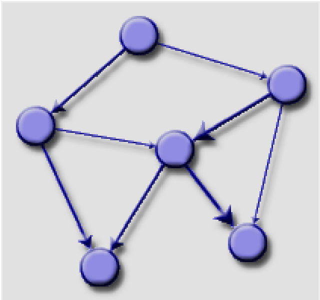


構造方程式モデリングによる 因果構造探索

清水昌平

大阪大学　産業科学研究所

Abstract

- Linear structural equation models (SEMs) often used to represent data generating processes of continuous variables

- Conventional methods based on conditional independences or Gaussianity cannot identify a number of data generating models

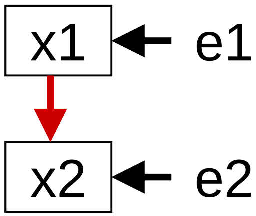
- Non-Gaussian methods can identify many more data generating models

Motivation

- Suppose that two observed variables x1 and x2 are linear mixtures of two independent latent variables e1 and e2

- You know that the mixing model (the data generating process) is either of the following two models (SEMs):

Model 1

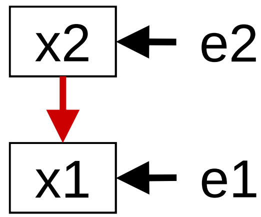


$$x_1 = e_1$$
$$x_2 = 0.8x_1 + e_2$$

or

Model 2

$$x_2 = e_2$$
$$x_1 = 0.8x_2 + e_1$$



- You want to estimate which SEM is the actual data generating process based on a random sample X from p(x1,x2)

Motivation (continued)

- Conventional methods cannot do this
  - Because they only use conditional independence relations between x1 and x2 or assume Gaussianity (even if data is non-Gaussian)

- Non-Gaussian data is encountered in many applications:
  - Neuroinformatics (Hyvarinen et al., 2001); Bioinformatics (Sogawa et al., 2010); Social sciences (Micceri, 1989); Economics (Moneta et al., 2010)

- A new approach: Non-Gaussian approach
  - Non-Gaussianity allows to identify the data generating process: which of the two SEMs generated the data

Background:
Structural equation models (SEMs)

Background: Structural equation models

(Bollen, 1989; Pearl, 2000)

- Structural equation models (SEMs) integrate many multivariate statistical models
  - path analysis, factor analysis, autoregressive models, analysis of variance etc.

- SEMs can be used to represent data generating processes

- SEMs can be used to represent causal relations under some additional assumptions including a definition of causal effect

Background: Structural equation models

(Bollen, 1989; Pearl, 2000)

- Structural equation models (SEMs) integrate many multivariate statistical models
  - path analysis, factor analysis, autoregressive models, analysis of variance etc.

- SEMs can be used to represent data generating processes

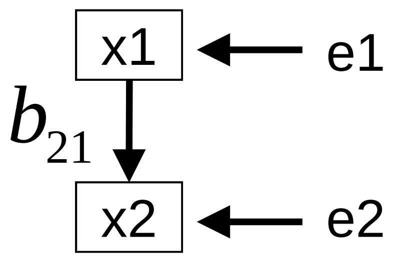
- SEMs can be used to represent causal relations under some additional assumptions including a definition of causal effect

A SEM represent a data generating process: Example

- A linear SEM (Bentler & Weeks, 1983)

$$x_1 = e_1$$

$$x_2 = b_{21} x_1 + e_2$$



- The SEM reads:
  - The value of x1 is determined by that of e1
  - The value of x2 is determined by that of x1(=e1) and that of e2

The latent variables e1 and e2 are called exogenous variables

(Bollen, 1989)

- An exogenous variable is a variable that is determined outside of the model: e1 and e2

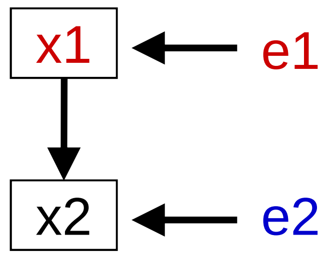
- Two kinds of exogenous variables:
  1. Exogenous observed variables: x1=e1
  2. (latent) Disturbances: e2

Structural equations:

$$x_1 = e_1$$

$$x_2 = b_{21} x_1 + e_2$$

Graph:



Basic linear SEM: DAG-SEM

(Bollen 1989)

- A linear acyclic SEM with independent exogenous variables:
  - Graphically represented by a directed acyclic graph (DAG)
  - Observed random vector x is linearly modeled by

$$x_i = \sum_{j \neq i} b_{ij} x_j + e_i \quad \text{or} \quad \mathbf{x} = \mathbf{B}\mathbf{x} + \mathbf{e}$$

where

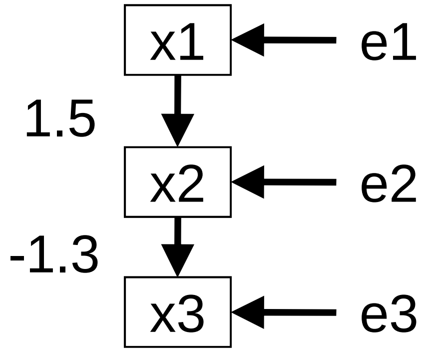
- The e_i are exogenous variables with non-zero variance and are mutually independent
- The 'path-coefficient' matrix $\mathbf{B} = [b_{ij}]$ is constant
- Acyclicity

Example

- Structural equations:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 0 & 0 \\ 1.5 & 0 & 0 \\ 0 & -1.3 & 0 \end{bmatrix}}_{\mathbf{B}} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix}$$

- Associated DAG:
  - x1 is a parent of x2
  - x1 is an ancestor of x3
  - x2 is a child of x1
  - x3 is a descendant of x1

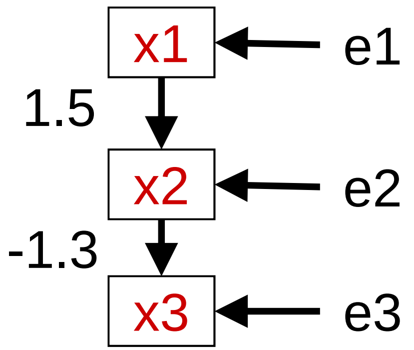


Assumption of acyclicity

- Acyclicity ensures existence of an ordering of variables xi that permutes matrix B lower-triangular with zeros on the diagonal

- In the example, the ordering is x1<x2<x3

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 0 & 0 \\ 1.5 & 0 & 0 \\ 0 & -1.3 & 0 \end{bmatrix}}_{\mathbf{B}} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix}$$

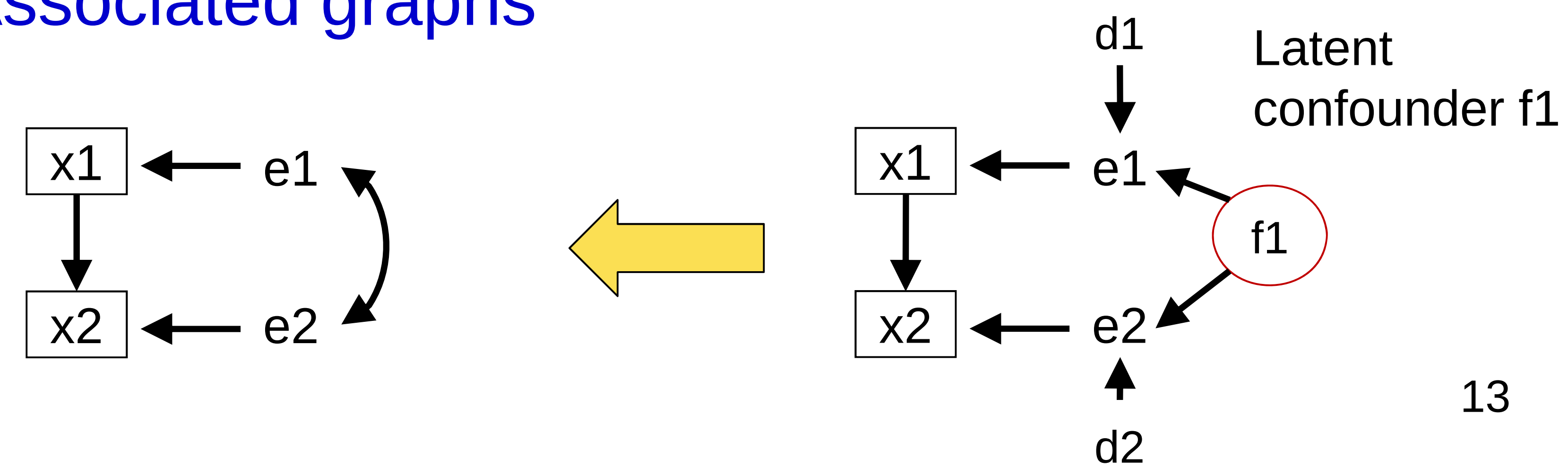


Dependent exogenous variables result from latent variables

- A latent variable (confounder) f1 introduces dependency between e1 and e2:

$$x_1 = e_1 \qquad e_1 = 2f_1 + d_1$$
$$x_2 = 3x_1 + e_2 \quad + \quad e_2 = 3f_1 + d_2$$

- Associated graphs



Background: Structural equation models

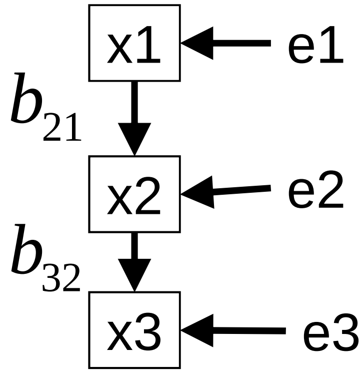
(Bollen, 1989; Pearl, 2000)

- Structural equation models (SEMs) integrate many multivariate statistical models
  - path analysis, factor analysis, autoregressive models, analysis of variance etc.
- SEMs can be used to represent data generating processes
- SEMs can be used to represent causal relations under some additional assumptions including a definition of causal effect

Can give a causal meaning to each path-coefficient b_{ij} (Pearl, 2000)

- Example:

$$x_1 = e_1$$
$$x_2 = b_{21} x_1 + e_2$$
$$x_3 = b_{32} x_2 + e_3$$



- Causal meaning of b32:
  - If you change x2 from c0 to c1, then E(x3) will change by b32(c1-c0)

`Change x2 from c0 to c1'

- It means `Fix x2 at c0 regardless of the variables that previously determined x2 (do(x2=c0)) and change the constant from c0 to c1' (Pearl, 2000)
- In the SEM, `Replace the function determining x2 with a constant c0':

$$x_1 = e_1$$
$$x_2 = b_{21}x_1 + e_2$$
$$x_3 = \qquad b_{32}x_2 + e_3$$

⇒

Mutilated model with do(x2=c0):

$$x_1 = e_1$$
$$x_2 = c_0$$
$$x_3 = \qquad b_{32}x_2 + e_3$$

(Assumption of invariance: the other functions don't change)

Average causal effect: Definition (Rubin, 1973; Pearl, 2000)

- Average causal effect of x2 on x3 when changing x2 from c0 to c1:

$$E(x_3 \mid do(x_2 = c1)) - E(x_3 \mid do(x_2 = c0))$$
$$= b_{32}(c_1 - c_0)$$

Computed based on the mutilated models with do(x2=c1) or do(x2=c0)

- Causal meaning of b32:
  – If you change x2 from c0 to c1,
    then E(x3) will change by b32(c1-c0)

Conducting experiments with random assignment is a method to estimate average causal effects

- Random assignment of participants into two groups:
  - Experimental group with x2 = c1
  - Control group with x2 = c0
- This fixes x2 to c1 or c0 regardless of the variables that determines x2

$$E\left(x_3 \mid do\left(x_2 = c_1\right)\right) - E\left(x_3 \mid do\left(x_2 = c_0\right)\right)$$

E(x3) for the exp. group - E(x3) for the control group

Estimate average causal effects using non-experimental data

- Experiments with random assignment often difficult to conduct: costs, ethics, etc.

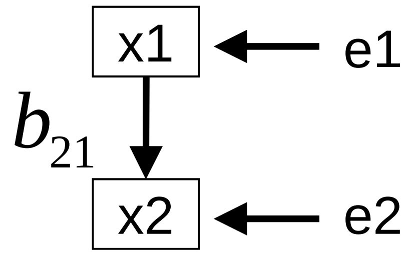
- Sometimes, average causal effects can be estimated using non-experimental data (Pearl, 2000; Spirtes et al., 2000)

- Non-Gaussian methods presented in this tutorial can be used to estimate average causal effects based on non-experimental data

Basic problem setup: Learning DAG-SEMs

- Assume that data X is randomly sampled from a DAG-SEM (linear, acyclic, no latent confounder):

$$\mathbf{x} = \mathbf{B}\mathbf{x} + \mathbf{e}$$



- Goal: Estimate the path-coefficient matrix B observing data X only!
  - Understand the data generating process
  - Compute average causal effects

Structural equation models (SEMs): Summary

- SEMs can be used to represent data generating processes
- If you accept that all the model assumptions are reasonable in your data, you can analyze the data generating process
- If you further accept the definition of causal effect, you can analyze causal relations
- SEMs help you to elaborate your theory
- Basic model is DAG-SEM
  - Various extensions extensively discussed

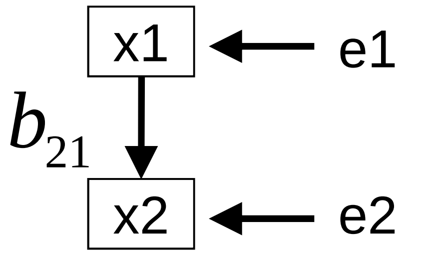
Problems:
Identifiability problems of conventional methods

Under what conditions B is identifiable?

- Identifiable = uniquely determined or estimated from p(x)

- DAG-SEM:

$$\mathbf{x} = \mathbf{B}\mathbf{x} + \mathbf{e}$$



  – B and p(e) induce p(x)
  – If p(x) are different for different B, then B is uniquely estimated

Conventional estimation principle: Causal Markov condition

- If the data generating model is a DAG-SEM, causal Markov condition holds :
  - Each observed variable x_i is independent of its non-descendants in the DAG conditional on its parents (Pearl & Verma, 1991) :

$$p(\mathbf{x}) = \prod_{i=1}^{p} p(x_i \mid \text{parents of } x_i)$$

Conventional methods based on causal Markov condition

- Methods based on conditional independencies (Spirtes & Glymour, 1991)
  - Many DAG-SEMs give same set of conditional independences and equally fit data

- Scoring methods based on Gaussianity (Chickering, 2002)
  - Many DAG-SEMs give the same Gaussian distribution and the same scores like BIC

- Often path-coefficient matrix B is not uniquely determined or estimated

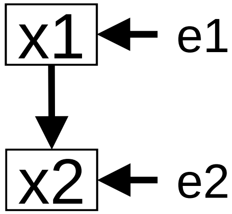
Example

- Two models with Gaussian e1 and e2:

Model 1:

$$x_2 = 0.8x_1 + e_2$$

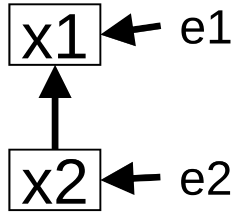
$$\mathbf{B} = \begin{bmatrix} 0 & 0 \\ 0.8 & 0 \end{bmatrix}$$



Model 2:

$$x_1 = 0.8x_2 + e_1$$

$$\mathbf{B} = \begin{bmatrix} 0 & 0.8 \\ 0 & 0 \end{bmatrix}$$



E(e1)=E(e2)=0, var(x1) = var(x2)=1

- Both introduce no conditional independence:

$$\mathrm{cov}\left(x_1, x_2\right) = 0.8 \neq 0$$

- Both induce same Gaussian distribution:

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}\right)$$

A solution:
Non-Gaussian approach

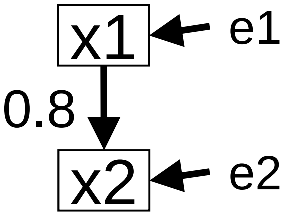
A new direction: Non-Gaussian approach

- Non-Gaussian data in many applications:
  - Neuroinformatics (Hyvarinen et al., 2001); Bioinformatics (Sogawa et al., 2010); Social sciences (Micceri, 1989); Economics (Moneta et al., 2010)
- Utilize non-Gaussianity for model identification (Bentler, 1983)
- The path-coefficient matrix B is uniquely estimated from p(x) if e_i are non-Gaussian
  - (Shimizu, Hoyer, Hyvarinen & Kerminen, JMLR, 2006)

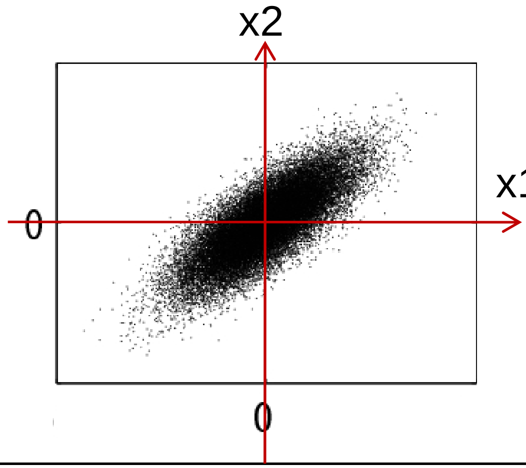
Illustrative example: Gaussian vs non-Gaussian

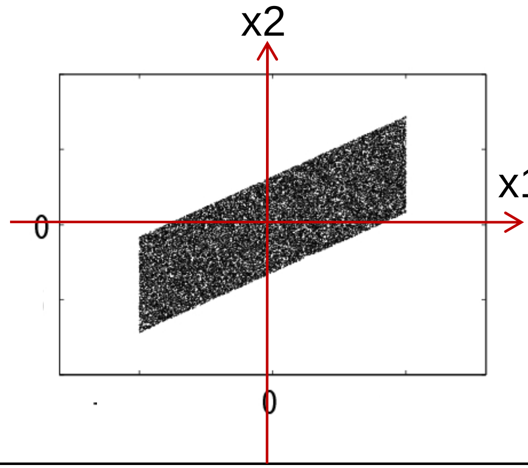
Gaussian | Non-Gaussian (uniform)

Model 1:

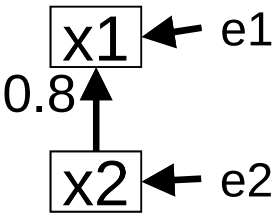


$$\mathbf{B} = \begin{bmatrix} 0 & 0 \\ 0.8 & 0 \end{bmatrix}$$

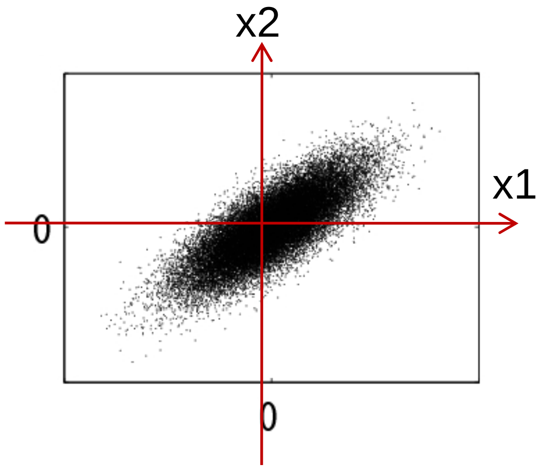


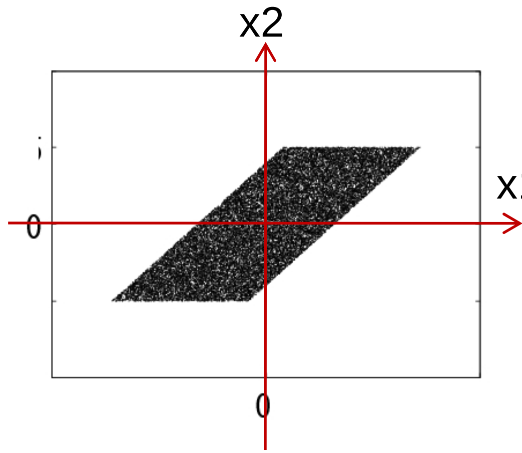


Model 2:



$$\mathbf{B} = \begin{bmatrix} 0 & 0.8 \\ 0 & 0 \end{bmatrix}$$





E(e1)=E(e2)=0
Var(x1) = var(x2)=1

Linear Non-Gaussian Acyclic Model: LiNGAM

(Shimizu, Hyvarinen, Hoyer & Kerminen, JMLR, 2006)

- An identifiable non-Gaussian version of DAG-SEM:

$$x_i = \sum_{j \neq i} b_{ij} x_j + e_i \quad \text{or} \quad \mathbf{x} = \mathbf{B}\mathbf{x} + \mathbf{e}$$

where

– The e_i are

- exogenous variables with non-zero variance
- (non-degenerate) non-Gaussian
- mutually independent

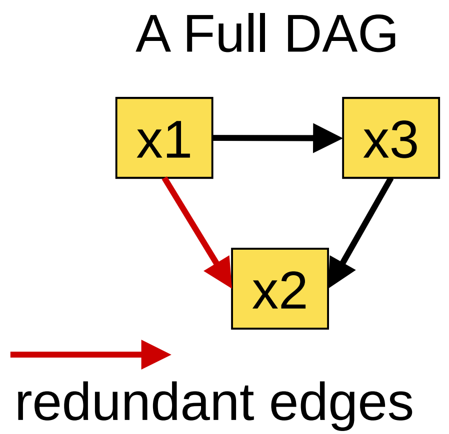
– Acyclicity

Two estimation algorithms

- LiNGAM algorithm
  (Shimizu, Hoyer, Hyvarinen & Kerminen, JMLR, 2006)
- DirectLiNGAM algorithm
  (Shimizu, Hyvarinen, Kawahara & Washio, UAI09)

- Both estimates an ordering of variables that makes the path-coefficient matrix B to be lower-triangular
  – Acyclicity ensures existence of such an ordering

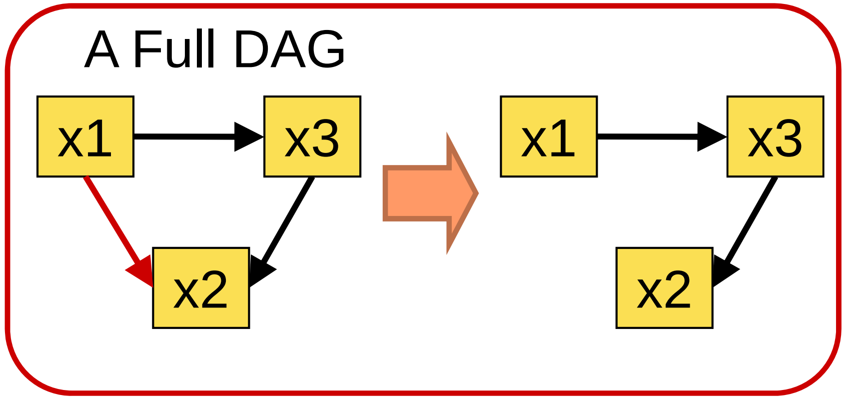
$$\mathbf{x}_{perm} = \underbrace{\begin{bmatrix} \ddots & \mathbf{O} \\ & \ddots \end{bmatrix}}_{\mathbf{B}_{perm}} \mathbf{x}_{perm} + \mathbf{e}_{perm}$$



Once such an ordering is found…

- Many existing methods can do:
  - Pruning the remaining path-coefficients
    - sparse methods like weighted lasso (Zou, 2006)
  - Finding significant path-coefficients
    - Testing, bootstrapping (Shimizu et al., 2006; Hyvarinen 2010)

$$\mathbf{x}_{perm} = \underbrace{\begin{bmatrix} * & & \mathbf{O} \\ 0 & * & \\ * & 0 & * \end{bmatrix}}_{\mathbf{B}_{perm}} \mathbf{x}_{perm} + \mathbf{e}_{perm}$$



まとめ

- Use of non-Gaussianity in linear SEMs is useful for model identification

- Non-Gaussian data encountered in many applications

- The non-Gaussian approach can be a good option

- Links to codes and papers:
  http://homepage.mac.com/shoheishimizu/lingampapers.html