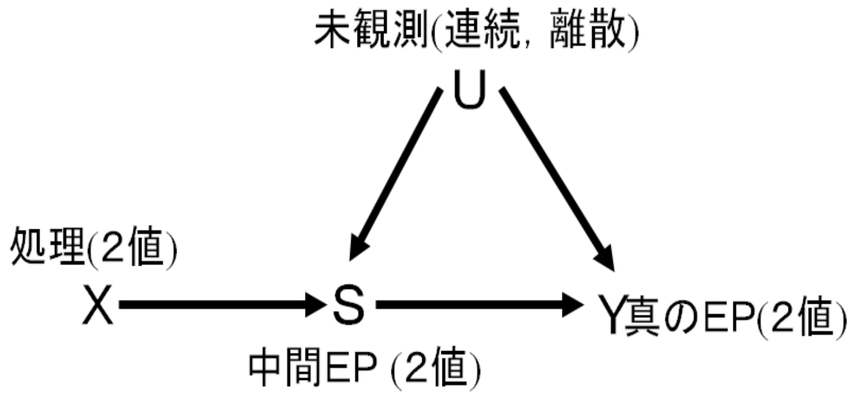
中間エンドポイントによる因果効果の評価

黒木　学　データ科学研究系　准教授

はじめに

ある処理を施したときに本来ならば評価しなければならない真のエンドポイント(真のEP)が時間的な制約や倫理的な問題を理由に観測することが難しくなることがあります．この問題を解決するために，処理から中間エンドポイント(中間EP)への因果効果についてはランダム化試験に基づいて評価することにし，それと代替EPと真のEPの間の関連情報と組み合わせて因果効果を評価することを考えます．



潜在反応アプローチ

直感的には：対象者に対してある外的操作をした際に現れる反応を，その対象者が持つ特徴と決定論的に結びつけたうえで，ランダムサンプリングなどの統計的要素を付加して因果効果の定量的評価を行うアプローチ

対象者Uに外的操作X=xを行ったときのYの反応:$Y_x(U)$
XからYへの因果リスク差：$\Pr(y_{1x1})-\Pr(y_{1x0})$
一致性(Consistency)：“$X=x \Rightarrow Y_x=Y$”

グラフから読みとれる仮定

XがSを通してYに影響する（排除規定）：$Y_{xs}=Y_s$
ランダム化：$(Y_{s0}, Y_{s1}, S_{x0}, S_{x1}) \perp X$

Surrogate Paradox

		u_1		u_0	
		s_1	s_0	s_1	s_0
y_1	x_1	0.686	0.007	0.232	0.000
	x_0	0.014	0.672	0.288	0.000
y_0	x_1	0.007	0.000	0.005	0.063
	x_0	0.000	0.014	0.006	0.006

因果効果
XからSへの因果リスク差=0.622
SからYへの因果リスク差= 0.301
XからYへの因果リスク差=−0.049

定式化

準備

XとSの関係

	weaker	normal	special	stronger
sicker	q_{00}	q_{01}	q_{02}	q_{03}
normal	q_{10}	q_{11}	q_{12}	q_{13}
special	q_{20}	q_{21}	q_{22}	q_{23}
healthier	q_{30}	q_{31}	q_{32}	q_{33}

SとYの関係

$$\left.\begin{array}{ll}\mathrm{pr}(y_0,s_0|x_0)=q_{00}+q_{01}+q_{10}+q_{11}, & \mathrm{pr}(y_0,s_1|x_0)=q_{20}+q_{22}+q_{30}+q_{32}\\ \mathrm{pr}(y_1,s_0|x_0)=q_{02}+q_{03}+q_{12}+q_{13}, & \mathrm{pr}(y_1,s_1|x_0)=q_{21}+q_{23}+q_{31}+q_{33}\\ \mathrm{pr}(y_0,s_0|x_1)=q_{00}+q_{01}+q_{20}+q_{21}, & \mathrm{pr}(y_0,s_1|x_1)=q_{10}+q_{12}+q_{30}+q_{32}\\ \mathrm{pr}(y_1,s_0|x_1)=q_{02}+q_{03}+q_{22}+q_{23}, & \mathrm{pr}(y_1,s_1|x_1)=q_{11}+q_{13}+q_{31}+q_{33}\end{array}\right\}$$

最適化

制約条件

$$\left.\begin{array}{l}\mathrm{pr}(y_0,s_0)=\mathrm{pr}(y_0,s_0|x_1)\mathrm{pr}(x_1)+\mathrm{pr}(y_0,s_0|x_0)\mathrm{pr}(x_0)\\ \mathrm{pr}(y_0,s_1)=\mathrm{pr}(y_0,s_1|x_1)\mathrm{pr}(x_1)+\mathrm{pr}(y_0,s_1|x_0)\mathrm{pr}(x_0)\\ \mathrm{pr}(y_1,s_0)=\mathrm{pr}(y_1,s_0|x_1)\mathrm{pr}(x_1)+\mathrm{pr}(y_1,s_0|x_0)\mathrm{pr}(x_0)\\ \mathrm{pr}(y_1,s_1)=\mathrm{pr}(y_1,s_1|x_1)\mathrm{pr}(x_1)+\mathrm{pr}(y_1,s_1|x_0)\mathrm{pr}(x_0)\\ \mathrm{pr}(s_0|x_1)=\mathrm{pr}(y_0,s_0|x_1)+\mathrm{pr}(y_1,s_0|x_1)\\ \mathrm{pr}(s_1|x_1)=\mathrm{pr}(y_0,s_1|x_1)+\mathrm{pr}(y_1,s_1|x_1)\\ \mathrm{pr}(s_0|x_0)=\mathrm{pr}(y_0,s_0|x_0)+\mathrm{pr}(y_1,s_0|x_0)\\ \mathrm{pr}(s_1|x_0)=\mathrm{pr}(y_0,s_1|x_0)+\mathrm{pr}(y_1,s_1|x_0)\end{array}\right\}.$$

目的関数

$$\mathrm{pr}(y_1|x_1)-\mathrm{pr}(y_1|x_0) \;=\; q_{11}+q_{22}-q_{12}-q_{21}$$

適用例（Walter et al.,2012）

インターフェロンαとラミブジンの併用療法はラミブジン単体療法よりもB型肝炎治療として有効なの？

B型肝炎の有無を調べるのには時間がかかるので、中間EPとして“HBeAgセロコンバージョン”を観測(投与から52週)

データ

Walter et al.(2012)

	y_1	y_0	
s_1	248	21	269
s_0	170	38	208

Schalm et al.(2000)

	s_1	s_0	
x_1	20	48	68
x_0	14	66	80

		平均	標準誤差
仮定なし	下限	−0.250(−0.269)	0.034
	上限	0.228(0.229)	0.028
単調性	下限	−0.154(−0.173)	0.028
	上限	0.147(0.147)	0.023

結果

仮定なし

下限

$$\max\left\{\begin{array}{c}\frac{\mathrm{pr}(y_0,s_1)-\mathrm{pr}(s_1|x_1)}{\mathrm{pr}(x_0)}-\frac{\mathrm{pr}(y_0,s_0)}{\mathrm{pr}(x_1)}\\ \frac{\mathrm{pr}(y_0,s_0)-\mathrm{pr}(s_0|x_1)}{\mathrm{pr}(x_0)}-\frac{\mathrm{pr}(y_0,s_1)}{\mathrm{pr}(x_1)}\\ -\mathrm{pr}(s_1|x_1)-\frac{\mathrm{pr}(y_1,s_1)}{\mathrm{pr}(x_0)}\\ -\mathrm{pr}(s_0|x_1)-\frac{\mathrm{pr}(y_1,s_0)}{\mathrm{pr}(x_0)}\\ -\mathrm{pr}(s_1|x_0)-\frac{\mathrm{pr}(y_0,s_1)}{\mathrm{pr}(x_1)}\\ -\mathrm{pr}(s_0|x_0)-\frac{\mathrm{pr}(y_0,s_0)}{\mathrm{pr}(x_1)}\\ -\frac{\mathrm{pr}(y_1)}{\mathrm{pr}(x_0)}\\ -\frac{\mathrm{pr}(y_0)}{\mathrm{pr}(x_1)}\end{array}\right\}$$

上限

$$\min\left\{\begin{array}{c}\frac{\mathrm{pr}(s_1|x_0)-\mathrm{pr}(y_0,s_1)}{\mathrm{pr}(x_1)}+\frac{\mathrm{pr}(y_0,s_0)}{\mathrm{pr}(x_0)}\\ \frac{\mathrm{pr}(s_0|x_0)-\mathrm{pr}(y_0,s_0)}{\mathrm{pr}(x_1)}+\frac{\mathrm{pr}(y_0,s_1)}{\mathrm{pr}(x_0)}\\ \mathrm{pr}(s_1|x_1)+\frac{\mathrm{pr}(y_0,s_1)}{\mathrm{pr}(x_0)}\\ \mathrm{pr}(s_0|x_1)+\frac{\mathrm{pr}(y_0,s_0)}{\mathrm{pr}(x_0)}\\ \mathrm{pr}(s_0|x_0)+\frac{\mathrm{pr}(y_1,s_0)}{\mathrm{pr}(x_1)}\\ \mathrm{pr}(s_1|x_0)+\frac{\mathrm{pr}(y_1,s_1)}{\mathrm{pr}(x_1)}\\ \frac{\mathrm{pr}(y_1)}{\mathrm{pr}(x_1)}\\ \frac{\mathrm{pr}(y_0)}{\mathrm{pr}(x_0)}\end{array}\right\}$$

処理から中間EPへの因果リスク差では抑えられない(他の因果尺度で抑えられる)

単調性

処理は中間EPに対して非負の因果効果を持つ

下限

$$\max\left\{\begin{array}{c}\frac{\mathrm{pr}(y_0,s_0)-\mathrm{pr}(s_0|x_1)}{\mathrm{pr}(x_0)}-\frac{\mathrm{pr}(y_0,s_1)}{\mathrm{pr}(x_1)}\\ \mathrm{pr}(s_1|x_0)-\mathrm{pr}(s_1|x_1)\\ -\frac{\mathrm{pr}(y_1,s_0)}{\mathrm{pr}(x_0)}\\ -\frac{\mathrm{pr}(y_0,s_1)}{\mathrm{pr}(x_1)}\end{array}\right\}$$

上限

$$\min\left\{\begin{array}{c}\frac{\mathrm{pr}(s_1|x_0)-\mathrm{pr}(y_0,s_1)}{\mathrm{pr}(x_1)}+\frac{\mathrm{pr}(y_0,s_0)}{\mathrm{pr}(x_0)}\\ \mathrm{pr}(s_1|x_1)-\mathrm{pr}(s_1|x_0)\\ \frac{\mathrm{pr}(y_1,s_1)}{\mathrm{pr}(x_1)}\\ \frac{\mathrm{pr}(y_0,s_0)}{\mathrm{pr}(x_0)}\end{array}\right\}$$

処理から中間EPへの因果リスク差で抑えられる