MLCS Journal Club 27.7.21

Latent PCMCI

Moritz Haas

University of Tübingen

machine learning in climate science

Content

Recap: PCMCI

1

- Fundamental Idea
- Examples

LPCMCI

2

- Setting
- Idea
- Algorithm

Conclusion

3

- Advances
- Weaknesses

I. Recap PCMCI:

Fundamental Idea and Example

Recap: PCMCI

Goal: learn causal graph of given variables with conditional independence tests

- FullCI has low detection power and estimates low effect size (too many conditionals)
- PC (Peter-Clarke) has high false positive rate for high autocorrelation (broken iid assumption)

→ PCMCI:

- First PC: Remove irrelevant conditions → learn superset of parents $\hat{\mathscr{P}}(X_j)$ for each variable
- Then MCI (Momentary Conditional Independence):

$$\text{MCI:} \quad X^i_{t-\tau} \perp\!\!\!\perp X^j_t \mid \widehat{\mathscr{P}}(X^j_t) \setminus \{X^i_{t-\tau}\}, \widehat{\mathscr{P}}(X^i_{t-\tau}).$$

Establish cond. independence

Reduce false positive rate

Source: https://advances.sciencemag.org/content/5/11/eaau4996

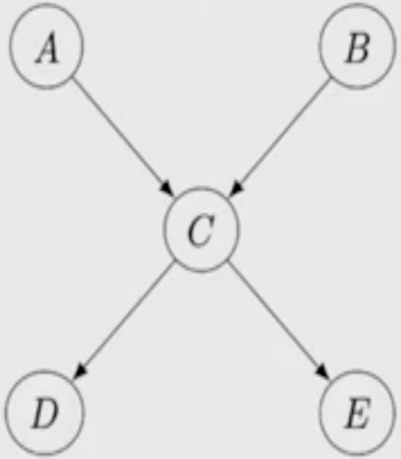
PC Example

Start with complete undirected graph

Three steps:

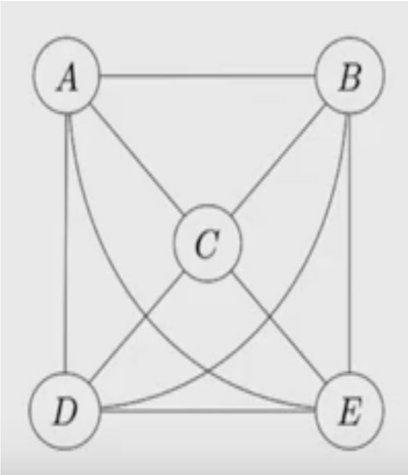
1. Identify the skeleton
2. Identify immoralities and orient them
3. Orient qualifying edges that are incident on colliders

True graph

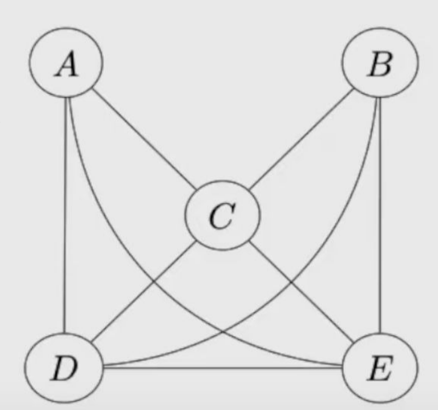


Source: https://www.youtube.com/watch?v=o2A61bJ0UCw

PC Example: Identify the skeleton

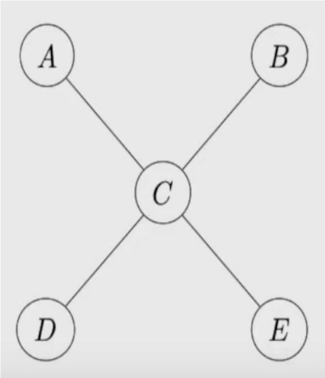


$$A \perp B \mid \{\} \longrightarrow$$



$$X \perp Y \mid \{C\} \longrightarrow$$

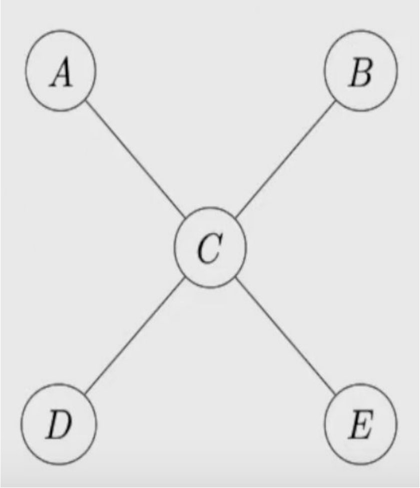
$$\forall\, other\, pairs\, (X, Y)$$

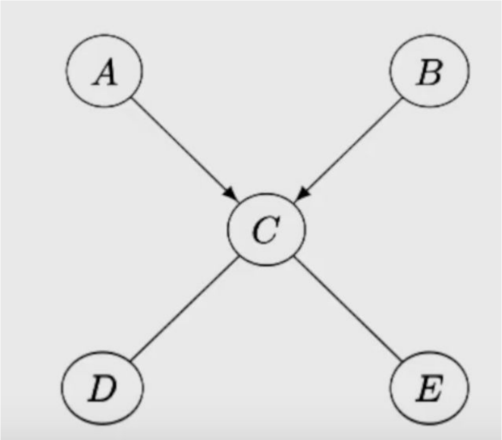


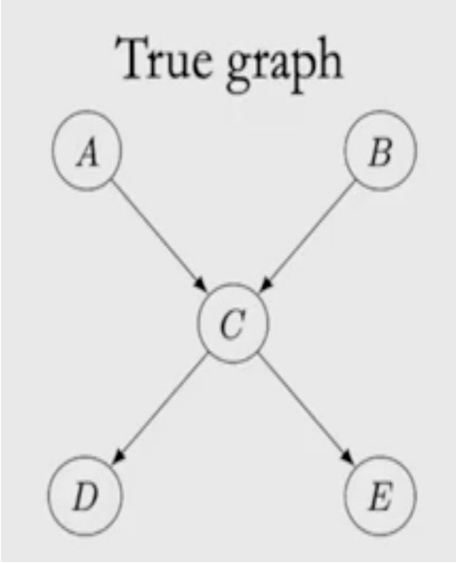
PC Example: Orient Edges

$A \perp B \mid \{\}$ and $X \perp Y \mid \{C\}$ $\forall\, other\, pairs\, (X, Y)$

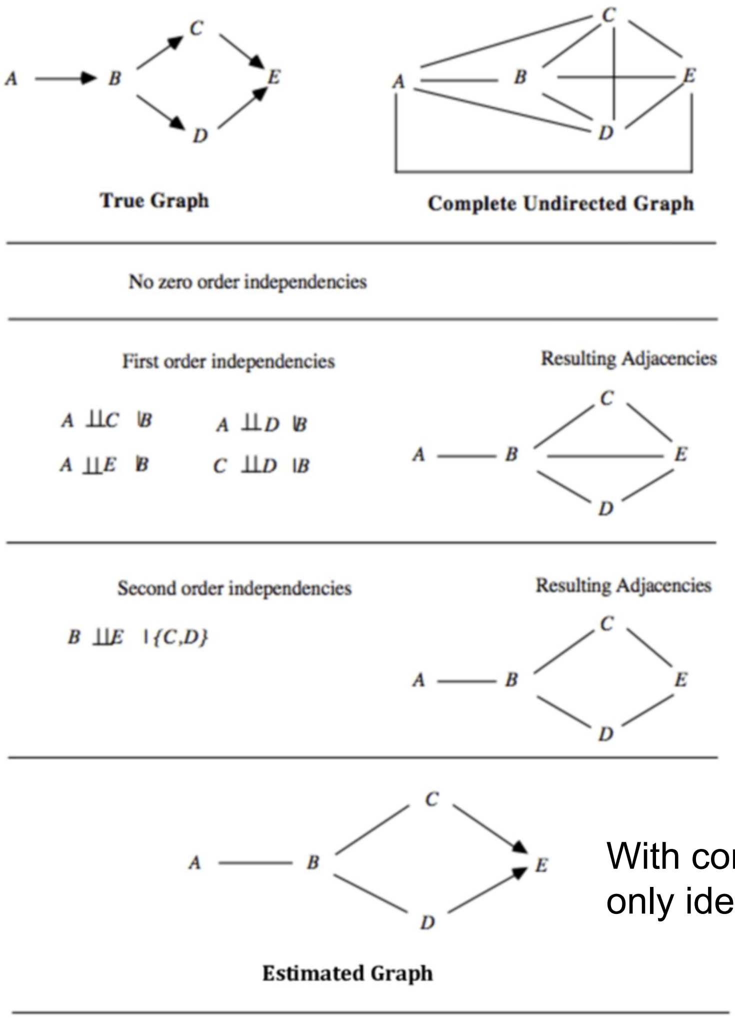
Look at X - Y - Z, where $X \perp Z \mid$ S and Y not in S







We have found all colliders



With conditional independence tests we can only identify the Markov equivalence class

Source: https://www.researchgate.net/figure/An-example-for-PC-algorithm-Figure-from-Spirtes-et-al-2000-Chapter-5_fig7_321828295

I. Latent PCMCI:

Setting, Idea and Algorithm

LPCMCI: Setting

<u>Assumptions:</u> Faithfulness, Causal Markov Condition, Causal Stationarity

$$V_t^j = f_j(pa(V_t^j), \eta_t^j) \quad \text{with } j = 1, \ldots, \tilde{N} \, .$$

Only observe $\{X_1, \ldots, X_N\} \subseteq \{V_1, \ldots, V_{\tilde{N}}\}$

$\mathscr{M}(\mathscr{G})$ Maximal Ancestral Graph (MAG) after marginalizing over unobserved variables and all t'<t - τ_{max}

$\mathscr{P}(\mathscr{G})$ Maximally informative Partial Ancestral Graph, which can be obtained by running FCI orientation rules (enforce time order and repeating adj.)

→ May include contemp. links and unobs. confounders

LPCMCI: Idea

Detection power of true link depends on

- Sample size
- CI tests' sign. level
- CI tests' estim. dimension
- Effect size $\min_S I(A, B|\, S)$ over used cond. sets S

Problem often: low signal-to-noise ratio (e.g. autocorrelation) due to unfortunate choice of conditioning sets leading to low effect size,
Leads to missing links, which lead to false positives and wrong orientations

→ Carefully choose conditioning sets:

- Discard known non-ancestors
- Include known ancestors ALWAYS

by entangling edge removal and orientation phase → multiple iterations k, collecting knowledge

Algorithm 1 LPCMCI

Require: Time series dataset $\mathbf{X} = \{\mathbf{X}^1, \dots, \mathbf{X}^N\}$, maximal considered time lag $\tau_{\max}$, significance level α, CI test $\mathrm{CI}(X, Y, \mathcal{S})$, non-negative integer k

1: Initialize $\mathcal{C}(\mathcal{G})$ as complete graph with $X^i_{t-\tau} \circ\!\xrightarrow{L} X^j_t$ $(0 < \tau \le \tau_{\max})$ and $X^i_{t-\tau} \circ\!\overset{?}{-}\!\circ X^j_t$ $(\tau = 0)$
2: **for** $0 \le l \le k-1$ **do**
3: Remove edges and apply orientations using Algorithm $\boxed{\text{S2}}$
4: Repeat line 1, orient edges as $X^i_{t-\tau} \xrightarrow{?} X^j_t$ if $X^i_{t-\tau} \overset{*}{\rightarrow} X^j_t$ was in $\mathcal{C}(\mathcal{G})$ after line 3
5: Remove edges and apply orientations using Algorithm $\boxed{\text{S2}}$
6: Remove edges and apply orientations using Algorithm $\boxed{\text{S3}}$
7: **return** PAG $\mathcal{C}(\mathcal{G}) = \mathcal{P}(\mathcal{G}) = \mathcal{P}(\mathcal{G})^{\tau_{\max}}_{statAO}$

After line 5, all links are correctly removed for pairs of variables where one is an ancestor of the other

- Introduce node ordering and 'middle marks' to save what is known about the link
- There exist several orientation rules for orienting and deleting edges
- Iterate until all middle marks have been removed or reached hyperparam. k

→ In practice for finite samples, larger k does not imply better results :((see river example)

Algorithm S2 Ancestral removal phase

Require: LPCMCI-PAG $\mathcal{C}(\mathcal{G})$, memory of minimal test statistic values $I^{\min}(\cdot,\cdot)$, memory of separating sets $\mathrm{SepSet}(\cdot,\cdot)$, time series dataset $\mathbf{X} = \{\mathbf{X}^1, \dots, \mathbf{X}^N\}$, maximal considered time lag $\tau_{\max}$, significance level α, CI test $\mathrm{CI}(X, Y, \mathcal{S})$

1: **repeat** starting with $p = 0$
2: **for** $-1 \le m \le \tau_{\max}$ **do**
3: **for all** ordered pairs of variables $(X^i_{t-\tau}, X^j_t)$ adjacent in $\mathcal{C}(\mathcal{G})$ with $X^i_{t-\tau} < X^j_t$ **do**
4: **if** $(m = -1$ and $i \ne j)$ or $(m \ge 0$ and $\tau \ne m$ or $i = j)$ **then** continue with next pair
5: $\mathcal{S}_{def} = pa(\{X^i_{t-\tau}, X^j_t\}, \mathcal{C}(\mathcal{G}))$
6: **if** the middle mark is '?' or 'L' **then**
7: $\mathcal{S}_{search} = apds_t(X^j_t, X^i_{t-\tau}, \mathcal{C}(\mathcal{G})) \setminus \mathcal{S}_{def}$, ordered according to $I^{\min}(X^j_t, \cdot)$
8: **if** $|\mathcal{S}_{search}| < p$ **then** update middle mark with 'R' according to Lemma $\boxed{\text{S8}}$
9: **for all** subsets $\mathcal{S} \subseteq \mathcal{S}_{search}$ with $|\mathcal{S}| = p$ **do**
10: $(p\text{-value}, I) \leftarrow \mathrm{CI}(X^i_{t-\tau}, X^j_t, \mathcal{S} \cup \mathcal{S}_{def})$
11: $I^{\min}(X^i_{t-\tau}, X^j_t) = I^{\min}(X^j_t, X^i_{t-\tau}) = \min(|I|, I^{\min}(X^i_{t-\tau}, X^j_t))$
12: **if** p-value $> \alpha$ **then**
13: mark edge for removal, add $\mathcal{S} \cup \mathcal{S}_{def}$ to $\mathrm{SepSet}(X^i_{t-\tau}, X^j_t)$
14: break innermost for-loop
15: repeat lines 6 - 14 with $X^i_{t-\tau}$ and X^j_t as well as 'R' and 'L' swapped
16: remove all edges that are marked for removal from $\mathcal{C}(\mathcal{G})$
17: **if** any edge has been removed in line 16 **then**
18: run Alg. $\boxed{\text{S4}}$ using [APR, MMR, $\mathcal{R}8'$, $\mathcal{R}2'$, $\mathcal{R}1'$, $\mathcal{R}9'$, $\mathcal{R}10'$], orient lagged links only
19: let $p = 0$
20: **else** increase p to $p+1$
21: **until** there are no other middle marks than '!' or '' (empty)
22: run Alg. $\boxed{\text{S4}}$ using [APR, MMR, $\mathcal{R}8'$, $\mathcal{R}2'$, $\mathcal{R}1'$, $\mathcal{R}0'd$, $\mathcal{R}0'c$, $\mathcal{R}3'$, $\mathcal{R}4$, $\mathcal{R}9'$, $\mathcal{R}10'$, $\mathcal{R}0'b$, $\mathcal{R}0'a$]
23: **return** $\mathcal{C}(\mathcal{G})$, $I^{\min}(\cdot,\cdot)$, $\mathrm{SepSet}(\cdot,\cdot)$

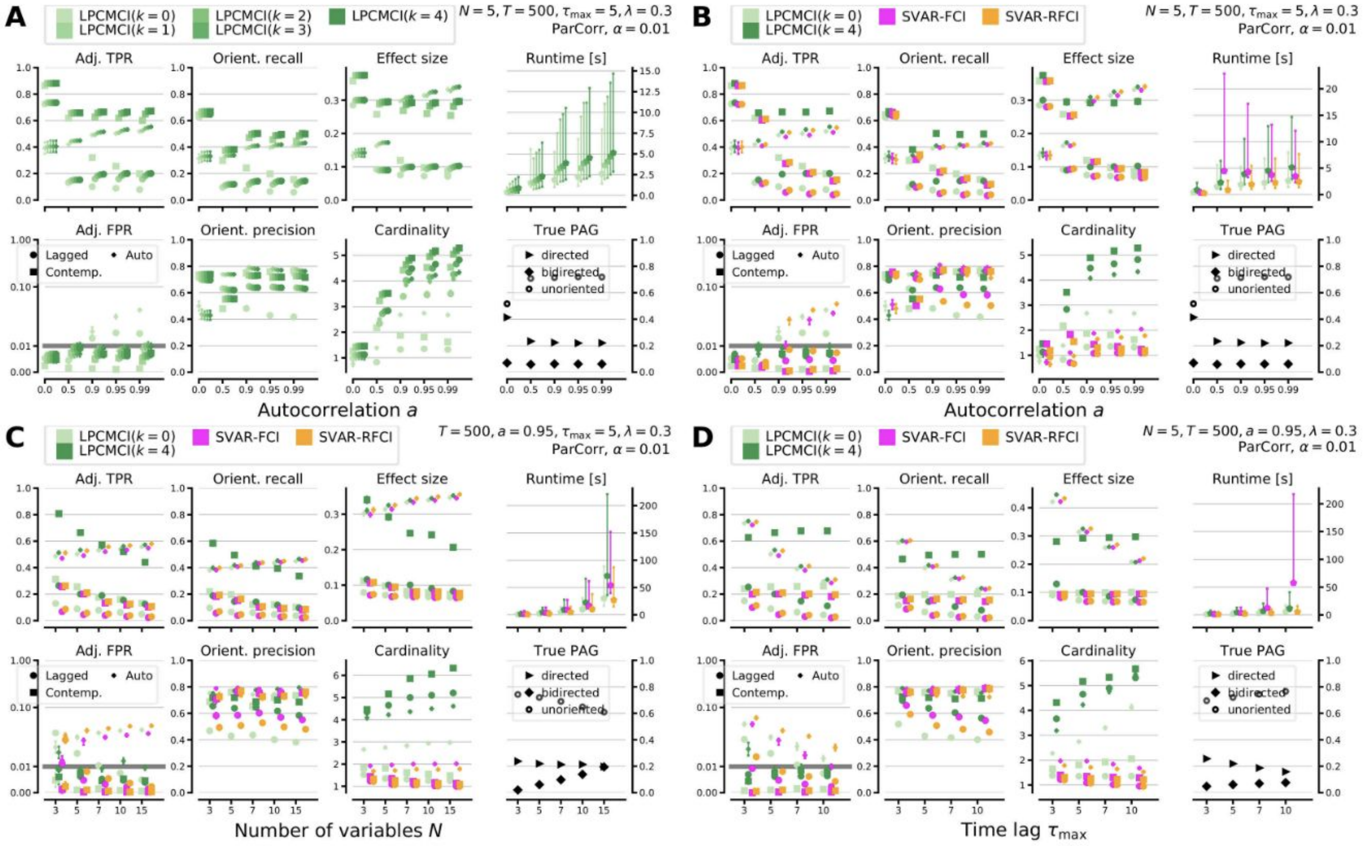


Figure 2: Results of numerical experiments for (**A**) LPCMCI(k) for different k, LPCMCI compared to SVAR-FCI and SVAR-RFCI for (**B**) varying autocorrelation, for (**C**) number of variables N, and for (**D**) maximum time lag $\tau_{\max}$ (other parameters indicated in upper right of each panel).

A) Improves with larger k if assumptions hold

B-D) LPCMCI better for contemp. links for k large enough

Controls FPR

Conclusion: LPCMCI

- Like PCMCI for setting with contemporaneous links and confounders

  → More general and realistic
- Idea: Always condition on known parents
- Uses middle marks and complicated link orientation and removal rules
- Higher recall for highly autocorr. time series
- order-independent, returns optimal PAG for infinite sample size
- additional hyperparam. k (number of iterations)
- no finite sample guarantees (conv. rate arbitrarily slow for almost unfaithful distr.)

LPCMCI: https://arxiv.org/abs/2007.01884

PCMCI: https://advances.sciencemag.org/content/5/11/eaau4996

Thank you!

machine learning in climate science

Important Realisations

- Weak autocorrelation, better recall
- Correct false positive rates in MCI step
- Assumptions: Causal Markov Condition, Faithfulness, causal stationarity, no contemporaneous causal links, no hidden variables
  Then: consistent
- LPCMCI: Causal Markov Condition, Faithfulness, causal stationarity
- Findings of non-causality hold for weaker assumptions (faithfulness, test flexible enough), thus more reliable
- Fast PC: only choose conditions with highest association
- Choose tau_max large, only longer runtime, not higher dim.
- Small alpha means few false positives, but less true positives
- Convergence rate to consistency can be made arbitrarily slow by almost unfaithful distributions
- Autocorrelation breaks iid assumption in tests, thus in ParCorr t distr. Has fewer degrees of freedom hence get more false positives: adjust the degrees of freedom in some way, using pre-whitening, or by block-shuffling. While these approaches help to some extent for the simple bivariate case, they fail in the multivariate case that is relevant for causal discovery https://arxiv.org/abs/1407.0742 (MCI good option)