構造方程式モデルによるデータ生成過程の学習
– 特に非ガウス性の利用

清水昌平 (大阪大学)

Abstract

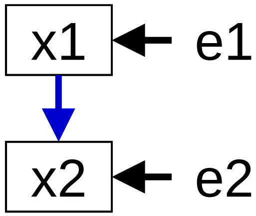
- 構造方程式モデルはデータ生成過程のモデル
  - 反実仮想＋構造方程式モデル→因果分析
- データ生成過程が未知の時に、
  構造方程式モデルを学習する方法を概観する
  (連続変数の場合)
- まずは、基本となる線形モデル
- 最後、非線形の方法にも触れる

イントロ (1/2)

- データ行列 X が次のどちらかのデータ生成過程からランダムに生成されたとしよう $(b_{21}, b_{12} \neq 0)$:

モデル 1:

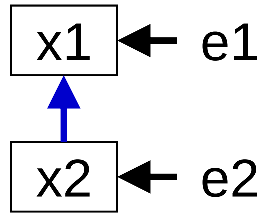
$$x_1 = e_1$$
$$x_2 = b_{21} x_1 + e_2$$



or

モデル 2:

$$x_1 = b_{12} x_2 + e_1$$
$$x_2 = e_2$$



ここで e_1 と e_2 は独立な潜在変数 (外的影響, かく乱項、誤差)

- データ行列Xのみを用いて、データXを生成したのがモデル1 なのか モデル2 なのかを同定したい

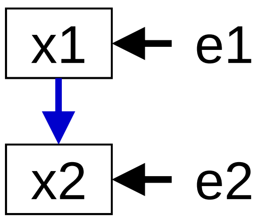
イントロ (2/2)

- 「同定はできない」と長らく思われていた
- 実は、「ほとんどの場合に同定可能」ということが最近分かってきた (Shimizu et al., 2006)
  - **e1**と**e2**がガウス分布に従うとダメ
- 非線形＋加法誤差でもポジティブな結果 (Hoyer et al., 2009; Zhang & Hyvarinen, 2009)

モデル 3:

$$x_1 = e_1$$

$$x_2 = f_2(x_1) + e_2$$

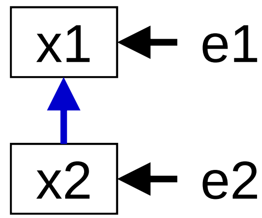


or

モデル 4:

$$x_1 = f(x_2) + e_1$$

$$x_2 = e_2$$



因果効果: モデル1が正しいとき

(cf. Pearl, 2000)

- x_1を定数 c から d へ変化させたときのx_2 への因果効果
  = E(x_2 | 母集団の全員のx_1 を強制的に d にする)
  - E(x_2 | ・・・ を強制的に c にする)

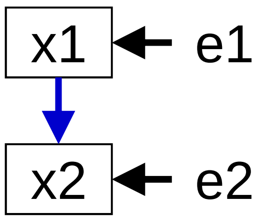
$$= E(b_{21}\, d + e_2) - E(b_{21}\, c + e_2)$$

$$= \underline{b_{21}(d - c)}$$

モデル1:

$$x_1 = e_1$$

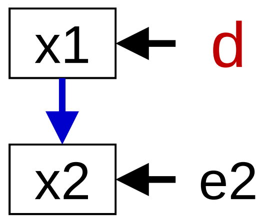
$$x_2 = b_{21} x_1 + e_2$$



モデル1':

$$x_1 = \underline{d}$$

$$x_2 = b_{21}\, \underline{d} + e_2$$



因果効果: モデル2が正しいとき

(cf. Pearl, 2000)

- x_1を定数 c から d へ変化させたときのx_2への因果効果

  = E(x_2 | 母集団の全員のx_1を強制的に d にする)

  - E(x_2 | ・・・ を強制的に c にする)

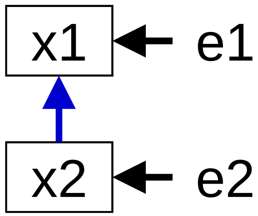
  = $E(e_2) - E(e_2)$

  = $\underline{0}$

モデル2:

$$x_1 = b_{12} x_2 + e_1$$

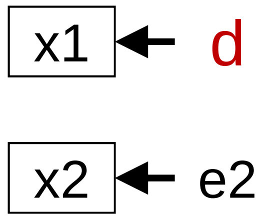
$$x_2 = e_2$$



モデル2':

$$x_1 = \underline{d}$$

$$x_2 = e_2$$

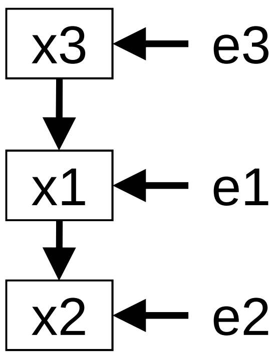


問題の定式化

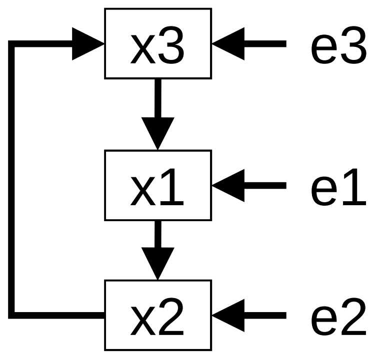
Basic problem setup (1/3)

- 仮定: 連続な観測変数 x_i のデータ生成過程が、グラフィカルには非巡回有効グラフ (DAG)である
  - ループがない

非巡回有向グラフの例 (DAG):



巡回有向グラフの例:



(□(四角)で囲まれているのは観測変数)

Basic problem setup (2/3)

- さらに、x_i の線形関係を仮定すると
  線形**非巡回**構造方程式モデル (Wright, 1921; Bollen, 1989):

$$x_i = \sum_{j:\, x_i\text{の親}} b_{ij} x_j + e_i \qquad \text{or} \qquad \mathbf{x} = \mathbf{B}\mathbf{x} + \mathbf{e}$$

  – e_i は、モデル内で規定されない(外的)連続な潜在変数:
    ここでは、**外的影響**と呼ぶ (かく乱変数、誤差変数).
  – e_i は、非ゼロの分散を持ち、互いに**独立**

例

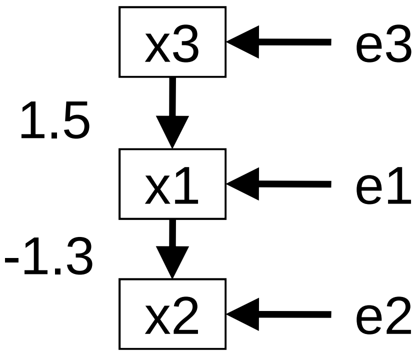
- 3変数の場合:

$$x_1 = 1.5x_3 + e_1$$
$$x_2 = -1.3x_1 + e_2$$
$$x_3 = e_3$$

or

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

- **B**のゼロ/非ゼロパターンが、1つのDAGに対応する:



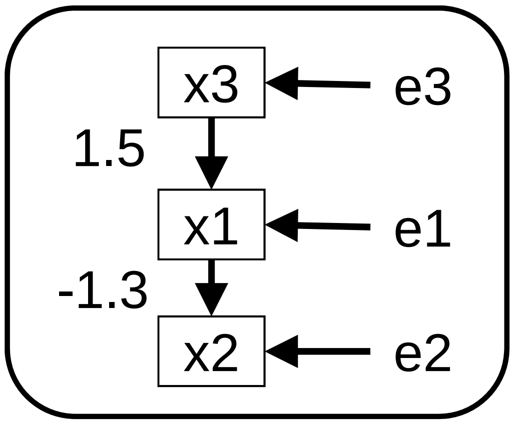
$b_{ij} = 0 \Rightarrow x_j$ から x_i に有向辺がない

$b_{ij} \neq 0 \Rightarrow x_j$ から x_i に有向辺がある

非巡回性の仮定

- 非巡回の場合は、パス係数行列Bを下三角にするような変数 x_i の順序が必ず存在する (Bollen, 1989).

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



そんな順序は例えば:

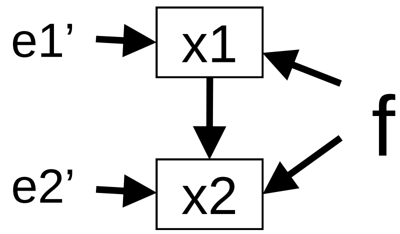
$x_3 < x_1 < x_2$.

x_3 は、x_1, x_2 の祖先でもよいが,

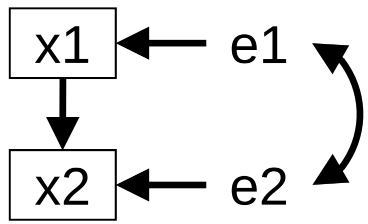
逆はない.

外的影響の独立性の仮定

- 「未観測交絡変数がない」ことを意味する (Spirtes et al. 2000)
  - 未観測交絡変数とは、2つ以上の観測変数の親であるような潜在変数:



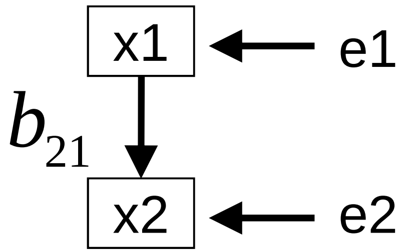
- 未観測交絡変数があると、外的影響が従属



Basic problem setup (3/3): 線形構造方程式モデルの学習

- 仮定 : データ行列 **X** は、モデルからランダムに生成される:

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



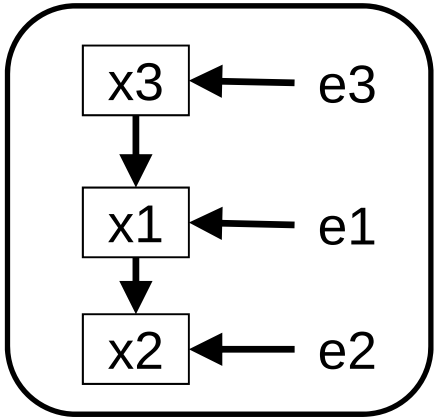
- **Goal:** データ行列**X**の情報のみを使って、パス係数行列 **B** を推定する!
  - **B**のゼロ/非ゼロパターンが、1つのDAGに対応する

推定原理

伝統的な推定原理: Causal Markov condition

- もしデータ生成過程が、(線形)非巡回構造方程式モデルなら、Causal Markov condition が成り立つ:
  - 各観測変数 x_i は、親で条件付けると非子孫と独立 (Pearl & Verma, 1991)：

$$p(\mathbf{x}) = \prod_{i=1}^{p} p\left(x_i \mid x_i\text{の親}\right)$$



- ノンパラメトリックの場合でも使える原理だが、一意に同定できないモデルが多い

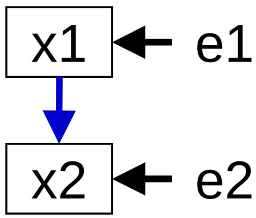
例

- 有向辺の向きが反対の2つのモデル:

モデル 1:

$$x_1 = e_1$$

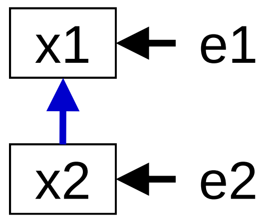
$$x_2 = 0.8x_1 + e_2$$



モデル 2:

$$x_1 = 0.8x_2 + e_1$$

$$x_2 = e_2$$



$$E(e_1) = E(e_2) = 0, \mathrm{var}(x_1) = \mathrm{var}(x_2) = 1$$

- どちらのモデルでも、**(条件付き)独立になる変数は****ない**:

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

- Causal Markov condition では、2つのモデルを区別できない

新しい推定原理：外的影響の独立性

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

- 「外的影響 e_i が独立である」ことを利用
  - 単に無相関ではなく

- 線形構造方程式モデル＋非ガウス性(LiNGAMモデル):

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

  - 外的影響 e_i は**非ガウス**かつ**互いに独立**

- モデル つまり Bを一意に同定できる

推定アルゴリズム

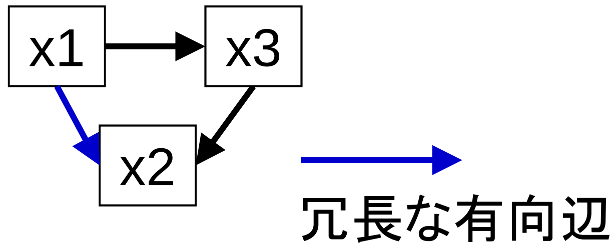
DirectLiNGAM アルゴリズム

(Shimizu et al., 2009; Sogawa et al., 2010; Inazumi et al., 2010)

- パス係数行列Bが下三角になるような変数x_iの順序を推定する(非巡回にするような順序)

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

A full DAG



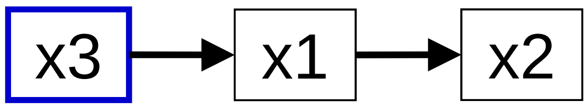
- 必要なら、枝刈りは既存の方法(e.g.,スパース正則化)が適用可能 (Zou, 2006; Hyvarinen et al. 2010)

Basic idea (1/2) : 外生変数は、正しい順序のトップに来れる

- 外生変数 x_j は、親のいない変数 (Bollen, 1989)
  - ここでは x_3
  - **パス係数行列 Bの対応する行の成分は全てゼロ**

- 外生変数はパス係数行列Bを下三角にするような変数順序のトップに来れる

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



Basic idea (2/2): 外生変数 x_3 の成分を取り除く

- 他の変数 $x_i\ (i=1,2)$ を外生変数 x_3 に回帰して残差 $r_i^{(3)}\ (i=1,2)$ を計算する
  - 残差 $r_1^{(3)}$ と $r_2^{(3)}$ も、LiNGAMモデルを形成する
  - 残差の順序は、元の観測変数の順序と同じ

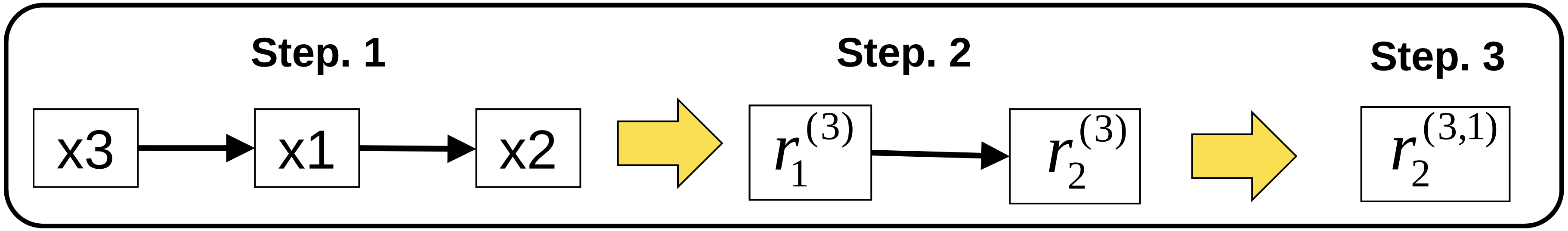
$$\begin{bmatrix} x_3 \\ x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 1.5 & 0 & 0 \\ 0 & -1.3 & 0 \end{bmatrix} \begin{bmatrix} x_3 \\ x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} e_3 \\ e_1 \\ e_2 \end{bmatrix} \Rightarrow \begin{bmatrix} r_1^{(3)} \\ r_2^{(3)} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ -1.3 & 0 \end{bmatrix} \begin{bmatrix} r_1^{(3)} \\ r_2^{(3)} \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

$$\boxed{\text{x3}} \rightarrow \boxed{\text{x1}} \rightarrow \boxed{\text{x2}} \qquad \boxed{r_1^{(3)}} \rightarrow \boxed{r_2^{(3)}}$$

- 残差 $r_1^{(3)}$ が外生なので、x_1 はトップから2番目に来れる

DirectLiNGAMの流れ

- 全ての変数が順序づけられるまで、繰り返し「外生」変数を見つける:
  1. 外生変数を見つける: x_3
     - x_3 を順序のトップに置く
     - 回帰して、x_3 の成分を取り除く
  2. 外生的な残差を見つける: $r_1^{(3)}$
     - x_1 を順序のトップから2番目に置く
     - 回帰して、$r_1^{(3)}$の成分を取り除く
  3. x_2 を順序のトップから3番目に置いて、終了
     推定された順序は $x_3 < x_1 < x_2$



外生変数の同定

- 外的影響が非ガウスの場合

- Lemma 1: x_j は その残差 $r_i^{(j)} = x_i - \frac{\mathrm{cov}(x_i, x_j)}{\mathrm{var}(x_j)} x_j$ のどれとも独立 (i は j 以外全部) $\Leftrightarrow$ x_j は外生変数

- 残差と最も独立な変数を見つけることによって、外生変数を同定できる
  - 独立性の指標 (Bach & Jordan 2002 etc.)
  - ペアワイズに評価して和をとる

外生変数の同定（2変数の場合）

i) $x_1 (= e_1)$ は外生変数

$$x_1 = e_1$$

$$x_2 = b_{21} x_1 + e_2 \quad (b_{21} \neq 0)$$

x_2 を x_1 に回帰して,

$$r_2^{(1)} = x_2 - \frac{\mathrm{cov}(x_2, x_1)}{\mathrm{var}(x_1)} x_1$$

$$= x_2 - b_{21} x_1$$

$$= e_2$$

x_1 と $r_2^{(1)}$ は独立

ii) x_1 は外生変数でない

$$x_1 = b_{12} x_2 + e_1 \quad (b_{12} \neq 0)$$

$$x_2 = e_2$$

x_2を x_1 に回帰して,

$$r_2^{(1)} = x_2 - \frac{\mathrm{cov}(x_2, x_1)}{\mathrm{var}(x_1)} x_1$$

$$= \left\{ 1 - \frac{b_{12}\,\mathrm{cov}(x_2, x_1)}{\mathrm{var}(x_1)} \right\} x_2 - \frac{b_{12}\,\mathrm{var}(x_2)}{\mathrm{var}(x_1)} e_1$$

x_1 と $r_2^{(1)}$ は独立でない

Darmois-Skitovitch' theorem

(Darmois, 1953; Skitovitch, 1953)

Darmois-Skitovitch' theorem:

変数 x_1 と x_2 を次のように定義する:

$$x_1 = \sum_{j=1}^{p} a_{1j} e_j, \quad x_2 = \sum_{j=1}^{p} a_{2j} e_j$$

ここで e_j は独立な確率変数.

もし $a_{1i} a_{2i} \neq 0$ となるような非ガウスな e_i があれば、
x_1 と x_2 は独立でない

ii) x_1 は外生変数でない

$$x_1 = b_{12} x_2 + 1 \cdot e_1 \quad (b_{12} \neq 0)$$

$$x_2 = e_2$$

x_2 を x_1 に回帰して,

$$r_2^{(1)} = x_2 - \frac{\mathrm{cov}(x_2, x_1)}{\mathrm{var}(x_1)} x_1$$

$$= \left\{ 1 - \frac{b_{12} \mathrm{cov}(x_2, x_1)}{\mathrm{var}(x_1)} \right\} x_2 - \frac{b_{12} \mathrm{var}(x_2)}{\mathrm{var}(x_1)} e_1$$

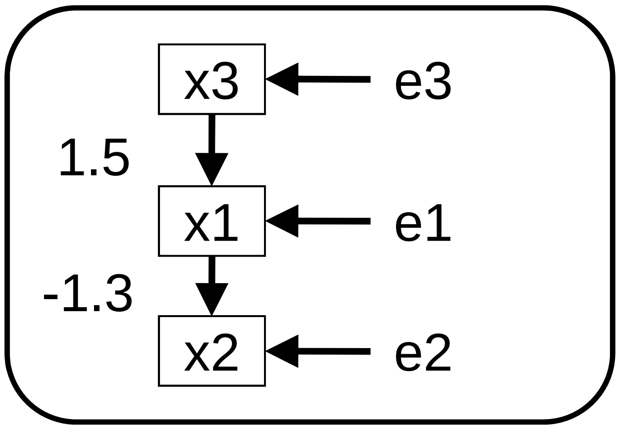
x_1 と $r_2^{(1)}$ は独立でない

DirectLiNGAMの大事な性質

- DirectLiNGAM は次の2つを繰り返す:
  - 単回帰
  - 各変数と残差の独立性評価
- Algorithmic parameters がない
  - ステップサイズ、初期値、収束基準
- データがモデルに厳密に従っていれば、
  変数の数と同じステップ数で
  正しい解に収束することを保証できる
  (収束はいつもする)

Causal Markov condition との関係は？

- 次の2つの性質は同値:
  (Zhang & Hyvarinen, ECML09; Hyvarinen et al., JMLR, 2010)
  1. 外的影響 e_i が互いに独立
  2. Causal Markov condition が成り立つ
     ＋ 各変数 x_i の親と外的影響 e_i が独立



非線形の場合

非線形＋加法の外的影響

- 「非線形＋加法の外的影響」のモデル:
  - – DAG; 未観測交絡変数なし

1. $$x_i = f_i\left(x_i \text{ の親} \right) + e_i$$ -- Hoyer et al. (NIPS08)
2. $$x_i = f_{i,2}^{-1}\left(f_{i,1}\left(x_i \text{ の親} \right) + e_i \right)$$ -- Zhang et al. (UAI09)

- 「外的影響の独立性」の推定原理が使える
- 2変数の場合、いくつかの非線形性と外的影響の分布の組み合わせを除いて、一意に同定可能

ノンパラメトリック

- DAG + 未観測交絡変数なし:
  PC algorithm (Spirtes & Glymour, 1991)

$$x_i = f_i\left(x_i \text{ の親}_i , e_i \right)$$

- 「外的影響の独立性」の推定原理は使えない

- 多くの場合、一意に同定できないが、関数形について事前知識がないのであれば、現状これ

おわりに

- データ生成過程が未知の時に、構造方程式モデルを学習する方法を概観した
- 線形モデルとノンパラトリックモデルの性質は結構わかってきている
  - その間: 非線形＋加法の外的影響?
  - ナナメ?: 線形＋各種拡長
    - 非巡回 → 巡回 (Lacerda et al., UAI2008)
    - 単一母集団 → 異質な母集団の混合 (Shimizu et al., ICONIP2007)
    - i.i.d. サンプリング → 時間構造 (Hyvarinen et al, JMLR, 2010)
    - 未観測交絡変数なし→ あり (Hoyer et al., IJAR, 2008; Kawahara et al. 2010)