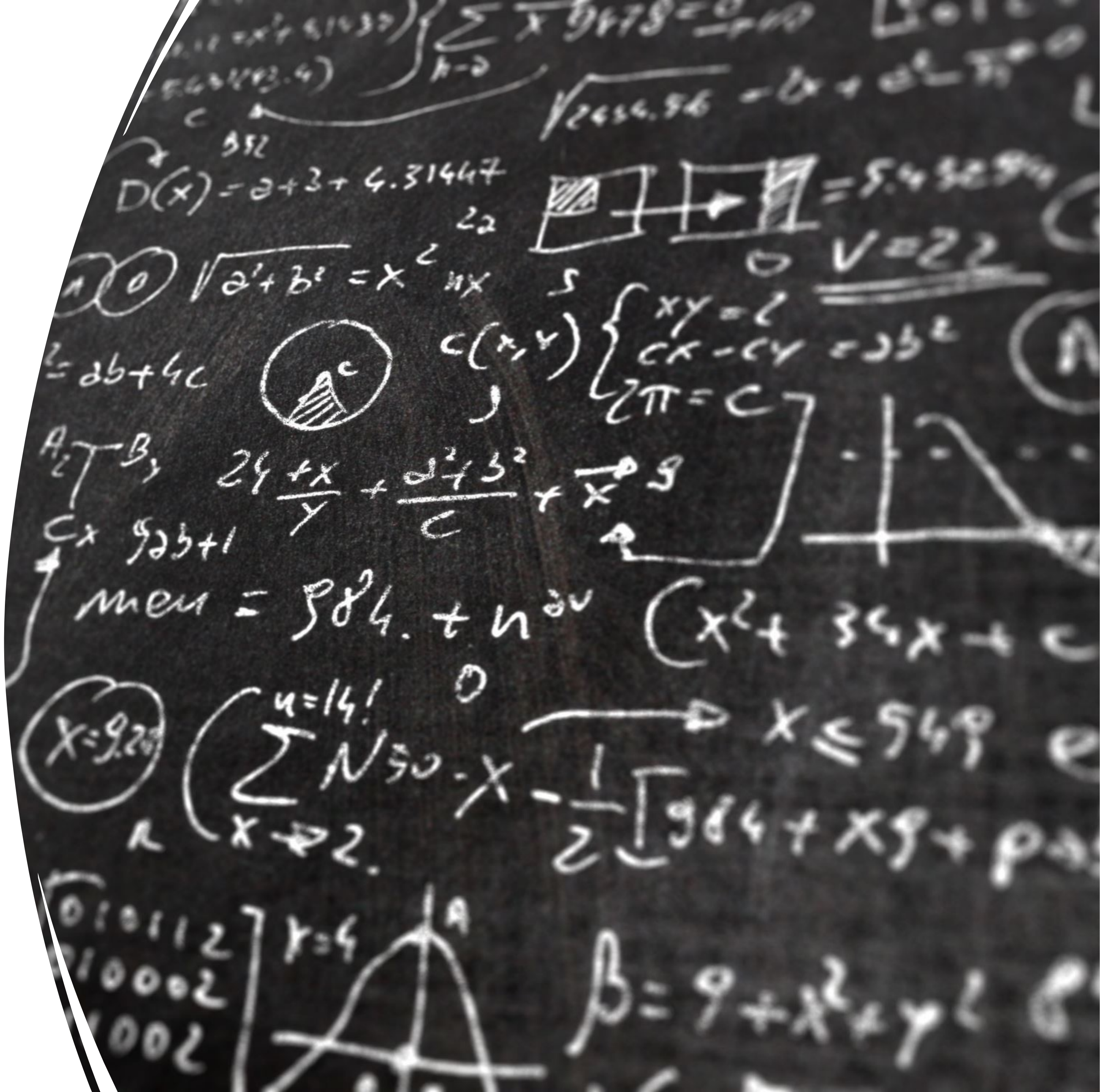
Poster for Causal DAG

Chang Li

Contents

- Introduction of DAG
- PC Algorithm
- Markov equivalence class (MEC)



DAGs

- A DAG is a graph that provides a visual representation of causal relationships among a set of variables.
- These causal relationships are either known to be true of more commonly are only assumed to be true.
- The direction o f the arrow is the direction of causation: A -> B means A causes B
- A = acyclic (no sequence of arrows forms a closed loop, which would be backwards causation)
- ***They can be used to determine if a given pair of variables are independent***.

Unblocked paths

- An unblocked front-door path from X to Y starts like this X ->
- An unblocked back-door path from X to Y starts like this X <-

- Unblocked front-door paths from X to Y are causal
- Unblocked back-door paths from X to Y are confonding

Blocking and Independence rules

- A path is blocked by a variable C if it looks like this
  - -> C <-
  - C is called a “collider”
  - This path is unblocked conditional on C
- A path is not blocked by a variable C if it looks like an of these
  - ->C->
  - <-C->
  - <-C<-
  - These paths are blocked conditional on C
- Two variables are correlated (also called “dependent" or “not dependent”) if there is one or more unblocked path between them
- Here “conditional on C” effectively mean include in a regression model

PC algorithm

- Start with complete undirected graph
- For each X and Y, see if $X \perp Y$; if so, remove their edge
- For each X and Y which are still connected, and each third variable Z_1, see if $X \perp Y|\, Z$; if so, remove the edge between X and Y.
- For each X and Y which are still connected, and each third and fourth variable $Z_1$and Z_2, see if $X \perp Y|\, Z_1, Z_2$; if so, remove the edge between X and Y.
- ...
- For each X and Y which are still connected, see if $X \perp Y|all\ the\ p-2\ other\ variables$; if so, remove the edge between X and Y.

- ***PC algorithm assumes no latent variables and selection variables***.

Conditional independence

- In the above PC algorithm, the conditional independence oracle is replaced by a statistical test for conditional independence.
- For situations without hidden variables and under some further conditions it has been shown that the PC algorithm using statistical tests instead of an independence oracle is computationally feasible and consistent even for very high-dimensional sparse DAGs

pcalg in R

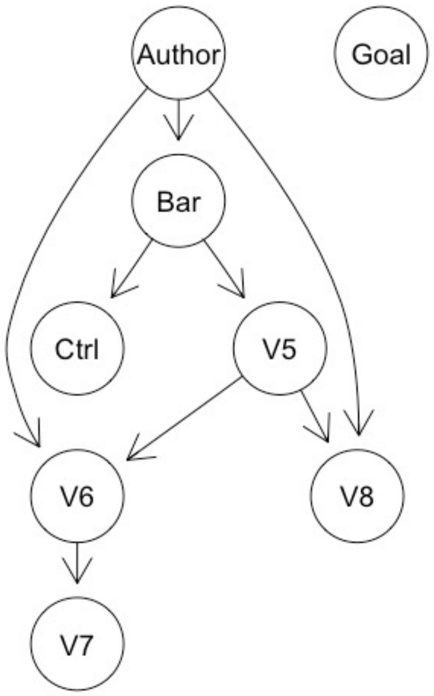
- Simulation

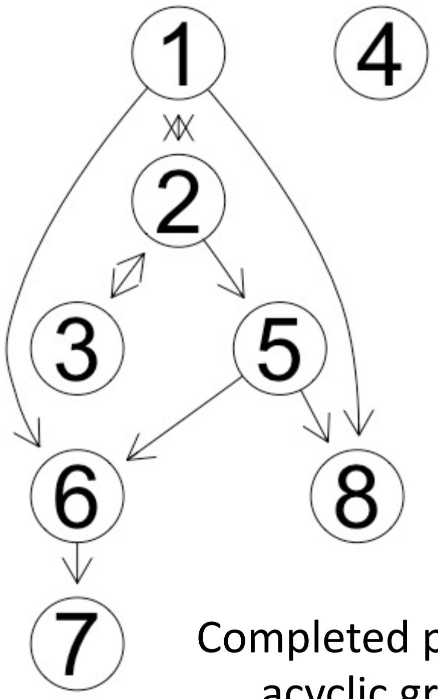
```
library(pcalg)
data("gmG")

suffStat <- list(C = cor(gmG8$x), n = nrow(gmG$x))
pc.gmG <- pc(suffStat, indepTest = gaussCItest, p = ncol(gmG8$x), alpha = 0.005)
stopifnot(require(Rgraphviz))
par(mfrow = c(1,2))
plot(gmG8$g, main = "")
plot(pc.gmG, main = "")
```

- The dataset in pcalg is an example dataset with p=8 continuous variables with Gaussian noise and n=5000 observations.

pcalg in R





Completed partially directed acyclic graph (CPDAG)

- Kalisch, Mächler, Hauser Colombo, Bühlmann Maathuis. More Causal Inference with Graphical Models in R Package Pcalg

Markov equivalence class (MEC)

- It was shown that two DAGs encode the same conditional independence statements if and only if the corresponding DAGs have the same skeleton and the same v-structures.
- Such DAGs are called Markov-equivalent.
- In this way, the space of DAGs can be partitioned into equivalence information. Conversely, if given a conditional independence oracle, on can only determine a DAG up to its equivalence class.
- Therefore, the PC algorithm cannot determine the DAG uniquely, but only the corresponding equivalence class of the DAG.

Removing Assumptions

- No assumed causal sufficiency: FCI algorithm (Spirtes et al., 2001)
- No assumed acyclicity: CCD algorithm (Richardson, 1996)
- ...