Causal discovery of cyclic nonlinear relations

Cyclic non-linear structure equation models with additive normal noise

Mario Nagase, Yutaka Kano
Division of Mathematical Science, Graduate School of Engineering Science
Osaka University

1 Introduction

Identifying causal structure gives a significant impact in many scientific disciplines. For example, knowing causal cellular signaling network can help to develop new drugs, knowing consumers network can help to improve marketing strategies and knowing causal mechanism behind policies can help to make better political decisions and so on. Several methods have been proposed and developed in the past decades to learn the causal structure. Before introducing our proposal, we shall summarize such methods and their problems under the following simple setting with additive noise models. Let X and Y be observed variables of interest, and one can consider the following three types of relationships:

$$\text{(a) } Y = f(X) + e_Y, \quad \text{(b) } X = g(Y) + e_X, \quad \text{(c) } \begin{cases} Y = f(X) + e_Y \\ X = g(Y) + e_X \end{cases},$$

where f and g are functions which connect the two variables and e_Y and e_X are additive error terms. It is well-known that the model (a) is not distinguishable from the model (b), and also the model (c) is not identifiable, provided that both f and g are linear functions and the error terms are assumed to be normally distributed [1]. Under the assumption of linearity and normality, many authors have pointed out the difficulty of identifiability to hold for many causal structures even with Directed Acyclic Graph (DAG) [2]. Cyclic causal structures are more difficult to be identified.

2 Models

Let (y, x) be random variables involved in the model. A cyclic structural equation model (SEM) has the form of

$$\begin{cases} y = f(x) + e_Y \\ x = g(y) + e_X \end{cases} \Longleftrightarrow \begin{cases} e_Y = y - f(x) \\ e_X = x - g(y) \end{cases}, \text{where } |f'(x)g'(y)| < 1. \tag{1}$$

The condition $|f'(x)g'(y)| < 1$ is necessary, because the model assumes that the set of the random variables (y, x) is yielded at the equilibrium state in the corresponding recurrence system $\begin{cases} y_{i+1} = f(x_i) + e_Y \\ x_{i+1} = g(y_i) + e_X \end{cases}$, where (y_i, x_i) are the ith sequence point, assuming that i) the error term (e_Y, e_X) does not change over time, ii) $\lim_{i\to\infty} y_i < \infty$, and $\lim_{i\to\infty} x_i < \infty$, and iii) $\lim_{i\to\infty}(y_i, x_i)$ is uniquely specified given (e_Y, e_X). The Jacobian matrix associated with Eq.(1) is

$$\begin{bmatrix} \dfrac{\partial y}{\partial e_Y} & \dfrac{\partial y}{\partial e_X} \\ \dfrac{\partial x}{\partial e_Y} & \dfrac{\partial x}{\partial e_X} \end{bmatrix} = \begin{bmatrix} \dfrac{\partial e_Y}{\partial y} & \dfrac{\partial e_Y}{\partial x} \\ \dfrac{\partial e_X}{\partial y} & \dfrac{\partial e_X}{\partial x} \end{bmatrix}^{-1} = \frac{1}{1 - f'(x)g'(y)} \begin{bmatrix} 1 & f'(x) \\ g'(y) & 1 \end{bmatrix},$$

and the Jacobian is given as $1 - f'(x)g'(y) := J(x, y) \neq 0$, since $|f'(x)g'(y)| < 1$, then the invertiability $(y, x) \leftrightarrow (e_Y, e_X)$ is guaranteed. We write the inverse function of (1) as $\begin{bmatrix} e_Y & e_X \end{bmatrix}^T = \boldsymbol{e} = \boldsymbol{e}(y, x)$. Suppose that the error term $\boldsymbol{e}$ follows a multivariate normal distribution $\mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$ with the probability density function as $p(e_Y, e_X; \boldsymbol{\Sigma}) = \frac{1}{2\pi\sqrt{|\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}\boldsymbol{e}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{e}\right)$, and also the *pdf* of (y, x) is obtained as $q(y, x; f, g, \boldsymbol{\Sigma}) = \frac{|J(y,x)|}{2\pi\sqrt{|\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}\boldsymbol{e}(y, x)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{e}(y, x)\right)$.

3 The model identifiability

Let $(f_1(x), g_1(y), \boldsymbol{\Sigma}_1)$ and $(f_2(x), g_2(y), \boldsymbol{\Sigma}_2)$ be two sets of functions and the variance and covariance matrices of cyclic SEMs as in (1):

$$\begin{cases} y = f_1(x) + e_Y \\ x = g_1(y) + e_X \end{cases}, \quad \boldsymbol{e} = (e_Y, e_X)^T \sim N_2(0, \boldsymbol{\Sigma}_1), \qquad \begin{cases} y = f_2(x) + e_Y^* \\ x = g_2(y) + e_X^* \end{cases}, \quad \boldsymbol{e}^* = (e_Y^*, e_X^*)^T \sim N_2(0, \boldsymbol{\Sigma}_2).$$

The mappings $(y, x) \mapsto (e_Y, e_X)$ and $(y, x) \mapsto (e_Y^*, e_X^*)$ are assumed to be one to one and invertible, i.e., $|f_1'(x)g_1'(y)| < 0$ and $|f_2'(x)g_2'(y)| < 0$. The model identifiability,

$$q(y, x; f_1, g_1, \mathbf{\Sigma}_1) = q(y, x; f_2, g_2, \mathbf{\Sigma}_2) \ \ \forall (y, x) \implies f_1 = f_2, g_1 = g_2, \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2,$$

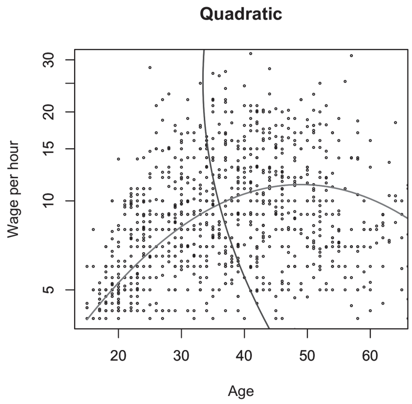
assuming the functions f and g are finite polynomial functions, will be proved in the talk.

4 Simulation study

We will show the fact that the causal structure can be completely recovered with the following setting using synthetic datasets.

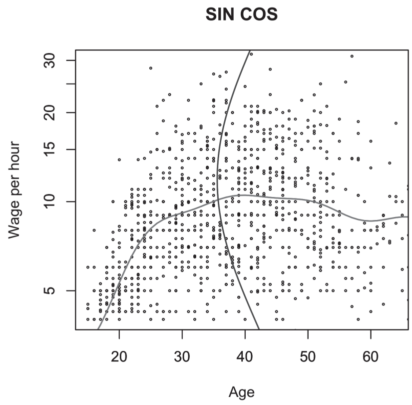
- Polynomial functions: $\begin{cases} y = a_0 + a_1 x + a_2 x^2 + \cdots + a_p x^p + e_Y \\ x = b_0 + b_1 y + b_2 y^2 + \cdots + b_q y^q + e_X \end{cases}$, $\boldsymbol{e} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$
- sin and cos functions: $\begin{cases} y = a_0 + \sum_{i=1}^{p} \Big(a_{2i-1} \sin(ix) + a_{2i} \cos(ix) \Big) + e_Y \\ x = b_0 + \sum_{j=1}^{q} \Big(b_{2j-1} \sin(jy) + b_{2j} \cos(jy) \Big) + e_X \end{cases}$, $\boldsymbol{e} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$

And those models are applied to real world datasets with trivial causal structure. For example, we applied the models to the dataset ”Census Income (KDD) dataset” which contains weighted census data extracted from the 1994 and 1995 current population surveys conducted by the U.S. Census Bureau [3], and the results show quite a reasonable causal structure, age affects wage and the wage is highest in the age of around 40 and 50 while wage has almost no impact on age, was obtained as shown in Figure 1.



[1]

Quadratic model, $n = 1,000$



[2]

Sin-Cos model, $n = 1,000$

Figure 1: Red lines and blue lines represent the estimated functions f and g, respectively

5 Conclusion and discussion

Most surprising part might be that the causal structures were nicely recovered and/or discovered from the datasets which do not appear to have causal structure at a glance. The model with sin-cos functions may have applicability to more practical situations, since it can cover versatile shapes of functions as the values of the complexity parameters p and/or q increase, and one can specify the complexity parameters by some criterion such as AIC. And also the model seems to be able to deal with the error covariance which is supposed to represent the existence of unobserved confounders. It might allow users to assess if confounders exist by applying the model.

References

[1] Nagase, M. and Kano, Y., Identifiability of nonrecursive structural equation models., *Statistics & Probability Letters* **122**, pp.109-117 (2017)

[2] Richardson, T., A discovery algorithm for directed cyclic graphs., *In Proceedings of the Twelfth international conference on Uncertainty in artificial intelligence*, Morgan Kaufmann Publishers Inc., pp.454-461 (1996)

[3] Mooij, J., Janzing, D., Peters, J., and Scholkopf, B., Regression by dependence minimization and its application to causal inference in additive noise models, *In Proceedings of the 26th annual international conference on machine learning*, pp. 745-752. ACM, (2009)