LiNGAM

石垣ゼミ3年
名執祐矢

今までの要点

・相関関係があっても因果関係があるとは限らない
（＝擬似相関）
・相関係数の値によって、因果関係の大きさが測れない

→背後に「未観測共通原因」の存在

→擬似相関を懐柔し、データから因果関係まで推測するための
方法を研究するのが「統計的因果探索」

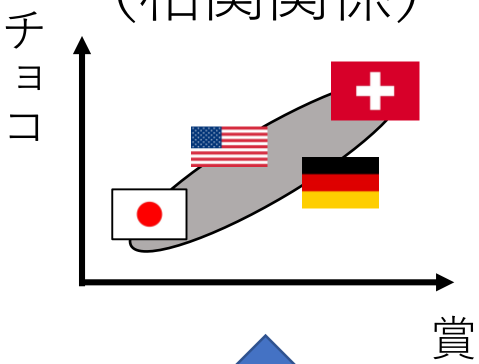
今までの要点（擬似相関）

チョコレートをたくさん食べている国ほど受賞者が多い
（相関関係）

・四角で囲まれている変数：観測変数
（データが収集される変数）

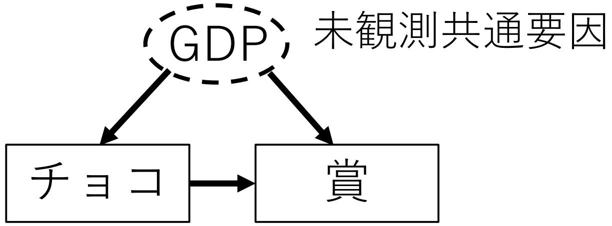
・点線の楕円で囲まれている変数
：未観測変数（データが収集されない変数）

・矢印の有無：因果関係の有無
（始点が原因、終点が結果の変数）

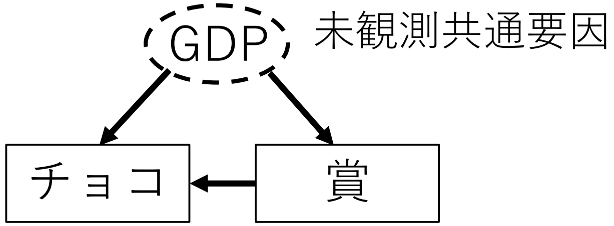


相関係数　0.79
P値　＜　0.0001

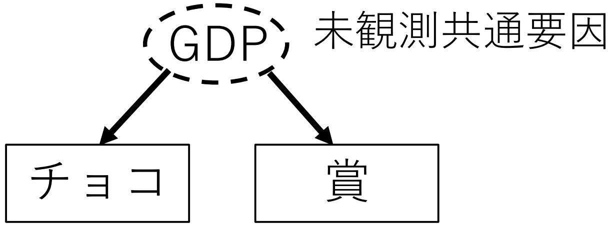
複数の因果関係が
同じ相関関係を与えうる



or

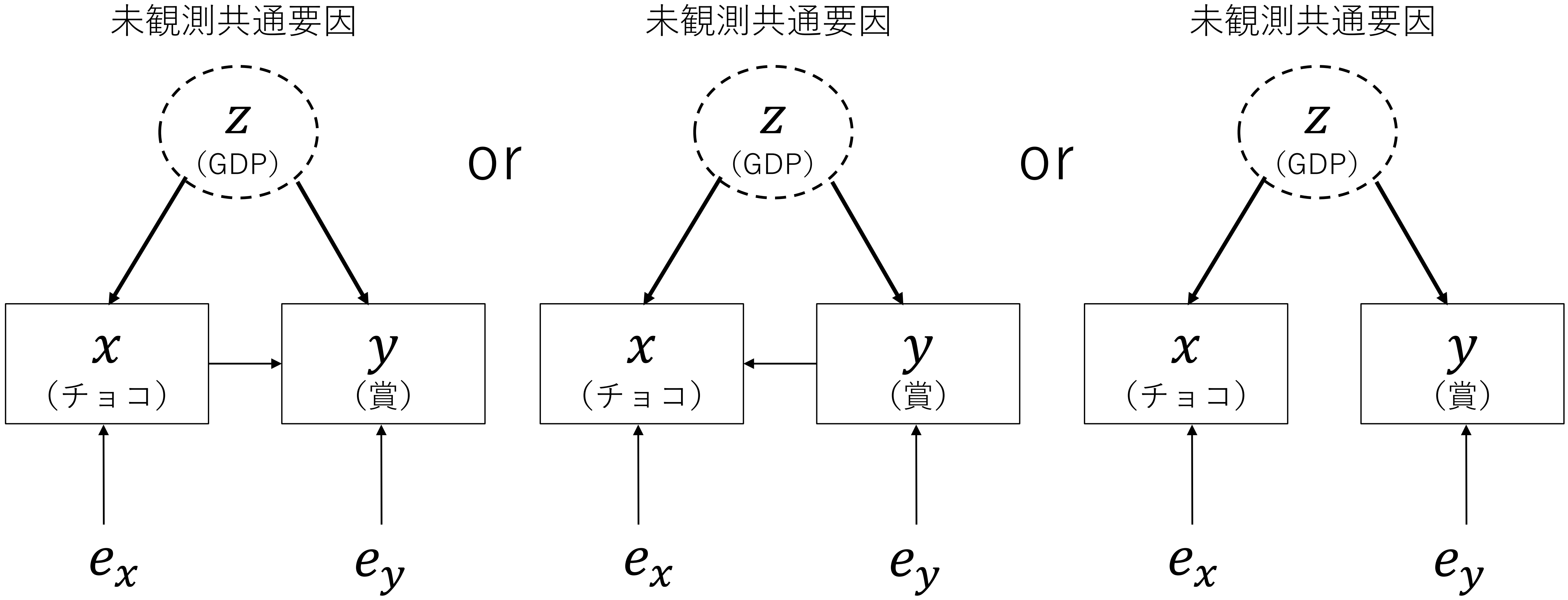


or



チョコレートをたくさん食べさせれば受賞者が増えるのか？
（因果関係）

今までの要点（擬似相関）


未観測共通要因
z
（GDP）
x
（チョコ）
y
（賞）
e_x
e_y
or
未観測共通要因
z
（GDP）
x
（チョコ）
y
（賞）
e_x
e_y
or
未観測共通要因
z
（GDP）
x
（チョコ）
y
（賞）
e_x
e_y

今までの要点（平均因果効果）

・変数xから変数yへの因果効果の大きさを定量化
→平均的な差を評価する

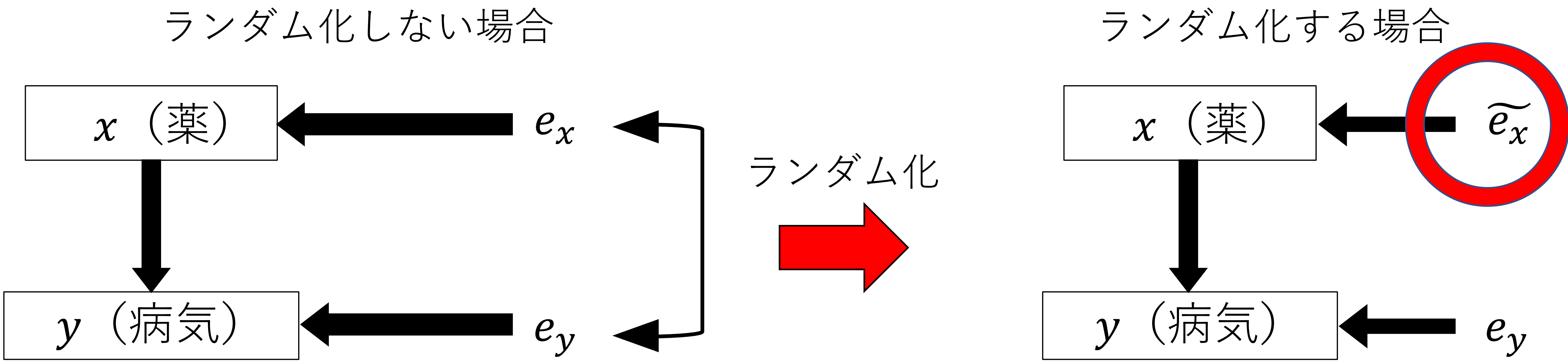
$$E(y|do(x=d)) - E(y|do(x=c))$$（←平均因果効果）

：$M_{X=d}$と$M_{X=c}$におけるyの期待値の差

※$do(x=c)$:ほかのどの変数がどんな値をとろうとも、変数xの値を定数cにとる（＝介入）

※$M_{x=c}$: xに介入して、その値をcに定めた構造方程式モデル

今までの要点（ランダム化実験）



・仮定1（事前知識）：時間的先行性

→ありうる因果の向きが決まる（病気→薬はありえない）

・仮定2：ランダム化

→誤差変数が独立になる：未観測共通原因がない

今までの要点（ランダム化実験）

ランダム化実験：一般的に、

$$p(y|do(x=c)) = p(y|x=c)$$

∵誤差変数$\widetilde{e_x}$とe_yが独立

$$p(y|x=1) \neq p(y|x=0)$$

→この集団において、薬を飲むかどうかが、

病気が治るかどうかの原因となる

（※ただし、倫理的・技術的観点からランダム化実験が困難な場合が多い

例：「抑うつ気分」と「睡眠障害」）

今までの要点（因果探索の基本問題）

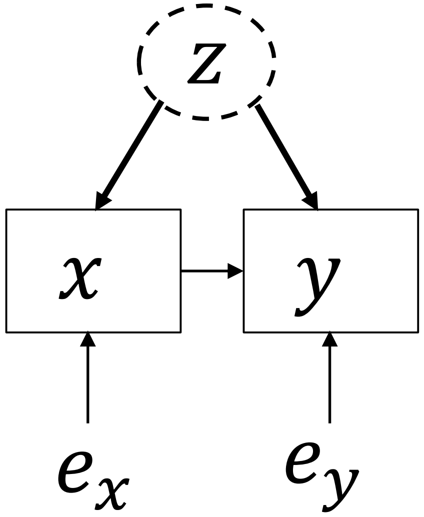
・x, y：観測変数　　・z：未観測共通要因　　・e_x, e_y：誤差変数

モデルA：
$$\begin{cases} x = f_x(z, e_x) \\ y = f_y(x, z, e_y) \\ p(z, e_x, e_y) = p(z)p(e_x)p(e_y) \end{cases}$$

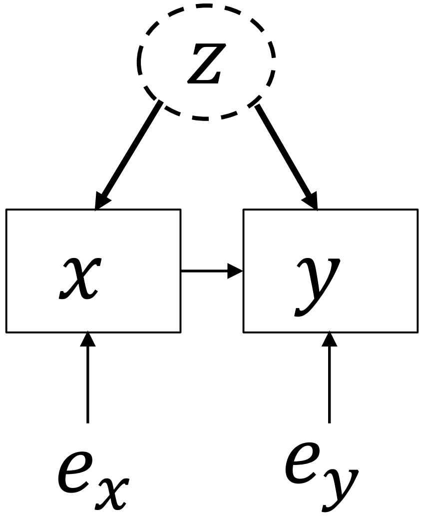
モデルB：
$$\begin{cases} x = f_x(y, z, e_x) \\ y = f_y(z, e_y) \\ p(z, e_x, e_y) = p(z)p(e_x)p(e_y) \end{cases}$$

モデルC：
$$\begin{cases} x = f_x(z, e_x) \\ y = f_y(z, e_y) \\ p(z, e_x, e_y) = p(z)p(e_x)p(e_y) \end{cases}$$

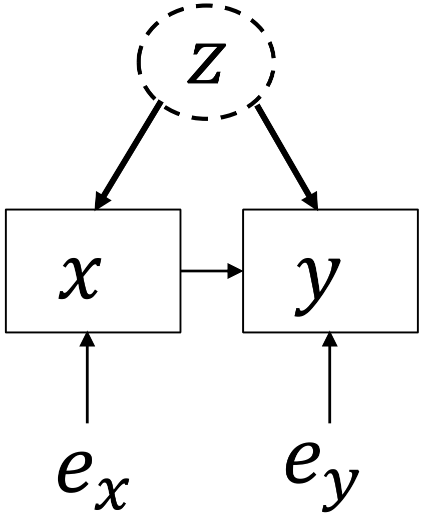
未観測共通要因



未観測共通要因



未観測共通要因



今までの要点（因果探索の基本問題）

性質1：自律性（介入してもほかは変わらない）

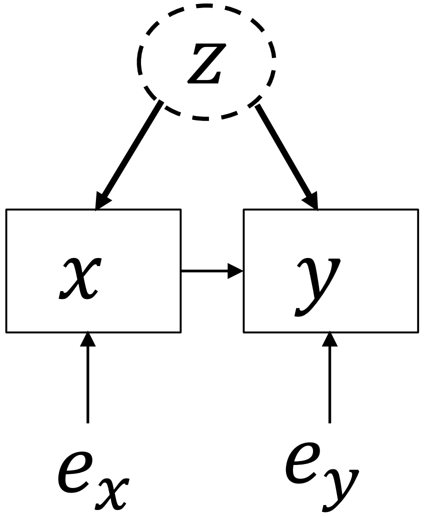
性質2：関数形と外生変数の分布が決まれば内生変数の分布が決まる

モデルA：$\begin{cases} x = f_x(z, e_x) \\ y = f_y(x, z, e_y) \\ p(z, e_x, e_y) = p(z)p(e_x)p(e_y) \end{cases}$

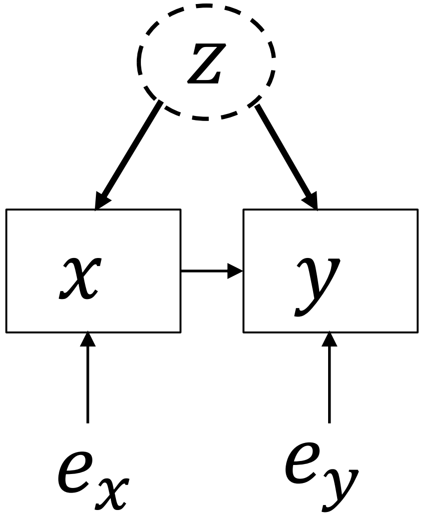
モデルB：$\begin{cases} x = f_x(y, z, e_x) \\ y = f_y(z, e_y) \\ p(z, e_x, e_y) = p(z)p(e_x)p(e_y) \end{cases}$

モデルC：$\begin{cases} x = f_x(z, e_x) \\ y = f_y(z, e_y) \\ p(z, e_x, e_y) = p(z)p(e_x)p(e_y) \end{cases}$

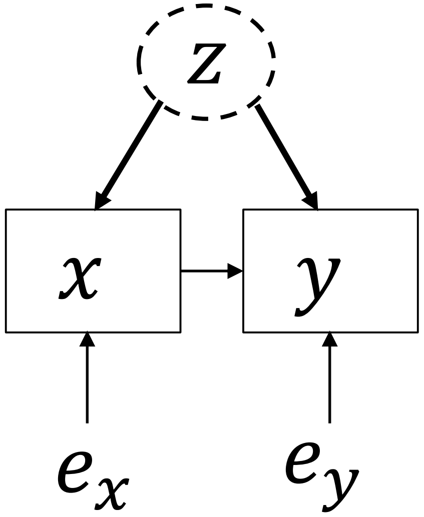
未観測共通要因



未観測共通要因



未観測共通要因



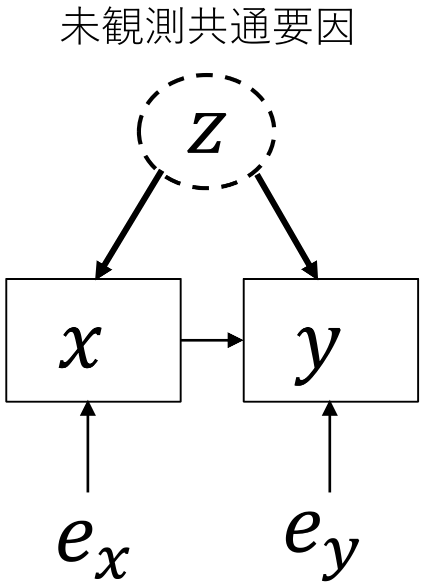
今までの要点（因果探索の基本問題）

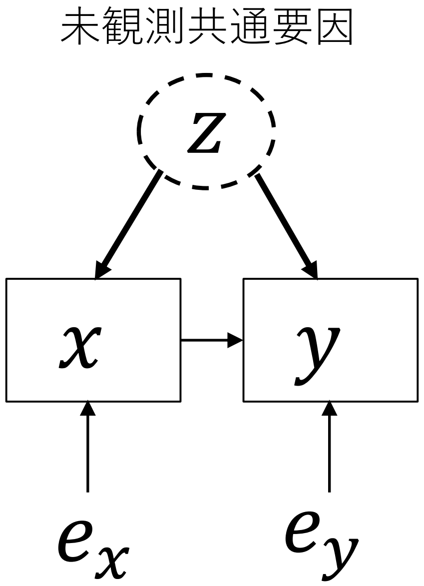
→統計的因果探索：「『関数形』と『外生変数の分布』にどのような仮定が成り立てば、
元の因果グラフをどの程度推測できるのか」を明らかにする

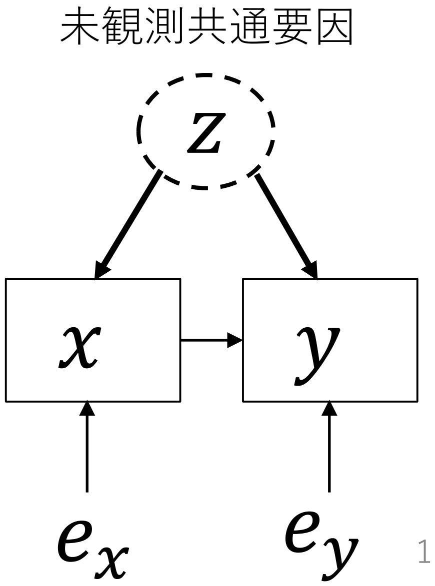
モデルA：
$$\begin{cases} x = f_x(z, e_x) \\ y = f_y(x, z, e_y) \\ p(z, e_x, e_y) = p(z)p(e_x)p(e_y) \end{cases}$$

モデルB：
$$\begin{cases} x = f_x(y, z, e_x) \\ y = f_y(z, e_y) \\ p(z, e_x, e_y) = p(z)p(e_x)p(e_y) \end{cases}$$

モデルC：
$$\begin{cases} x = f_x(z, e_x) \\ y = f_y(z, e_y) \\ p(z, e_x, e_y) = p(z)p(e_x)p(e_y) \end{cases}$$







今までの要点（3つのアプローチ）

①ノンパラメトリック・アプローチ
関数形：仮定なし　　　　　外生変数の分布：仮定なし
（ただし、因果関係は一方向であり双方向ではない）
→3つのモデルのうち、どのモデルか推測できない

②パラメトリック・アプローチ
関数形：（主に）線形　　　　　外生変数の分布：（主に）ガウス分布
→3つのモデルのうち、どのモデルか推測できない

③セミパラメトリック・アプローチ（特にLiNGAMアプローチ）
関数形：線形　　　　　外生変数の分布：非ガウス（連続）分布
→3つのモデルのうち、どのモデルか推測できる

・ガウス分布（正規分布）
：平均と分散共分散行列が定まれば分布形は一意に決まる

・非ガウス分布
：歪度・尖度などのパラメータをもつ
→情報量が多い

今までの要点（識別可能性）

・因果グラフの構造が異なれば、観測変数の分布が必ず異なる
→因果グラフは「識別可能」である

↑観測変数の分布の違いを手がかりに元の因果グラフを復元する、という前提からして重要

今までの要点（識別可能性）

・未観測共通原因がないと仮定

→「共通原因にあたる変数はすべて観測されており、かつ、変数間の全ての因果関係が未知である」とする

・各変数$x_i (i = 1, \ldots, p)$の生成過程

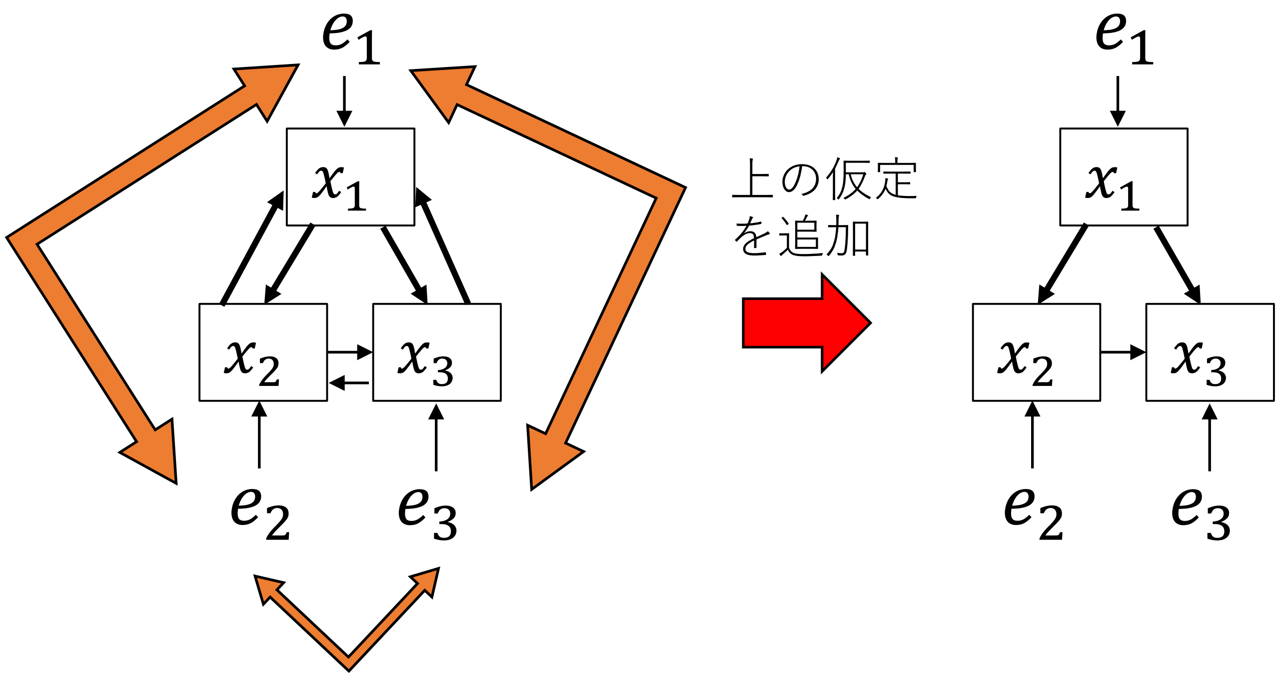
→ $x_i = f_i\left(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_p, e_i\right) \quad (i = 1, \ldots, p)$ （※入力時にx_iを含まない）

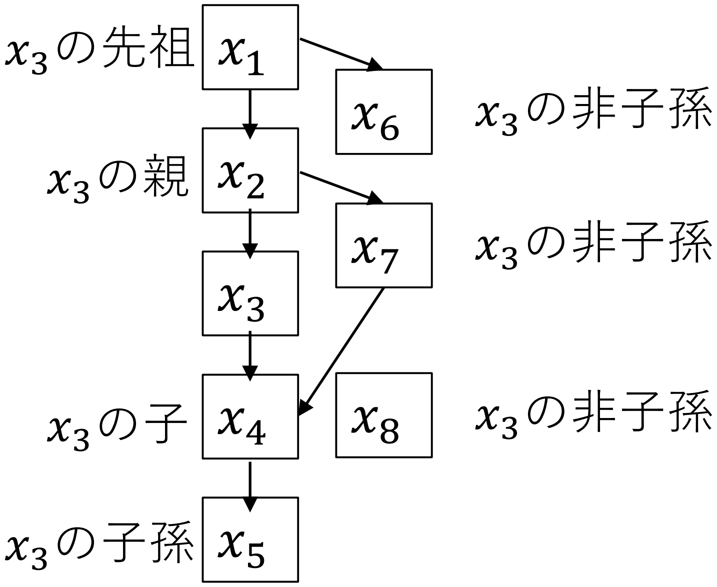
今までの要点（識別可能性）

・さらに因果関係の非巡回性・誤差変数の独立性を仮定

→　$x_i = f_i(pa(x_i), e_i) \quad (i = 1, \ldots, p)$

・$pa(x_i)$：観測変数x_iの親にあたる観測変数の集合





今までの要点（識別可能性）

・さらに誤差変数が連続変数、関数形が線形だと仮定。

→ $x_i = \sum_{x_j \in pa(x_i)} b_{ij} x_j + e_i \quad (i = 1, \ldots, p)$

今までの要点（識別可能性）

・ノンパラメトリック・アプローチ
→有向辺の向きを絞りきれない（詳細は省略）
→因果グラフが識別可能ではない

・パラメトリック・アプローチ
→同じ観測変数の分布を与えるような構造方程式モデルが複数ある（詳細は省略）
→因果グラフが識別可能ではない

・セミパラメトリック・アプローチ（特にLiNGAMアプローチ）
→モデルが異なる場合に、観測変数の分布が必ず異なる
→観測変数の分布からモデルを一意に推測できる場合がある（識別可能性あり）

LiNGAMモデル

LiNGAMモデル：セミパラメトリックアプローチのうち、因果グラフが識別可能なモデル

→説明のために、「独立成分分析」を紹介

独立成分分析

・独立成分分析：未観測変数の値が混ざり合って、観測変数の値が生成されると考える

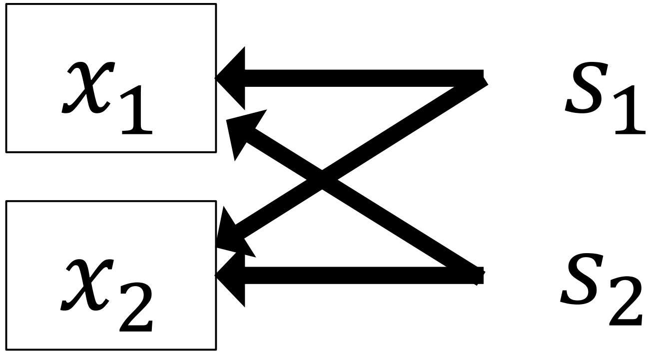
独立成分分析モデル

構造方程式

$$x_1 = a_{11}s_1 + a_{12}s_2$$
$$x_2 = a_{21}s_1 + a_{22}s_2$$

x_1, x_2：観測変数
$a_{11}, a_{12}, a_{21}, a_{22}$：係数（定数）
s_1, s_2：未観測の連続変数

因果グラフ



行列表現

$$\underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{x} = \underbrace{\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} s_1 \\ s_2 \end{bmatrix}}_{s}$$

独立成分分析モデル

◎独立成分分析の目的
①モデルから生成されるデータ行列を用いて係数を推定
②未観測変数を復元

◎独立成分分析の特徴
①未観測変数（s_1, s_2）が独立
②かつ、非ガウス連続分布に従う

→独立性を利用して、未観測変数を復元

独立成分分析モデル

◎定義

一般に、p個の観測変数$x_i(i = 1, \ldots p; p \geq 2)$の独立成分分析モデルは、

$$x_i = \sum_{j=1}^{q} a_{ij} s_j \quad (i = 1, \ldots, p)$$

と表せる。

独立成分分析モデル

◎定義

行列を用いると、

$$x = As$$

と表せる。

A:混合行列（独立成分がどのように混ざり観測変数になるかを表す）

→Aは、列の順序と尺度を除いて識別可能

行列表現

$$\underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{x} = \underbrace{\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} s_1 \\ s_2 \end{bmatrix}}_{s}$$

独立成分分析モデル

$$\mathrm{A_{ICA}} = ADP$$

$\mathrm{A_{ICA}}$：混合行列Aと、列の順序と尺度だけ
が異なる可能性のある行列
P：$q \times q$の置換行列(Aの列の順序を変えるかもしれない)
D：$q \times q$の対角行列（Aの列の尺度を変えるかもしれない）

※P, Dは未知

独立成分分析モデル

$$\underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{x} = \underbrace{\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} s_1 \\ s_2 \end{bmatrix}}_{s}$$

Aの1列目と2列目を交換

$$\underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{x} = \underbrace{\begin{bmatrix} a_{12} & a_{11} \\ a_{22} & a_{21} \end{bmatrix}}_{A\,\text{置換後}} \underbrace{\begin{bmatrix} s_2 \\ s_1 \end{bmatrix}}_{s\,\text{置換後}}$$

s_1, s_2を置換

$$\underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{x} = \underbrace{\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} s_1 \\ s_2 \end{bmatrix}}_{s}$$

Aの1列目をc倍

$$\underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{x} = \underbrace{\begin{bmatrix} a_{11} \times c & a_{12} \\ a_{21} \times c & a_{22} \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} s_{1/c} \\ s_2 \end{bmatrix}}_{s}$$

s_1,を1/c倍

独立成分分析モデル

→次のような行列とベクトルの組であれば、独立成分分析の
　仮定を満たす

$$\mathrm{A_{ICA}} = ADP$$
$$s_{ICA} = P^T D^{-1} s$$

$$\begin{aligned}(x &= As \\ &= \mathrm{ADP}P^T D^{-1} s \\ &= (ADP)(P^T D^{-1} s) \\ &= A_{ICA} S_{ICA})\end{aligned}$$

独立成分分析モデル

◎データから混合行列Aを推定（※混合行列Aは正方行列であると仮定）

→観測変数ベクトルxをp×pの行列Wで線形変換してつくったベクトル

$y = Wx$によって、独立変数ベクトルsを推定することを考える

W = A^{-1}となれば、

$$s = Wx(= A^{-1}As = s)$$

によって、独立成分ベクトルsを復元できる

独立成分分析モデル

（続き）

ベクトルyの成分y_j（$j = 1, \dots, p$）の独立性が最大になるようなWを探す

∵ベクトルyで推定しようとしている独立成分ベクトルsの成分$s_j (j = 1, \dots, p)$が独立

独立成分分析モデル

（続き）

独立性の指標：相互情報量

・非負の値をとる

・変数が独立のときに限り、相互情報量は０

→相互情報量が最小になるようにWを推定

独立成分分析モデル

（続き）

ベクトルyの相互情報量

$$I(y) = \left\{ \sum_{j=1}^{q} H\left(y_j\right) \right\} - H(y)$$

$\rightarrow W_{ICA} = PDW = (PDA^{-1})$

LiNGAMモデル

LiNGAMモデル：セミパラメトリックアプローチのうち、因果グラフが識別可能なモデル

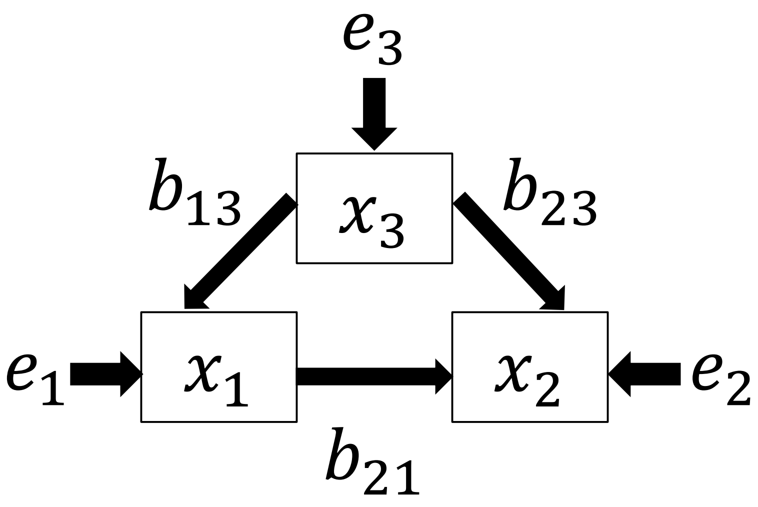
→係数行列Bを観測変数の分布$p(x)$に基づいて一意に推定可能

構造方程式モデル

$$x_i = \sum_{j \neq i} b_{ij} x_j + e_i$$

行列表現

$$x = Bx + e$$



因果グラフ

ー線形

ー非巡回

ー誤差変数e_i

・非ガウス連続分布

・独立（未観測共通原因なし）

LiNGAMモデル

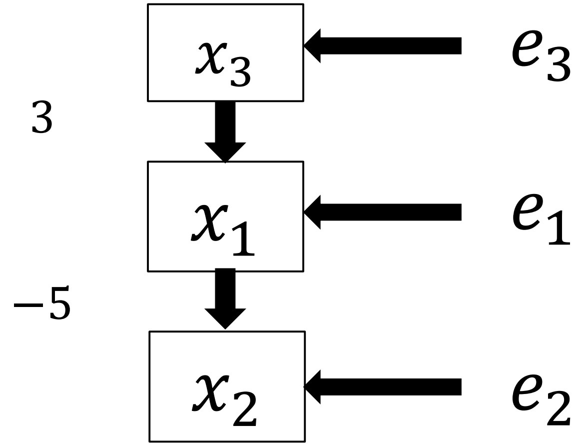
・観測変数の「因果的順序」
：その順序に従って変数を並び替えると、「順番が後の変数」が「順番が後の変数」の原因になることがない順序

→上のような観測変数$x_1, x_2, \ldots, x_p$の順序を
$k(1), k(2), \ldots, k(p)$と表す。

LiNGAMモデル

- 構造方程式モデルと因果グラフ

$$x_1 = 3x_3 + e_1$$
$$x_2 = -5x_1 + e_2$$
$$x_3 = e_3$$



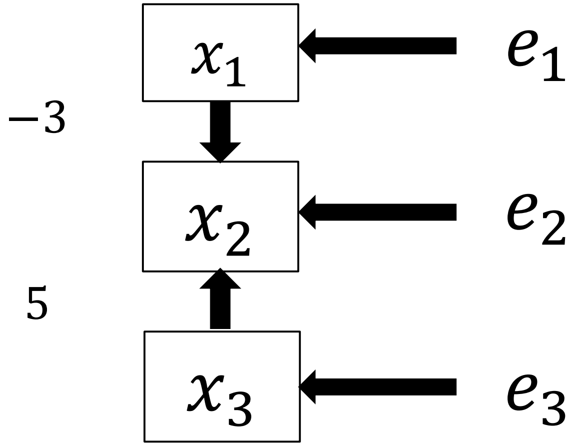
- 因果順序$k(3) = 1, k(1) = 2, k(2) = 3$で並び替え

$$\underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}}_{x} = \underbrace{\begin{bmatrix} 0 & 0 & 3 \\ -5 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}}_{B} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}}_{x} + \underbrace{\begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix}}_{e} \qquad \underbrace{\begin{bmatrix} x_3 \\ x_1 \\ x_2 \end{bmatrix}}_{x_{後}} = \underbrace{\begin{bmatrix} 0 & 0 & 0 \\ 3 & 0 & 0 \\ 0 & -5 & 0 \end{bmatrix}}_{B_{後}} \underbrace{\begin{bmatrix} x_3 \\ x_1 \\ x_2 \end{bmatrix}}_{x_{後}} + \underbrace{\begin{bmatrix} e_3 \\ e_1 \\ e_2 \end{bmatrix}}_{e_{後}}$$

LiNGAMモデル

- 構造方程式モデルと因果グラフ

$$x_1 = e_1$$
$$x_2 = -3x_1 + 5x_3 + e_2$$
$$x_3 = e_3$$



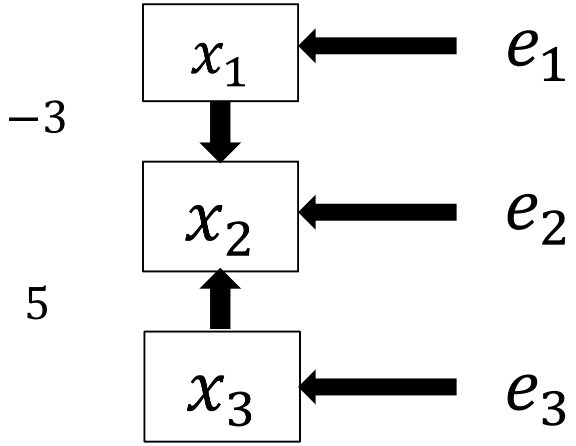
- 因果順序$k(3) = 1, k(1) = 2, k(2) = 3$で並び替え

$$\underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}}_{x} = \underbrace{\begin{bmatrix} 0 & 0 & 0 \\ -3 & 0 & 5 \\ 0 & 0 & 0 \end{bmatrix}}_{B} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}}_{x} + \underbrace{\begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix}}_{e} \qquad \underbrace{\begin{bmatrix} x_3 \\ x_1 \\ x_2 \end{bmatrix}}_{x_{後}} = \underbrace{\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 5 & -3 & 0 \end{bmatrix}}_{B_{後}} \underbrace{\begin{bmatrix} x_3 \\ x_1 \\ x_2 \end{bmatrix}}_{x_{後}} + \underbrace{\begin{bmatrix} e_3 \\ e_1 \\ e_2 \end{bmatrix}}_{e_{後}}$$

LiNGAMモデル

・構造方程式モデルと因果グラフ

$$x_1 = e_1$$
$$x_2 = -3x_1 + 5x_3 + e_2$$
$$x_3 = e_3$$



・因果順序$k(3) = 1, k(1) = 2, k(2) = 3$で並び替え

$$\underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}}_{x} = \underbrace{\begin{bmatrix} 0 & 0 & 0 \\ -3 & 0 & 5 \\ 0 & 0 & 0 \end{bmatrix}}_{B} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}}_{x} + \underbrace{\begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix}}_{e} \qquad \underbrace{\begin{bmatrix} x_1 \\ x_3 \\ x_2 \end{bmatrix}}_{x_{後}} = \underbrace{\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 5 & -3 & 0 \end{bmatrix}}_{B_{後}} \underbrace{\begin{bmatrix} x_1 \\ x_3 \\ x_2 \end{bmatrix}}_{x_{後}} + \underbrace{\begin{bmatrix} e_1 \\ e_3 \\ e_2 \end{bmatrix}}_{e_{後}}$$

LiNGAMモデル

$x = Bx + e$

⬇ 観測変数ベクトルxについて解く

$(I - B)x = e$

⬇ 行列$I - B$の逆行列を両辺にかける

$(I - B)^{-1}(\mathrm{I} - \mathrm{B})\mathrm{x} = (I - B)^{-1}e$

⬇ $(I - B)^{-1} = A$とおく

$x = (I - B)^{-1}e$
$\quad = Ae$

- （p次元ベクトル）x：観測変数ベクトル
- （$p \times p$の行列）B：係数行列
- （p次元ベクトル)e：誤差変数ベクトル

※誤差変数ベクトルeの成分$e_i(i = 1, \dots, p)$は独立、かつ、それぞれ非ガウス連続分布に従う

LiNGAMモデル

- LiNGAMモデル

$$x_1 = e_1$$
$$x_2 = b_{21}x_1 + e_2$$

- 行列表現

$$\underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{x} = \underbrace{\begin{bmatrix} 0 & 0 \\ b_{21} & 0 \end{bmatrix}}_{B} \underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{x} + \underbrace{\begin{bmatrix} e_1 \\ e_2 \end{bmatrix}}_{e}$$

LiNGAMモデル

$$\underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{x} = \underbrace{\left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ b_{21} & 0 \end{bmatrix} \right)^{-1}}_{(I-B)^{-1}} \underbrace{\begin{bmatrix} e_1 \\ e_2 \end{bmatrix}}_{e}$$

$$= \underbrace{\begin{bmatrix} 1 & 0 \\ -b_{21} & 1 \end{bmatrix}^{-1}}_{W^{-1}} \underbrace{\begin{bmatrix} e_1 \\ e_2 \end{bmatrix}}_{e}$$

$$= \underbrace{\begin{bmatrix} 1 & 0 \\ b_{21} & 1 \end{bmatrix}^{-1}}_{A} \underbrace{\begin{bmatrix} e_1 \\ e_2 \end{bmatrix}}_{e}$$

ここより

$$W = \begin{bmatrix} 1 & 0 \\ -b_{21} & 1 \end{bmatrix}$$

LiNGAMモデル

対角行列

$$D=\begin{bmatrix} d_{11} & 0 \\ 0 & d_{22} \end{bmatrix} \quad DW=\begin{bmatrix} d_{11} & 0 \\ -d_{22}b_{21} & d_{22} \end{bmatrix}$$

置換行列

$$\mathrm{P}_{12}=\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \qquad P_{12}DW=\begin{bmatrix} -d_{22}b_{21} & d_{22} \\ d_{11} & 0 \end{bmatrix}$$

LiNGAMモデル

$$\mathrm{W_{ICA}} = P_{12}DW$$
$$= \begin{bmatrix} -d_{22}b_{21} & d_{22} \\ d_{11} & 0 \end{bmatrix}$$

←行の順序が正しくない

$$\tilde{P}\mathrm{W_{ICA}} = \tilde{P}P_{12}DW$$
$$= \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} -d_{22}b_{21} & d_{22} \\ d_{11} & 0 \end{bmatrix}$$
$$= \begin{bmatrix} d_{11} & 0 \\ -d_{22}b_{21} & d_{22} \end{bmatrix}$$