

Title	統計的因果探索アルゴリズム“LINGAM”を活用した専攻分野別の博士課程進学に関する研究
Author(s)	高山, 正行; 小柴, 等; 前田, 高志 ニコラス; 三内, 顕義; 清水, 昌平; 星野, 利彦
Citation	年次学術大会講演要旨集, 37: 192-197
Issue Date	2022-10-29
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/18520
Rights	本著作物は研究・イノベーション学会の許可のもとに掲載するものです。This material is posted here with permission of the Japan Society for Research Policy and Innovation Management.
Description	一般講演要旨

統計的因果探索アルゴリズム“LiNGAM”を活用した 専攻分野別の博士課程進学に関する研究

○ 高山 正行 (NISTEP, MEXT), 小柴 等 (NISTEP), 前田 高志 ニコラス (NISTEP, 東京電機大学),
三内 顕義 (NISTEP, 理研 AIP), 清水 昌平 (NISTEP, 理研 AIP, 滋賀大学), 星野 利彦 (NISTEP, QST)

1 はじめに

我が国の科学技術・イノベーション政策の中核をなす研究力強化の文脈において、若手研究者支援の議論は欠くことのできない位置を占めている。ここで若手研究者支援の枠組みの中でも、特に若手研究者のエントリーポイントとも言える博士（後期）課程^{*1}への進学支援に目を向けると、これまでも様々な施策が適用されてきているものの、博士課程進学率は長らく減少傾向にある。そのため、「博士課程進学率の向上」という目標に対して、既存の政策的手段をこれまでの規模を維持して継続するのみでは、博士課程進学率のV字回復は見込めない。今後V字回復を実現するには、修士課程学生の博士課程進学の阻害要因を乗り越えるべく、必要な政策的手段の洗い出しと規模間に関する定量的な特定が改めて必要となる。つまり、EBPM (Evidence-Based Policy Making) の観点からも、博士課程進学に関連すると思われる要因や手段となる各種政策要素と博士課程進学率の間の因果関係についての定量的な分析・理解が重要である。

ところで近年、大量のデータを用いて様々な分析や予測を行う、いわゆるデータサイエンスが盛り上がりを見せている。政策科学の領域においても、EBPMの文脈とも相まってこうした数理手法・モデルの活用の可能性が改めて認識されつつある[高山 21a]。上述した博士課程進学率に関する政策研究においても、経済的支援や研究環境、アカデミックキャリアパスに関する各種変数と博士課程進学率との関係について、統計的因果探索手法 LiNGAM (Linear Non-Gaussian Acyclic Model)[Shimizu06, Shimizu11, 清水 17]を活用した試行的分析がなされている[高山 21b, 高山 22a, 高山 22b]。

一方、この因果探索の試行的分析では今後の課題として、

- 各種変数による影響の表出の遅れを考慮した、時系列データの分析
- 分野ごとの特色を踏まえた因果探索と比較

等を指摘している[高山 22a, 高山 22b]。

本研究では、これらの問題点を克服すべく、専攻分野別での博士課程進学率・経済的支援・研究環境・アカデミックキャリアパスに関する各種データを構築し、さらに LiNGAM を用いて遅延効果も考慮した因果グラフの推定・比較を試みた。本稿では、その計算方法・結果について議論する。

2 研究手法

2.1 データセット構成にあたっての変数選定の考え方

本研究の特徴は博士課程進学に関するいくつかの要素について、1. 専攻分野別に、2. 遅延も考慮して、因果探索する点にある。

ここで本研究において構成するデータセットの専攻分野の分類は学校基本調査等の分類を念頭に、各分類が占める割合なども考慮して、1. 人文・社会科学、2. 理学、3. 工学、4. 農学、5. 保健、6. その他、の6区分とし、これに「全体」を加えて計7区分で設定した。

変数については先行研究をベースに設定した。一般に博士課程進学率に関連する要因・政策手段には様々なものが考えられる。本領域において DirectLiNGAM を試行的に活用した研究では、定量的に計測できる要因・政策手段として経済的支援や研究環境、アカデミックキャリアパスに関する変数をいくつか挙げ、それらの統計値を活用してデータセットを構築・分析している[高山 21b, 高山 22a, 高山 22b]。今回の分野別分析を行うにあたって、この考え方に準じて変数を選定した。ただし、分野別に値をアサインすることが困難な部分もあるため、一部変数について代替や削除も行った。

以下、(A) 経済的支援、(B) キャリアパス、(C) 研究環境の3つの観点から説明する。

^{*1} 本稿では、単純に「博士課程」という。

表 1: 本研究における暫定的分類と DC1 分類の対応。

DC1 分野分類	本研究での分野分類						
	人文・ 社会科学	理学	工学	農学	保健	その他	全体
人文学	1/2					1/2	2/2
社会科学	1/2					1/2	2/2
数物系科学		1/3	1/3			1/3	3/3
化学		1/3	1/3			1/3	3/3
生物系科学		1/4	1/4		1/4	1/4	4/4
情報学		1/3	1/3			1/3	3/3
工学系科学			1/2			1/2	2/2
農学・環境学				1/2		1/2	2/2
医歯薬学					1/2	1/2	2/2

* 採択者数の和をとる際の重率

(A) 経済的支援に関わる変数

国による経済的支援のうち、博士課程進学前に受給が決まるものとしては、まず日本学術振興会 (JSPS) の特別研究員 DC1 が挙げられる。他にもグローバル COE やリーディング大学院、卓越大学院といった文部科学省の事業が挙げられるが、これらは各大学が設計したプログラム単位での採択であり、そのプログラムのコンセプトは様々で、どの分野に紐づいたものか、一つ一つ特定することは困難である。よって今回は、

- 博士進学の前年度の DC1 採択者数

のみを変数として採用した。

一方で、DC1 に関する分野分類と上述の学校基本調査等の分野分類とは必ずしも一対一対応していない。例えば、DC1 における「情報」分野は、本研究の分類でいう「理学」にも「工学」にも属しうる。そこで DC1 の採択者数を本研究の分野分類に対応させるに当たって、その対応・計算のための重率を暫定的に表 1 の通り整理した。例えば、本稿でいう「人文・社会科学」における DC1 の採択者数は、DC1 での「人文学」採択者数に重率 $1/2 \cdot 2$ を乗じたものと、「社会科学」採択者数に重率 $1/2$ を乗じたものの和としている。厳密には、例えば「数物系科学」分野での DC1 採択者のうち何割が、本研究でいう「理学」・「工学」分野に該当するものかをそれぞれ数え上げる必要があるが、今回は試行的分析として簡易的な対応付けを採用した。

(B) キャリアパスに関わる変数

^{*2} 人文学が「人文・社会科学」と「その他」で 2 回登場していることにより、全分野で暫定的に同じ重率 $1/2$ としている。表 1 に記載されている他の重率も同様の考え方である。

本研究では、先に挙げた先行研究 [高山 21b, 高山 22a, 高山 22b] と同様、

- 博士課程修了直後の大学教員としての就職率
- 博士課程修了直後のポストドクターとしての就職率

を変数として採用し、年・分野別で算出した。なお、ポストドク就職者数について、2011 年度以前はポストドクター等の定義が定まっていない状態でもあり、学校基本調査の本体調査では集計されていないかった。そこで本研究では、報告されている他の補完的な調査結果を適宜キャリブレーションしながら活用し、さらに全く調査がなされていない年度については、前後の直近の年度の 2 値から線形補完で処理した。

(C) 研究環境に関する変数

研究環境について、

- 大学の研究本務者一人当たりの基盤的経費
- 研究時間割合

を採用した。ただし、「大学の研究本務者一人当たりの基盤的経費」については、分野別に切り出すことが不可能なため、先行研究 [高山 21b, 高山 22a, 高山 22b] を直接踏襲することができない。そこで本研究では基盤的経費の代わりに、総務省の科学技術研究調査から大学の研究開発費を採用した。また、研究時間割合については“大学等におけるフルタイム換算データに関する調査 (FTE 調査)”を用いた。ただし FTE 調査は現状 5 年ごとの調査のため、1 年ごとに得られる他の変数に合わせて、各年度の値は線形補完で埋めて処理した。

以上 6 つの変数について、2006 年度～2020 年度の年度単位で 15 点、データを収集した。

2.2 遅延を考慮した因果探索のためのアルゴリズムとデータセットの構築

本研究のように、時系列データを対象に遅延 (lag) を考慮した因果探索を行うためのアルゴリズムとして VARLiNGAM (Vector AutoRegression-LiNGAM) が挙げられる [Hyvarinen10]。この方法は、すべての変数について、任意の lag までを考慮して時系列の因果探索を実行できる点で優れている。ただし、この lag の設定には制限もある。時系列データのデータ点数を N 、設定する最大の lag に相当する時系列の点数を $\tau_{\max}$ 、変数の数を p とおくと、すべての変数について $\tau_{\max}$ までのすべての lag を考慮して、回帰分析を正しく実行し各因果

表 2: 構成したデータセットの変数および事前知識の一覧。

変数 (単位)	変数名	事前知識
博士課程進学率	x_0	$x_6 \sim x_{11}$ の影響なし
DC1 採択者数 (人)	x_1	$x_6 \sim x_{11}$ の影響なし
研究開発費 (億円) ^{*1}	x_2	$x_6 \sim x_{11}$ の影響なし
研究時間割合	x_3	$x_6 \sim x_{11}$ の影響なし
大学教員就職率 ^{*2}	x_4	$x_6 \sim x_{11}$ の影響なし
ポスドク就職率 ^{*2}	x_5	$x_6 \sim x_{11}$ の影響なし
博士課程進学率_L1	x_6	
DC1 採択者数_L1 (人)	x_7	外生変数
研究開発費_L1 (億円) ^{*1}	x_8	外生変数
研究時間割合_L1	x_9	
大学教員就職率_L1 ^{*2}	x_{10}	
ポスドク就職率_L1 ^{*2}	x_{11}	

*1: 一人当たり, *2: 博士修了直後
_L1 は lag 1, すなわち1年先の値

係数の最適解を一意に定めるには,

$$N - \tau_{\max} > p(\tau_{\max} + 1) \quad (1)$$

を満たす必要がある。しかし, 本研究で構築しているデータセットにおいては変数の数 $p = 6$ に対して, データ点数 $N = 15$ と非常に少なく, 最大 lag $\tau_{\max} = 1$ であり, 2 点以上の lag を VARLiNGAM で扱うには, p を 4 以下まで小さくせざるを得ない。一方で, このように $\tau_{\max}$ を大きくしようとして p を小さくすることで, 必要な変数を取り入れられなくなる可能性もある。

そこで本研究では, データセットを工夫することで VARLiNGAM の代わりに, DirectLiNGAM で時系列の因果探索を実施した。具体的には, 2.1 節に列挙した各変数について, 見たい lag の分だけ時間を先にずらした新たなデータを作成し, データセットに追加した。

τ だけ先にずらした変数の数を $p(\tau)$ とすると, 式 (1) に対応する, 因果探索が正しく実行される必要条件是,

$$N - \tau_{\max} > \sum_{\tau=0}^{\tau_{\max}} p(\tau) \quad (2)$$

のように表現される。式 (1) の右辺は関与しうる変数すべてについて, $\tau_{\max}$ までの lag を考慮することによるものであるのに対し, 式 (2) は $\tau_{\max}$ までの各 lag において, 影響を調べる変数を任意に決められるので, 式 (2) の場合の方が変数を残しやすくなり, またより大きな $\tau_{\max}$ をとることも可能となる。

表 2 に, 実際に本研究で構築したデータセットの概要をまとめた。本研究では試行的に, 6 つすべての変数について $\tau_{\max} = 1$ までの範囲とし, 1 年先の変数を加えている。データの点数は, 2006 年度～2019 年度の年度単位で 14 点となっている。

■計算に関する諸条件 本研究における DirectLiNGAM での因果探索にあたり, 各変数の因果関係に関する事前知識 (prior knowledge) について, 表 2 に併せて示した。

x_7 (DC1 採択者数_L1) ・ x_8 (一人当たりの研究開発費_L1) については, 政府の予算額によって定まるものであるため, 政府の意思決定により任意に定められる外生変数とし, 他の変数からは何も影響を受けないものとした。また, $x_6 \sim x_{11} \rightarrow x_0 \sim x_5$ は未来から過去への影響となるため, これらのパスは禁止した。

また, 先行研究 [高山 21b, 高山 22a, 高山 22b] を踏襲し, 本研究でも積の構造的因果モデルを導入し, DirectLiNGAM を適用した。データセットに対しては, 分野ごとに, 対数をとってから各分野での平均値を引く操作を実施している。

3 計算結果

■「全体」の区分に対する因果探索の結果 表 2 に基づいて構成したデータセットのうち, 分野を問わない「全体」の区分に対する DirectLiNGAM に対する DirectLiNGAM での計算結果を図 1 に示した。この計算結果において, 特に x_6 (博士課程進学率_L1) に対する影響を中心に考えると, 特徴的な点は, 以下の通りである。

- 係数が 0.1 以上の正の影響を与えている箇所は, x_2 (一人当たり研究開発費) ・ x_3 (研究時間割合) ・ x_7 (DC1 採択者数_L1) となっている。
- 係数が 0.1 以下の負の影響を与えている箇所は, x_4 (大学教員就職割合) ・ x_8 (一人当たり研究開発費_L1) となっている。
- x_6 (博士課程進学率_L1) は 1 年前の自分自身 (つまり x_0 (博士課程進学率)) の影響, つまり自己回帰の存在可能性が示唆される計算結果となっている。

■各分野での因果探索の結果における博士課程進学率_L1 への影響比較 続いて, 分野別のデータに対して同様に因果探索を行った結果のうち, x_6 (博士課程進学率_L1) への影響について比較する。表 3 では, x_6 の 1 年前の変数である $x_0 \sim x_5$, そして x_6 と同じ年度の変数である $x_7 \sim x_{11}$ からの影響に関する係数の計算結果を示した。

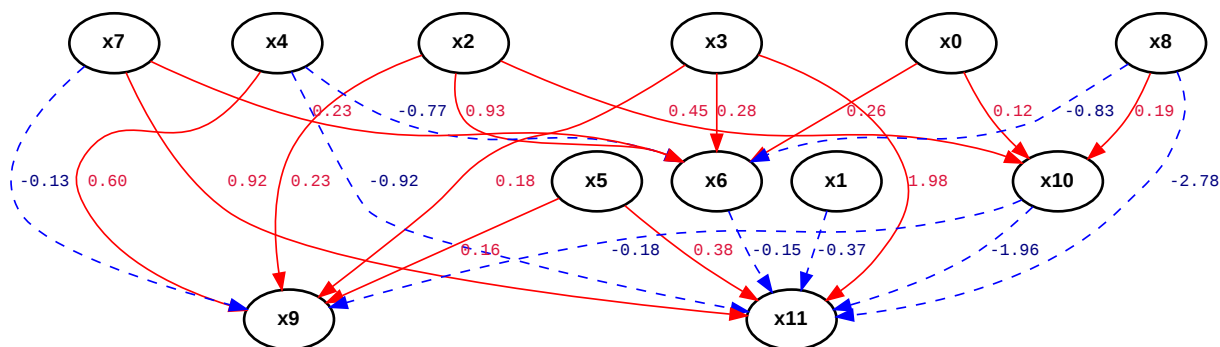


図 1: 表 2 に基づいて構成されるデータセットのうち、分野を問わない「全体」の区分に対する DirectLiNGAM での計算結果。係数の絶対値が 0.1 以上のもののみを表示し、有向辺の色は係数の符号に対応し、赤の実線が正、青の破線が負。x0~x11 のラベルと変数の対応については表 2 を参照。

表 3: 各分野での計算結果のうち、x0~x5, x7~x11 から x6 への影響。係数の絶対値が 0.1 以上のもののみを表示。

変数	全体	人文・社会科学	理学	工学	農学	保健	その他
x0	+0.26		+0.55	-1.13			-0.43
x1				+0.44			-0.19
x2	+0.93	-0.21	+0.25	-1.83	+0.87		-1.09
x3				+6.39			+18.89
x4	-0.77	+0.35	-0.12	-1.70			+0.42
x5				+0.28	+0.20		+0.80
x7	+0.23		+0.12	+0.72			+0.44
x8	-0.83	-1.42					-8.42
x9		+1.22					-17.23
x10							-0.13
x11			+0.12		+0.29		+0.90

表 4: 各分野での計算結果のうち、x0~x5, x7~x11 から x6 への影響。係数の絶対値が 0.1 以上となるブートストラップ確率。

変数	全体	人文・社会科学	理学	工学	農学	保健	その他
x0	0.62	0.67	0.72	0.66	0.75	0.77	0.70
x1	0.56	0.59	0.38	0.66	0.29	0.62	0.72
x2	0.89	0.83	0.81	0.88	0.89	0.79	0.85
x3	0.65	0.73	0.65	0.71	0.73	0.61	0.68
x4	0.81	0.86	0.47	0.83	0.83	0.73	0.75
x5	0.58	0.48	0.45	0.47	0.56	0.55	0.71
x7	0.67	0.64	0.57	0.61	0.68	0.66	0.75
x8	0.88	0.79	0.70	0.85	0.81	0.67	0.82
x9	0.25	0.42	0.42	0.39	0.36	0.38	0.24
x10	0.41	0.44	0.29	0.26	0.31	0.45	0.37
x11	0.25	0.27	0.21	0.24	0.37	0.28	0.38

特徴的な点は以下の通りである。

- 同時刻の変数 (x7~x11) よりも 1 年前の変数 x0~x5 の方が影響を与えやすい (係数の絶対値が 0.1 以上)。
- 1 年前の変数の中でも、特に x2 (一人当たり研究開発費)・x4 (大学教員就職割合) は多くの分野で x6 に影響を与えやすいが、その影響の正負は分野によっても異なる可能性がある。
- DC1 については、x1 (DC1) よりも x7 (DC1_L1) の方が x6 に影響を与えやすい。また、全体・理学・工学においていずれも x7 から正の影響を受けている。
- 保健分野については、x6 に係数の絶対値 0.1 以上の影響を与えるような変数は見受けられない。
- 一方「その他」の分野については、全変数からの x6 への影響が表れているものの、極端に大きい係数が吐き出されているものもあるなど、挙動が安定していない。

■ブートストラップ確率に関する比較 ここまで、データセット全体を用いた各分野での時系列因果探索の結果を示した。しかしながら x6 (博士課程進学率_L1) への他の変数からの影響の有無について、統計的信頼性の評価も重要である。そこで、DirectLiNGAM とブートストラップ法を組み合わせた手法 [Komatsu10] により、各変数が直接、係数の絶対値が 0.1 以上となるレベルで x6 に影響を与えるブートストラップ確率を評価^{*3}し、表 4 に示した。

表 3 とも見比べると、特徴的な点は、以下の通りである。

- 表 3 で比較的どの分野でも見られた x2 → x6 のパスについては、ブートストラップ確率もおおむね 0.8 を超えており、x4 → x6 についても理学を除いて 0.7 を超えている。
- x6 と同年度の変数のうち、x9 (研究時間割合_L1)、x10

^{*3} サンプルング回数は 1,000 回

(大学教員就職割合_L1) については、どの分野でもブートストラップ確率が 0.5 を下回っているのに対し、これらの 1 年前に対応する x_3 (研究時間割合), x_4 (大学教員就職割合) については、どの分野でもブートストラップ確率が 0.6 を超えている。

- x_5 (ポストドク就職割合), x_{11} (ポストドク就職割合_L1) については、多くの分野でブートストラップ確率がそれぞれ 0.6, 0.3 を下回るなど、他の変数に比べて x_6 に影響する確率が低いことが示唆された。

■ $x_2 \rightarrow x_6$ に関する分野別ヒストグラムの比較 表 3 における $x_2 \rightarrow x_6$ の効果の表れ方、特に正負の影響の違いをより詳しく見るため、ブートストラップ法で 1,000 回計算した結果のうち、図 2 には、 $x_2 \rightarrow x_6$ の係数の計算結果をヒストグラム表示した。

分野によるヒストグラムの主な特徴は以下の通りである。

- 表 3 の通り、データセット全体での計算でも正の値が出力されていた全体・理学・農学の分野において、ブートストラップでの計算結果でもほとんどが正の値をとるものとなっている。
- 人文・社会科学については、データセット全体では負の値が出力されており、図 2 でもかなり幅広・まばらではあるものの、0 または負の値を返す場合が多い。
- 一方で工学については、データセット全体では負の値 (-1.8) が出力されているが、ブートストラップでの計算結果では、1.0 付近および -2.0 付近に大・小の山が観察できる。データセット全体での計算結果はこの小さい方の山に対応するものと考えられる。

4 考察

4.1 考察～定性的な因果推論

上述の計算結果に基づいて、各種政策要素が博士課程進学率に与える影響について考察する。

■ $\tau = 0$ と $\tau = 1$ での効果の比較 先述の通り、表 3・4 の結果から、博士課程進学率については全体的に同じ年度 ($\tau = 0$) の変数からの影響よりも、1 年前 ($\tau = 1$) の変数からの影響の方が出やすいことが示唆される。博士課程進学率の向上に向けて、これらの各種変数が政策的に操作される場合には即時的な効果が期待されるわけではなく、少なくとも 1 年のタイムラグを見込むべきである、ということが改めてわかる。

■ $x_2 \cdot x_4$ からの影響 x_2 については、多くの場合 x_6 に正の影響を与えやすいという結果が示されている。これは定性的には、一人当たりの研究開発費が増えることで研究成果を上げやすくなる、もしくは研究費からの経済的支援が見込まれることで博士課程進学へのインセンティブになる、というストーリーが考えられる。一方で、 x_4 については、 x_6 に負の影響を与えるという結果が複数見受けられる。これは、大学教員のポストが埋まっていくことで、将来的にアカデミアでの若手ポストにつまらなくなると感じ、博士課程進学へのインセンティブになっている、というストーリーが考えられる。

■人文・社会科学分野と保健分野の特殊性 表 3 の通り、人文・社会科学分野については、 $x_2 \cdot x_4$ からの x_6 への影響が、他の分野と符号が異なる。また保健分野では x_6 への他の変数からの影響が一切出ないという結果になっている。これらの結果には、この 2 分野の特殊性が反映されている可能性がある。具体的には、人文・社会科学分野では、研究開発に直接費やす金額が小さい傾向にあるため、研究開発費用の増額がインセンティブになりにくい可能性がある。また、保健分野では、医師が症例について学位論文を執筆することも多く、他分野と博士課程進学に関する考え方が異なる可能性が影響していると考えられる。

4.2 本研究に関する課題

ここまで DirectLiNGAM を用いた分野別の計算の結果について、その統計的信頼性とこれらに基づく因果推論について述べたが、以下の点が課題となる。

- 変数の数 12 に対して、データ点数が 14 と変数の数と同程度となっている点。この結果のみに基づいた因果関係の断定は困難。
- 研究時間割合による博士課程進学率への影響はほとんど見られず、先行研究 [高山 21b, 高山 22a, 高山 22b] と異なるという点に関する説明。一方、これらの研究では、基盤的経費との強い関係の可能性とともに論じられているが、本研究においては基盤的経費の代わりに研究開発費を変数としており、その影響を議論する必要もある。
- $\tau \geq 2$ での影響に関する議論への拡大。本研究で影響が比較的考えにくいとされた $x_9 \sim x_{11}$ を外すなど、変数を絞り込みながら $\tau \geq 2$ の影響を調べるのが期待される。これは、博士課程進学率への影響にどれだけのタイムラグがありうるかを見込んだ上で、各変数を政策的に操作していくうえでも重要で

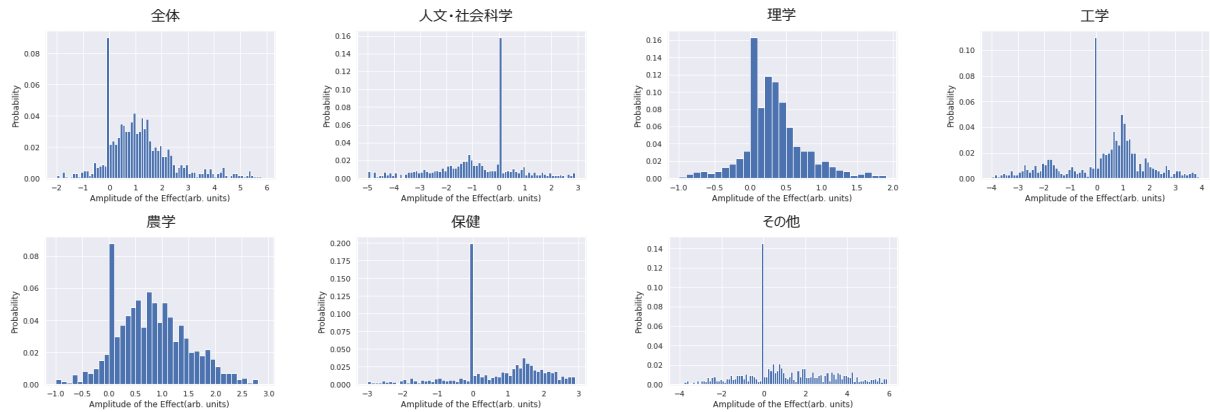


図 2: サンプル回数 1,000 回でブートストラップ法で計算したときの, $x_2 \rightarrow x_6$ の係数の値に関するヒストグラム。横軸はグラフの形状がわかる範囲で, 結果に応じてスケールしているが, bin は 0.1 で統一している。また, 0~0.1 におけるパルス的な振る舞いは, DirectLiNGAM の計算プロセスにおける枝刈りで, データ点のサンプリングによっては因果係数 0 と出力されることがあるためである。

ある。

- 自己回帰の存在の妥当性に関する考察。表 3・4 でも見たように, 博士課程進学率に関する自己回帰現象 $x_0 \rightarrow x_6$ が見られることがある。時系列データを扱う以上, artifact として自己回帰が観察されることも多いが, 博士課程進学率については, 前年度に先輩が博士課程に進んだことがモチベーションとなり, ある人の博士課程進学に繋がる, というストーリーもあり得る。

5 総括

本稿では, 研究力強化・若手研究者支援に関する EBPM に向けて, LiNGAM を用いて博士課程進学率に関する因果関係について, 専攻分野別に, 遅延も考慮して試行的な因果探索を行った。

今後は, 本研究に基づき, 統計的信頼性の課題の克服や, 厳密な因果関係の解明の拡大等が期待される。

参考文献

- [Hyvarinen10] A. Hyvärinen, K. Zhang, S. Shimizu, and P. O. Hoyer, Estimation of a structural vector autoregression model using non-gaussianity, *Journal of Machine Learning Research*, 11: 1709–1731, 2010. <https://www.jmlr.org/papers/volume11/hyvarinen10a/hyvarinen10a.pdf>
- [Hyvarinen13] A. Hyvärinen and S. M. Smith. Pairwise likelihood ratios for estimation of non-Gaussian structural equation models. *Journal of Machine Learning Research*, 14:111–152, 2013. <https://jmlr.org/papers/v14/hyvarinen13a.html>
- [Komatsu10] Yusuke Komatsu, Shohei Shimizu, and Hidetoshi Shimodaira: Assessing statistical reliability of LiNGAM via multiscale bootstrap. *In Proc. International Conference on Artificial Neural*

Networks (ICANN2010), Thessaloniki, Greece, pp.309–314, 2010. https://doi.org/10.1007/978-3-642-15825-4_40

- [Shimizu06] Shohei Shimizu, Patrik O. Hoyer, Aapo Hyvärinen, and Antti Kerminen: A linear non-gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, 7:2003-2030, 2006. <https://www.cs.helsinki.fi/group/neuroinf/lingam/JMLR06.pdf>

- [Shimizu11] S. Shimizu, T. Inazumi, Y. Sogawa, A. Hyvärinen, Y. Kawahara, T. Washio, P. O. Hoyer and K. Bollen.: DirectLiNGAM: A direct method for learning a linear non-Gaussian structural equation model. *Journal of Machine Learning Research*, 12(Apr): 1225–1248, 2011. <https://dl.acm.org/doi/10.5555/1953048.2021040>

- [加藤 09] 加藤真紀, 角田英之: 日本の理工系修士学生の進路決定に関する意識調査. 文部科学省 科学技術政策研究所 調査資料 (Research Material), No.165, 2009. <http://hdl.handle.net/11035/895>

- [治部 21] 治部眞里, 星野利彦: 修士課程 (6 年制学科を含む) 在籍者を起点とした追跡調査 (2020 年度修了 (卒業) 者及び修了 (卒業) 予定者に関する報告). 文部科学省 科学技術・学術政策研究所 調査資料 (Research Material), No.310, 2021. <https://doi.org/10.15108/rm310>

- [清水 17] 清水昌平.: 統計的因果探索. 講談社 機械学習プロフェッショナルシリーズ, 2017.

- [高山 21a] 高山正行, 小柴等, 前田高志ニコラス, 三内顕義, 清水昌平, 星野利彦: EBPM と統計的因果探索・数理モデルの利活用. 研究イノベーション学会 第 36 回年次学術大会 (予稿集), 公演番号 2G02, 2021.

- [高山 21b] 高山正行, 小柴等, 前田高志ニコラス, 三内顕義, 清水昌平, 星野利彦: 統計的因果探索アルゴリズム “LiNGAM” を用いた若手研究者支援政策に関する研究. 研究イノベーション学会 第 36 回年次学術大会 (予稿集), 公演番号 2G03, 2021.

- [高山 22a] 高山正行: 科学技術・イノベーション政策研究への統計的因果探索の利活用の新たな可能性 - 博士課程進学率に関する政策論を例として -. 科学技術・学術政策研究所 第 14 回政策研究レビューセミナー, 2022. <https://www.youtube.com/watch?v=GMSWn2abRXw>

- [高山 22b] 高山正行, 小柴等, 前田高志ニコラス, 三内顕義, 清水昌平, 星野利彦: 博士課程進学率に関する因果モデルの構築: 統計的因果探索アルゴリズム “LiNGAM” による試行的分析. *Jxiv*, 2022. (preprint) <https://doi.org/10.51094/jxiv.1>