Estimating and visualizing high-dimensional dependence structures with the PC-algorithm

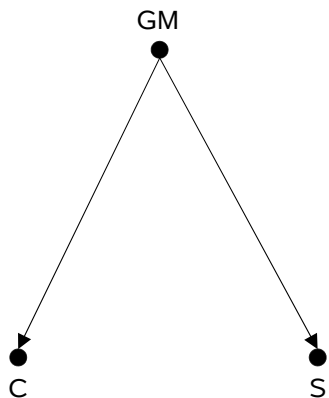Markus Kalisch

ETH Zürich

Switzerland

Overview

- Directed Acyclic Graph (DAG) and its skeleton
- The PC-algorithm for finding the skeleton is consistent
- R-package: pcalg

Directed Acyclic Graphs (DAGs)

- Nodes: Random Variables
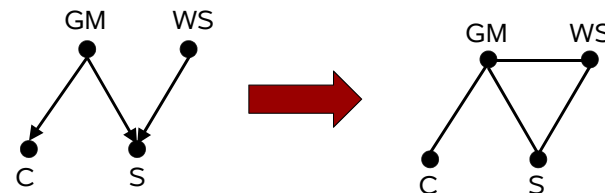- Edges: Some Dependence
- Recursive factorization:
  f(GM,C,S) =
  f(GM) f(C|GM) f(S|GM)

Directed Global Markov Property

- DAG implies conditional independence relations
- $C \perp S | GM \iff$ C,S are separated by GM in $\left(G_{An(C \cup S \cup GM)}\right)^m$
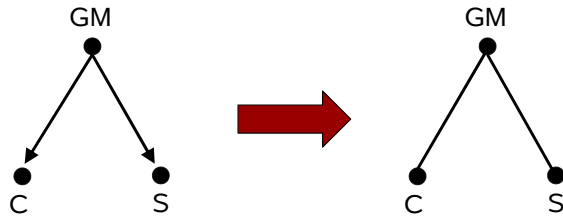

- Ancestral set
- Moralize
- Drop directions

Skeleton of a DAG

- Ignore directions of arrows
- Edge between two nodes A and B $\Longleftrightarrow$
  A, B are dependent given every subset of
  remaining nodes

PC-algorithm for finding the skeleton

Form complete graph G
l = -1
repeat
       l=l+1
       repeat
              select ordered pair of adjacent nodes A,B in G
              select neighborhood N of A with size l (if possible)
              delete edge A,B in G if A,B are cond. indep.
              given N
       until all ordered pairs have been tested
until all neighborhoods are of size smaller than l

Main Result

- Test Cond. Indep. Relations $A \perp B|S$
  consistently

- PC-Algorithm is consistent
  (more detailed results when using
  assumptions)

R-package: pcalg

- Estimate the skeleton given a data matrix
- Visualize the estimated skeleton