XA 章　さらに奥地へ！：潜在結果モデルとメカニズム的理解の繋がり

本書の本編第 2 章ではバックドア基準と因果ダイアグラム、本編第 3 章では無作為化と潜在結果モデルについて学びました。オンライン補遺の本章ではそれらの章の内容をさらに掘り下げ、データの背後にある生成メカニズムと潜在結果モデルとの関係を見ていきます。それらが奥の方で繋がっていることが見えてくると、統計的因果推論と、研究対象の質的知識が、どういった補完関係にあるのかを理論的に理解できるようになります。
　本章では、

(I) 因果ダイアグラムは構造的因果モデルの枠組みで表現できる
(II) 潜在結果モデルは構造的因果モデルの枠組みで表現できる

こと、次いで、これら(I)(II)からの帰結として

(III)因果ダイアグラムと潜在結果モデルはシームレスに繋がっている（どちらも構造的因果モデルの枠組みの中で捉えられる）

ことを見ていきます[1]。

最初にちょっとお断りしておきますが、本章の内容は（オンライン補遺に回されていることからもお察しのとおり）いささかマニアックです。表現としても$f()$や$g()$などの記号が多く出てくることもあり、抽象的な内容になります。また、「何故こういうことを考えるのか」がそもそもピンとこない人も多いかもしれません。たとえば、普段から現象のメカニズム的な側面にも興味があるデータ解析者にとっては、「潜在結果モデルとメカニズム的理解の繋がり」という論点の重要さはほぼ自明かもしれません。その一方で、「メカニズムには興味がない、あるいは、むしろメカニズムを無視できることがクール」という感覚のデータ解析者には、その論点の意味するところ自体がピンとこないかもしれません。もし後者の場合には、現時点では適当に読み飛ばし、いつか時が満ちたときに再読していただければと思います。人との出会いもそうですが、理論との出会いにも適切なタイミングがあります。この補遺の章のファイルはいつでもダウンロードできますので、無理せず、それぞれのタイミングで紐解いていただければと思います。

では、まず「(I) 因果ダイアグラムは構造的因果モデルで表現できる」ことを見ていきましょう。

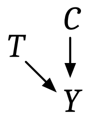
XA.1 構造的因果モデル：因果ダイアグラムと関数的表現の繋がり

本編の第 2 章では、因果ダイアグラムを用いて、データの生成メカニズムを議論してきました。本章では例として、以下の因果ダイアグラムを考えてみます。

---

[1] この章の内容は、黒木学(2017)『構造的因果モデルの基礎』の内容（の主に第 6 章）を基に、分析対象の総合的理解に関する筆者自身の興味関心と表現方法を織り交ぜて書かれたものです。数学的に厳密な説明や、構造的因果モデルの詳細については当該書籍の記述をぜひご参照ください。

図 XA.1　T，Y，C の 3 要因間の関係を表す因果ダイアグラムの例



このダイアグラムは、要因間での影響の入出力関係を表したものです。この関係は、数式では

$$Y = f(T, C) \quad \text{式 XA.1}$$

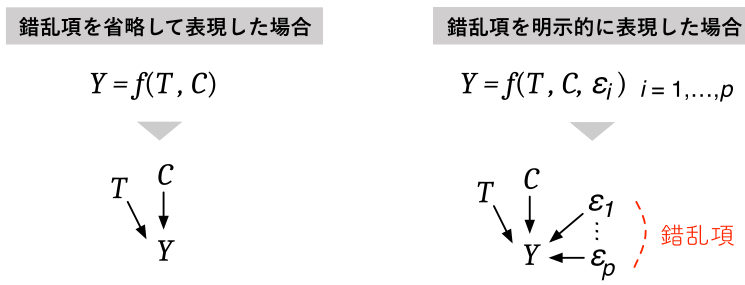
と表現できます。この式は「YはTとCの値によって決定論的に一意に定まる（YはTとCの関数である）」ことを意味しています。ここで$f(\,)$はその関係性の具体的な形を規定するものです。例えば「YはTとCの線形結合により一意に定まる」という関係性の場合には、その関数の具体的な中身はたとえば $Y = f(T, C) = \alpha T + \beta C$となります（本章の議論は$f(\,)$の具体的な形には依存しないので、関係性の具体的な形は明示しないまま、「何が何の関数であるか」という入出力の構造のみに焦点を当てて話を進めていきます）。逆の方向から見ると、式 XA.1 の関数の入出力関係をグラフィカルに表現したものが図 XA.1 の因果ダイアグラムである、とも言えます。

もしTとC以外の要因を錯乱項として明示的に考慮すると、

$$Y = f(T, C) + \epsilon_i, \quad i = 1, \ldots, p \quad \text{式 XA.2}$$

と記述できます。ここでは錯乱項として p 個の要因があることが表現されています。以下の図 XA.2 左の因果ダイアグラムを上記の式の表現としたときには、撹乱項は表記上省略されていることになります。省略しないで撹乱項まで明示的に記述すると、図 XA.2 右の形になります。煩雑さを避けるために、ここからは撹乱項は考慮せず、「結果Yは、他の明示的に考慮されている変数から決定論的に一意に定まる」という前提のもとで説明していきます[2]。

図 XA.2　構造的因果モデルとその因果ダイアグラムによる表現

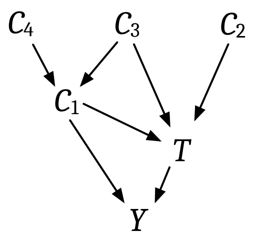


[2] そもそもデータを「決定論的モデルで記述すること」の含意や、「錯乱項」「誤差項」とは何を表しているのか、については XB 章で改めて議論します。

一般に、データ生成メカニズムを式 XA.1 や式 XA.2 のような関数関係の形で表したものを「構造的因果モデル（Structual Causal Model; SCM)」あるいは「関数因果モデル（functional causal model)」と呼びます（BOX XA.1)。ここで重要な点として、構造的因果モデルにおける等号記号（＝）は単に式の両辺が等しいことを意味するわけでなく、代入関係（←）を表していることが挙げられます。つまりここでの等式は、右辺にある変数が入力、左辺にある変数が出力となり、右側の変数に値が与えられることによってはじめて左辺にある変数の値が決まることを表しています[3]。

上の例は入出力の階層が２つだけの例ですが、図 XA.1.3 のように階層が３つある例もみてみましょう。

図 XA.3　階層が 3 つある因果ダイアグラムの例



この図については：

$$\begin{cases} C_1 = f_3(C_3, C_4) \\ T = f_2(C_2, C_3, C_1) \\ Y = f_1(T, C_1) \end{cases} \qquad \text{式 XA.3}$$

の代入を伴う構造方程式の形で表せます。ここで、T, C_1に右辺の式を代入すると、この式はさらに

$$Y = f_1(f_2(f_3(C_3, C_4), C_2, C_3), f_3(C_3, C_4)) \qquad \text{式 XA.4}$$

と表せます。尚、さらに上記のf_1, f_2, f_3から構成される関数をまとめて一つの関数$g()$として表現すると、

$$Y = g(C_2, C_3, C_4) \qquad \text{式 XA.5}$$

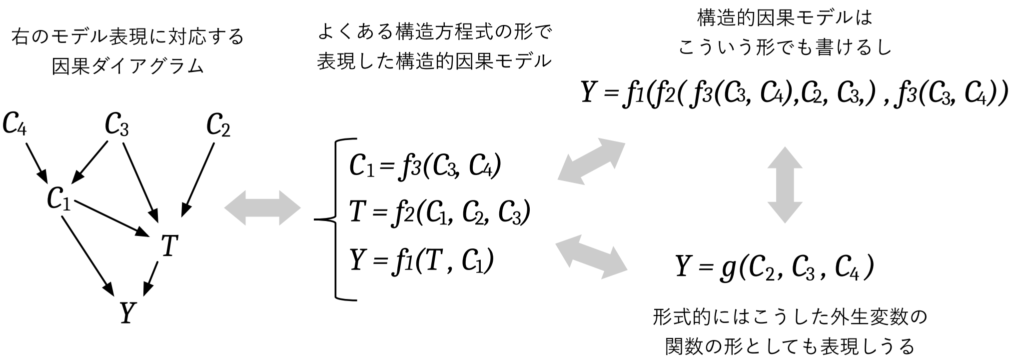
として、Yを、外生変数であるC_2, C_3, C_4からなる関数としても表現できます。ここで「外生変数」とは、構造方程式の形で表現したときに、左辺に現れない変数のことを指します[4]。因果ダイアグラムで説明すると、その変数に刺さる矢線をもたない（="親"をもたない）変数が「外生変数」です。これらの表現の対応関係をまとめてみると図 XA.4 のようになります。このように、要因間の入出力関係は、因果ダイアグラムで表すことも、幾つかの異なるやり方の数式的表現により記述することもできます。

---

[3] この等号記号が等価関係ではなく代入関係を表現していることが、「因果関係」を扱う構造方程式モデルと「確率関係」を扱うベイジアンネットワークの間の本質的な違いとなります。

[4] 「外生変数」には幾つかの定義の仕方がありますが、ここではひとまずこのように定義します。

図 XA.4　同一の因果構造を異なる表現方法で表した例



ここではひとまず、要因間の入出力関係の情報が明示化されたままである

$$Y \;=\; f_1(f_2(f_3(C_3, C_4), C_2, C_3), f_3(C_3, C_4)) \qquad \text{式 XA.4・再掲}$$

の式の形で考えていきます。

まずは、この式が意味することを考えてみましょう。これは概念的には、C_2, C_3, C_4の値を与えると、Yはある特定の値に定まることを意味しています。原理的には、構造方程式モデルはこうした外生変数の関数の形で表現することが可能です。上記の表現では、関数$f_1(\,)$, $f_2(\,)$, $f_3(\,)$の具体的な形は特定しないまま、要因間での入出力関係の構造的な側面だけを一般的形式のまま扱っています。バックドア基準の章で見てきたように、関数の具体的な形は分からないままでも、こうした入出力関係の構造が分かれば、識別可能性（バイアスのない因果効果の推定が可能かどうか）の議論は可能です。

以上、「因果ダイアグラムは構造的因果モデルによって表現できる」ことを見てきました。では、その逆の「構造的因果モデルは因果ダイアグラムによって表現できる」も真でしょうか？ 理論的には、因果ダイアグラムで表現できる現象の範囲よりも、構造的因果モデルで数式として表現できる範囲の方が広いことが知られています[5]。つまり、「因果ダイアグラムで扱えるものの限界」が「構造的因果モデルの扱えるものの限界」ではありません。

一般に、Judea Pearl による統計的因果推論の理論体系の本質はその「因果ダイアグラムの使用」にあると誤解されていることが多いかもしれません。しかし実は、その理論的な本質は「生成メカニズムの構造的因果モデルによる定式化」 --- つまり因果関係を式 XA.4 のように、要因間の関数関係の連なりとして定式化すること --- の方にあります。そして、構造的因果モデルは「関数の（ループをもたない）入出力関係の連なりとして書けるもの」はほぼ全てその枠組みの範囲内で扱えるため、かなりの広大なクラスのモデルがその範囲に

[5] 因果ダイアグラムは集団レベルの定性的因果関係（変数間の条件付き独立性など）を記述したものであるため、個体レベルの因果的仮定（個体間での因果効果の異質性など）を表現するのは基本的に難しい部分があります。一方、構造的因果モデルは個体間の異質性を表現できます。構造的因果モデルと因果ダイアグラムの関係に関する詳細は黒木(2017)の第6章を参照ください。

含まれることになります[6]。この観点から見ると、因果ダイアグラムとは、より根源的なモデルである「構造的因果モデル」の中の「視覚化／GUI ツール」にすぎない ---けれども、直感的な理解をもたらす有益で重要なツールである --- という位置づけのものと理解できます。

BOX XA.1 構造的因果モデルの数学的定義

構造的因果モデルの数学的な表現や定義に本書では深入りしませんが、少しだけ紹介しておきたいと思います。黒木(2017)では、構造的因果モデルを次の式 3.1 で定義しています（以下、同書の p70 からの引用)。

> 非巡回的有向グラフGとその頂点に対応する確率変数の集合$V = \{X_1, ..., X_p\}$が与えられている。グラフGが確率変数間の関数関係を
>
> X_i=$g_i(pa(X_i), \epsilon_i)$, $i = 1, ..., p$ （式 3.1）
>
> なる形に規定し、確率変数がこの関数関係にしたがって自律的でかつ定常的に生成されるとき、Gを因果ダイアグラムという。ここで、錯乱項ϵ_1, . . . , ϵ_pは互いに独立であるとする。また、$pa(X_i)$は因果ダイアグラムGにおけるX_iの直接的原因(direct cause)と解釈される。

この式 3.1 は、ある変数（X_i）を、「その親となる変数群（$pa(X_i)$）」と「錯乱項」の関数として表現したものであり、全ての変数（$i = 1, ..., p$）をそうした表現で定式化したものが構造的因果モデルとなります。また、式 3.1 のこの変数間の構造はたまたま生じた偶発的なものでなく、「確率変数がこの関数関係にしたがって自律的でかつ定常的に生成」される生成メカニズムをモデル化した表現であることが明記されています。

XA.2 潜在結果モデルを構造的因果モデルの枠組みで表現する

XA.2.1 潜在結果を潜在的因果モデルの形で表現する

ここまで、「(I) 因果ダイアグラムは構造的因果モデルの枠組みで表現できる」、しかし「構造的因果モデルは因果ダイアグラムによって表現できる」は必ずしも真ではないことを見てきました。ではここからは「(II) 潜在結果モデルは構造的因果モデルで表現できる」ことを見ていきましょう。

まずは、個体レベルでの潜在結果を構造的因果モデルの形式で表してみます。ネコの「ぴかそ」にノミの駆除剤を与えた場合の駆除までの日数の潜在結果を

[6] 例えば、関数の連なりとして記述される iterative なシミュレーションモデルの全ては、構造的因果モデルの枠組み内で（原理的には）記述することができます。また、時系列の状態空間モデルなども（過去の値に現在の値が遡って影響を与えたり、全く同時刻にループ的相互作用が生じるようなかなり特殊なモデリングがされている場合を除けば）基本的には構造的因果モデルで表現できます。

$$Y\left(\text{ぴかそ}\right)^{\mathrm{if}(T)} = f_{\text{ぴかそ}}(\mathrm{if}(T), C_1^{\text{ぴかそ}}, \ldots, C_K^{\text{ぴかそ}}) \quad \text{（式 XA.6）}$$

の形式で表してみます。ここで、$C_1^i, \ldots, C_K^i$は、個体iにおける潜在結果$Y(i)^{\mathrm{if}(T)}$に影響を与えうる全ての変数の値を表します[7]。$C_1^{\text{ぴかそ}}, \ldots, C_K^{\text{ぴかそ}}$は、「ぴかそ」における各変数の値を表しており、それらの値にもとづき、潜在結果となる$Y\left(\text{ぴかそ}\right)^{\mathrm{if}(T)}$の値が定まります。また、$f(\,)$中の$\mathrm{if}(T)$は「もし$T$に任意の値（処置）を与えたら」という意味です[8]。

このとき、もし「ぴかそ」に投薬しないとき（$T = 0$）と投薬したとき（$T = 1$）の潜在結果まで書くと：

$$Y\left(\text{ぴかそ}\right)^{\mathrm{if}(T)} = f_{\text{ぴかそ}}(\mathrm{if}(T), C_1^{\text{ぴかそ}}, \ldots, C_K^{\text{ぴかそ}})$$

もし$T = 0$のとき：$Y\left(\text{ぴかそ}\right)^{\mathrm{if}(T=0)} = f_{\text{ぴかそ}}(T = 0, C_1^{\text{ぴかそ}}, \ldots, C_K^{\text{ぴかそ}})$

もし$T = 1$のとき：$Y\left(\text{ぴかそ}\right)^{\mathrm{if}(T=1)} = f_{\text{ぴかそ}}(T = 1, C_1^{\text{ぴかそ}}, \ldots, C_K^{\text{ぴかそ}})$

となります。ここで、潜在結果における処置Tの割り付けはあくまで「もしも」の世界の話なので、潜在結果に影響を与える各変数の「$C_1^{\text{ぴかそ}}, \ldots, C_k^{\text{ぴかそ}}$」の値は処置$T$の値によらず同一になります。もしそうでなければ、「もしも」を仮想しただけで、割り付けに関係ない個体の性質が変化してしまうことになり、潜在結果モデルの元々の想定と矛盾してしまいます[9]。別の言い方をすると、潜在結果モデルを構造的因果モデルで表現したときには、「ぴかそ」の個体としての同一性（処置$T = 0$を受ける個体も$T = 1$を受ける個体も同じく「ぴかそ」であることの担保）は、「$C_1^{\text{ぴかそ}}, \ldots, C_K^{\text{ぴかそ}}$」の諸変数の同一性により表現されています[10]（図 XA.5）。

---

[7] ここで「全ての変数」としては、式 XA.5 の表現でしたときのように、Yの祖先となる全ての外生変数という形で表現することができます。実際に全ての変数を知ることは当然に困難ですが、ここではあくまで概念的な話としてください。

[8] この$f(\,)$中の$\mathrm{if}(T)$の表現は本書での説明の便宜のための表現であり、一般的な表現ではないことにはご留意ください。また、SUTVA 条件（第 8 章）が成り立つとき、$f(\,)$の中の$\mathrm{if}(T)$は単にTに置き換えても同じ意味になりますが、ここでは潜在結果と繋がりの説明のため$\mathrm{if}(T)$の表記のまま話を進めます。

[9] 別の言い方をすると、SUTVA 条件が非成立となります。

[10] ここで個体の同一性が諸変数の同一性によって表現されうることは、Rubin の潜在結果モデルの枠組みにおいて treatment-unit の同一性が基底的な（同一性の根拠の内在的な説明がその体系内には存在しない）仮定であることと対照的です。詳しくは、本書本編の BOX8.11、および黒木(2017)の第 6 章を参照。

図 XA.5　構造的因果モデルにおいて、個体の同一性は何によって表現されるか？

「もしT=0のときのぴかそ」と「もしT=1のときのぴかそ」は本当に
同一の「ぴかそ」という個体なのか？

$$Y(\text{ぴかそ})^{\mathrm{if}(T)} = \mathrm{f}_{\text{ぴかそ}}(\mathrm{if}(T), c_1, \ldots, c_k)$$

「もしT=0」のとき

$$Y(\text{ぴかそ})^{\mathrm{if}(T=0)} = \mathrm{f}_{\text{ぴかそ}}(T=0, c_1, \ldots, c_k)$$

「もしT=1」のとき

$$Y(\text{ぴかそ})^{\mathrm{if}(T=1)} = \mathrm{f}_{\text{ぴかそ}}(T=1, c_1, \ldots, c_k)$$

個体の同一性は、潜在結果を構成する「関数f」と「共変量cの値」が
どちらの場合でも変わらないことによって表現されうる

逆に言うと、もしfやcが異なるとき、これらの「ぴかそ」を果たして同一個体とみなしてよいのか？
それともそれらは厳密には別の個体なのか？という問いが生じる

XA.2.2 潜在結果モデルにおいて「同一の個体である／ない」とはどういうことか

第 3 章で見たとおり、潜在結果モデルの枠組みにおいて「ぴかそ」にノミの駆除剤を与えたときの因果効果は、

$$Y\left(\text{ぴかそ}\right)^{\mathrm{if}(T=1)} - Y\left(\text{ぴかそ}\right)^{\mathrm{if}(T=0)}$$

と、潜在結果の二つの項の差分として表されます。そして、**この二つの項は同時には観測できない**というのが「因果推論の根本問題」でした。

ではここで、ノミの駆除剤を与えていない別の「だり」という猫の潜在結果との差分の式を考えてみましょう。

$$Y\left(\text{ぴかそ}\right)^{\mathrm{if}(T=1)} - Y\left(\text{だり}\right)^{\mathrm{if}(T=0)} \qquad \text{式 XA.7}$$

ここでは「ぴかそ」に投薬した場合でも、投薬していない「だり」の反応は観測できます。しかし、これらの差分となる上式 SI.2.2.1 は、もちろん「ぴかそ」に対する因果効果とは等しくなりません。なぜなら、$Y\left(\text{ぴかそ}\right)^{\mathrm{if}(T=0)}$と$Y\left(\text{だり}\right)^{\mathrm{if}(T=0)}$は一般に、同一であると期待できないからです。

さて、少し哲学的な趣の話となりますが、ここで「『ぴかそ』以外は『ぴかそ』ではない」ということはどういうことでしょうか。「$Y\left(\text{ぴかそ}\right)^{\mathrm{if}(T=0)}$と$Y\left(\text{だり}\right)^{\mathrm{if}(T=0)}$は一般に同一でない」の意味内容を少し踏み込んで考えてみましょう。どちらも構造的因果モデルの形で書き下してみます。

$$Y\left(ぴかそ\right)^{\mathrm{if}(T=0)} = f_{ぴかそ}(T = 0, C_1^{ぴかそ}, \dots, C_K^{ぴかそ})$$

$$Y\left(だり\right)^{\mathrm{if}(T=0)} = f_{だり}(T = 0, C_1^{だり}, \dots, C_K^{だり})$$

この二者の潜在結果の違いは何により生じているのでしょうか？そのありうる要因は２つあります（図 XA.6)。１つ目は、要因間の繋がりありようを表す関数である$f_{ぴかそ}()$と$f_{だり}()$における違いです。$f_{ぴかそ}()$と$f_{だり}()$は、それぞれの結果の背後にある生成メカニズムの表現となっており、その「メカニズムにおける同等性」がない場合には「$Y\left(ぴかそ\right)^{\mathrm{if}(T=0)}$と$Y\left(だり\right)^{\mathrm{if}(T=0)}$は同等である」とは一般に期待できません。

では、$f_{ぴかそ}()$と$f_{だり}()$とみなせる --- すなわち、結果の背後にあるメカニズムについては実質的に同等とみなせる場合を考えてみます。この場合でも、潜在結果を決定しうる要因群がとる状態（値）が異なる場合、つまり上式中のいずれかの共変量c_k^iにおいて異なる値をもつ場合（「潜在結果を決定しうる共変量の状態における同等性」がない）には、「$Y\left(ぴかそ\right)^{\mathrm{if}(T=0)}$と$Y\left(だり\right)^{\mathrm{if}(T=0)}$は同等である」とは一般に期待できません。

図 XA.6　「個体間で潜在結果が異なること」の構造的因果モデル的な解釈

なぜ「もしT=1のときのぴかそ」と「もしT=1のときのだり」の結果Yは異なるのか？

$Y(ぴかそ)^{\mathrm{if}(T=1)} = \mathrm{f}_{ぴかそ}(T=1, c_1, \dots, c_k)$　　$Y(だり)^{\mathrm{if}(T=1)} = \mathrm{f}_{だり}(T=0, c_1, \dots, c_k)$

「個体間で結果Yが異なる」ということは
潜在結果を構成する「関数f」and/or「共変量c」の値が異なることに対応する

このように、異なる個体の潜在結果が「同等であるとはみなせない」ことの理由には、潜在結果の背後にある「メカニズムにおける同等性がない」場合と、「潜在結果を決定しうる共変量の状態における同等性がない」場合の二通りがありえます。これは逆に言うと、これらの同等性が期待できるときには、それらの個体の潜在結果の同等性も期待できる、ということになります。

XA.2.3 個体レベルでの潜在結果を代替できる条件を考える

ではここまでの議論の続きとして、「ある個体iの潜在結果を、他の個体jの潜在結果で代替できる条件」を考えてみましょう。ここでは、ある双子（かずや、たつや）の例を考えてみます。

かずやの潜在結果$Y\left(かずや\right)^{\mathrm{if}(T)}$を、たつやの潜在結果$Y\left(たつや\right)^{\mathrm{if}(T)}$で代替できる条件として、潜在結果を生み出すメカニズムにおける一定の同等性が成り立つ必要があります。つまり、$f_{かずや}() = f_{たつや}()$とみなせる

ことが必要となります。さらに、潜在結果を決定する諸要因がとる値についても、一定の同等性が必要となります。つまり、$C_1^{かずや},\ldots,C_K^{かずや}=C_1^{たつや},\ldots,C_K^{たつや}$とみなせることも必要となります。もしこれらの同等性が成立するときには、かずやの潜在結果$Y\left(かずや\right)^{\mathrm{if}(T)}$をたつやの潜在結果$Y\left(たつや\right)^{\mathrm{if}(T)}$で代替できると期待できます[11]。

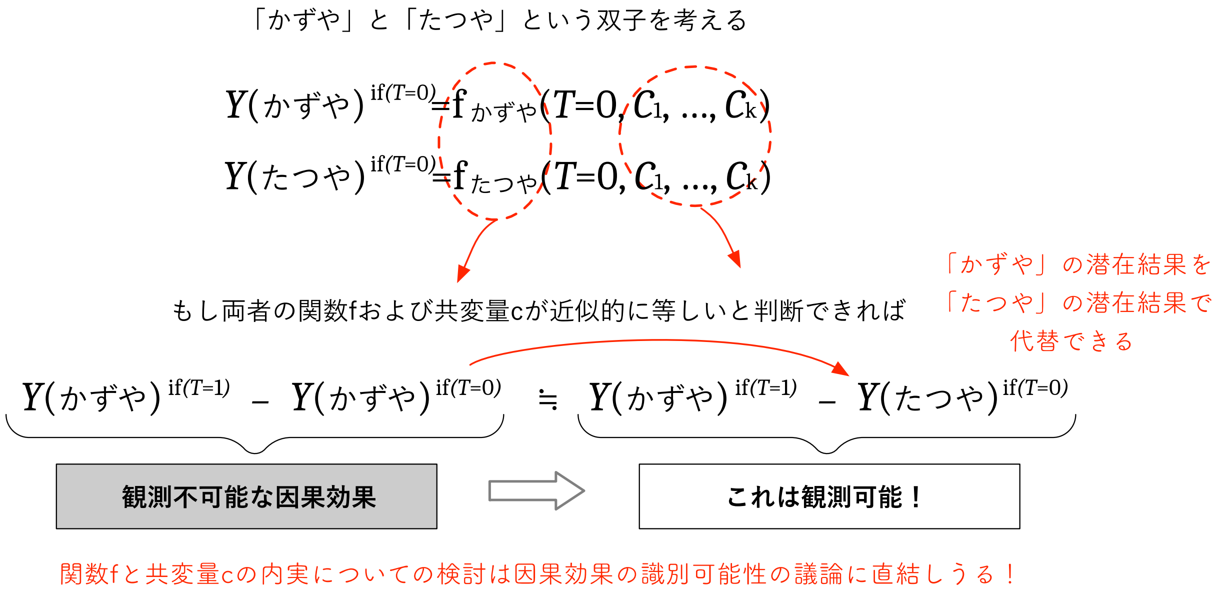
これらの同等性が成立するかは対象についてのドメイン知識によって判断することになります。たとえば、一卵性双生児は他人同士のペアと比べれば、多くの生理学的・社会的な要因において、類似したメカニズムや要因の値をもっていると一般に期待できます。そのため、他人同士の潜在結果を比較するときよりも、一卵性の双子の間での比較の方が、潜在結果の同等性が成り立つ可能性はずっと高そうです。

このことは「潜在結果の背後にあるメカニズムや共変量における同等性」が異なる個体の間で成り立つとみなせるときには、ある個体での反事実的な潜在結果（たとえば、$Y\left(かずや\right)^{\mathrm{if}(T=0)}|T=1$）を、別の個体の観測可能な潜在結果（たとえば、$Y\left(たつや\right)^{\mathrm{if}(T=0)}|T=0$）で代替しうることを意味します（図 XA.2.7)。別の言い方をすると、「潜在結果を構成するものの内実を考えること」によっても、因果推論の根本問題を”迂回”できる余地があるということです。このように、潜在結果モデルでは潜在結果の構成要素はブラックボックスとして扱われますが、そのブラックボックスの中を積極的に考察することも、因果効果の識別可能性の検討へと連続的に繋がりうるものです[12]。

[11] とは言え、「かずや」を「たつや」で”代替できると期待する”ことが果たして本当に正当なのかは良く分かりません。（われわれは喪われた可能世界を”代替”によって回復できるのでしょうか？）

[12] 尚、ここでの「個体レベルでの代替可能性」の議論は、「集団レベルでの外的妥当性／移設可能性」の話とも繋がってきます。第 9 章での外的妥当性／移設可能性の議論の中で、集団レベルでの話としてもこの「潜在結果の背後にある関数や共変量における同等性」の議論がされています。

図 XA.7　潜在結果の内実について考えることは識別可能性の議論に直結しうる



XA.2.4 集団の潜在結果と平均因果効果を構造的因果モデルで記述する

ではここからは、「集団レベルでの平均潜在結果」を構造的因果モデルで記述していきます。個体の潜在結果を考える上では、個体iがもつ生成メカニズムf_iと、潜在結果を決めるK個の要因であるc_k^iを考えました。ある集団 A の潜在結果を考える上では、その集団の個体が共通してもつメカニズムf_Aと、集団 A における要因C_k^Aの分布$P(C_k^A)$を考える必要があります。

ここでは説明のため、潜在結果の集団平均値$E\left[{Y_A}^{\text{if}(T=0)}\right]$に着目して考えていきます（これ以降、煩雑さを避けるために集団 A を表す添字の A は省略しますが、以下の式における$f(\,), C_j, Y^{\text{if}(T=0)}$はあくまで特定の集団 A について定義されたものであり、これらの概念が集団 A 以外にもそのままの形で適用できるか否かの判断には、本来は一般化可能性や移設可能性の検討が必要になるものであることは覚えておいてください）。N匹（$i = 1, \ldots, N$）からなる猫のサンプル集団 A の全個体に対して投薬したとき（$T = 1$）の潜在結果の平均値を、構造的因果モデル（$C_1, \ldots, C_K$と Tの関数）の形で表すと

$$E\left[Y^{\text{if}(T=1)}\right] = \left(\frac{1}{N}\right) \Sigma_{i=1}^{N} \left[Y(i)^{\text{if}(T=1)}\right] = \left(\frac{1}{N}\right) \Sigma_{i=1}^{N} \left[f(\text{if}(T=1), C_1^i, \ldots, C_K^i)\right]$$

となります。この右辺は、サンプル集団に含まれている猫の潜在結果の平均をとったものです。第 3 章で見てきたように、一般にこの潜在結果の平均$E\left[Y^{\text{if}(T=1)}\right]$は、$T = 1$で条件づけたときに観測される反応の平均となる$E[Y|T = 1]$ ($= E[f(T = 1, C_1^i, \ldots, C_K^i)|T = 1]$)とは異なります。つまり、式であらわすと一般に $E\left[Y^{\text{if}(T=1)}\right] \neq E[Y^{if(T=1)}|T = 1]$ です。なぜなら、「集団において$T = 1$で条件付けられたサブグループ」と「条件づけられていない集団全体」の間で共変量$C_1, \ldots, C_K$の分布が同等になっている（バランシングが達成されている）とは、一般には期待できないからです。

逆に言うと、もし（第 3 章でみた無作為化などにより）「$T = 1$で条件付けられたサブ集団」と「集団全体」の間

で「潜在反応に影響を与えうる共変量$C_1,\ldots,C_K$」の分布が同等になっているときには、サブグループと集団全体における潜在結果の分布も同等であると考えられるため

$$E[f(\mathrm{if}(T=1),C_1,\ldots,C_K)] = E[f(\mathrm{if}(T=1),C_1,\ldots,C_K)|T=1]$$
$$E[f(\mathrm{if}(T=0),C_1,\ldots,C_K)] = E[f(\mathrm{if}(T=0),C_1,\ldots,C_K)|T=0]$$

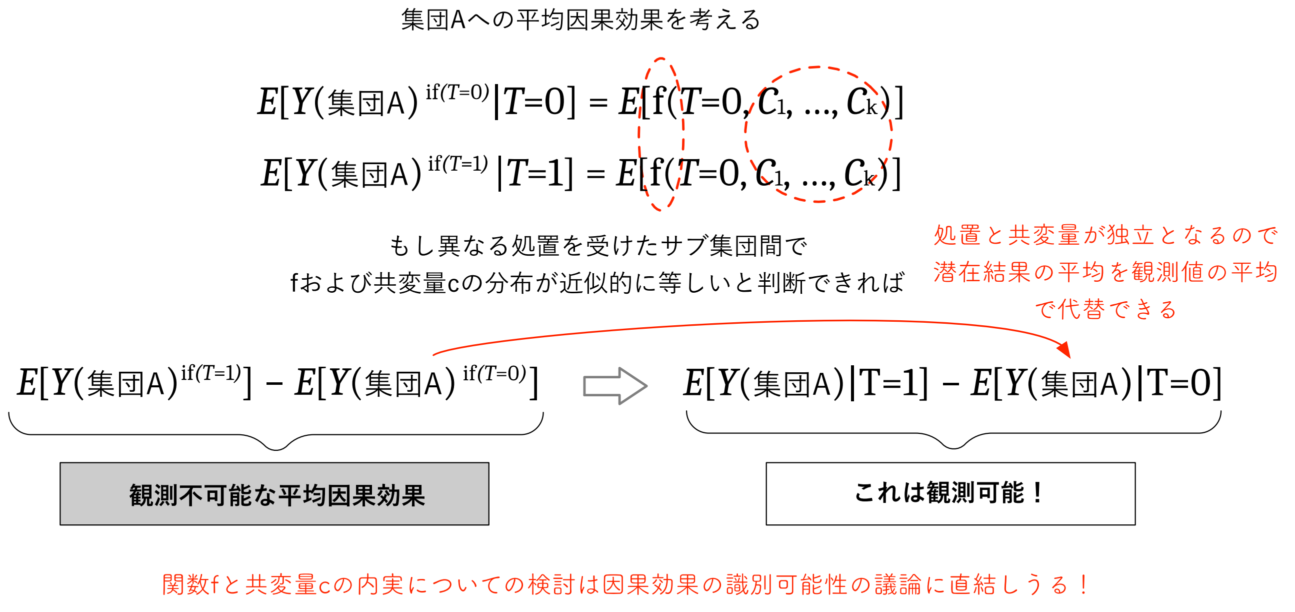
となります。つまり、

$$E[Y^{\mathrm{if}(T=1)}] = E[Y^{\mathrm{if}(T=1)}|T=1] = E[Y|T=1]$$
$$E[Y^{\mathrm{if}(T=0)}] = E[Y^{\mathrm{if}(T=0)}|T=0] = E[Y|T=0]$$

が成り立ちます[13]。このとき、第 3 章で見たように、潜在結果で表現される平均因果効果$E[Y^{\mathrm{if}(T=1)}] - E[Y^{\mathrm{if}(T=0)}]$は、観察された結果 Yの平均値の差（$E[Y|T=1] - E[Y|T=0]$）を用いて推定可能となります。

このように、潜在結果モデルは、仮想的な割付に対して特定の値を返す関数を想定することで、構造的因果モデルを用いて表現できます。こうした視点に立つと、「メカニズム（関数）の同等性」および「異なる処置を受けるサブグループ間での$C_1,\ldots,C_K$の分布の同等性」に基づいて、「平均因果効果をバイアスなく推定できるかどうか（識別可能性）」を議論する理論的な見通しがあることも見えてきます（第 8 章参照）。

図 XA.8　構造的因果モデルで構成した潜在結果から集団レベルでの反事実の代替可能性を眺める



13　ここでは SUTVA 条件が仮定されています。

XA.3 関数因果モデルと潜在結果モデルと因果ダイアグラムの繋がり

さて、ここまでで

（I）因果ダイアグラムは構造的因果モデルで表現できる

（II）潜在結果モデルは構造的因果モデルで表現できる

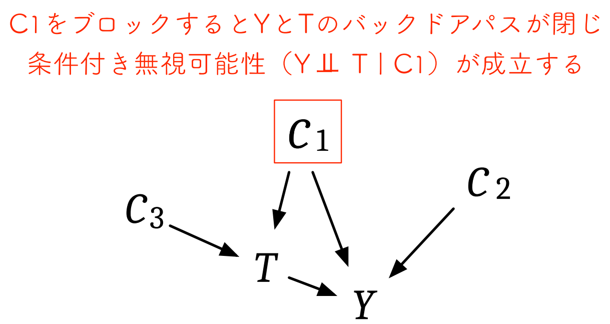
ことを見てきました。では、ここから「(III)因果ダイアグラムと潜在結果モデルはシームレスに繋がっている（どちらも構造的因果モデルの枠組みの中で捉えられる)」ことを見ていきましょう。

例として、以下の構造的因果モデルを考えてみます。

$$\begin{cases} T = g(C_1, C_3) \\ Y = f(T, C_1, C_2) \end{cases} \qquad \text{（式 XA.8）}$$

このモデルを因果ダイアグラムとして表現すると、以下の図のようになります。

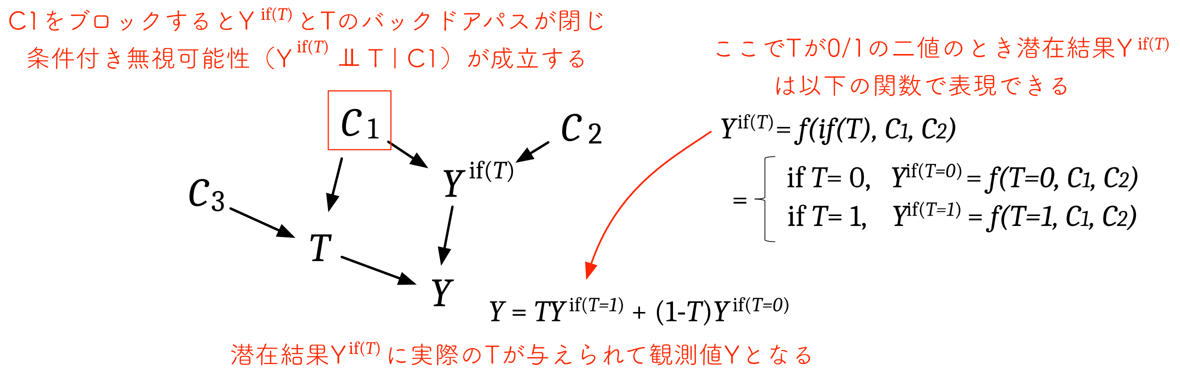
図 XA.9　式 XA.8 の構造的因果モデルの因果ダイアグラムによる表現



ここで、潜在結果$Y^{\mathrm{if}(T)}$を、この因果ダイアグラムを改変して表現してみましょう。観察値 Yは、潜在結果$Y^{\mathrm{if}(T)}$と実際の処置 Tの値から定まるため、因果ダイアグラム上で以下のように表してみましょう[14]。

[14] 潜在結果を因果ダイアグラムで表すやり方には今のところ標準的な記法はないので、この記法はあくまで本書の中での便宜的な記法としてご理解ください。

図 XA.10　式 XA.8 を潜在結果も含めた因果ダイアグラムにより表現した例



この図から、以下の交換可能性が成り立つ条件を考えてみましょう。

$$P(Y^{\mathrm{if}(T)}|T) = P(Y^{\mathrm{if}(T)})$$

これは Tの値と$Y^{\mathrm{if}(T)}$が独立であるということなので、$Y^{\mathrm{if}(T)}$が処置 Tの子孫ではないことと、因果ダイアグラムにおいて Tの値と$Y^{\mathrm{if}(T)}$の間に開きっぱなしのバックドアパスがないことが、上記条件が成り立つ条件になります（図 XA.10）。このダイアグラムの構造から、変数C_1の追加によりバックドアパスがブロックされると、以下の条件付き交換可能性が成り立つことになります。

$$P(Y^{\mathrm{if}(T)}|T, C_1) = P(Y^{\mathrm{if}(T)}|C_1)$$

このように、因果ダイアグラムの観点から見ると、バックドア基準を満たす変数をブロックすることで潜在結果における条件付き無視可能条件が満たされることがわかります。

こうして潜在結果を構造的因果モデルで表現することにより、潜在結果モデルとバックドア基準の話がつながりました。潜在結果モデルの枠組みではそのメカニズムはブラックボックスとされることが多いですが、それは潜在結果を考える上での論理的な必然というわけではありません。そのブラックボックスの内部をさらに「関数の入出力関係の連なり」である構造的因果モデルで表現することで、バイアスのない推定のために必要な条件（識別可能性）をより具体的に検討しうる道が開けることがあります。また、構造的因果モデルの観点から潜在結果モデルと因果ダイアグラムを眺めることにより、両者の話が奥の方でちゃんと繋がっていることが理解できます。これらのことは、ある現象について、潜在結果モデルの枠組みで考えることと、その現象のメカニズム的な内実（因果構造）を考えることは、全くもって相反するものではないこと意味します。こうした理解は、「分析対象に内在するメカニズムや状態についての質的・量的な理解」と「分析対象における因果効果の統計的推定」を両輪とした「分析対象の総合的な理解」を目指す上では非常に重要なものとなります。

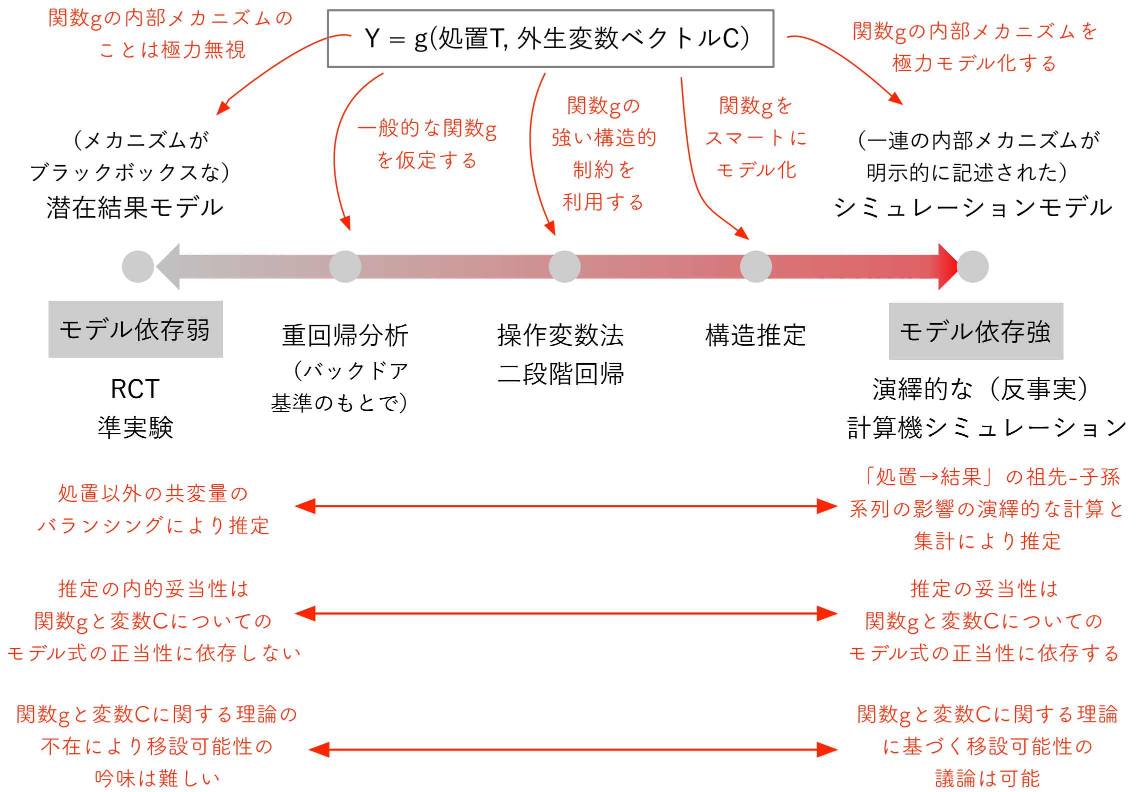
BOX XA.2 モデル依存性の強さから見る因果効果の推定アプローチの連続的な理解

さて、本章でみてきたように、構造的因果モデルは式 XA.15 のような、結果 Yが外生変数（$C_1,...,C_K$）の関数として生成される形としても表現できます。

$$Y = g(C_1,...,C_K) \qquad \text{式 XA.1.5・再掲}$$

これは極めて一般的な表現であり、たとえば温暖化ガス等による気温の変化を計算する気候変動シミュレーションモデルのような非常に大規模なモデルでさえも、原理的には「外生変数のセットを与えると、気温を返す構造的因果モデル（*気温* = *g*(*パラメータベクトル*)）として表現できます[15]。ここで、大規模な気候変動シミュレーションも「ある条件下での反事実状況の推定（what-if simulations)」を行うアルゴリズムであることを考えれば、一種の因果効果推定のアプローチとして捉えることができます[16]。こうした観点から、モデルへの依存性の度合いの観点から、広い意味での因果効果推定のアプローチを整理してみたのが以下の図 SB1.3.1 です。（尚、以下の整理はあくまで筆者による整理であり、一般的に広く受け入れられている見解ではないことにご注意ください。）

図 XA.11 モデル依存性の観点から見た因果効果推定アプローチの連続性



---

[15] たとえ確率的なシミュレーションであっても、乱数の種を入力パラメータする外生変数とみなせば、それも含めて決定論的な構造的因果モデルの枠組みの中で原理的には記述可能です。

[16] たとえば、近年の研究例としては、「CO_2濃度が産業革命以後に増加した（この）可能性世界」と「CO_2濃度が変化しなかった（反事実的な）可能性世界」において、「大雨による洪水災害が起こる頻度」をそれぞれシミュレーションにより計算し、それらの可能性世界間での差分を取ることにより、「CO_2濃度増加→洪水災害発生」の因果的影響を推定する event-attribution と呼ばれるアプローチの研究が行われています。

この図では、モデルへの依存性が低いものが左側、依存性が強いものが右側に配置されています。ここで一番左側に位置するのは、RCTや準実験などの、背景にある生成メカニズムをブラックボックスとして扱う系統の潜在結果モデルに基づく手法です。これらの手法での因果効果の計算の際には、生成メカニズムの情報を一切用いずに、処置以外の共変量をバランシングすることにより因果効果を推定します。こうした手法では、上記の式XA.1.5のモデル式（関数gや変数C）の正当性に依存せずに”model-free”での因果効果ができることが大きな利点です。一方、関数gや変数Cの理論的想定を欠くことは、移設可能性を把握する上では、何の手がかりもないことを意味しています。

反対側の一番右側には、例えばスパコンでぶん回すような気候変動シミュレーションのような、生成メカニズムを一連の数式で構成した演繹的なシミュレーションモデルによる因果効果の推定手法を位置づけられるでしょう。処置と外生変数からの効果伝搬を演繹的に計算・集計することにより因果効果（反事実的条件下における結果Y）が推定されます。この手法は”full-of-model”での推定となり、理論モデル式（関数gや変数C）と入力する外生変数の値（パラメータ値）の正当性により推定の妥当性が決まります。一般に、モデル式の正当性を示すことは原理的に簡単なことではないため[17]、こうしたモデル依存の推定には常に一定以上の不確実性が伴います。一方で、移設可能性については、関数gや変数Cの理論的想定に基づく一定の検討は可能です。

そして両極の中間的な手法としては、操作変数法やフロントドア基準を用いた手法などの、背景にある生成メカニズムによる（局所的な）制約の情報を利用して因果効果を推定する手法を位置づけられると考えられます。そしてその左側には、バックドア基準を満たした重回帰モデル等の統計モデルによる推定手法を、その右側には、経済学における構造推定などの理論モデルを、理念的には位置づけることができそうです[18]。

この図XA.11は、因果効果の推定様式を式XA.5のような形式を通して眺めてみると、そのスコープの中には、さまざまな因果推論の手法を位置づけうることを示しています。本書で扱っている対象は統計的因果推論ですが、この図に示されている統計的ではない手法も視野に入れることにより、対象をより深く分析するための新たな道が拓ける局面もあります[19]。

---

[17] そもそも「何が示されたらモデル式の正当性が示されたことになるのか」の基準設定自体が難しいという事情があります。モデルの正当性に関する議論は松王政浩著『科学哲学からのメッセージ:因果・実在・価値をめぐる科学との接点』を参照ください。

[18] 同じ名前のモデルやアプローチとして括られるもの（たとえば「構造推定」）であっても、その複雑さにはかなり幅がありうるため、この辺りの左右の位置づけには異論はあるかもしれません。

[19] 例えば、水田用農薬が赤トンボの個体群の増減に与える影響を分析対象とした筆者らの研究では、重回帰分析を用いた統計的因果推論による研究（Nakanishi et al. 2022 Science of the Total Environment 787, 147526）と、赤トンボの個体群動態の数理モデルを用いたシミュレーションによる研究（Nakanishi et al. 2020 Science of the Total Environment 703, 134499）の両面からアプローチを行っています。