1, do:
  - Compute $PS(U|V) = PS(\sum_{Xi \setminus Q} \prod_{\text{bucket } i} Kj)$ and add it to the largest-index variable bucket.

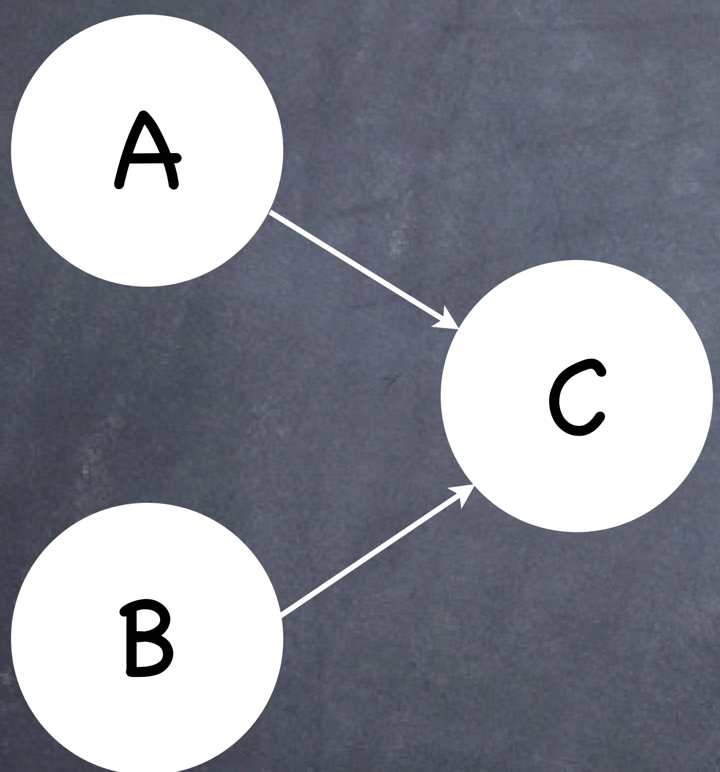
Example

- G = ({A,B,C},{(A,C),(B,C)})
- K = {K(A),K(B),K(C|A,B)}
- max K(c) = max PS(c) ?



Example

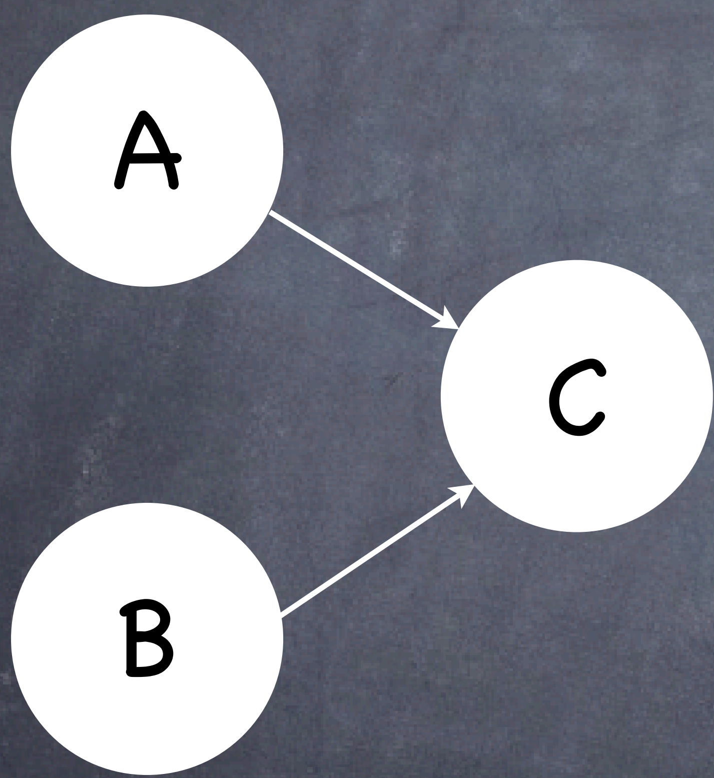
ordering: B, C, A



initialization:

bucket B	bucket C	bucket A
K(B)		K(c|A,B) K(A)

Example



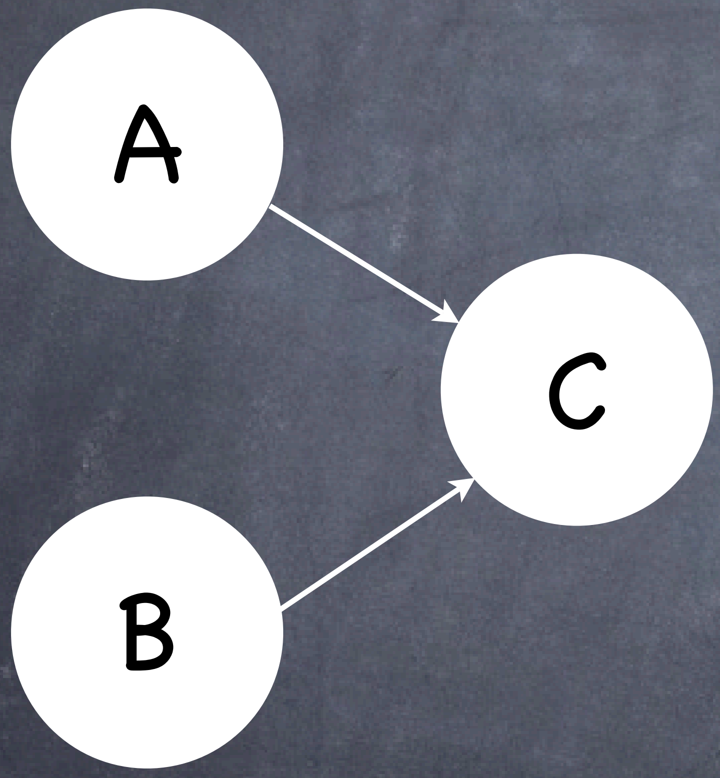
ordering: B, C, A

recursion 1:

$PS(c|B) = PS(\sum_A K(A) K(c|A,B))$

K(B)	PS(c|B)	K(c|A,B) K(A)
bucket B	bucket C	bucket A

Example



ordering: A, B, C

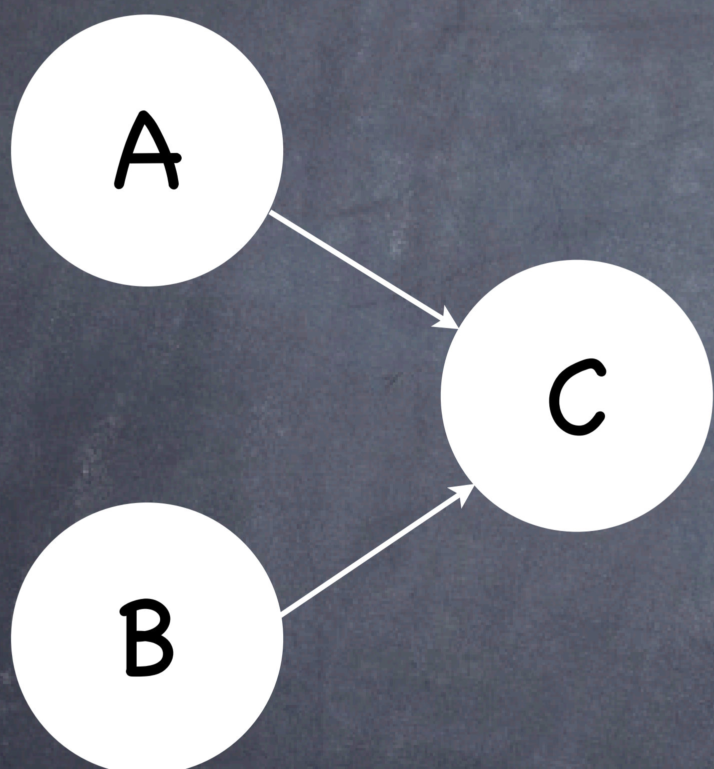
recursion 2:

$\varnothing$

bucket B	bucket C	bucket A
K(B) PS(c|B)	PS(c|B)	K(c|A,B) K(A)

Example

ordering: A, B, C

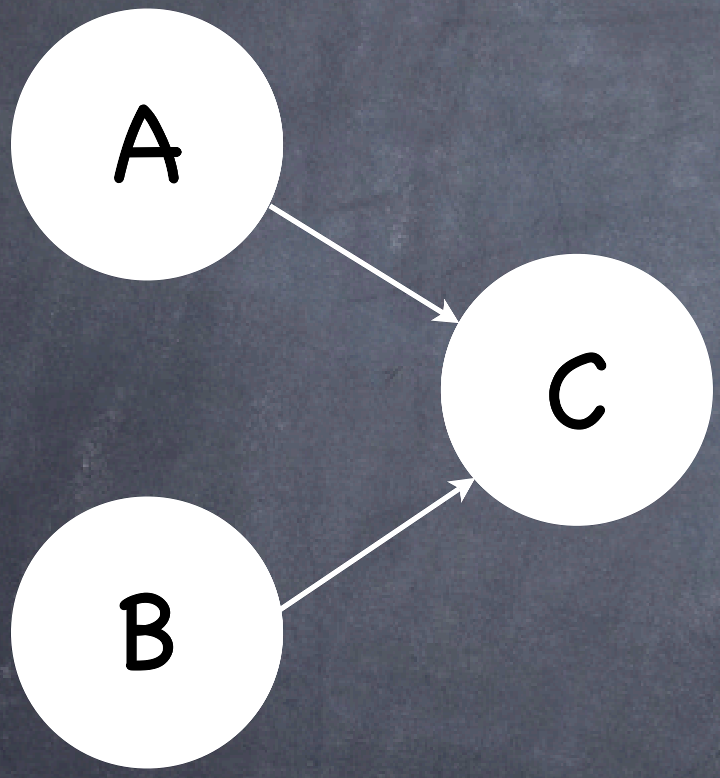


recursion 3:

bucket B	bucket C	bucket A
$K(B)$ $PS(c\|B)$	$PS(C\|B)$	$K(C\|A,B)$ $K(A)$

$$PS(c) = PS\left(\sum_B K(B)PS(c|B)\right)$$

Example



- Final computation:
  - $PS(c) = PS(\Sigma_B K(B) PS(\Sigma_A K(A) PS(c|A,B)))$
    $= PS(\Sigma_B K(B) \Sigma_A K(A) K(c|A,B))$
  - $|PS(c)| \leq |K(B)| \times |K(A)| \times |K(C|A,B)|$

Inference

- Worst-case running time is still exponential.
- Preliminary experiments show algorithm to be efficient in practice.
- Conjecture: algorithm has polynomial expected running time (based on properties of Pareto sets).

ε-Pareto Dominance

- We say that a potential P1(X) ε-Pareto dominates a potential P2(X) iff
  - $P1(x) \geq (1+\varepsilon)P2(x)$, for all x,
  - $\varepsilon > 0$.
- Notation: $P1(X) >_{\varepsilon} P2(X)$.
- P1(X) almost dominates P2(X).

ε-Pareto Set

- Given a credal set K(X)
- The ε-Pareto set $PS_ε(X)$ is a subset of K(X) such that for each potential P'(X) in K(X) not in $PS_ε(X)$ there is some potential in $PS_ε(X)$ that ε-Pareto dominates P'(X).
  - $PS_ε(X)$:= { P(X) in K(X): there is no P'(X) in K(X) such that $P'(X) >_ε P(X)$ for all P(X)}.
  - $|PS_ε(X)| \leq |K(X)|$.

Results

- Let $K(X)=K(Y)K(Z)$.
- **Theorem**. There is a $PS_\varepsilon(X)$ that is size polynomial in $|K(Y)|$ and $|K(Z)|$ and in $1/\varepsilon$ (but not in $|X|$).
- Given a credal set, we can construct an ε-pareto set in polynomial time.

Bucket Elimination

- Assume an ordering X1,...,Xn of the variables in X.
- Start with an ordered partition bucket 1,...,bucket n of K.
  - bucket i contains all credal sets whose highest variable is Xi **instantiated** at q.
- Recursion: for i := n to 1, do:
  - Compute $PS_\varepsilon(U|V) = PS_\varepsilon(\sum_{Xi \backslash Q} \prod_{\text{bucket } i} Kj)$ and add it to the largest-index variable bucket.

Results

- **Theorem**. Bucket elimination with ε-Pareto set propagation is an FPTAS.
- $\max p(q) \leq (1+\varepsilon) \max PS_{\varepsilon}(q)$

Future

- Experiments (w/ both exact and approx.).
- Complexity results for exact.
- Selecting good orderings (not like the Bayesian case).
- Different queries (maximality, e-admissibility, maximin).

Questions?