

**データ生成過程の学習：
因果推論・特徴選択へのアプローチ**

- 発表者 -

星野崇宏 (名古屋大学)
清水昌平 (大阪大学)
河原吉信 (大阪大学)
鈴木大慈 (東京大学)

データ生成過程：変数の決定関係

- ある母集団では、
観測変数 x_1 は、**観測変数** x_2 と**それ以外の潜在的要因** e_1 を引数とする**関数** f で決まるとしよう：

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

- もし線形なら：

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

-- 構造方程式モデル (Wright, 1921; Pearl, 2000) --

データ生成過程：変数の決定関係

- ある母集団では、
観測変数 x_1 は、**観測変数** x_2 と**それ以外の潜在的要因** e_1 を引数とする**関数** f で決まるとしよう：

$$x_1 := \underline{f}(x_2, e_1)$$

- もし線形なら：

$$x_1 := \underline{b}_{12}x_2 + e_1$$

-- 構造方程式モデル (Wright, 1921; Pearl, 2000) --

データ生成過程：変数の決定関係

- 例えば、あるインフルエンザ患者の集団では、「**治癒までにかかる日数**」は、「**投薬の有無**」と「**それ以外の要因**」を引数とする**関数 f**で決まるとしよう(狩野, 2002):

治癒日数 = f (投薬の有無, それ以外の要因)

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

-- 構造方程式モデル (Wright, 1921; Pearl, 2000) --

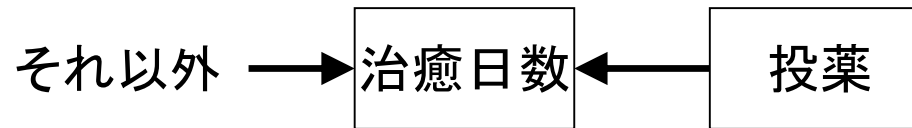
データ生成過程(変数の決定関係)を グラフで表現

- 構造方程式

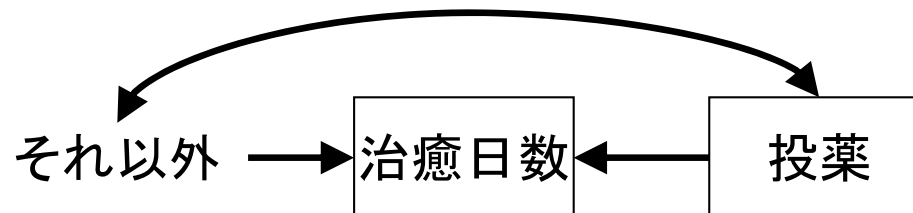
治癒日数 = f (投薬の有無, それ以外の要因)

- グラフ表現

- 「投薬の有無」と「それ以外の要因」が独立なら:



- 独立「でない」かもしれないなら:



データ生成過程：変数の決定関係

- 決定関数を表す構造方程式:

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

治癒日数 = f (投薬, それ以外の要因)

**Q. これら構造方程式は因果関係を表している？
どんな追加の仮定が必要？**

A. 主な考え方の1つが反実仮想モデル: 星野さん

因果分析で推定したいこと：介入効果

- 介入効果：**患者に投薬したら、投薬しなかった場合と比べて治癒日数はどのくらい短く or 長くなるか**
- シンプルな推定方法：無作為化実験
 - 患者それぞれに、投薬するかしないかを無作為に割り当てる

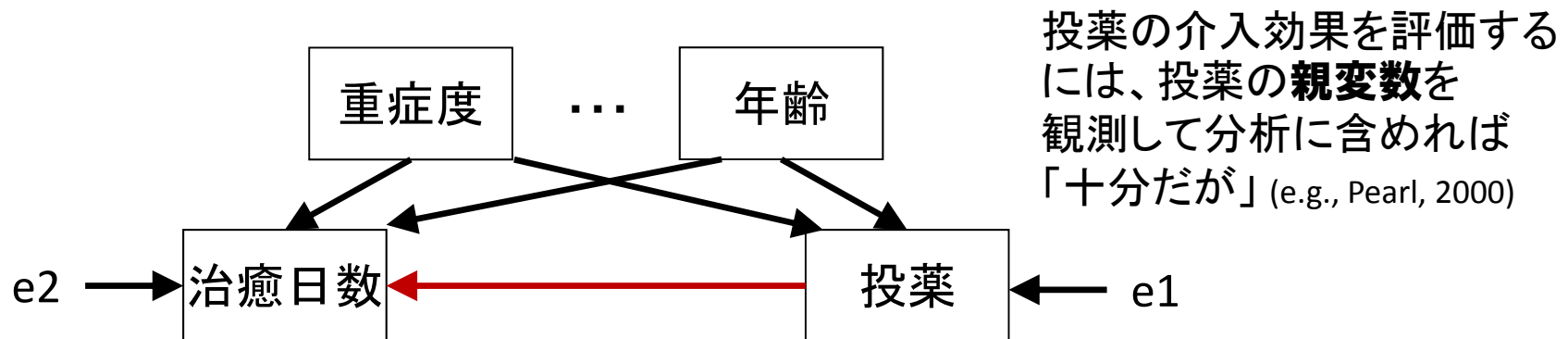
Q. 無作為化実験が行えないとき、
(重症度に応じて、投薬するかしないかを**選ぶ**とき)
介入効果を推定できる場合があるか？

**A. ある。「投薬」と「治癒日数」以外の変数(e.g.,重症度)を
分析に含める必要: 星野さん**
- 適切な変数を選択: 特徴選択

データ生成過程の学習：因果推論・特徴選択へのアプローチ

適切な変数を選ぶには、 データ生成過程に関する知識が必要

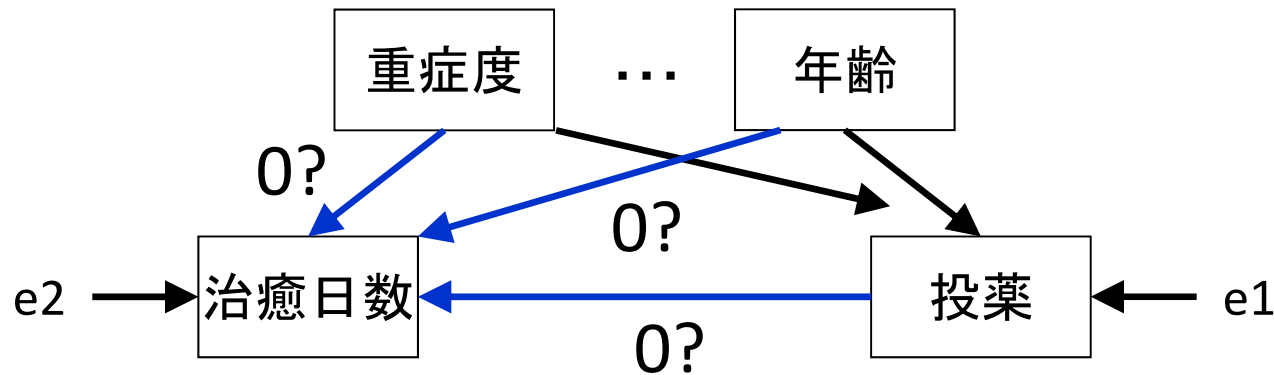
- 1つの方法は、**事前知識**からグラフを描くこと



Q. では、データ生成過程が未知のときは、
どのような仮定の下で、どの程度データ生成過程
を推定できるか: **清水**

治癒日数を決定する変数(要因)はたくさんある

- 親変数(直接決定する変数)はどれ?: 特徴選択



$$\text{治癒日数} = b_{11} \times \text{投薬} + b_{12} \times \text{重症度} + b_{13} \times \text{年齢} + \dots + e_2$$

スパース正則化:

劣モジュラ性を利用し、L1からL0へ : 河原さん

因果分析と(通常の)予測は違う…けど

- **因果**: 薬の量を**変化**させることで、
治癒日数はどのくらい**変化**する？(動的)
- **予測**: 薬を飲んだとき、治癒日数はどのくらいかな？(静的)

Q. どう「予測」するとよいか (L1, L2, or ?)は、
データ生成過程にどのように依存する？: 鈴木さん
- Multiple Kernel Learningを題材に

データ生成過程の学習：因果推論・特徴選択へのアプローチ

では、星野さんのトークからスタートです。