統計的機械学習　（応用計量分析２）第7回

構造方程式モデルとバックドア基準（参考pdf 10章）

決定森と深層学習を用いた手法

- 因果木は出力をτとし分割学習用と予測値推定用にデータ分割した決定木
- 因果森は因果木をベース学習器とするランダム森
  - ランダム森はデータをサブサンプリングして学習した決定木を平均したもの
- BARTはベイズ推論を決定森モデルに適用したもの
- 深層学習を用いるとIPMによる正則化を敵対的学習で実現可
  - 重み付けとは異なるアプローチ
- 小テストとコーディングを含む記述課題を出しました（11/11締切）

（参考）敵対的ドメイン適応の理論

出欠コード
XXXX

（観測できない）x による損失の差を $\sup$ で置き換える

- 理論
  - 表現抽出器 ϕ が逆関数を持つなら、一様分布上の損失 MSE^u が、データ分布上の損失 MSE と分布間IPMを用いて上から抑えられる
  - → MSE + IPM を最小化すれば MSE^u が抑えられる

CounterFactual error

$$\epsilon_{CF}(h,\Phi) \leq (1-u)\epsilon_F^{t=1}(h,\Phi) + u\epsilon_F^{t=0}(h,\Phi) + B_\Phi \cdot IPM_G\left(p_\Phi^{t=1}, p_\Phi^{t=0}\right),$$

- Φ 上の真のモデル h と損失 L の合成
  - $$\ell_{h,\Phi}(x,t) = \int_{\mathcal{Y}} L(Y_t, h(\Phi(x),t))p(Y_t|x)dY_t$$
  - が関数クラスGに入るように定数(パラメタ) B_Φ を決めれば、その未知の関数 $\ell_{h,\Phi}$ に関して最悪 sup をとれば上界になる

Lemma A4 (Lemma 1, main text)**.** *Let $\Phi : \mathcal{X} \to \mathcal{R}$ be an invertible representation with Ψ its inverse. Let $p_\Phi^{t=1}, p_\Phi^{t=0}$ be defined as in Definition A3. Let $u = p(t=1)$. Let G be a family of functions $g : \mathcal{R} \to \mathbb{R}$, and denote by $IPM_{\mathrm{G}}(\cdot,\cdot)$ the integral probability metric induced by G. Let $h : \mathcal{R} \times \{0,1\} \to \mathcal{Y}$ be an hypothesis. Assume there exists a constant $B_\Phi > 0$, such that for $t = 0, 1$, the function $g_{\Phi,h}(r,t) := \frac{1}{B_\Phi} \cdot \ell_{h,\Phi}(\Psi(r),t) \in \mathrm{G}$. Then we have:*

$$\epsilon_{CF}(h,\Phi) \leq (1-u)\epsilon_F^{t=1}(h,\Phi) + u\epsilon_F^{t=0}(h,\Phi) + B_\Phi \cdot IPM_{\mathrm{G}}\left(p_\Phi^{t=1}, p_\Phi^{t=0}\right). \quad (6)$$

Proof.

$$\epsilon_{CF}(h,\Phi) - \left[(1-u)\cdot\epsilon_F^{t=1}(h,\Phi) + u\cdot\epsilon_F^{t=0}(h,\Phi)\right] =$$
$$\left[(1-u)\cdot\epsilon_{CF}^{t=1}(h,\Phi) + u\cdot\epsilon_{CF}^{t=0}(h,\Phi)\right] - \left[(1-u)\cdot\epsilon_F^{t=1}(h,\Phi) + u\cdot\epsilon_F^{t=0}(h,\Phi)\right] =$$
$$(1-u)\cdot\left[\epsilon_{CF}^{t=1}(h,\Phi) - \epsilon_F^{t=1}(h,\Phi)\right] + u\cdot\left[\epsilon_{CF}^{t=0}(h,\Phi) - \epsilon_F^{t=0}(h,\Phi)\right] = \quad (7)$$
$$(1-u)\int_{\mathcal{X}}\ell_{h,\Phi}(x,1)\left(p^{t=0}(x) - p^{t=1}(x)\right)dx + u\int_{\mathcal{X}}\ell_{h,\Phi}(x,0)\left(p^{t=1}(x) - p^{t=0}(x)\right)dx = \quad (8)$$
$$(1-u)\int_{\mathcal{R}}\ell_{h,\Phi}(\Psi(r),1)\left(p_\Phi^{t=0}(r) - p_\Phi^{t=1}(r)\right)dr + u\int_{\mathcal{R}}\ell_{h,\Phi}(\Psi(r),0)\left(p_\Phi^{t=1}(r) - p_\Phi^{t=0}(r)\right)dr =$$
$$B_\Phi\cdot(1-u)\int_{\mathcal{R}}\underline{\frac{1}{B_\Phi}\ell_{h,\Phi}(\Psi(r),1)}\left(p_\Phi^{t=0}(r) - p_\Phi^{t=1}(r)\right)dr + B_\Phi\cdot u\int_{\mathcal{R}}\underline{\frac{1}{B_\Phi}\ell_{h,\Phi}(\Psi(r),0)}\left(p_\Phi^{t=1}(r) - p_\Phi^{t=0}(r)\right)dr \leq \quad (9)$$
$$B_\Phi\cdot(1-u)\sup_{\underline{g\in\mathrm{G}}}\left|\int_{\mathcal{R}}g(r)\left(p_\Phi^{t=0}(r) - p_\Phi^{t=1}(r)\right)dr\right| + B_\Phi\cdot u\sup_{\underline{g\in\mathrm{G}}}\left|\int_{\mathcal{R}}g(r)\left(p_\Phi^{t=1}(r) - p_\Phi^{t=0}(r)\right)dr\right| = \quad (10)$$
$$B_\Phi\cdot\mathrm{IPM}_{\mathrm{G}}(p_\Phi^{t=0}, p_\Phi^{t=1}). \quad (11)$$

本日の内容

出欠コード

XXXX

構造的因果モデルとその活用法を学ぶ

- 1. ガイダンス・因果推論と機械学習の概論
- 2. 意思決定理論（期待効用理論）の復習、因果推論との関係
- 3. 潜在結果モデルに基づく因果推論の枠組み
- 4. 平均因果効果の推定法
- 5. 条件付き平均因果効果（CATE）の推定法１：メタ学習器
  - CATEの推定法２：二重機械学習
- 6. CATEの推定法３：決定木と決定森
  - 深層学習に基づく方法
- **7. 構造方程式モデルとバックドア基準**
- 8. 因果探索
- 9. 発展的な因果推論手法：フロントドア調整、操作変数法、回帰不連続デザイン、代理変数法
- 10. 続き
- 11. 発展的な意思決定理論
- 12. 強化学習
- 13. オフライン強化学習
- 14. バンディット
- 15. まとめ

構造的因果モデル（Structural Causal Model；SCM）の導入

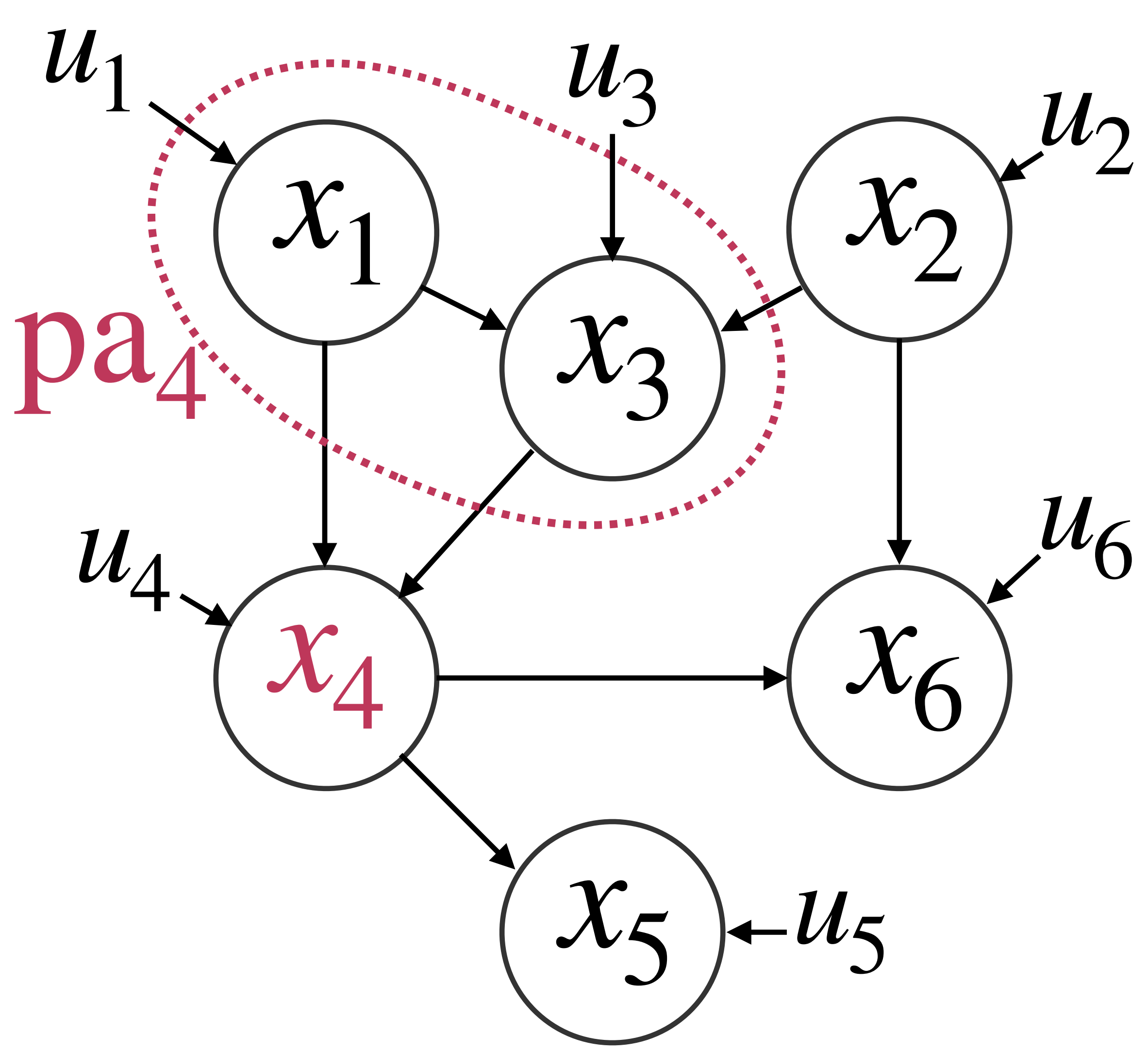
出欠コード
XXXX

個々の変数間の因果関係を表現する方法
因果の流れをグラフで、関係を関数で、不確実性を外生変数で表す

- **因果構造** G：変数間の因果の流れを表す
  - 有向非巡回グラフ (Directed Acyclic Graph; **DAG**) で表す
    - ※巡回を認めるモデルもあるが、「因果」というと非巡回性の仮定を含意することが多い
  - 各変数 x_i は、入ってくる辺で繋がる変数群（親）pa_i のみで条件付けられる

$$p(x_1, x_2, \cdots) = \prod_i p(x_i \mid \mathrm{pa}_i)$$

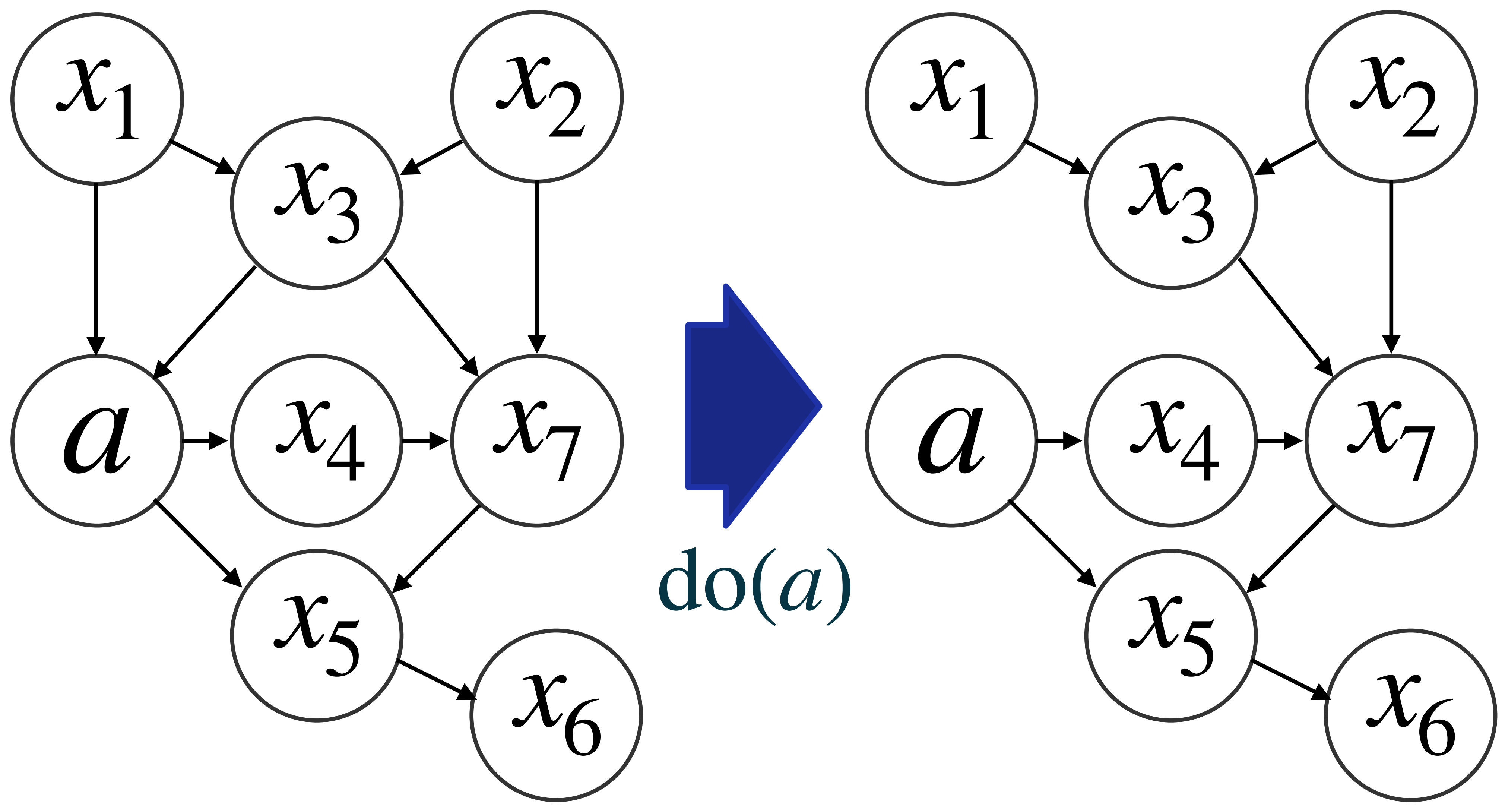
- **因果モデル** M：関数と外生変数（ノイズ）によるモデル化
  - $x_i = f_i(\mathrm{pa}_i, u_i)$ （例: $\beta_i^\top x_{\mathrm{pa}_i} + u_i$ ）
    - 線形モデルを仮定すると、まとめて $x = Bx + u$（Bは三角行列）と書ける
  - $p(u_1, u_2, \cdots) = \prod_i p(u_i)$ （外生変数は独立）
    - ※依存を考慮するモデル化もあるが、特に断らない限り独立を仮定することが多い
  - $M = \left(\{f_i\}_i, \{p(u_i)\}_i\right)$



因果的介入とdo演算子

介入とは、因果構造を部分的に変更し分布を変化させること do演算子は介入の内容を表し、介入後の分布を指定する

- SCMにおける介入=因果モデルの変更
  - 典型的には特定の変数の値の固定
  - $$\begin{cases} \vdots \\ x_{j-1} = f_{j-1}(\mathrm{pa}_{j-1}, \varepsilon_{j-1}) \\ x_j \quad = x' \\ x_{j+1} = f_{j+1}(\mathrm{pa}_{j+1}, \varepsilon_{j+1}) \\ \vdots \end{cases}$$
  - **変更は局所的に留まる**と仮定
    - x_j を変更しても $f_{j'}$ （$j' \neq j$）は不変 ⇒ 介入前の分布の知識が転移可能
- do演算子 $\mathrm{do}(x_j = x')$ は介入を表す
  - 分布の条件部につける：$p(x \mid \mathrm{do}(x_j = x'))$ または $p(x \mid \mathrm{do}(x_j))$
  - 介入後の分布： $p(x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_J \mid \mathrm{do}(x_j)) = \Pi^J_{j' \neq j} p(x_{j'} \mid \mathrm{pa}_{j'})$
- 潜在結果モデルとの対応： $p(y \mid x, \mathrm{do}(a = \alpha)) = p\left(y_{a=\alpha} \mid x\right)$ （xはaより上流）



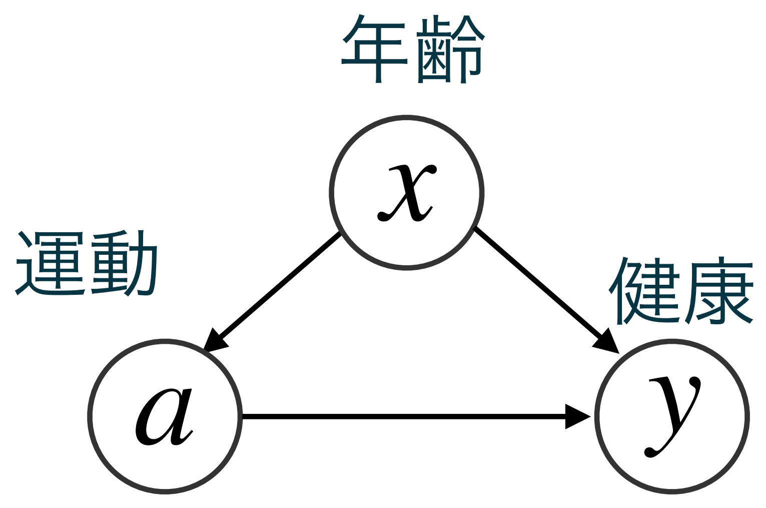
介入前　　　介入後

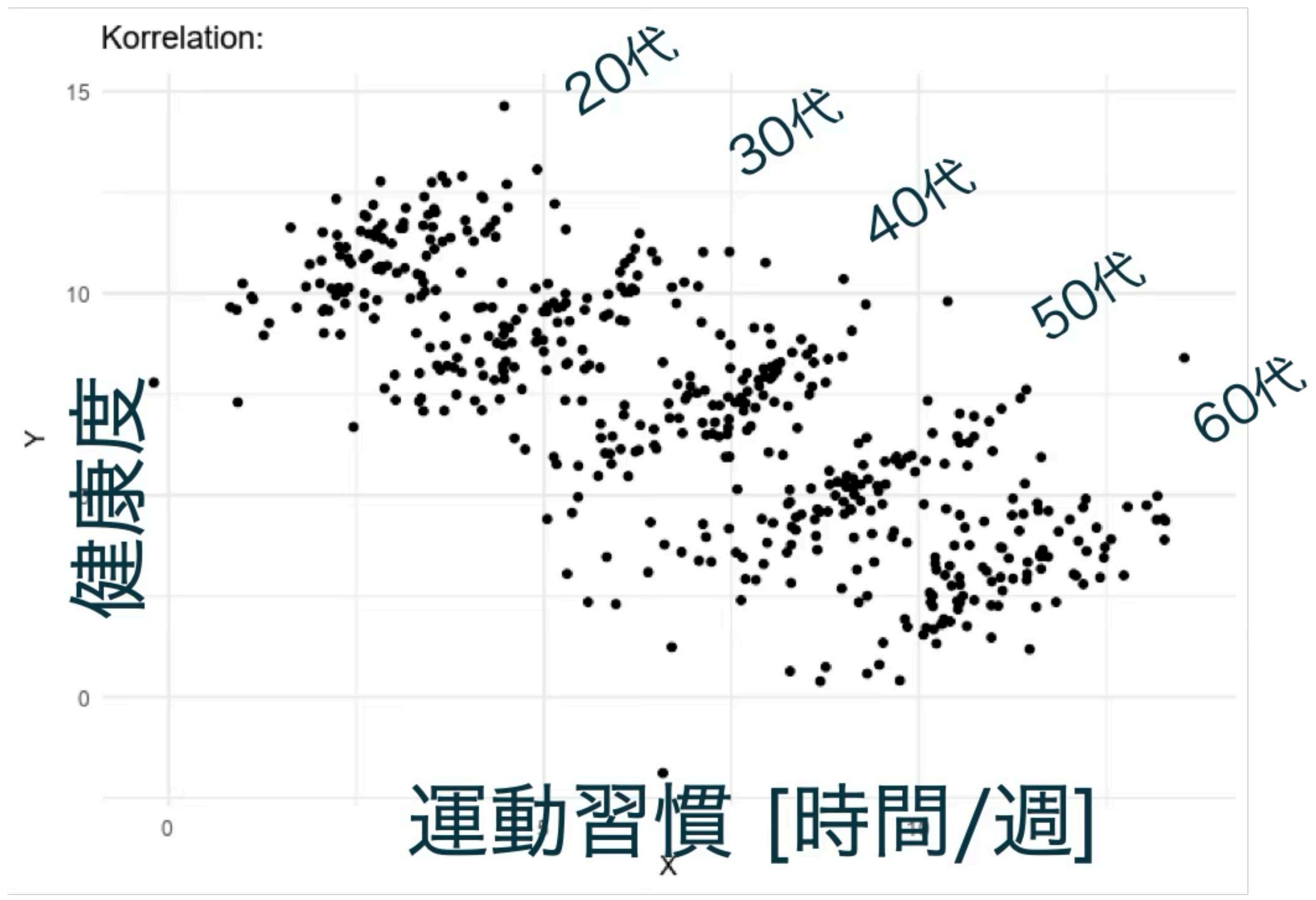
介入は入射する辺の削除

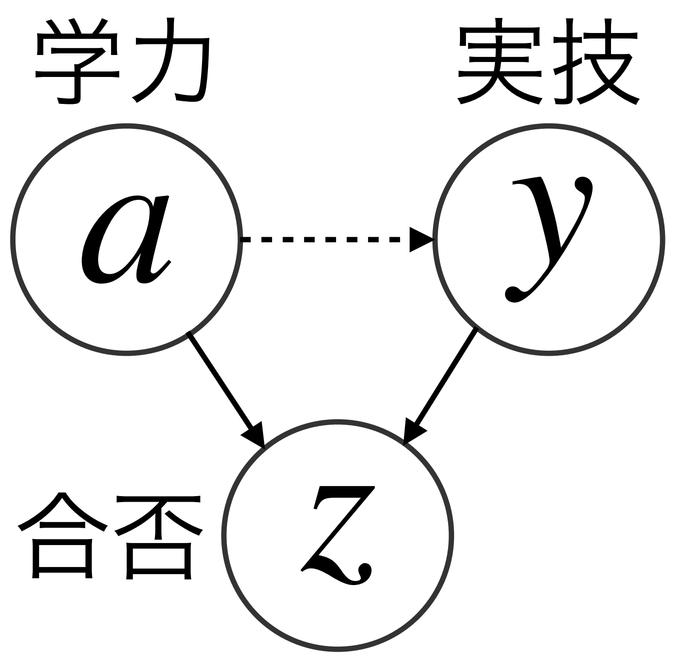
合流点：調整（層別化）するべきでない変数

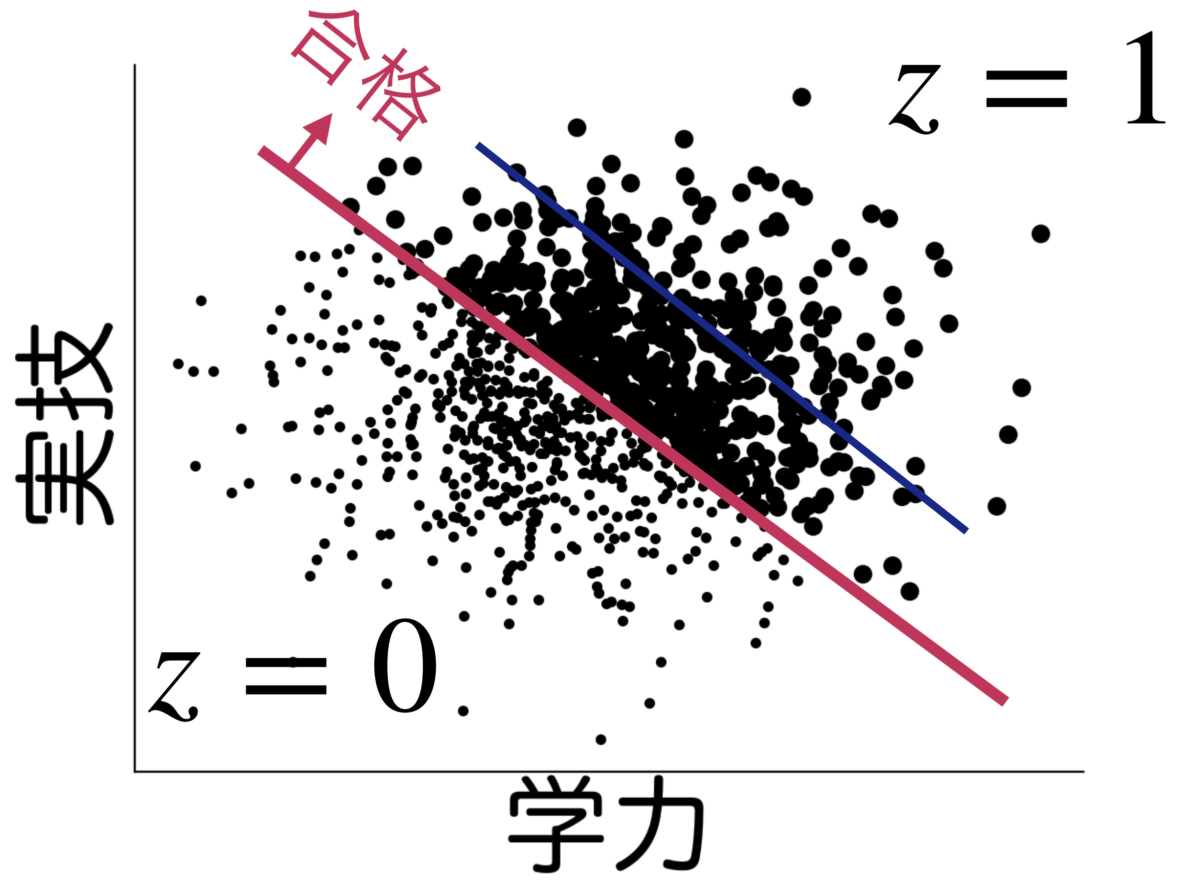
あらゆる変数を調整（層別化等）すれば良いわけではない
合流点を固定すると上流変数間に擬似相関

- 無視可能性のために、できるだけ多くの変数を取得して調整すればいいのか？ → No
- 学力＋実技で選別された合格者の学力と実技の分布は負の相関
  - **選択バイアス**と呼ばれる
- 共通の下流変数（合流点）で層別化すると擬似相関が生まれる
  - 予め層別(選択)されたデータになっていないかも注意が必要
- どの変数を調整すればよいのか？
  →**バックドア基準**を満たすように変数を選ぶ









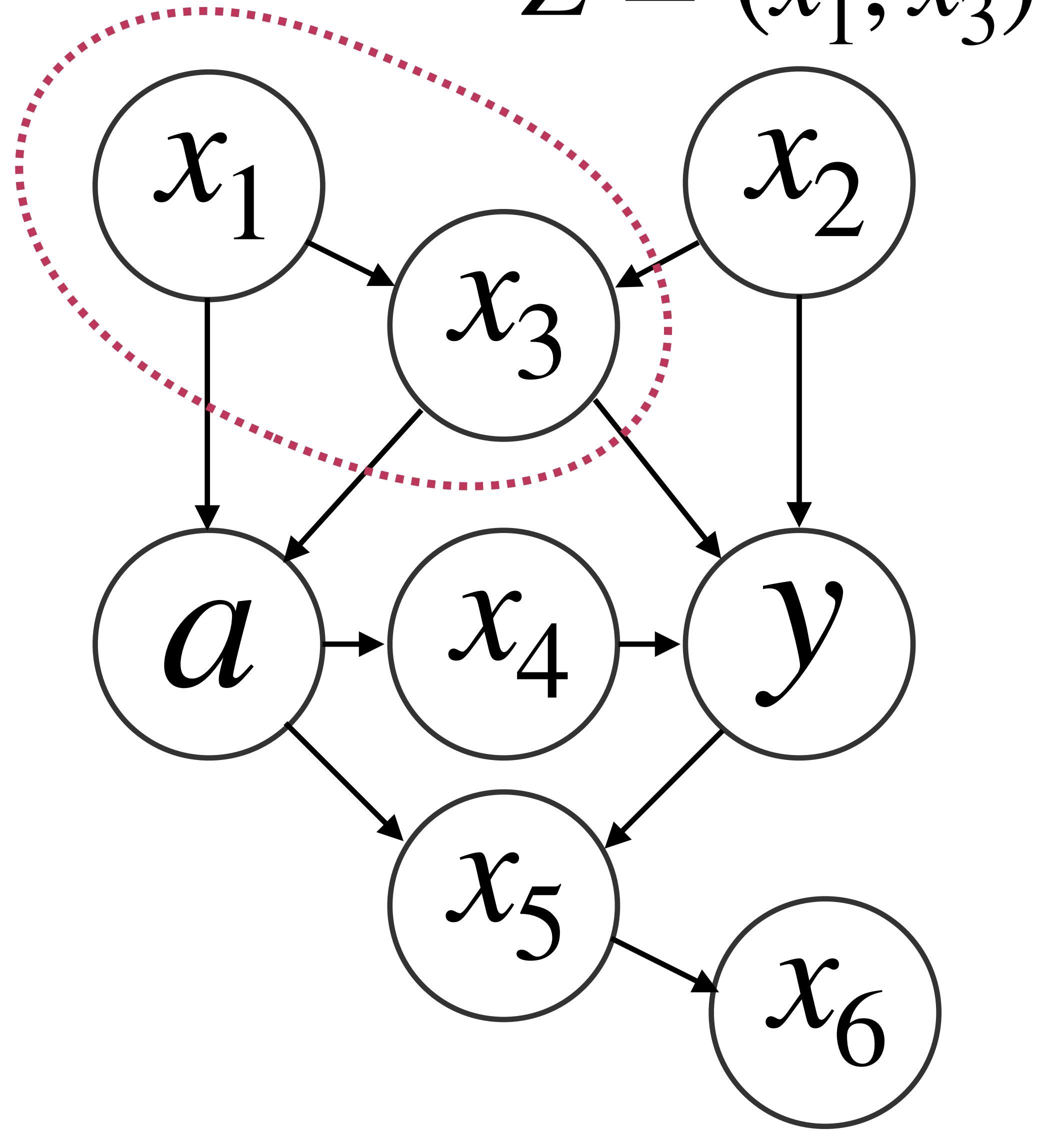
バックドア基準

因果構造の情報を用いて調整する変数集合を選ぶ基準

- 因果ダイアグラムGにおいて、aからyへの有向パスがあるとする。次の２条件を満たすとき、**変数集合 Z は順序対 (a, y) についてバックドア基準を満たす**という
  - (B1) aからZへの有向パスがない（行動より下流の変数を含まない）
  - (B2) aに入るパスを含む、aとy を結ぶパス（**バックドアパス**）において、Zがaとyを**有向分離**（ブロック）する
- ただし、a-y間の<u>全ての</u>パス p に対してZが以下の条件の<u>いずれか</u>を満たすとき、Zはaとyを**有向分離する**という
  - 鎖 i→m→j またはフォーク i←m→j を含み、m は Z に含まれる
  - 合流点 i→m←j を含み、m及びその子孫は Z に含まれない
- このとき、以下のバックドア調整定理が成立
  - $P(y \mid \mathrm{do}(a)) = \sum_{z} P(y \mid a, Z) P(Z)$
- Zの選び方は複数ありうるが、最も調整しやすい最小の集合を選ぶとよい

バックドア基準を満たすZの例

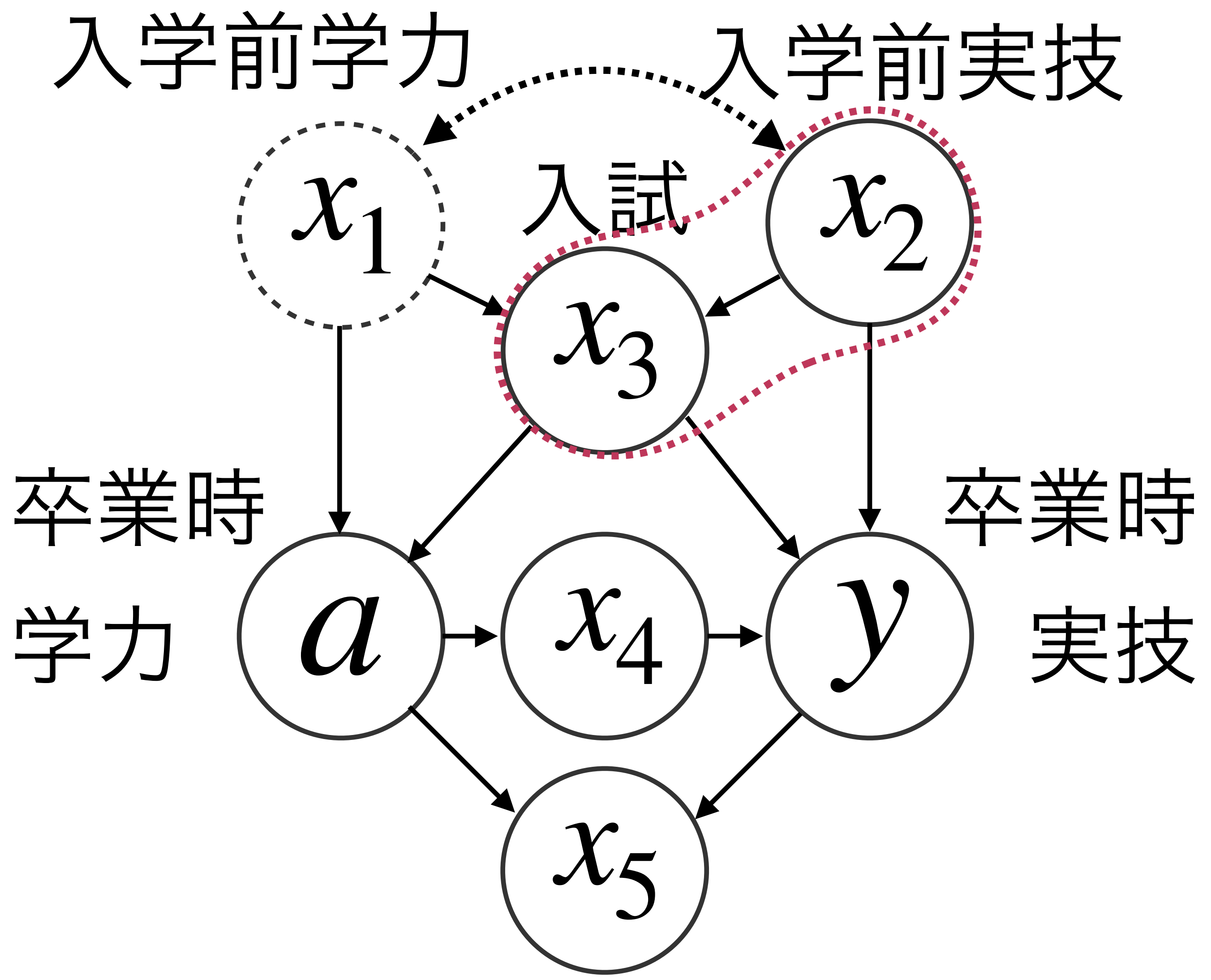
$Z = (x_1, x_3)$



バックドア基準を満たす簡易的な変数の選び方

共通原因は調整、行動より下流は調整せず、合流点の調整による擬似相関に注意

- バックドア基準を満たす調整変数群の選び方
  1. 行動 a よりも下流の変数は調整しない
  2. 行動 a と 結果 y の両方に影響を与える交絡変数mがある場合、その変数mか、またはその因果パス
     $a \leftarrow \ldots \leftarrow m \rightarrow \ldots \rightarrow y$
     上にある変数のいずれかを調整する
     ⇒ a の親が全て観測されていれば $Z = \mathrm{pa}_a$ でＯＫ
  3. 合流点(及びその子孫)を調整すると上流変数間に擬似相関が発生、擬似的な交絡になることに注意し、擬似相関も含めたパス上でいずれかの変数を調整



調整基準を満たす調整変数集合を出力してくれる

- 調整基準：バックドア基準よりも精緻
  - バックドア調整定理の必要十分条件を与える
- DAGitty (R言語) は調整変数集合 Z を出力可能
  - 可能な Z の列挙も可能

```
In [ ]: install.packages("dagitty")
```

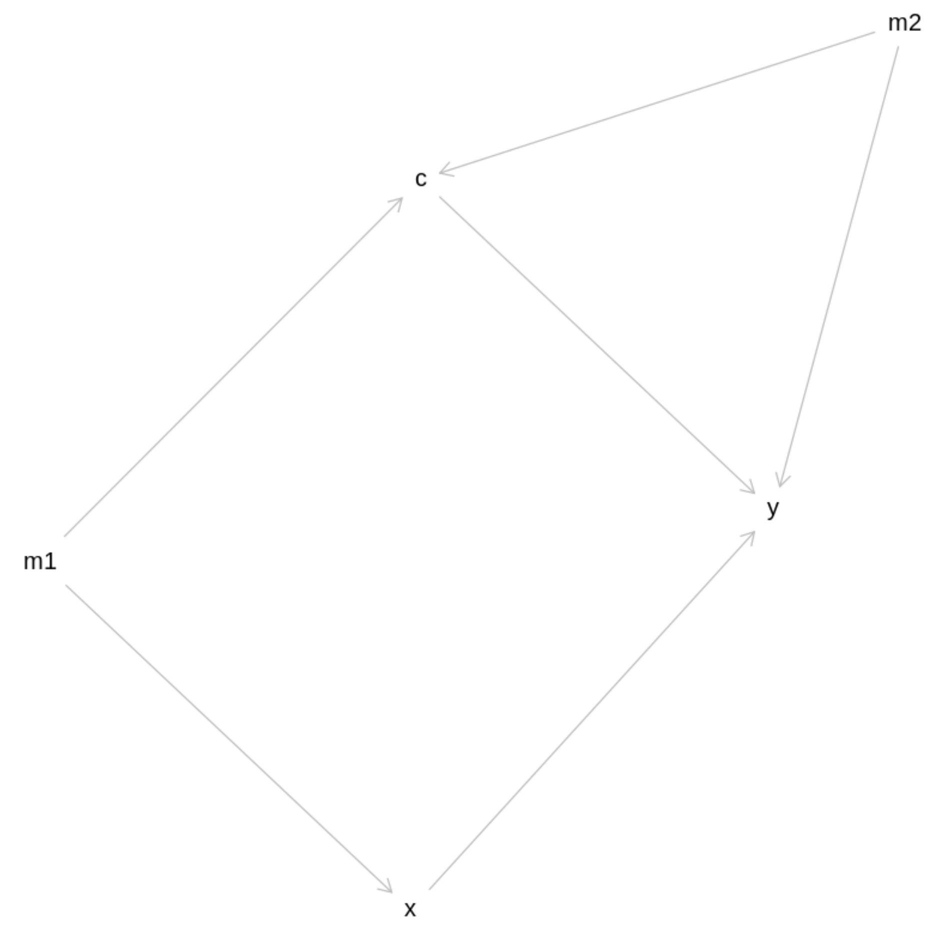
Installing package into '/usr/local/lib/R/site-library'
(as 'lib' is unspecified)

```
In [ ]: library(dagitty)
```

```
In [ ]: g <- dagitty( "dag{ x -> y ; x <- m1 -> c <- m2 -> y <- c}" )
        # g <- dagitty( "dag{ x -> y ; x <- m1 -> c <- m2 -> y <- c; d->c; c->e ; d->e ; e->f; f->y ;
        f->x; m1->e }" )
```

```
In [52]: plot(g)
```

Plot coordinates for graph not supplied! Generating coordinates, see ?coordinates for ho
w to set your own.




```
In [53]: adjustmentSets( g, "x", "y" )
```

{ c, m2 }
{ m1 }

再現用Colab

因果の定義の哲学

出欠コード
XXXX

因果とは何か自体が科学哲学的な議論対象
現在は要素還元的な定義は置いておいて、「発見して使える」へ

- 規則性説（デイヴィッド・ヒューム）
  - 「数学以外の真理の追求は経験と観察に基づく」
    - 今日まで太陽が登ったからといって明日も同じとは言い切れない
    - あらゆる因果も真理とは限らず、単に経験から抽出した規則性である
  - 反論：擬似相関は規則性があるが因果ではない
- 反事実条件説（デイヴィッド・ルイス）
  - 「種に水をやらなかったら発芽しなかったはず」
    - 現実とは異なっていた場合を仮定し、ありえた別の結果を推論する
  - 反論：別の世界がどうなっていたか特定できるのか。現実に観測できない形而上学では
- プロセス説（ウェスリー・サモン）
  - 「因果は時空間的に隣接した事象の連鎖」
    - 特に、保存量の交換が行われるプロセス（保存量説）
  - 反論：欠落因果「ブレーキを踏まなかったから事故が起きた」を端的に表現できない
- 介入説（ジェームズ・ウッドワード）
  - 「同様の状況において、原因に介入すれば結果が変化する」
    - 個別の事象ではなく、繰り返し起きる状況xを同一視することで比較可能に

まとめ

SCMは潜在結果モデルとは異なるもう1つの因果の定式化
因果構造がわかればバックドア基準により調整すべき変数がわかる

- 構造的因果モデルは変数ごとのノイズと関数で変数間の因果関係を表す
- 介入は因果構造および構造方程式の局所的な変更と表せる
- 因果の合流点で条件づけると非因果的相関が生まれる（選択バイアス）
- 調整すべき変数集合を選択するためバックドア基準を用いる
  - 介入対象変数aに入ってくる矢印を含む結果yとの間のパスを遮断するように調整する変数集合Zを選択
  - 合流点をZに含めるとその上流変数間に非因果的相関の経路（バックドアパス）が新たに開くことに注意