Identifiability of Gaussian Linear Structural Equation Models

Gunwoong Park

University of Seoul 02/12/2019

University of Seoul

Graph Notation

- Directed acyclic graph G:

  - $\triangleright$ $G = (V, E)$.
  - $\triangleright$ V : a set of nodes, e.g. $V = \{1, 2, 3\}$.
  - $\triangleright$ E : a set of directed edges, $E = \{(1, 2), (2, 3)\}$.
  - $\triangleright$ π: (causal) ordering, $\pi = (1, 2, 3)$.

Factorization (Lauritzen, 1996)

$$f_G(X_1, X_2, ..., X_p) = \prod_{j=1}^{p} f_j\left(X_j \mid X_{\mathrm{Pa}(j)}\right),$$

where $f_j(X_j \mid X_{\mathrm{Pa}(j)})$ refers to the conditional distribution of node X_j given its parents.

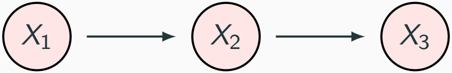
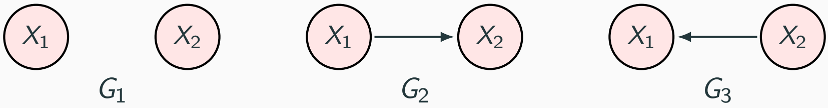


Figure 1: 3-node Chain Graph with the ordering (X_1, X_2, X_3)

$$f(X_1, X_2, X_3) = f_1(X_1)\ f_2(X_2 \mid X_1)\ f_3(X_3 \mid X_1).$$

Model Identifiability

Is it possible to recover a graph from data? **Partially Yes.**



- We can distinguish G_2 and G_3 from G_1 using dependence.
- We cannot identify the direction of an edge. Hence, we cannot distinguish G_2 from G_3.

Model Identifiability

How to recover a graph? **Require more information.**

- Linear non-Gaussian SEMs (Shimizu et al.,2006).
- Nonlinear additive noise models (hoyer et al., 2009; mooij et al., 2009; peters et al., 2012)
- Gaussian linear SEMs with the known error variances (Peters and Bühlmann, 2014, Loh and Bühlmann 2014).
- Exponential family DAG models (Park and Raskutti, 2015, 2018; Park and Park 2019)
- Linear SEMs with the unknown error variances (Peters et al., 2014; Ghoshal and Horino, 2017, 2018)

Identifiability of Gaussian SEMs

Gaussian Linear SEM

- Form of Gaussian Linear SEMs:
  - ▷ Node-wide: $X_j = \sum_{k \in \mathrm{Pa}(j)} \beta_{kj} X_k + \epsilon_j \quad \forall j = 1, \ldots, p,$
  - ▷ Matrix:

$$(X_1, X_2, ..., X_p)^T = B_0 + B(X_1, ..., X_p)^T + (\epsilon_1, ..., \epsilon_p)^T$$

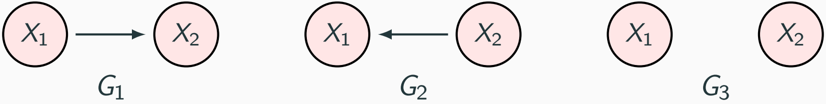
- Assumptions:
  - ▷ Gaussian Noise: $\epsilon_j \sim^{iid} N(0, \sigma_j^2)$ with $\sigma_j^2 > 0$.
  - ▷ Causal Minimality: $\beta_{kj} \neq 0$ for all $k \in \mathrm{Pa}(j)$, otherwise $\beta_{jk} = 0$.
  - ▷ Causal Sufficiency: all variables are observed.

- Joint Distribution:

$$f(G) = \frac{1}{\sqrt{(2\pi)^p \det(\Theta^{-1})}} \exp\Big(-\frac{1}{2}(x_1, ..., x_p)\Theta(x_1, ..., x_p)^T\Big).$$

Motivations of Identifiability: Ordering Recovery

Is it possible to recover a graph from a Gaussian SEM? Yes.



For G_1,

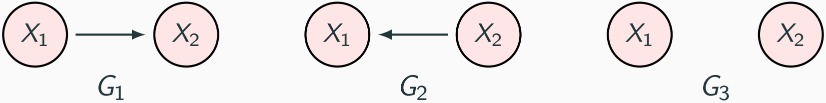
- Forward Selection:

$$\begin{aligned}\mathrm{Var}(X_2) &= \mathbb{E}(\mathrm{Var}(X_2 \mid X_1)) + \mathrm{Var}(\mathbb{E}(X_2 \mid X_1)) \\ &= \sigma_2^2 + \beta_1^2\sigma_1^2 > \sigma_1^2 = \mathrm{Var}(X_1),\end{aligned}$$

- Backward Elimination:

$$\begin{aligned}\mathbb{E}(\mathrm{Var}(X_1 \mid X_2)) &= \mathrm{Var}(X_1) - \mathrm{Var}(\mathbb{E}(X_1 \mid X_2)) \\ &= \sigma_1^2 - \beta_1^2\sigma_1^4/(\beta_1^2\sigma_1^2 + \sigma_2^2) < \sigma_2^2 = \mathbb{E}(\mathrm{Var}(X_2 \mid X_1)).\end{aligned}$$

Motivations of Identifiability: Parent Recovery



Suppose that $\sigma_1^2 < \sigma_2^2$. Then, for G_3,

$$\mathrm{Var}(X_2) = \sigma_2^2 > \sigma_1^2 = \mathrm{Var}(X_1).$$

Under the minimality condition,

$$G_1 : X_1 \not\perp\!\!\!\perp X_2 \quad \text{and} \quad G_3 : X_1 \perp\!\!\!\perp X_2$$

Theorem: Identifiability

Let $P(G)$ be generated from a Gaussian linear SEM.

- Π_G is a set of true orderings of graph G,
- $X_{1,2,..,j} = \{X_{\pi_1}, X_{\pi_2}, ..., X_{\pi_j}\}$,
- For any node $m \in V$, let $j = \pi_m$ and $k \in V \setminus \mathrm{Nd}(j)$.

Theorem

The DAG G is uniquely identifiable, if there exists $\pi \in \Pi_G$ satisfying either of the two following conditions

(A) **Forward Selection:**
$$\sigma_j^2 < \sigma_k^2 + \mathbb{E}(\mathrm{Var}(\mathbb{E}(X_k \mid X_{\mathrm{Pa}(k)}) \mid X_{1:(j-1)})), \text{ or}$$

(B) **Backward Elimination:**
$$\sigma_j^2 > \sigma_\ell^2 + \mathbb{E}(\mathrm{Var}(\mathbb{E}(X_\ell \mid X_{1:(j-1)}) \mid X_{\mathrm{Pa}(\ell)})),$$

Prior Identifiability Assumptions for Gaussian SEMs

Identifiability Assumption (Ghoshal and Horino, 2018)

Let $P(X)$ be generated from a Gaussian linear SEM with directed acyclic graph G. For any $m \in \{1, 2, ..., p\}$, let $j = \pi_m$ and $\ell \in \mathrm{De}(j)$.

$$\frac{1}{\sigma_j^2} < \frac{1}{\sigma_\ell^2} + \sum_{k \in \mathrm{Ch}(\ell)} \frac{\beta_{\ell k}^2}{\sigma_k^2}.$$

- It is strictly milder than the equal variance assumption (Peters and Bühlmann, 2014).
- It is equivalent to our assumption of (B) Backward Elimination.

Algorithm

Input: n i.i.d. samples from a Gaussian Linear SEM, $X^{1:n}$.

Output: Estimated causal graph, $\widehat{G} = (V, \widehat{E})$.

Step (1): Ordering Estimation using the OLS.

Step (2): Parents Estimation using independence tests.

Note that

$$\mathrm{Var}(X_j \mid X_S) = X_j^T (I - Proj_S) X_j$$

Simulation Settings

Simulations

100 realizations of a p-node random **Gaussian** linear SEM.

Homogeneous Error

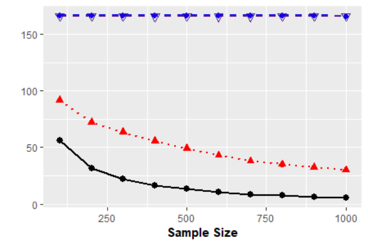
- $\sigma_j^2 = 1$ for all $j \in V$.
- $\beta_{jk} \in [-2, -0.5] \cup [0.5, 2]$.
- $p \in \{20, 50\}$.
- $n \in \{100, 200, ..., 1000\}$

Heterogeneous Error

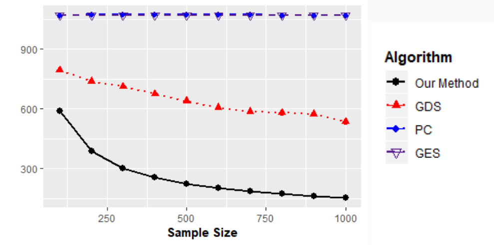
- $\sigma_j^2 \in [1, 3]$
- $\beta_{jk} \in [-2, -1] \cup [1, 2]$.
- $p \in \{20, 50\}$.
- $n \in \{100, 200, ..., 1000\}$

- Evaluation Measure: Hamming Distance, $|(E \cup \widehat{E} - E \cap \widehat{E})|$.

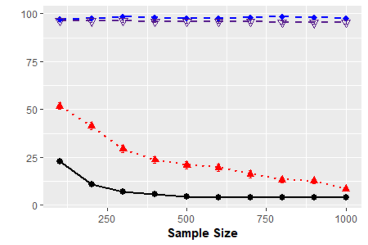
Average Hamming Distances



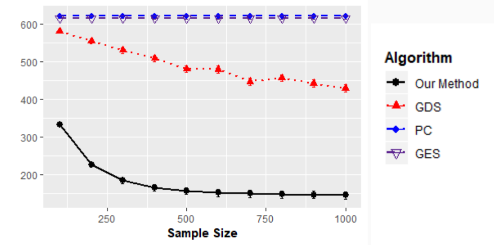
(a) Homo: $p = 20$



(b) Homo: $p = 50$



(c) Hetero: $p = 20$



(d) Hetero: $p = 50$

Future Works

Future Works

Future Works

- Gaussian Linear SEMs with Dependent Errors.
- Non-Gaussian Linear SEMs.
- Non-Linear SEMs.
- High-Dimensional Learning.
- Robust Learning.

References

- Asish Ghoshal and Jean Honorio. Learning identifiable gaussian bayesian networks in polynomial time and sample complexity. In Advances in Neural Information Processing Systems, pages 6457–6466
- Asish Ghoshal and Jean Honorio. Learning linear structural equation models in polynomial time and sample complexity.
- Jonas Peters and Peter Bühlmann. Identifiability of gaussian structural equation models with equal error variances. Biometrika, 101(1):219–228, 2014.

Thank you!