

因果ダイアグラムでの変数への介入と矢線への介入

宮川 雅 巳*, 黒田 健 成**

Intervention Effects on Variable and Arrow in Causal Diagram

Masami Miyakawa* and Kensei Kuroda**

因果ダイアグラムとそれに対応した線形構造方程式モデルに基づく因果推論のための回帰分析について考察する。処理変数に対する無条件介入効果と条件付き介入効果を一般的に定式化し、これに対する識別可能条件を与える。線形構造方程式モデルのもとで、平均への介入効果と分散への介入効果を明示する。次に、共変量から中間特性への矢線に対する介入効果を定式化するとともに、その識別可能条件を与える。線形構造方程式モデルのもとで、反応変数の分散に対する矢線への介入効果を明示し、最適な介入方式を導く。これらの結果は、因果推論のための回帰分析における変数選択基準において有用である。

This article deals with regression analysis for causal inference based on a causal diagram and the corresponding linear structural equation models. Both unconditional and conditional intervention effects are formulated and their identifiable conditions are shown. Intervention effects not only on mean of a characteristic but also on the variance are presented under linear structural equation models. Intervention effect on arrows from covariate to intermediate variable is formulated and the optimal intervention is given. These results are useful for selection of variable in regression analysis for causal inference.

Key Words and Phrases: back-door criterion, d -separation criterion, intervention effect, linear structural equation model, path coefficient

1. はじめに

統計的品質管理での回帰分析の使われ方について、以前から強い疑問を抱いていた。品質管理の課題は、品質特性の値が理想値に近づく設計および製造条件を見出し、それを安定して実現することである。この目的からして、観察研究から求めた回帰式に対しても、品質改善のためには純粋な受動的予測にとどまらず、たとえば

- 有意な説明変数を調整（故意に変更）したときの目的変数の平均への効果
- 有意な説明変数を制御（故意に固定）したときの目的変数の分散への効果

を定量的に推定したい。新たに開発された設計条件下で製品が市場クレームを引き起こすことを事前に予測し、実際その通りになったとき、予測としては完璧でも品質保証としては落第である。クレームが予測されたときにはそれを回避する対策が求められる。

しかし、受動的に観察した要因変数と特性変数のデータから、要因変数のばらつきを小さく

* 東京工業大学大学院社会理工学研究科：〒152-8552 東京都目黒区大岡山 2-12-1
 E-mail: miyakawa.m.aa@m.titech.ac.jp

** 東京工業大学大学院社会理工学研究科：〒152-8552 東京都目黒区大岡山 2-12-1

したときの特性変数への効果を推測するのは難しい問題である。それは、相関・回帰分析の背後にある統計的関連モデルのもとでは、この推測が不可能であるからである。この推測が可能になるためには、統計的因果モデルに立脚した統計分析手法が必要である。

このような統計的因果モデルの起源は、Wright (1923) によって創始されたパス解析である。パス解析では、ある要因変数を人為的に1単位変化させたときの特性変数の期待される変化量を総合効果と呼び、この推定方法を論じた。1990年代になって、パス解析は因果ダイアグラムと呼ばれる統計的因果モデルによって精緻化された。これらは、たとえば Pearl (1998), Spirtes et al. (1998) の研究にその流れを見ることができる。本論文はこの流れに沿っている。また、Pearl (1995) は、ある要因変数 X を外的操作によりある値 x に固定したときの特性変数 Y の分布を $f(y|set(X=x))$ として与え、これを介入効果と呼んだ。この精緻化には非巡回的有向グラフで表現される確率モデル、すなわちグラフィカルモデルが貢献した。これらの研究成果は Pearl (2000) にまとめられている。

ある要因変数に外的操作を施したときの特性変数の分散について初めて定式化したのは黒木・宮川 (1999a) である。黒木・宮川が与えた分散への介入効果とは、 $f(y|set(X=x))$ での Y の分散であり、線形構造方程式のもとで具体的な形を示している。よって、要因変数の値をある値に固定したときの特性変数の分散を求めたことになる。しかし、現実の対策を考えると、介入可能な要因変数とは反事実的仮定が可能な処理変数に限定される。ここで、反事実的仮定が可能とは、ある個体について観測された処理条件とは異なる条件をとることが潜在的に可能であり、そのような状況の仮定に意味のある場合を意味する。一方で、興味ある特性変数に影響する要因変数には処理変数以外にも、個体に固有な共変量や処理変数から影響を受け得る中間特性がある。個体に固有な属性とは、たとえば人であれば性別や年齢で、これらについては反事実的仮定に意味が乏しい。Cox (1992) は、個体属性に加えて外的要因を共変量にみなすことが妥当な場合があることを指摘している。一方、中間特性は制御不能な結果系特性である。これら共変量や中間特性に対して直接介入することは難しい。

いま、共変量から中間特性に矢線があり、さらにこの中間特性から特性変数へ矢線がある場合を考えよう。この場合、中間特性の変動が特性変数の変動を引き起こしているのだが、その中間特性の変動の原因はその親である共変量の変動である。そこで、共変量から中間特性への影響の度合いを人為的に変更することを考える。このような変更が可能なのは、いわゆるタグメソッドにおけるパラメータ設計の多数の事例によって実証されている。たとえば、1953年に伊奈製陶(現 INAX)で行われたタイル実験の背景では、トンネルキルン内部の温度の不均一性が焼成後のタイル寸法の変動を生んでいた。これに対して実験的に様々な条件を設定したところ、ある添加剤を加えることで、キルン内部の温度変動がタイル寸法に与える影響を大幅に減少させることに成功したのである(田口 (1976), 17.1節)。このような操作は、変数でなく矢線に介入することとみなせる。黒田 他 (2006) は、この矢線への介入を定式化し、その識別可能条件を与えるとともに、線形構造方程式モデルのもとで具体的な推定法を明示した。本稿では、Pearl らが発展させた変数への介入効果に加えて、この矢線への介入効果についても併せてレビューし、品質管理など諸分野における因果推論のための回帰分析について考察する。

2. 動機付けとなった事例

奥野 他 (1986) の事例1「ボディ塗装条件の設定」で報告されている内容に基づいて本稿の問題意識を説明する。この事例では、自動車のボディ塗装工程において、塗着率を高め、かつ、安定化させる工程条件を設定するために、次の変数についてデータが採取されている。

データ数は $n=38$ で、要因変数として

- 塗料条件：希釈率 (x_1), 粘度 (x_2), 塗料温度 (x_8)
- 吹付条件：ガンスピード (x_3), 吹付距離 (x_4), 霧化エア－圧 (x_5), パターン幅 (x_6), 吐出量 (x_7)
- 環境条件：温度 (x_9), 湿度 (x_{10})

が測定され、特性変数として

- 塗着率 (y), 中心膜圧, 膜圧幅

が測定されている。以下の解析では、特性変数として塗着率のみを対象とする。

奥野 他 (1986, p. 47, p. 49) の記述によれば、取り上げた 10 個の要因変数には、制御可能性という観点から程度の異なるものが混在している。希釈率 (x_1) と吹付距離 (x_4) が最も制御しやすい要因で、粘度 (x_2), ガンスピード (x_3), 霧化エア－圧 (x_5), およびパターン幅 (x_6) がある程度制御できる要因とある。一方、塗料温度 (x_8) と吐出量 (x_7) は結果系として変動する量で、温度 (x_9), 湿度 (x_{10}) は制御していない環境条件である。このデータに対する奥野らの解析結果は次のようである (p. 57)。変数増減法により選択した回帰式は

$$\hat{y} = 80.39 - 0.933x_4 - 4.381x_6 + 0.032x_7 - 0.448x_8$$

で、この回帰式に対する解釈として

- ① 吹付距離 (x_4) を 10 cm 遠ざけると (他の条件は一定) 塗着率は、9.33%減る。
- ② パターン幅 (x_6) が 1 cm 広ければ (他の条件は一定) 塗着率は、4.38%減る。
- ③ 吐出量 (x_7) が 100 g 多ければ (他の条件は一定) 塗着率は、3.2%増える。
- ④ 塗料温度 (x_8) が 10°C 高ければ (他の条件は一定) 塗着率は、4.48%減る

を与えている。

この記述はやや誤解を招く内容だと考える。まず、塗料温度 (x_8) と吐出量 (x_7) は結果系として変動する量だから、直接の調整の対象にならない。吐出量について言えば「多くすると」でなく「多ければ」としているので、著者らも制御が難しいものであることを明らかに認識している。しかし、他の条件を一定にしたうえでの「多ければ」という状況を想定することに意味が乏しい。また、吹付距離 (x_4) は確かに制御しやすい変数であるが、ここで求めた偏回帰係数は直接効果であって、「1 cm 遠ざけた」ときの効果を表す総合効果ではない。このような適用例はこの文献に限るものでなく、1980 年代までは品質管理の分野では通常の事例であったと考える。

説明変数の制御可能性の違いや説明変数間の因果関係を考慮はしていても反映されていないこのような解析に対して、黒木・宮川 (1999a) はグラフィカルモデリングによって、上記 11 個の変数に対する因果ダイアグラムを作成した。それを図 1 に示す。図 1 の矢線にはパス係数を併記した (すべての変数を平均 0, 分散 1 に基準化している)。なお、温度と湿度の間の無向辺についた係数は標本相関係数である。

この因果ダイアグラムから推定される制御可能な (ある程度制御可能なものも含めて) 変数の総合効果は、4 節で与える結果を用いると

希釈率 (x_1): 0.065, 粘度 (x_2): -0.116, ガンスピード (x_3): 0.052

吹付距離 (x_4): -0.503, エア－圧 (x_5): -0.167, パターン幅 (x_6): -0.465

である。ちなみに吹付距離 (x_4) の標準偏回帰係数は -0.636 で、上の総合効果とは一致しない。それは、吹付距離 (x_4) から塗着率への有向道で、エア－圧 (x_5) やパターン幅 (x_6) の介在するものがあるからである。

一方で、図 1 の因果ダイアグラムは奥野らの回帰分析の結果と整合する。それは、奥野らの

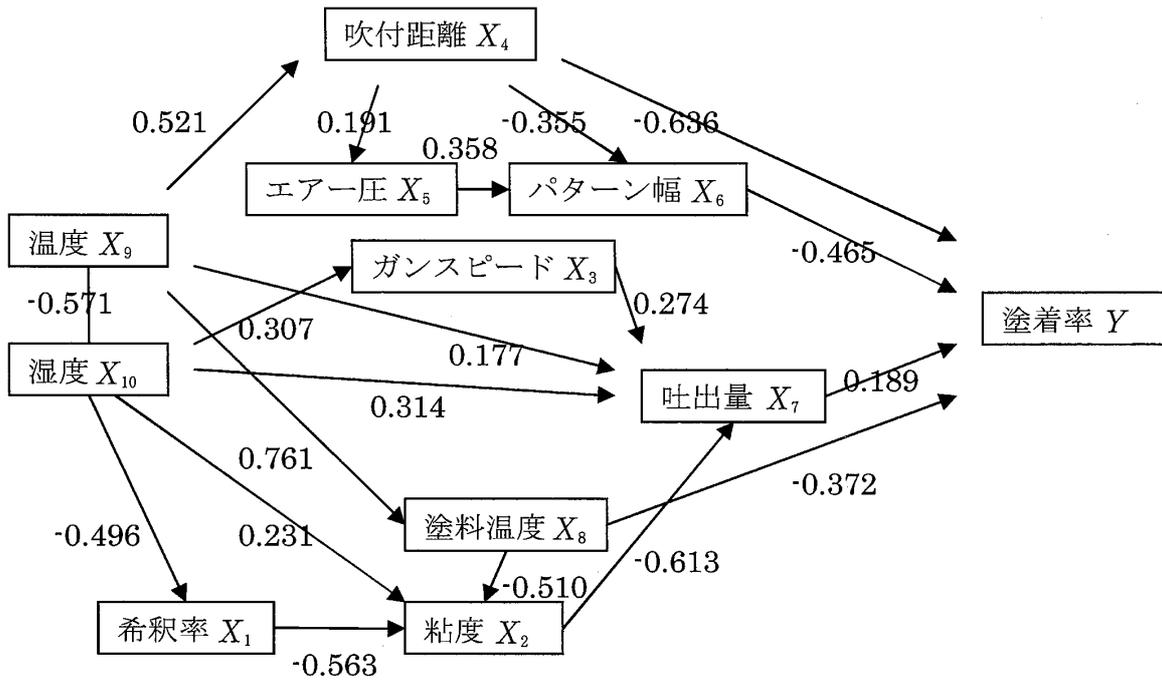


図1 「ボディ塗装条件の設定」に対する因果ダイアグラムとパス係数

回帰分析で選択された説明変数はすべて塗着率の親であるという点である。親を与えれば、子とその非子孫は条件付き独立になり、非子孫は説明変数に選択されなくなる。

ところで、図1を見ると、結果系の中間特性である塗料温度と吐出量から、特性変数である塗着率に向かって矢線が出ている。よって、これらの変動が塗着率の変動の要因になっている。しかし、これらはいずれも直接制御することが困難である。いま、塗料温度に注目すれば、塗料温度には環境条件である温度が効いている。ここでは温度も制御が困難な変数である。そこで、塗料の保管方式等を改善することで、温度の変動が塗料温度の変動に及ぼす影響を緩和できたとする。これはすなわち、温度から塗料温度への矢線に介入し、そのパス係数を人為的に変更することの意味する。このとき、最終特性変数である塗着率の変動はどれくらい小さくなるだろうか。また、その値はどのようにすれば推定できるのだろうか。

6節で与える結果を用いると、温度から塗料温度へ向かう矢線への介入効果は、誤差項に正規分布を仮定した線形構造方程式モデルのもとで、次のように推定できる。ここでは、図1の因果ダイアグラムが正しいとする。このもとでの11個の変数の相関係数行列の推定値は表1に示すものになる(黒木・宮川, 1999b)。

まず、温度から塗料温度へのパス係数を人為的介入により0に変更出来たときの塗着率 y の分散を求める。図1では、(塗料温度, 塗着率)に対して温度がバックドア基準(3節に定義を与える)を満たしているので、塗着率 y を目的変数、塗料温度 x_8 と温度 x_9 を説明変数にした回帰モデルを想定し、回帰係数を最小2乗推定すると

$$\hat{y} = -0.3125x_8 - 0.2371x_9 \quad (2.1)$$

を得る。このときの残差分散 $\sigma_{y \cdot x_8 x_9}^2$ の推定値は0.7333である。また、塗料温度を目的変数、温度を説明変数にした単回帰モデルでの残差分散 $\sigma_{x_8 \cdot x_9}^2$ の推定値は0.4209である。

6節に与える公式により、温度から塗料温度へのパス係数を0に変更したときの y の分散は、これらの値から

表1 図1の因果ダイアグラムに基づく相関行列の最尤推定値

| | x_1 | x_2 | x_3 | x_4 | x_5 | x_6 | x_7 | x_8 | x_9 | x_{10} | y |
|----------|--------|--------|--------|--------|--------|--------|--------|--------|--------|----------|-------|
| x_1 | 1.000 | | | | | | | | | | |
| x_2 | -0.736 | 1.000 | | | | | | | | | |
| x_3 | -0.152 | 0.210 | 1.000 | | | | | | | | |
| x_4 | 0.148 | -0.331 | -0.091 | 1.000 | | | | | | | |
| x_5 | 0.028 | -0.063 | -0.017 | 0.191 | 1.000 | | | | | | |
| x_6 | -0.043 | 0.095 | 0.026 | -0.286 | 0.291 | 1.000 | | | | | |
| x_7 | 0.324 | -0.479 | 0.195 | 0.184 | 0.035 | -0.053 | 1.000 | | | | |
| x_8 | 0.216 | -0.684 | -0.134 | 0.397 | 0.076 | -0.114 | 0.396 | 1.000 | | | |
| x_9 | 0.286 | -0.635 | -0.175 | 0.521 | 0.099 | -0.149 | 0.353 | 0.761 | 1.000 | | |
| x_{10} | -0.496 | 0.684 | 0.307 | -0.298 | -0.057 | 0.085 | -0.146 | -0.435 | -0.571 | 1.000 | |
| y | -0.091 | 0.326 | 0.134 | -0.614 | -0.277 | -0.250 | -0.044 | -0.493 | -0.475 | 0.283 | 1.000 |

$$0.7333 + (-0.3125)^2 \times 0.4209 + (-0.2371)^2 \times 1 = 0.8306$$

と推定される。また、 y の分散が最小になるような温度から塗料温度へのパス係数を求めると

$$-(-0.2371) / (-0.3125) = -0.7590$$

となり、このときの分散は

$$0.7333 + (-0.3125)^2 \times 0.4209 = 0.7744$$

である。これは温度が上がると塗料温度は下がるという関係なので、このような設定は困難かもしれないが、 y の分散がパス係数の $(-0.7590, 1)$ の範囲で増加関数になるという事実は対策を講じるうえで有益な示唆を与えるものといえる。

なお、温度からは吹付距離へも矢線がある。吹付距離は制御しやすい変数であるが、仮に中間特性とみなして温度からこれへのパス係数を最適化すると、その値は -0.4232 となり、そのときの y の分散は 0.7744 であり、上述の値と一致する。これは偶然でなく、ある条件のもとでは理論的に一致するのである。これについても6節で述べる。

3. 変数への介入効果

まず、因果ダイアグラムの定義を与える。

定義 3.1 (Pearl, 1995) 非巡回的有向グラフ $G=(V, E)$ と、その頂点に対応する確率変数の集合 $V=\{X_1, X_2, \dots, X_p\}$ が与えられている。グラフ G が変数間の因果的關係を

$$X_i = g_i(pa(X_i), \varepsilon_i) \quad (i=1, 2, \dots, p) \quad (3.1)$$

となる形で規定し、各変数がこの因果的關係にしたがって生成されるとき、グラフ G を因果

ダイアグラムという。ここに、 $pa(X)$ は X の親で、誤差変数 $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p$ は互いに独立とする。

確率変数間の因果的關係が (3.1) 式によって規定されたとき、その同時分布においてはグラフ G にしたがう逐次的因数分解

$$f_v(x_1, x_2, \dots, x_p) = \prod_{i=1}^p f_{i \cdot pa}(x_i | pa(x_i)) \quad (3.2)$$

が成立する (Pearl, 1995)。

因果ダイアグラムで成立しているマルコフ性を過不足なく記述するには、次の有向分離の概念が有用である。

定義 3.2 (Geiger et al., 1990) 非巡回的有向グラフ $G=(V, E)$ を考える。2つの頂点 α と β を結ぶすべての道のそれぞれについて、 $\{\alpha, \beta\}$ と排反な頂点集合 S が次の条件のいずれかを満たすとき、 S は α と β を有向分離 (d -separate) するという。

- 1) α と β を結ぶ道上の合流点で、その合流点とその子孫が S に含まれないものがある。
- 2) α と β を結ぶ非合流点で S に含まれるものがある。

なお、 α と β を結ぶ道がないときは、空集合が α と β を有向分離するという。

このとき、次のマルコフ性が成立する。

定理 3.1 (Geiger et al., 1990) 因果ダイアグラムにおいて、 $\{X_i, X_j\}$ と排反な変数集合 S が X_i と X_j を有向分離していれば、 S を与えたときに X_i と X_j は条件付き独立である。

次に、Pearl (1995) は、因果ダイアグラムが与えられたとき、ある変数に対して、それへの有向道をもつ変数には操作をせず、その変数自体を外的操作によってある値に固定する行為を介入 (intervention) と定義した。そして、変数 X に介入したときの別の変数 Y への因果的効果を介入効果と呼び、その数学的定義を次のように与えた。

定義 3.3 (Pearl, 1995) 頂点集合を $V=\{X, Y, Z_1, Z_2, \dots, Z_p\}$ とする因果ダイアグラム G において

$$f(y | set(X=x)) = \int \dots \int \frac{f_v(x, y, z_1, z_2, \dots, z_p)}{f_{x \cdot pa}(x | pa(x))} dz_1 \dots dz_p \quad (3.3)$$

を X の Y への介入効果という。

左辺に現れる $set(X=x)$ はセット・オペレーションと呼ばれ、介入によって X の値を x に固定したことを意味する。介入効果を因果ダイアグラムで解釈すると、 X へ向かう矢線を G よりすべて取り除いたグラフにおいて X の値を x に固定したときの Y の周辺密度関数である。

一方、Pearl and Robins (1995) は、ある変数 X に介入するときに、 X の値を既に観測されている変数集合 W の関数 $h(W)$ に設定する介入を条件付き介入と呼び、次のような定義を与えた。

定義 3.4 (Pearl and Robins, 1995) 因果ダイアグラム G において、頂点集合を $V=\{X, Y\} \cup S$ とし、 $\{X, Y\}$ と S とは排反であるとする。 X のすべての非子孫からなる変数集合のある部分集合 $W(\subset S)$ による X への条件付き介入を行うとする。このとき

$$f(y|\text{set}(X=h(W))) = \int \frac{f_v(y, h, s)}{f_{x,pa}(h(W)|pa(x))} ds \quad (3.4)$$

を W による X の Y への条件付き介入効果という。 $\text{set}(X=h(W))$ は外的操作により X の値を $h(W)$ に設定することを意味する。

頂点集合 $V = \{X, Y, Z_1, Z_2, \dots, Z_p\}$ の部分集合 $A = \{X, Y\} \cup S$ において、 A の周辺密度関数によって $f(y|\text{set}(X=x))$ が表現されるとき、 X の Y への介入効果は変数集合 A において識別可能であるという。識別可能となるひとつの十分条件としてバックドア基準がある。

定義 3.5 (Pearl, 1995) 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 つの条件を満たす頂点集合 S は (X, Y) に対してバックドア基準を満たすという。

- 1) X から S の任意の要素へ有向道がない。
- 2) X から出る矢線を因果ダイアグラム G より除いたグラフにおいて、 S は X と Y とを有向分離する。

バックドア基準を満たす頂点集合は必ず存在し、一般には複数存在する。また、因果ダイアグラム G より X から出る矢線を除いたグラフで、 X と Y を結ぶ非合流点のみからなる道がないとき、空集合はバックドア基準を満たす。

定理 3.2 (Pearl, 1995) 因果ダイアグラム G において、頂点集合 S が (X, Y) に対してバックドア基準を満たすならば、 X の Y への介入効果は $A = \{X, Y\} \cup S$ において識別可能であり、 $S = \{Z_1, Z_2, \dots, Z_r\}$ と記したとき介入効果は

$$f(y|\text{set}(X=x)) = \int \dots \int f_z(z_1, \dots, z_r) f_{Y \cdot XZ}(y|x, z_1, \dots, z_r) dz_1, \dots, dz_r \quad (3.5)$$

で与えられる。

条件付き介入効果でもバックドア基準が識別可能となるひとつの十分条件である。

定理 3.3 (黒木・宮川, 1999b) 因果ダイアグラム G において、変数集合 T が (X, Y) に対してバックドア基準を満たしているとする。 $W(C T)$ による X への条件付き介入を行うとする。このとき、 T が、 X, Y とともに観測されていれば、 W による X の Y への条件付き介入効果は識別可能であり、それは

$$f(y|\text{set}(X=h(W))) = \int f_{Y \cdot XT}(y|h(w), t) f_T(t) dt \quad (3.6)$$

で与えられる。

4. 線形構造方程式モデルでの変数への介入

ここでは、変数集合 V の要素がすべて連続変数の場合を考え、(3.1) 式に対応する基礎的モデルとして次の線形構造方程式モデル

$$X_i = \sum_{X_j \in pa(X_i)} \alpha_{ij} X_j + \varepsilon_i \quad (i=1, 2, \dots, p) \quad (4.1)$$

を想定する。ここに、 V の各要素は平均 0、分散 1 に基準化されているものとし、誤差変数 $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p$ は互いに独立に正規分布にしたがうものとする。右辺の母数 α_{ij} が X_j から X_i への直接的因果の大きさを示すパス係数である。線形構造方程式モデルを表現した因果ダイアグラ

ムはパスダイアグラムと呼ばれる。

観察研究で採取される統計データより、このパス係数を推定するひとつの方法は回帰分析である。このとき、因果推論のための回帰分析を合理的なものにするうえで、想定する回帰モデルの偏回帰係数が構造方程式モデルのパス係数に一致するための十分条件を明らかにしておく必要がある。ひとつの自明な十分条件は「ある変数を目的変数にとり、その変数の親のすべてを説明変数にした回帰モデルでの偏回帰係数は、対応するパス係数に等しい」である。

一方、因果仮説をより明確に記述するために、直接観測できない変数を含めてパスダイアグラムを作成することがある。その場合、ある変数のすべての親を観測できないことがある。そこで、パス係数を識別するための実質的な必要十分条件が求められる。その条件が次の単一結合基準である。

定義 4.1 (Pearl, 1998) 因果ダイアグラム G において X から Y への矢線があるとする。このとき、次の2つの条件を満たす頂点集合 S は (X, Y) に対して単一結合基準を満たすという。

- 1) Y から S の任意の要素へ有向道がない。
- 2) X から Y への矢線を因果ダイアグラム G より除いたグラフにおいて、 S は X と Y とを有向分離する。

このとき、次の命題が成り立つ。

定理 4.1 (Pearl, 1998) 非巡回的有向グラフであるパスダイアグラム G において X から Y への矢線があるとする。このとき、頂点集合 S が (X, Y) に対して単一結合基準を満たすならば、 Y を目的変数、 X と集合 S に含まれるすべての変数を説明変数にした回帰モデルでの X の偏回帰係数 $\beta_{yx \cdot S}$ は、 X から Y へのパス係数 α_{yx} に等しい。

一方で、処理変数にとっての主たる興味は直接効果よりも総合効果である。これが処理変数 X を人為的介入により1単位変化させたときの反応変数 Y への因果的効果を表すからである。パスダイアグラムにおいて、処理変数 X から反応変数 Y へ少なくともひとつの有向道（直接の矢線も含めて）があるとき、処理変数 X から反応変数 Y への総合効果は、これらの有向道にそったパス係数の積和で表される。よって、すべてのパス係数が推定できていれば、これらを積み上げることで総合効果は推定される。しかし、単一の回帰モデルで総合効果を推定したい場面もある。そのとき役立つのが前述のバックドア基準である。

定理 4.2 (Pearl, 1998) 非巡回的有向グラフであるパスダイアグラム G において X から Y への有向道があるとする。このとき、頂点集合 S が (X, Y) に対してバックドア基準を満たすならば、 Y を目的変数、 X と集合 S に含まれるすべての変数を説明変数にした回帰モデルでの X の偏回帰係数 $\beta_{yx \cdot S}$ は、 X から Y への総合効果に等しい。

次に、ある要因変数 X に介入したときの特性変数 Y の分散、すなわち、定義 3.3 に与えた介入効果 $f(y|\text{set}(X=x))$ での Y の分散について述べておく。この分散は、バックドア基準を満たす変数を説明変数に加えた回帰モデルの母数により具体的に表現できる。簡単のため、単一変数 Z が (X, Y) に対してバックドア基準を満たしているとする。すると回帰モデル

$$Y = \beta_{yx \cdot z} x + \beta_{yz \cdot x} z + \varepsilon_{y \cdot xz} \quad (4.2)$$

における x の偏回帰係数 $\beta_{yx \cdot z}$ は、定理 4.2 より、 X から Y への総合効果を表している。さらにこのとき、介入効果 $f(y|\text{set}(X=x))$ での Y の分散は

$$V(Y|\text{set}(X=x)) = \sigma_{y \cdot xz}^2 + \beta_{yz \cdot x}^2 \sigma_z^2 \quad (4.3)$$

と表現される (黒木・宮川, 1999a). ここに, $\sigma_{y \cdot xz}^2 = V(\varepsilon_{y \cdot xz})$ であり, σ_z^2 は Z の分散で, 上記基準化のもとでは 1 である. Z がベクトル変数の場合については, 黒木・宮川 (1999a) を参照されたい.

2節にとり上げた事例で, 具体的に求めてみよう. 図1において, 吹付距離 (x_4) は最も制御しやすい変数ということなので, これをある値に固定したとする. (吹付距離, 塗着率) に対してバックドア基準を満たす変数として吹付距離の親である温度 (x_9) をとり上げると, 表1に与えた相関行列より, (4.3) 式の値は 0.635 となる.

さらに, 条件付き介入をしたときの特性変数 Y の分散についても併せて考察しておこう. 本節では, 線形構造方程式モデルを仮定している. そこで, 定義 3.4 での条件付き介入の関数 $h(W)$ についても

$$h(W) = x_0 + a^T W \quad (4.4)$$

という線形構造を仮定し, これを線形条件付き介入と呼ぶ. さらに簡単のため, W が単一変数で, $\{Z, W\}$ が $\{X, Y\}$ に対してバックドア基準を満たすとする. すると, 回帰モデル

$$Y = \beta_{yx \cdot zw} x + \beta_{yz \cdot xw} z + \beta_{yw \cdot xw} w + \varepsilon_{y \cdot xzw} \quad (4.5)$$

を想定すれば, $X = x_0 + aW$ と線形条件付き介入したときの Y の分散は

$$V(Y|\text{set}(X=x_0+aW)) = \sigma_{y \cdot xzw}^2 + \beta_{yz \cdot xw}^2 \sigma_z^2 + (a\beta_{yx \cdot zw} + \beta_{yw \cdot xz} + \rho_{zw} \beta_{yz \cdot xw})^2 \sigma_w^2 \quad (4.6)$$

となる (黒木・宮川, 1999b). ここに ρ_{zw} は Z と W の相関係数である. ここで, (4.6) 式の右辺第3項が 0 になるように, 線形条件付き介入の係数 a を

$$a^* = -\frac{\beta_{yw \cdot xz} + \rho_{zw} \beta_{yz \cdot xw}}{\beta_{yx \cdot zw}} \quad (4.7)$$

とおけば, Y の分散は最小になる. (4.7) 式を最適な線形条件付き介入と呼ぶ. Z と W がベクトル変数の場合については, 黒木・宮川 (1999b) を参照されたい.

2節の例で最適な線形条件付き介入をしたときの挙動を具体的に見てみよう. 上の無条件介入のときと同様に, 吹付距離 (x_4) を処理変数 X として, これに条件付き介入する. 条件とする変数 (制御のための共変量) は吹付距離の非子孫であれば任意であり, ここでは塗料温度 (x_9) を W とする. また, (吹付距離, 塗着率) に対してバックドア基準を満たす変数として温度 (x_9) を識別のための共変量 Z とする. すると, 最適な線形条件付き介入をしたときの塗着率の分散は, (4.6) 式の右辺第1項と第2項の和で与えられ, 0.549 となる. 当然ながら, この値は上で求めた無条件介入したときの分散値 0.635 よりも小さい. この有意差の検討については今後の課題とする.

なお, ここで, 線形構造方程式モデルでの平均や分散への介入効果を回帰分析で推定することの意義を述べておく. 線形構造方程式モデルでのパス係数の推定は, 古典的には構造方程式の各式を回帰モデルとしたときの最小2乗推定で行われていた (たとえば Asher (1976) 3.1節). 一方, 最近では, 共分散構造分析で潜在変数がないという特別な場合として最尤推定が用いられることが多い. しかし, 共分散構造分析の枠組みで最尤推定を行うには, 先験的に作成した因果ダイアグラムに現れるすべての変数が観測されることが前提となる. これに対して,

本論では、処理変数から反応変数への介入効果を推定するうえで観測すべき共変量を因果ダイアグラムから選択する立場に立っている。ここで推定すべきパラメータは、処理変数から反応変数への総合効果であり、すべてのパス係数ではない。さらには、回帰分析の解析ツールは品質管理をはじめ多くの分野で十分に普及しているという事実がある。これらの点から、バックドア基準を利用した回帰分析での最小2乗推定に意義があると考えられる。

5. 矢線への介入効果

この節では黒田 他 (2006) が与えた矢線への介入効果について述べる。矢線への介入を (3.1) 式に与えた構造方程式の変更と定める。すなわち、与えられた因果ダイアグラムに対して、頂点 X_i に向かう矢線のすべてあるいは一部に介入するとは、対応する構造方程式 $X_i = g_i(\text{pa}(X_i), \varepsilon_i)$ を外的操作により別な $X_i = h(\text{pa}(X_i), \varepsilon_i)$ に変更することとする。これを $\text{intervene}(\text{pa}(X) \rightarrow X)$ と表記する。この構造方程式の変更に伴い、(3.2) 式の逐次的因数分解において $f_{i \cdot \text{pa}}(x_i | \text{pa}(x_i))$ が新たな条件付き密度関数 $f^{\text{New}}(x_i | \text{pa}(x_i))$ に変わる。これより、矢線への介入効果を次のように定義する。

定義 5.1 (黒田 他, 2006) 頂点集合を $V = \{X, Y\} \cup S$ とする因果ダイアグラム G において

$$f(y | \text{intervene}(\text{pa}(X) \rightarrow X)) = \int \frac{f_v(y, x, s)}{f_{X \cdot \text{pa}}(x | \text{pa}(x))} f^{\text{New}}(x | \text{pa}(x)) dx ds \quad (5.1)$$

を矢線 $(\text{pa}(X) \rightarrow X)$ への Y に対する介入効果という。ここに $f^{\text{New}}(x | \text{pa}(x))$ は外的操作で設定された $X = h(\text{pa}(X), \varepsilon)$ に対応する条件付き密度関数である。

矢線への介入効果は次のように表現できる。

定理 5.1 (黒田 他, 2006) 頂点集合を $V = \{X, Y\} \cup S$ とする因果ダイアグラム G において、 $\{X, Y\}$ と S は排反であるとする。矢線 $(\text{pa}(X) \rightarrow X)$ への Y に対する介入効果は

$$f(y | \text{intervene}(\text{pa}(X) \rightarrow X)) = \int \int \frac{f_v(y, h, s) f_{\varepsilon_x}(e)}{f_{X \cdot \text{pa}}(h | \text{pa}(x))} ds de \quad (5.2)$$

と表現できる。

次に矢線への介入効果に関する識別可能条件として次の命題を得る。

定理 5.2 (黒田 他, 2006) 因果ダイアグラム G において、 (X, Y) について変数集合 T がバックドア基準を満たすとき、 X, Y, T の同時分布、かつ、 ε_x の分布 $f_{\varepsilon_x}(e)$ が識別可能ならば、矢線 $(\text{pa}(X) \rightarrow X)$ への Y に対する介入効果は識別可能で

$$f(y | \text{intervene}(\text{pa}(X) \rightarrow X)) = \int \int f_{Y \cdot XT}(y | h, t) f_T(t) f_{\varepsilon_x}(e) dt de \quad (5.3)$$

で与えられる。

この定理において X, Y, T の同時分布が多変量正規分布に従えば、線形構造モデルにおいて誤差項の正規性が成り立つので次の系が導かれる。

系 5.1 (黒田 他, 2006) T が $\text{pa}(X)$ を含み、かつ、 X, Y, T の同時分布が多変量正規分布に従えば、矢線 $\text{pa}(X) \rightarrow X$ への介入効果は、 X, Y, T において識別可能である。

6. 線形構造方程式モデルでの矢線への介入

この節では、4節と同様に、(4.1)式の線形構造方程式モデルを想定する。この線形構造方程式モデルのもとでは、矢線への介入は単純で、矢線に対応するパス係数の変更他にない。いま、与えられたパスダイアグラムにおいて、 X から Y へ有向道があり、 X へ向かうひとつ以上の矢線があるとする。このとき、 X へ向かういくつかの矢線へ介入したときの Y の分散は、 Y を目的変数、 X と X のすべての親 Z を説明変数にした回帰モデルのパラメータにより次のように記述できる。まず、介入前の X から Y への総合効果を求めるための回帰モデルとして

$$Y = \beta_{yx \cdot z} x + \beta_{yz \cdot x}^T z + e_{y \cdot xz} \quad (6.1)$$

を設定する。ここで $\{Z\}$ が (X, Y) に対してバックドア基準を満たしていることに注意する。すると、矢線 $Z \rightarrow X$ へ介入、すなわち Z から X への矢線のパス係数ベクトルを α_{xz} から γ_{xz} に変更したときの Y の分散は、次の定理で記述される。

定理 6.1 (黒田 他, 2006) パスダイアグラム G において、変数集合 Z を X の親集合とする。 X, Y, Z が多変量正規分布に従うならば、矢線 $\text{pa}(X) \rightarrow X$ の介入効果をしたときの Y の分散は

$$V(Y|\text{intervene}(Z \rightarrow X)) = (\beta_{yx \cdot z} \gamma_{xz} + \beta_{yz \cdot x})^T \Sigma_z (\beta_{yx \cdot z} \gamma_{xz} + \beta_{yz \cdot x}) + \beta_{yx \cdot z}^2 \sigma_{x \cdot z}^2 + \sigma_{y \cdot xz}^2 \quad (6.2)$$

で与えられる。ここに、 Σ_z は Z の分散共分散行列である。

ここで、特性変数 Y の分散を最小にするような矢線への介入方式を考えると、定理 6.1より直ちに次の系を得る。

系 6.1 (黒田 他, 2006) パスダイアグラム G において、変数集合 Z が X の親集合とする。 X, Y, Z が多変量正規分布に従うならば、矢線へ介入したときの Y の分散は

$$\gamma_{xz}^T = -\frac{\beta_{yz \cdot x}^T}{\beta_{yx \cdot z}} \quad (6.3)$$

のときに最小となり、そのときの分散は

$$V(Y|\text{intervene}(Z \rightarrow X)) = \beta_{yx \cdot z}^2 \sigma_{x \cdot z}^2 + \sigma_{y \cdot xz}^2 \quad (6.4)$$

である。

この(6.4)式より、同一の親をもつ矢線への介入について次の命題が導かれる。

定理 6.2 (黒田 他, 2006) 因果ダイアグラム G において、複数の変数が同一の親をもち、かつ、それらから Y への有向道があるとする。このとき、その親からそれらの変数へのいずれかの矢線に対して(6.3)式の介入を行ったならば、介入後の Y の分散は選択した矢線に依らず一定である。

線形構造方程式モデルにおける矢線への介入の適用例は、既に2節で示している。そこでは、吹付距離が処理変数であるにもかかわらず、あえて温度から吹付距離への矢線へ最適介入した場合を想定し、そのときの塗着率の分散が0.7744であることを述べた。これは、この定理 6.2

の例示であるとともに、4節で求めた無条件介入、最適な線形条件付き介入をした場合との比較でもある。この場合は矢線への介入の効果は、これらに劣っていることが観察される。このとき、次の点に注意すべきである。矢線への介入では、一旦介入してパス係数の値を変更すれば、その後は特に工程に操作する必要はない。これに対して条件付き介入では、定期的な制御のための共変量 W を観測し、それに応じて後続する処理変数の値を $X=h(W)$ に設定するのであるから、その手間は矢線への介入とは比べものにならない。無条件介入は、一旦設定すればよいという点では矢線介入と同じであるが、工程での量的な処理変数を誤差なく一定の値に保つのは概して難しい。

7. おわりに

本稿では、生産工程の品質管理を題材にしながら、線形構造方程式モデルに基づく因果推論のための回帰分析についてのこれまでの研究成果を要約した。本稿の主張点は次のようにまとめられる。

- 1) 説明変数を操作可能な処理変数とそうではない共変量、中間特性に峻別することが重要である。総合効果、あるいはそれを一般化した変数への介入効果を求める意味がある変数は処理変数のみである。
- 2) このとき、特性変数の平均への効果を表す総合効果とともに、特性変数の分散への介入効果（条件付き介入を含めて）を求めることが品質管理では重要である。
- 3) 一方、直接操作不能な中間特性や共変量の変動も特性変数の変動要因になっているので、これらについては、そこへ向かう矢線への介入効果に利用価値がある。
- 4) 1) から 3) のいずれにおいても、興味ある介入効果を相関データから識別するための変数選択原理が重要である。最終品質特性を目的変数とした回帰分析において、介入対象となる変数（矢線の場合はその親と子）に加えてどの変数を説明変数に含めれば興味ある介入効果が識別できるか、という問題である。これについてはバックドア基準が本質的役割を果たす。

謝 辞

有益なコメントをしていただきました2名の査読者の先生方、および本稿執筆の機会を与えていただいた75周年記念事業委員会の方々に感謝します。

参 考 文 献

- Asher, H. B. (1976). *Causal Modeling*, Sage Publications, London. (広瀬弘忠訳 (1980) 『因果分析法』朝倉書店).
- Cox, D. R. (1992). Causality: some statistical aspects, *J. Roy. Statist. Soc., B*, **34**, 187-220 (with discussions).
- Geiger, D., Verma, T. S. and Pearl, J. (1990). Identifying independence in Bayesian networks, *Networks*, **20**, 507-534.
- 黒木 学, 宮川雅巳 (1999a). 「因果ダイアグラムにおける介入効果の推定と工程解析への応用」『品質』**29**, 237-247.
- 黒木 学, 宮川雅巳 (1999b). 「適応制御における条件付き介入効果の定式化とその推定」『品質』**29**, 407-486.
- 黒田健成, 宮川雅巳, 田中研太郎 (2006). 「因果ダイアグラムにおける矢線への介入効果の定式化とその応用」『応用統計学』**35**, 79-91.
- 宮川雅巳 (1997). 『グラフィカルモデリング』朝倉書店.
- 宮川雅巳 (2004). 『統計的因果推論—回帰分析の新しい枠組み—』朝倉書店.
- 奥野忠一, 片山善三郎, 上野長昭, 伊東哲二, 入倉則夫, 藤原信夫 (1986) 『工業における多変量データの解析』日科技連出版社.
- Pearl, J. (1995). Causal diagrams for empirical research, *Biometrika*, **82**, 669-709.
- Pearl, J. (1998). Graphs, causality, and structural equation models, *Sociological Methods and Research*, **27**, 226-284.
- Pearl, J. (2000). *Causality; Models, Reasoning, and Inference*, Cambridge University Press.
- Pearl, J. and Robins, J. M. (1995). Probabilistic evaluation of sequential plans from causal model with hidden variables.

- Proc. 11th Conf. On Uncertainty in AI*: San Francisco, Morgan Kaufmann, 444-453.
- Spirtes, P., Richardson, T., Meek, C., Scheines, R. and Glymour, C. (1998). Using path diagram as a structural equation modeling tool, *Sociological Methods and Research*, **27**, 182-225.
- 田口玄一 (1976). 『実験計画法 (上)』丸善.
- Wright, S. (1923). The theory of path coefficients: a reply to Nile's criticism, *Genetics*, **8**, 239-255.
- Wright, S. (1934). The method of path coefficients, *Ann. Math. Statist.*, **5**, 161-215.