

因果ダイアグラムにおけるバックドア/フロントドア基準について

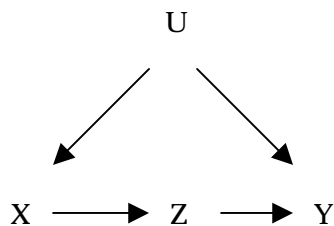
岡山理科大学 山本英二

1. 序

データに基づく統計的因果推論には、背景の知識を用いてデータから因果モデルを構築し、母数を推定していく立場と、前提となる因果モデルの下で関心のある因果指標をデータから推定する立場がある。近年、疫学の分野では後者の立場に立ち、観察研究で得られたデータから因果指標を推定出来るか、推定するか、の研究が発展してきている（松山 2001, 宮川・黒木 1999, Pearl 1995, 2000）ここでは Pearl によって導入された因果モデル：因果ダイアグラムの構造を前提として、因果効果(Causal Effect) の推定に関心がある時の問題を扱う。観察データにより、この因果効果が推定可能となるためには、因果ダイアグラムがある基準を満たせば良い。Pearl 1995 によってバックドア基準、フロントドア基準が与えられている。本論文ではこのバックドア基準、フロントドア基準の不充分性を指摘し、その補正を行う。本論文全体を通じて、Pearl 2000 p83 で与えられている、喫煙・肺ガンの例を拡張して用い、読者の理解の助けとすることにする。

2. 因果ダイアグラムと因果効果

例として4つの2値変量 X:喫煙 Y:肺ガン U:遺伝子型 Z:肺中タール とし、因果モデルは次のグラフで示されるとする。



見かけの喫煙・肺ガンの関連が遺伝子型が示す体質によって説明できるモデルである。即ち喫煙・肺ガンの因果効果は無いというモデルである。各変量間の矢線が因果関係を示し、巡回経路がないので、非巡回的有向グラフ DAG (Directed Acyclic Graphs) である。

因果関係を条件付き確率で表現したとき、次のマルコフ性が仮定できるとする。

$$p(z|x,u) = p(z|x)$$

$$p(y|x,z,u) = p(y|z,u)$$

このとき $p(x,y,z,u) = p(u)p(x|u)p(z|x)p(y|z,u)$ と分解される。マルコフ性を持った DAG を因果ダイアグラム (Causal Diagram) という。例は因果ダイアグラムとなって

いる。さて、X:喫煙とY:肺ガンの因果効果の定義である。この例ではU:遺伝子型が交絡候補因子であることを示している。実際、交絡候補因子の必要条件

(C1) U は非暴露 X=0 において Y のリスクである。

(C2) U は X と関連がある。

(C3) U は X から Y への因果パスの通過点ではない。

を満足している。ここで観察研究によって、(X,Y,U) が以下の分布を持つとする。

表 1

		p(x,u)	p(u x)	P(Y=1 x,u)	
X=0	U=0	0.48	0.96	0.063	p(x)=0.5 x=0,1 p(u)=0.5 u=0,1
X=1	U=0	0.02	0.04	0.179	
X=0	U=1	0.02	0.04	0.932	
X=1	U=1	0.48	0.96	0.887	

非暴露群における U の Y へのリスク比は $0.932/0.063=14.8$ であり、U と X のオッズ比は $(0.48 \times 0.48)/(0.02 \times 0.02)=576$ となり (C1)(C2) が満たされている。この表から見かけの X の Y へのリスク、リスク比は、 $P(Y=1|X=1)=0.825$, $P(Y=1|X=0)=0.0975$, $0.825/0.0975=8.46$ となり、強い因果関係を見かけ上示す。しかし、U で層別すれば、U=0 層で $0.179/0.063=2.84$, U=1 層で $0.887/0.932=0.95$ となり、2層で X は Y へ逆の効果を示している。このとき U は修飾因子である。

因果効果は反事実モデルによって定義される。対象集団に介入により暴露させたときの反応と暴露させないときの反応やその差、比を因果効果と定義する。実際には同一人物に暴露と非暴露の状況はあり得ないので、反事実モデルといわれる。ここで対象集団として観察研究における暴露群とする場合と、観察集団全体とする場合が考えられる。Pearl は後者の立場で因果効果を定義している。観察集団全体に介入して強制的に $X=x$ とすることは因果ダイアグラムで X へ入る矢線を落とすことに相当する。介入後の同時分布は

$$\frac{p(x, y, z, u)}{p(x|z)} = p(u)p(z|x)p(y|z, u)$$

となり因果効果は

$$p(y|\hat{x}) = \sum_{z,x} p(x, y, z, u) / p(x|u)$$

で定義される。

$$p(y|x) = \sum_{z,u} p(x, y, z, u) / p(x)$$

との違いに注意。Pearl2000 は $p(y|\hat{x}) = p(y|x)$ が成立するときを、非交絡と因果定義し、(C1)(C2)(C3) は交絡の関連基準であるとして区別している。この因果効果は $p(y|set(X=x))$, $p(y|do(X=x))$, $p(y|\tilde{x})$, $P(y_x)$ などの表記も使われている。この因果

効果 $p(y|\hat{x})$ は因果ダイアグラムが次のバックドア基準，フロントドア基準を持つとき，観察研究からの (X,Y,U) 又は (X,Y,Z) の同時分布，又は条件付き分布を与えることで推定可能となる．

3．バックドア基準

因果ダイアグラム G において， X から Y への有向道があるとする．このとき次の2条件を満たす変量集合 Z は (X,Y) についてバックドア基準を満たすという．

(B1) X から Z の要素への有向道がない．

(B2) X へ入る矢線を含む X と Y を結ぶ道（バックドアパス）において Z が X と Y を有効分離（ブロック）する．

ここで X と Y を結ぶすべての道について，変量集合 Z が次の条件のいずれかを満たすときに Z は X と Y を有効分離（ブロック）するという．

(S1) X と Y を結ぶ道に鎖 ($i \rightarrow m \rightarrow j$) またはフォーク ($i \leftarrow m \rightarrow j$) があるとき， m は Z に含まれる．

(S2) X と Y を結ぶ道に合流点 ($i \rightarrow m \leftarrow j$) があるとき，その合流点とその子孫は Z に含まれない．

Z は (X,Y) についてバックドア基準を満たすとき，次のバックドア調整定理が証明される．
$$p(y|\hat{x}) = \sum_z p(y|x,z)p(z)$$

例では U は X と Y を結ぶ道のフォークを形成しているので(B1)(B2)条件を満たし， U は (X,Y) についてバックドア基準を満たしている．そこでバックドア調整定理を適用すると X の Y への因果効果は

$$P(Y=1|do(X=1)) = 0.179 \times 0.5 + 0.887 \times 0.5 = 0.4525$$

$$P(Y=1|do(X=0)) = 0.063 \times 0.5 + 0.932 \times 0.5 = 0.4975$$

と求まる．観察集団全体に喫煙させる，させないの介入を行ったときの肺ガンリスクが計算される．

4．フロントドア基準

因果ダイアグラム G において， X から Y への有向道があるとする．このとき次の3条件を満たす変量集合 Z は (X,Y) についてフロントドア基準を満たすという．

(F1) X から Y へのすべての有向道を Z は遮断する．

(F2) X から Z へのバックドアパスが存在しない．

(F3) Z から Y へのすべてのバックドアパスは X によってブロックされる．

例では(F1) (F3)は満足されているが,(F2)が問題である . $X \leftarrow U \rightarrow Y \leftarrow Z$ の X から Z へのバックドアパスが存在する . Pearl2000 はこの例をフロントドア基準を満足する例にあげている . X から Z へのパスを考えるとときには Z の子孫は無視するとのルールを暗黙に入れていたとの解釈が Kuroki 1999, 2000 では行われている . このルールを用いるとして , Z が (X,Y) についてフロントドア基準を満たすとき , 次のフロントドア調整定理が証明される .

$$p(y|\hat{x}) = \sum_z p(z|x) \sum_{x'} p(y|x',z)p(x') = \sum_z p(y|\hat{z})p(z|\hat{x})$$

例では (X,Y,Z) が次の分布を持つとする .

表 2

		$p(x,z)$	$p(z x)$	$P(Y=1 x,z)$	
$X=0$	$Z=0$	0.475	0.95	0.1	$p(x)=0.5 \quad x=0,1$
$X=1$	$Z=0$	0.025	0.05	0.9	$p(z)=0.5 \quad z=0,1$
$X=0$	$Z=1$	0.025	0.05	0.05	
$X=1$	$Z=1$	0.475	0.95	0.85	

このとき $P(Y=1|X=1)=0.825$, $P(Y=1|X=0)=0.0975$ は (X,Y,U) の時と同じである . またフロントドア調整定理により ,

$$P(Y=1|do(X=1)) = 0.05 \times (0.10 \times 0.5 + 0.9 \times 0.5) + 0.95 \times (0.05 \times 0.5 + 0.85 \times 0.5) = 0.4525$$

$$P(Y=1|do(X=0)) = 0.95 \times (0.10 \times 0.5 + 0.9 \times 0.5) + 0.05 \times (0.05 \times 0.5 + 0.85 \times 0.5) = 0.4975$$

となりバックドア調整と同じになる .

5 . バックドア基準の補正

フロントドア基準は(F2) は明示的には不十分で誤解を生む . そこで , 以下のようにバックドア基準を補正する .

因果ダイアグラム G において , X から Y への有向道があるとする . このときサブ因果ダイアグラム G_Y (Y からでる矢線をカットしたグラフ) において 2 条件(B1) (B2)を満たす変数集合 Z を (X,Y) についてバックドア基準を満たすという .

この補正によってバックドア基準の内容に変化は起こらず , フロントドア基準のあいまいさが解消する . 因果律からはサブ因果ダイアグラム G_Y に制約を入れる合理性が指示される .

最後に (x,y,u,z) の分布から (x,y,u) , (x,y,z) の周辺分布に基づきバックドア調整とフロ

ントドア調整によって、同じ因果効果が推定値が与えられるべきであるが、このことは正しいのか、この証明が可能か、今後の課題としたい。

参考文献

1. 宮川,黒木 (1999). 因果ダイヤグラムにおける介入効果推定のための共変量選択. 応用統計学 28-3, 151-162.
2. 松山裕 (2001). 反事実モデルと因果効果の推定. 日本統計学会第69回チュートリアルセミナー:臨床試験を支える因果推論 11-26.
3. M. Kuroki and M. Miyagawa (1999). Identifiability criteria for causal effects of joint intertventions. JJSS 29-2, 105-117.
4. M.Kuroki (2000). Selection of post-treatment variables for estimating total effect from empirical research. JJSS 30-2, 115-128.
5. J. Pearl (1995). Causal diagrams for empirical research. Biometrika 82-4, 669-710.
6. J. Pearl (2000). Causality, Cambridge University Press.