Marginal consistency of constraint-based causal learning

Anna Roumpelaki, Giorgos Borboudakis, Sofia Triantafillou and Ioannis Tsamardinos
University of Crete, Greece.
Constraint-based Causal Learning

Measure $X_1, \ldots, X_8$

FCI
Constraint-based Causal Learning

Measure $X_1, \ldots, X_8$

<table>
<thead>
<tr>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$X_3$</th>
<th>$X_4$</th>
<th>$X_5$</th>
<th>$X_6$</th>
<th>$X_7$</th>
<th>$X_8$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

FCI
Constraint-based Causal Learning

Measure $X_1, ..., X_8$

Would you have learnt the same relationships if you had chosen a subset of the variables?

How important is the choice of variables in learning causal relationships?

Would you have learnt the same relationships if you had chosen a subset of the variables?
Maximal Ancestral Graphs

- Maximal Ancestral Graphs
  - Capture the conditional independencies of the joint probability $P$ over a set of variables under Causal Markov and Faithfulness conditions

- Directed edges denote causal ancestry.
- Bi-directed edges denote confounding.
- No directed cycles.
- No almost directed cycles $A \rightarrow B \rightarrow \cdots \rightarrow C \leftrightarrow B$.

- Are closed under marginalization.
  - $G = (V, E)$ is a MAG, $G_L = (V \setminus L, E')$ is also a MAG.

- Can also handle selection (not here).

![Diagram of Maximal Ancestral Graphs]

D causes E
confounded
Partially Oriented Ancestral Graphs

- Summarize pairwise features of a Markov equivalence class of MAGs.
- Circles denote ambiguous endpoints.
- Can be learnt from data using FCI.

- FCI:
  - Sensitive to error propagation.
  - Order-dependent (if you change the order of variables in the data set you get a different result).

- Extensions of FCI [4]:
  - Order independent (iFCI) [1]:
    - output does not depend on the order of the variables.
  - Conservative FCI (cFCI) [3]:
    - Makes additional tests of independence for more robust orientations.
    - Forgoes orientations that are not supported by all tests.
  - Majority rule (mFCI) [3]:
    - conservative FCI that orients ambiguous triplets based on majority of test results for each triplet.
What happens when you marginalize out variables?

Measure $X_1, \ldots, X_8$

<table>
<thead>
<tr>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$X_3$</th>
<th>$X_4$</th>
<th>$X_5$</th>
<th>$X_6$</th>
<th>$X_7$</th>
<th>$X_8$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

FCI
What happens when you marginalize out variables?

Measure $X_1, \ldots, X_8$

<table>
<thead>
<tr>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$X_3$</th>
<th>$X_4$</th>
<th>$X_5$</th>
<th>$X_6$</th>
<th>$X_7$</th>
<th>$X_8$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Marginal dataset without variables $X_2, X_5$

<table>
<thead>
<tr>
<th>$X_1$</th>
<th>$X_3$</th>
<th>$X_4$</th>
<th>$X_6$</th>
<th>$X_7$</th>
<th>$X_8$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
What happens when you marginalize out variables?

Measure $X_1, \ldots, X_8$

<table>
<thead>
<tr>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$X_3$</th>
<th>$X_4$</th>
<th>$X_5$</th>
<th>$X_6$</th>
<th>$X_7$</th>
<th>$X_8$</th>
</tr>
</thead>
</table>

Marginal dataset without variables $X_2, X_5$

<table>
<thead>
<tr>
<th>$X_1$</th>
<th>$X_3$</th>
<th>$X_4$</th>
<th>$X_6$</th>
<th>$X_7$</th>
<th>$X_8$</th>
</tr>
</thead>
</table>

FCI
Finding $NAn_c P$

- $(X, Y) \in NAn_c P$: $X$ is not an ancestor of $Y$ in all MAGs represented by $P$.
- $X$ has no potentially directed path to $Y$. 
Finding $N\text{Anc}_P$

- $(X,Y) \in N\text{Anc}_P$: $X$ is not an ancestor of $Y$ in all MAGs represented by $P$.
- $X$ has no potentially directed path to $Y$.
- $(D,A), (D,B), (D,C)$
Finding $\text{NAn}c_P$

$\text{PAG } P$

- $(X, Y) \in NAn_{c_P}: X$ is not an ancestor of $Y$ in all MAGs represented by $P$
- $X$ has no potentially directed path to $Y$.
- $(D, A), (D, B), (D, C)$
- $(E, A), (E, B), (E, C), (E, D)$
Finding $NAnc_P$

- $(X, Y) \in NAnc_P$: $X$ is not an ancestor of $Y$ in all MAGs represented by $P$

- $X$ has no potentially directed path to $Y$.

- $(D, A), (D, B), (D, C)$
- $(E, A), (E, B), (E, C), (E, D)$
- $(F, A), (F, B), (F, C), (F, D), (F, E)$
Finding $\text{Anc}_P$

- $(X, Y) \in N\text{Anc}_P$: $X$ is an ancestor of $Y$ in all MAGs represented by $P$.
- Take the transitive closure of directed edges in the PAG.
Finding $\text{Anc}_P$

- $(X, Y) \in N\text{Anc}_P$: $X$ is an ancestor of $Y$ in all MAGs represented by $P$.
- Take the transitive closure of directed edges in the PAG.
  - $(D, E), (D, F)$
  - $(E, F)$
Finding $\text{Anc}_P$

- $(X, Y) \in N\text{Anc}_P$: $X$ is an ancestor of $Y$ in all MAGs represented by $P$.
- Take the transitive closure of directed edges in the PAG.
  - $(D, E)$, $(D, F)$
  - $(E, F)$
Finding \( \text{Anc}_P \)

- \((X, Y) \in N\text{Anc}_P: X\) is an ancestor of \(Y\) in all MAGs represented by \(P\).
- Take the transitive closure of directed edges in the PAG.
  - \((D, E), (D, F)\)
  - \((E, F)\)

Some relations are not so obvious!

\(A\) causes \(D\) in all MAGs repr. by \(P\).
Finding $\text{Anc}_P$

**Theorem:**
If $(X,Y)$ is not in the transitive closure of PAG $P$, $X$ is an ancestor of $Y$ if and only if $\exists U,V, U\neq V$ such that:

1. There are uncovered potentially directed paths from $X$ to $Y$ via $U$ and $V$ in $P$, and
2. $<U,X,V>$ is an unshielded definite non-collider in $P$.
Finding $\text{Anc}_P$

- $(X, Y) \in \text{NAnc}_P$: $X$ is an ancestor of $Y$ in all MAGs represented by $P$.

- Take the transitive closure of directed edges in the PAG.
  
  - $(D, E), (D, F)$
  - $(E, F')$
  - $(A, D), (A, E), (A, F)$
Ambiguous relationships

- If $(X, Y) \notin A_n P \cup N A_n P$: Ambiguous relationship.

- If you marginalize out variables, (non) ancestral relationships can become ambiguous.
Marginal Consistency

Under perfect statistical knowledge:

- Ancestral relations in the marginal are ancestral in the original PAG (over $V \setminus L$)
  - $Anc(P[L]) \subseteq Anc(P)_{V \setminus L}$ ($d=0$, $e=0$)

- Non-ancestral relations in the marginal non-ancestral in the original PAG (over $V \setminus L$)
  - $NAnc(P[L]) \subseteq NAnc(P)_{V \setminus L}$ ($c=0$, $f=0$)

In reality:

- Ancestral relations in the marginal can be:
  - non-ancestral in the original PAG ($d$).
  - ambiguous in the original PAG ($e$)

- Non-ancestral relations in the marginal can be:
  - Ancestral in the original PAG ($c$).
  - ambiguous in the original PAG ($f$)
Experiments

- 50 DAGs of 20 variables.
- Graph density: 0.1 and 0.2.
- Continuous data sets with 1000 samples.
- 100 random marginals of 18 and 15 variables.
- Algorithms used (pcalg) [2]:
  - FCI
  - iFCI
  - mFCl
Results (Ancestral relationships)

- **18 variables**
  - iFCl: Ancestral in both.
  - FCl: Ancestral in marginal non-ancestral in original.
  - mFCl: Ancestral in marginal ambiguous in original.

- **15 variables**
  - iFCl: Ancestral in both.
  - FCl: Ancestral in marginal non-ancestral in original.
  - mFCl: Ancestral in marginal ambiguous in original.

Density 0.1

Density 0.2
Results (Non Ancestral Relationships)

- Non-Ancestral in both.
- Non-Ancestral in marginal, ancestral in original.
- Non-Ancestral in marginal ambiguous in original.

18 variables

15 variables

density 0.1

density 0.2
Conclusions

• Constraint-based methods are sensitive to marginalization.
• Consistency of causal predictions drops for denser networks/smaller marginals.
• Non-causal predictions are very consistent.
• Majority rule FCI outperforms other FCI variants.
Ranking based on marginal consistency

• Can you use marginal consistency to find more robust predictions?

• Are predictions that are frequent in the marginals more robust?

• Compare with bootstrapping.
Results

18 variables

15 variables

bootstrapping

density 0.1

density 0.2
Conclusion/Future work

- Ranking by marginal consistency can help identify robust causal predictions.
- Bootstrapping is more successful in ranking causal predictions (by a small margin).
- Ranking by marginals can become much faster by caching tests of independence.
- Try it for much larger data sets (number of variables).
- Combine with bootstrapping.
- Try to identify a marginal that maximizes marginal consistency.
- Try to identify a graph that is consistent for more marginals.

