Learning AMP Chain Graphs under Faithfulness

Jose M. Peña
ADIT, IDA, Linköping University, Sweden

PGM 2012, 19-21 September 2012

AMP Chain Graphs

- A graph G containing (possibly) directed and undirected edges is a chain graph (CG) if there is a partition $\mathcal{K} = \{K_1, \ldots, K_n\}$ of the nodes of G st
  - if $A \to B$ is in G, then $A \in K_i$ and $B \in K_j$ with $1 \leq i < j \leq n$, and
  - if $A - B$ is in G, then $A, B \in K_i$ with $1 \leq i \leq n$.
- A node B in a route ρ is called a head-no-tail node in ρ if $A \to B \leftarrow C$, $A \to B - C$, or $A - B \leftarrow C$ is a subroute of ρ (note that maybe $A = C$ in the first case).
- Given three disjoint subsets of nodes X, Y and Z, we say that X is separated from Y given Z in G when there is no route ρ between a node in X and a node in Y st
  - every head-no-tail node in ρ is in Z, and
  - every other node in ρ is not in Z.
- The independence model induced by a CG G, $I(G)$, is the set of separations in G.
- A probability distribution p is faithful to a CG G iff every independence in p corresponds to a separation in G and vice versa.

Why AMP Chain Graphs ?

- Present interpretation: Andersson-Madigan-Perlman (AMP) CGs.
- Other interpretations: Lauritzen-Wermuth-Frydenberg (LWF) CGs, and multivariate regression (MVR) CGs by Cox and Wermuth.
- Reason 1: No interpretation subsumes any other.
- For every AMP CG G, there is a probability distribution that is faihtful to G.
- Reason 2: Every AMP CG represents a probabilistic independence model (also true for LWF and MVR CGs).
- Every AMP CG G has associated a system of linear equations with normally distributed errors as follows:

  For every $K_i \in \mathcal{K}$

  * $K_i = \beta_i \, pa_G(K_i) + \epsilon_i$

  subject to the following constraints:

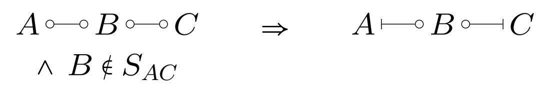
  * If $s \to t$ is not in G, then $(\beta_i)_{st} = 0$, and
  * $\epsilon_i \sim \mathcal{N}(0, \Sigma_i)$ st if $s - t$ is not in G, then $(\Sigma_i^{-1})_{st} = 0$.

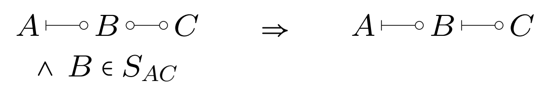
- Reason 3: In the Gaussian framework, AMP CGs specify a direct mode of data generation (also true for MVR CGs but not for LWF CGs).

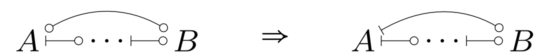
The Learning Algorithm

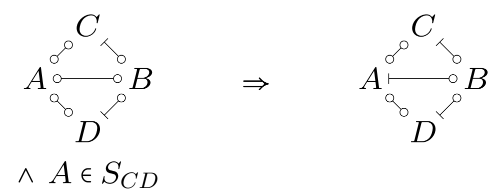
Input: A probability distribution p that is faithful to an unknown CG G.
Output: A CG H st $I(H) = I(G)$.

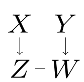
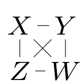
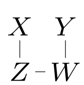
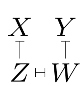
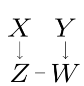
1 Let H denote the complete undirected graph
2 Set $l = 0$
3 Repeat while $l \leq |V| - 2$
4 For each ordered pair of nodes A and B in H st $A \in ad_H(B)$ and $|[ad_H(A) \cup ad_H(ad_H(A))] \setminus B| \geq l$
5 If there is some $S \subseteq [ad_H(A) \cup ad_H(ad_H(A))] \setminus B$ st $|S| = l$ and $A \perp_p B | S$ then
6 Set $S_{AB} = S_{BA} = S$
7 Remove the edge $A - B$ from H
8 Set $l = l + 1$
9 Apply the rules R1-R4 to H while possible
10 Replace every edge $\vdash$ ($\mapsto$) in H with $\rightarrow$ ($-$)

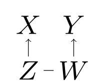
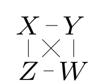
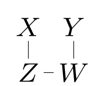
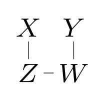
R1:


R2:


R3:


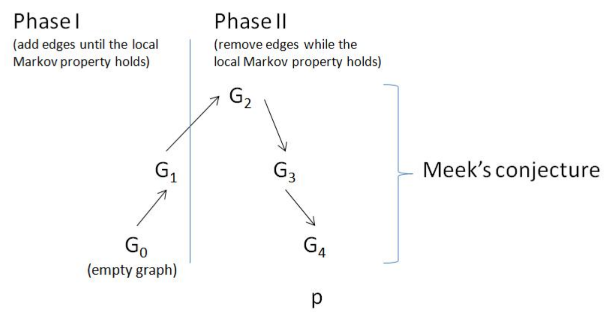
R4:


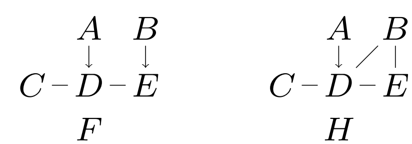
p, G		
H		by line 1
H		by line 5, $S_{XY} = S_{YX} = S_{XW} = S_{WX} = S_{YZ} = S_{ZY} = \emptyset$
H		by R1
H		by line 10

p, G		
H		by line 1
H		by line 5, $S_{XY} = S_{YX} = S_{XW} = S_{WX} = Z$, $S_{YZ} = S_{ZY} = W$
H		by line 10

Future Work

- Relax the faithfulness assumption ? Replace it with the composition property assumption ? This would compromise the development of correct and efficient score+search learning algorithms, because Meek's conjecture does not hold for AMP CGs (it does hold for LWF CGs) as the following example illustrates.





$I(F) \supseteq I(H)$

- Replace R3: $A \,\circ\!\!\!\multimap \cdots \multimap\!\!\circ\, B \Rightarrow A \vdash\!\circ \cdots \vdash\!\circ\, B$ by R3: $A \multimap B \multimap C \Rightarrow A \vdash\!\circ B \vdash\!\circ C$?

- Restrict R3 to chordless cycles ? Thanks to Reviewer 2.

- Marginal AMP CGs, which have undirected, directed and bidirected edges.