

線形反事実分析とその応用

黒木 学 データ科学研究系 准教授

はじめに

現在、ある状況におかれている部分母集団に対して、

「その部分母集団に限定して処理を行ったら結果はなるんだろう？」

と疑問に思ったことはないでしょうか？本ポスターでは、この問題を線形構造方程式モデルの立場からどのように定式化されるのか考えてみたいと思います。

データ生成過程 (1)

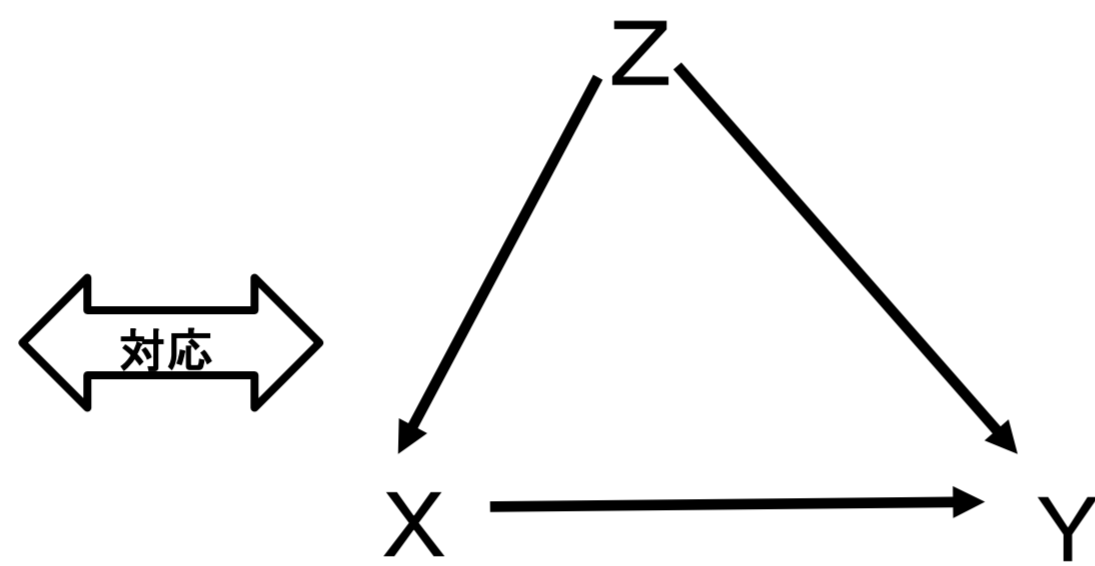
データは(広い意味で)自律的な物理的データ生成過程(線形構造方程式モデル)にしたがって生成

例: データ生成過程

$$Y = a_y + a_{yx} X + a_{yz} Z + e_y$$

$$X = a_x + a_{xz} Z + e_x$$

$$Z = a_z + e_z$$



因果ダイアグラム

e_x, e_y, e_z は独立に平均を0とする正規分布にしたがうと仮定
 a_x, a_y, a_z は定数項, a_{yx}, a_{yz}, a_{xz} はパス係数(直接効果)

外的操作 (2)

処理変数Xに外的操作“ $X=x$ ”を実施:
 \Rightarrow 線形構造方程式の中のXに関する方程式を $X=x$ に置き換える

$$Y = a_y + a_{yx} X + a_{yz} Z + e_y$$

$$X = a_x + a_{xz} Z + e_x \quad (1)$$

$$Z = a_z + e_z$$



$$Y = a_y + a_{yx} X + a_{yz} Z + e_y$$

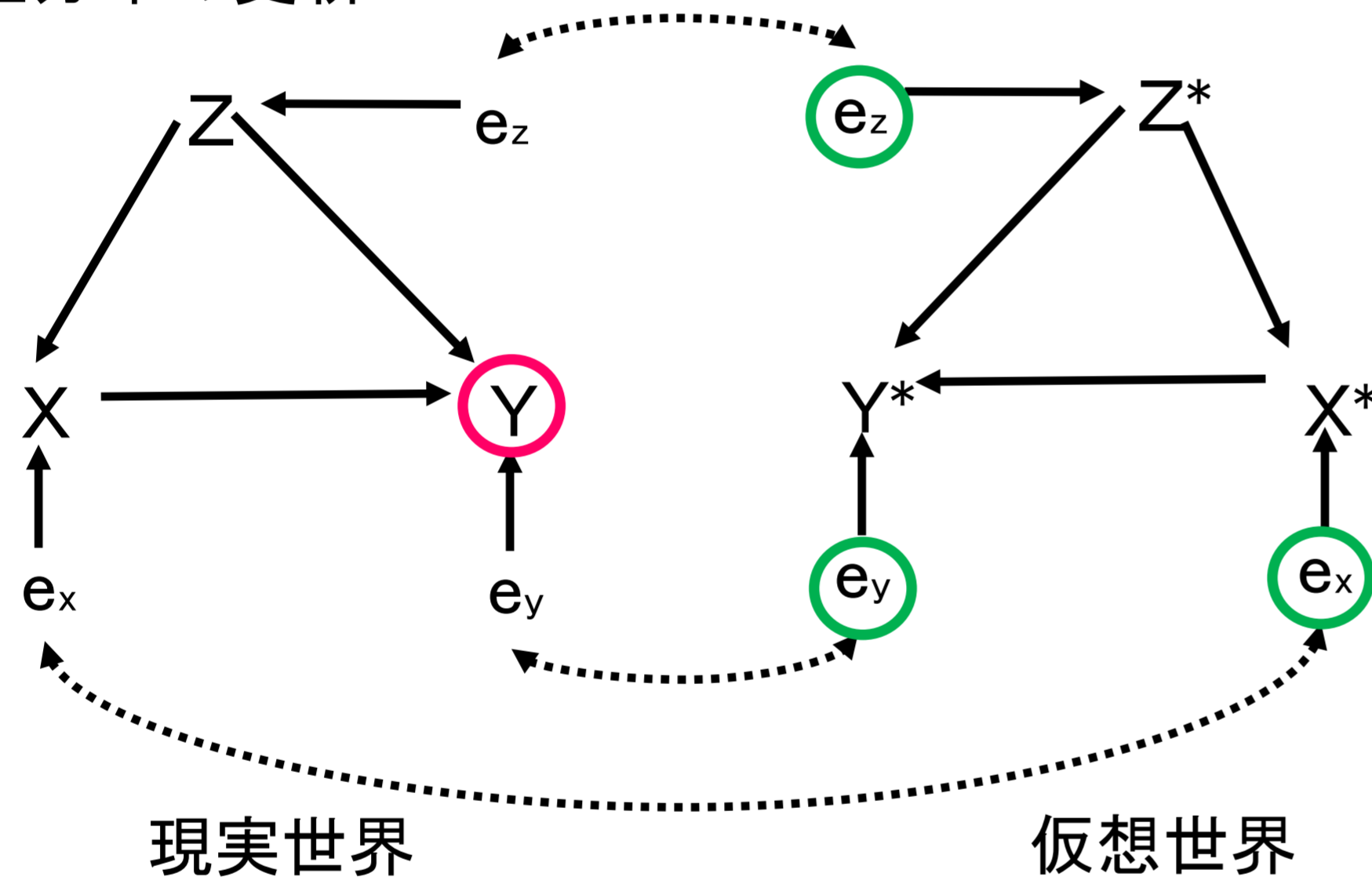
$$X = x \quad (2)$$

$$Z = a_z + e_z$$

(1)式にしたがって生成されたデータを使って、“ある条件H”のもとでの(2)式の挙動をしらべたい

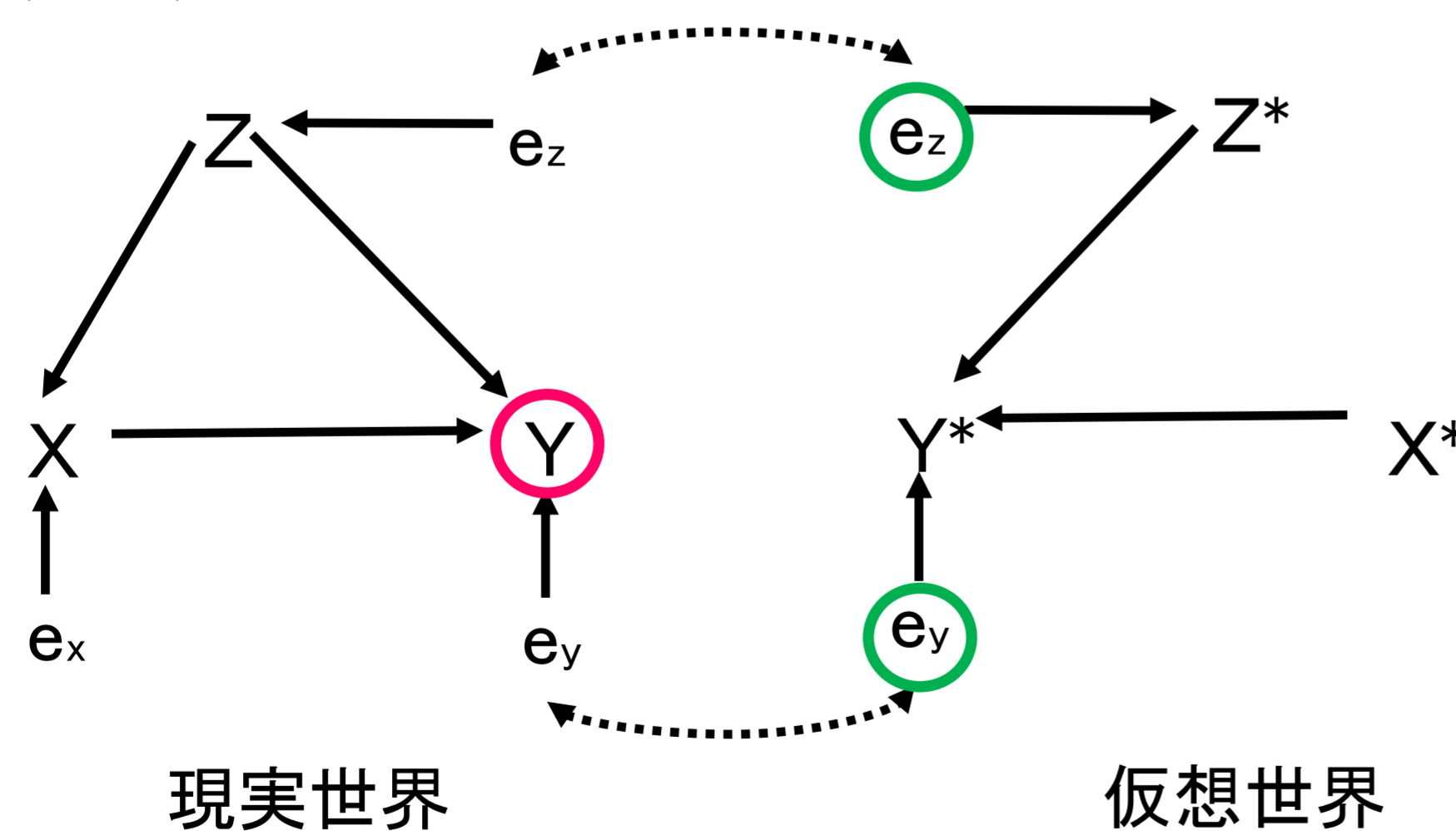
線形反事実分析: 双対ネットワークによる定式化 (3)

Step 1 (仮説形成): 現在の情報H(○に相当)による誤差分布の更新



情報Hがなければ2つの“世界”は同じだが、Hが得られると仮想世界における誤差分布(○に相当)が更新される

Step 2 (行動): Xに対する外的操作を実施



仮想世界において外的操作が実施される

Step 3 (予測): 仮想世界においてYの平均と分散を計算

品質管理への応用 (4)

自動車のボディ塗装工程における工程条件の設定(奥野他, 1986)

相関行列

	希釈率	湿度	塗着率
希釈率	1.000	-0.496	-0.091
湿度	-0.496	1.000	0.283
塗着率	-0.091	0.283	1.000

(Zは十分な交絡因子であることを仮定)

Y=yを満たす母集団について

$$E(Y^* | do(X^*=x), Y=y) = 0.065x + 0.0054y$$

$$Var(Y^* | do(X^*=x), Y=y) = 0.0042$$

条件つき平均・分散について

$$E(Y | X=x) = -0.091x, \quad Var(Y | X=x) = 0.992$$

母集団全体について

$$E(Y^* | do(X^*=x)) = 0.065x$$

$$Var(Y^* | do(X^*=x)) = 1.016$$

参考文献

Balke, A. and Pearl, J. (1995). Counterfactuals and policy analysis in structural models. *Uncertainty in Artificial Intelligence*, 46-54.