

Title	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
Author(s)	高山, 正行; 小柴, 等; 前田, 高志; 三内, 顕義; 清水, 昌平; 星野, 利彦
Citation	年次学術大会講演要旨集, 36: 758-763
Issue Date	2021-10-30
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/17798
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”を用いた 若手研究者支援政策に関する研究

○高山 正行 (NISTEP), 小柴 等 (NISTEP), 前田 高志 ニコラス (NISTEP, 理研 AIP, 東京大学),
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)

1 はじめに

近年の我が国の科学技術・イノベーション政策の大きな話題の一つである研究力については、様々な指標に基づき、我が国の現状と課題解決が議論されており、その中核的な課題として若手研究者支援が位置付けられている。例えば 2020 年 1 月には「研究力強化・若手研究者支援総合パッケージ」が示されており、2021 年度からの第 6 期科学技術・イノベーション基本計画にも、若手研究者支援に関する目標設定がなされている。

特に、「研究力強化・若手研究者支援総合パッケージ」では、若手研究者を中心に修士課程（または博士前期課程）学生から中堅・シニア研究者までの幅広い・切れ目のない支援と、それによる研究力強化を狙いとしており、そのために 2025 年を年限とした様々な測定指標と目標が掲げられている。

一方これらの目標には現状達成が容易とは言い難いものもある。例えば博士（後期）課程^{*1}進学率については V 字回復が目標とされているが、これまで年々減少傾向が続いている。また、大学本務教員の 40 歳未満割合を 2025 年度までに 3 割以上という目標についても、これまで単調減少となっており、また確率遷移モデルによる推計・シミュレーション [高山 21a] でも、達成にあたっては抜本的な政策改革が必要であるとしている。これらの目標達成に向けては、改めて各種政策要素の間の因果関係について定量的な理解が重要である。

そこで本研究では、若手研究者支援政策における各種政策要素の間の因果関係の究明を目標とし、その第一歩としてまずはその指標の一つでもある我が国の博士課程進学率について、統計的因果探索の手法 LiNGAM (Linear Non-Gaussian Acyclic Model) を適用し、統計的アプローチに基づいた因果グラフの推定を試みた。本稿

では、その計算方法・結果について議論する。

なお本稿は、同じく本年次大会で講演する「2G02 : EBPM と統計的因果探索・数理モデルの利活用」にて提唱するアプローチの一例を示すものであり、そちらの予稿 [高山 21b] も併せて参照されたい。

2 博士課程進学率に関する LiNGAM での分析

まず、分析手法である LiNGAM について紹介する。詳細は教科書 [清水 17] に譲るが、LiNGAM [Shimizu06] は統計的因果探索の手法の一つであり、線形性、因果グラフの非巡回性、誤差変数の非ガウス性を仮定し、連続値をとる変数を対象に、因果グラフを一意に識別可能にする手法である。また LiNGAM には目的や手法に応じていくつかのバリエーションが存在するが、本研究では Python の LiNGAM パッケージ^{*2}のうち、DirectLiNGAM と呼ばれる推定法を用いた [Shimizu11, Hyvarinen13]。

2.1 データセット構成の考え方と計算条件

■変数の選定とデータセットの構成 博士課程進学率に関連する政策指標・手段は様々考えられるが、本研究ではその出発点として科学技術・学術政策研究所 (NISTEP) における調査結果 [加藤 09, 治部 21] をもとに変数の選定を行った。2009 年の加藤らによる調査報告 [加藤 09] によれば、博士課程進学者・博士課程進学を真剣に検討したことがある就職者が進学を検討する際に重要と考えられる要件として、経済的支援、民間企業等での博士課程修了者の雇用の増加と処遇の改善、アカデミックポストへの就職機会の拡充や任期等の待遇改善、研究環境の充実等が上位に挙がっている。加藤らの調査から 10 年以上経ち、近年治部らによって行われた調査報告 [治部 21] でも定性的に類似した傾向の報告があることから、これらの要件の重要性自体は 10 年以上の間隔をおいても色褪せていないと推測される。定量化や統計の取得自体が難しいものもあるが、本研究ではまずこの

^{*1} 本稿では、単純に「博士課程」という用語で統一することとする。

^{*2} <https://github.com/cdt15/lingam> に公開されている。

中で統計調査で定量化がある程度なされている項目について、アンケートによる意識調査とは別途、統計情報からの定量的な因果関係の探索を行うこととした。変数は(A)経済的支援、(B)キャリアパス、(C)研究環境の3つの観点から選んだ。

(A) 経済的支援に関わる変数

国による経済的支援としては、博士課程進学前に受給が決まるものとして日本学術振興会の特別研究員 DC1 がまず挙げられ、採択されると月額 20 万円の給与に加え一定額の研究費が支給される。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、これらはプログラムによって支援の仕方も異なる。このように、経済的支援の性質も様々であることから本来は別々に考慮すべきだが、データ点数の限界を踏まえ変数を絞る必要があることから、本研究では以下の通り、

- 博士進学の前年度の DC1 採択者数
- グローバル COE・リーディング大学院・卓越大学院の年度ごとの予算額合計

と大きくくり化した上で変数として採用した。

(B) キャリアパスに関わる変数

キャリアパス全体では高山らのシミュレーション [高山 21a] のように、アカデミアの中でもフェーズに応じたポストの変化、および民間企業-アカデミア間の転職等の各人材流動を考慮する必要がある。しかしここでは、議論の簡略化と変数を絞ることを優先し、また修士課程学生にとって博士課程進学を検討にあたっては、博士課程修了後の最初のキャリアパスが主要な検討材料になるものと仮定し、学校基本調査の結果に基づいて、

- 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

を計算し、変数として採用した。ただし、学校基本調査においてポストドク就職者数は 2011 年度以前については公開されていないため、2012 年度以降の平均値で埋めて処理した。

(C) 研究環境に関する変数

研究環境については、実験室における設備の共用等も含めて様々な観点があり、定量化が困難であるが、

表 1: 本研究で構成したデータセットの変数とそれぞれに課する事前知識の一覧

変数 (単位)	変数名	事前知識
博士課程進学率	x_0	
前年度 DC1 採択者数 (人)	x_1	外生変数
国全体の基盤的経費 (億円)	x_2	外生変数
大学研究本務者数 (人)	x_3	
一人当たりの基盤的経費 (億円)	x_4	$x_2 \div x_3$
研究時間割合	x_5	
博士修了直後の大学教員就職率	x_6	
博士修了直後のポストドク就職率	x_7	
DC1 以外の経済的支援 (億円)	x_8	外生変数

NISTEP が定点調査として実施している項目^{*3}を参考に、

- 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。基盤的経費は文部科学省にて多用されている議論に基づき、(国立大学法人運営費交付金等予算額 + 私立大学等経常費補助金)として計算し、この国全体での合計額と、総務省の科学技術研究調査に基づく大学の研究本務者数も変数に加え、一人当たりの値として算出した。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。

以上を踏まえ、表 1 には、本研究で構成したデータセットの変数をまとめた。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 本研究で用いている DirectLiNGAM のアルゴリズムでは、定義式や各学問領域の知見などに基づいた因果関係における事前知識 (prior

^{*3} <https://doi.org/10.15108/nr189>

knowledge) を前提とした計算も可能となっている。本研究においてもこの手法を採用し、各変数に関する事前知識も、表 1 に併せて示した。 x_1 (前年度 DC1 採択者数)・ x_2 (国全体の基盤的経費)・ x_8 (DC1 以外の経済的支援) については、政府の予算額として定まるものであることから、(決定までのプロセスは別途存在するものの) 政府の意思決定により定まる外生変数だとし、他の変数からは何も影響を受けないものとした。また、 x_4 (一人当たりの基盤的経費) については x_2 (国全体の基盤的経費) と x_3 (大学研究本務者数) に及ぼす影響はないものし、その他の変数からの影響がないものとした。

なお通常の LiNGAM では、通常の (和の) 構造的因果モデルを仮定しているため、変数間の関係は線形で表現されるが、今回は $x_2 \sim x_4$ のように積の関係が定義として入っていることもあり、既に線形結合で表現できないモデルとなっている。そこで本解析では、以下のような積の構造的因果モデルを導入し、LiNGAM を適用した。

$$x_i = \prod_{j \neq i} x_j^{b_{ij}} e^{e_i} \quad (1)$$

多変量解析においても、変数間の弾力性 (変化率の比が一定) が期待される場合に積のモデルが導入され、変数を指数で評価することがしばしばなされるが、式 (1) もこの考え方に準じている。この辺々について (自然) 対数をとることで、

$$\log x_i = \sum_{j \neq i} b_{ij} \log x_j + e_i \quad (2)$$

となり、そのまま LiNGAM を適用できる。なお、式 (2) を用いる場合は、以下の点に注意が必要である。

- 実数値としての分析のため、全ての変数が常に正の値をとることが条件となる。
- 式 (1)・(2) に見られるように、通常の LiNGAM とは異なり、誤差変数は変数の対数値について考え、非ガウス性を仮定する。
- 和の構造的因果モデルと同様、非巡回性と外生変数の独立性 (未観測共通要因が存在しないか、あっても影響は小さい) ことを仮定した分析となる。
- 通常の線形の範囲でのアプローチでは各変数の寄与 (係数) を対等な条件で比較するため、データセットの各変数について平均値を引いて標準偏差で除する「標準化」を施すが、この手法で現れる寄与は式 (1) の通り指数であり、対数をとってから平均値を引く

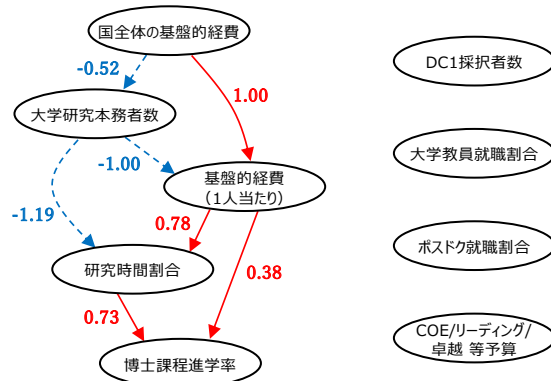


図 1: 表 1 に基づいて構成されるデータセット全体に対する DirectLiNGAM での計算結果。有向辺の色は係数の符号に対応し、赤の実線が正、青の破線が負。

操作*4を行えば、指数の比較という観点では十分である。

2.2 計算結果

■データセット全体に対する LiNGAM での計算結果 表 1 に基づいて構成したデータセット全体に対する 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徴的な点は、以下の通りである。

- 「国全体の基盤的経費 → 一人当たり基盤的経費」と「大学研究本務者数 → 一人当たり基盤的経費」の係数がそれぞれ 1.00 と -1.00 になっているのは、 $x_4 = x_2 \div x_3$ の辺々対数をとった結果と対応する。
- 博士課程進学率に直接寄与しているのは、「一人当たりの基盤的経費」と「研究時間割合」となった。
- 経済的支援に関する変数、キャリアパスに関する変数は、いずれもこの統計データセット上では博士課程進学率と因果関係があるとは言えなかった*5。
- 大学研究本務者数および一人当たり基盤的経費から研究時間割合への影響が示唆される。基盤的経費や研究時間割合は、若手研究者支援や研究力向上の重要課題としてこれまでばらばらに議論されてきたが、その一方でこれら二つの要素間の関係は、議論されてこなかった。それゆえこの結果は、若手研究者支援・研究力向上の政策的議論を深める上での新

*4 これは、対数をとる前に変数ごとにデータセットにおける相乗平均で除して無次元化してから対数をとるという操作に等しい。

*5 ただし、あくまで博士進学率の増減についてマクロスコープに見た因果関係を示唆するにすぎず、経済的支援やアカデミックポストの整備に意味がないとするものではない。調査報告果 [加藤 09, 治部 21] からも明らかな通り、現場のニーズとしては存在することを踏まえると、博士進学率の向上という目標達成に関わらず改善されることが望ましい。

表 2: ブートストラップ法によりサンプリング回数 1000 回で DirectLiNGAM を回した時の因果係数が非ゼロとなる確率 (上位 10 選)。

順位	直接的な因果関係	係数の符号	確率
1	$x_5 \rightarrow x_3$	-	55.6%
2	$x_2 \rightarrow x_4$	+	52.2%
3	$x_3 \rightarrow x_4$	-	47.5%
4	$x_5 \rightarrow x_0$	+	43.4%
5	$x_4 \rightarrow x_5$	+	41.9%
6	$x_3 \rightarrow x_5$	-	39.7%
7	$x_8 \rightarrow x_5$	+	21.7%
8	$x_4 \rightarrow x_0$	+	21.6%
9	$x_8 \rightarrow x_6$	+	19.1%
10	$x_2 \rightarrow x_3$	-	18.2%

しい知見の創出に繋がりを示唆である。

■ブートストラップ法による各因果に関する評価 データセット全体を用いて推定された因果グラフおよび各因果関係についての確からしさの評価の方法として、DirectLiNGAM とブートストラップ法を組み合わせた手法 [Komatsu10] がある。本研究においては、サンプリング回数 1000 回で試行したときに、直接的な因果関係が表れる割合であるブートストラップ確率上位 10 選について、表 2 のような結果^{*6}を得た。先述のデータセット全体に対する DirectLiNGAM の結果で現れた因果関係はいずれもこの上位 10 選に入っており、このデータセットの範囲では、データセット全体に対する DirectLiNGAM に表れた各因果関係の統計的信頼度はどれも 15% 以上であることがわかる。一方で例えば x_5 (研究時間割合) \rightarrow x_3 (大学研究本務者数) についてはデータセット全体での DirectLiNGAM の結果には表れなかったが、統計的信頼度としては 1 位であった。

図 2 には、代表的に x_4 (研究者一人あたりの基盤的経費) \rightarrow x_5 (研究時間割合) に関する係数の計算結果のうち非ゼロとなった場合についてヒストグラムとしてプロットした。ここでは 0.8 付近にひとつ大きいピークがあるだけでなく、1.3 付近にサブピークが生じているような構造となっている。この構造の意味するところは別途調査・考察を行う必要があるが、例えば必ずしも $x_4 \rightarrow x_5$ といった直接の因果関係以外にも、別の変数を介した間接的な

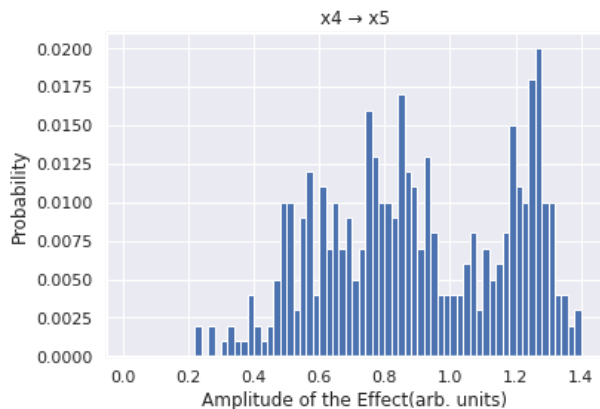


図 2: 表 2 のうち x_4 (研究者一人あたりの基盤的経費) \rightarrow x_5 (研究時間割合) について係数の計算結果のうち非ゼロとなった場合の結果についてヒストグラムで表示したもの。

因果関係が存在することで、こういったヒストグラムの構造となっている可能性がある。

2.3 考察～定性的な因果推論

上述の DirectLiNGAM の結果に基づいて、とり上げた各種政策要素間の因果関係を考察する。ここでは、政策研究の文脈から定性的な解釈の一例^{*7}として、 $x_4 \rightarrow x_5$ について簡単に考察する。

この $x_4 \rightarrow x_5$ が示す直接的な因果関係は、「研究本務者一人あたりの基盤的経費を増やす (減らす)」ことで「研究時間割合が増える (減る)」ということである。直観的には、例えば大学の研究者が使える基盤的経費を減らすと、その分競争的資金の確保が必要となり、その申請書の作成等に追われ、研究時間割合が減る、という説明が考えられる。しかし、この研究時間割合の引用元の FTE 調査では、競争的資金等の申請に係る文書等の作成時間は研究時間に含むと定義している。そのためこの説明は必ずしも成り立つとは言えず、また FTE 調査でも競争的資金等の申請に係る文書等の作成時間を詳らかにしたのは最新の 2018 年度調査のみであることから、この文書等の作成時間を除いた研究時間割合がどのように変化しているのか確認することができない。一方で、表 2 の通りこの直接的な因果関係の大きさを表す係数が非ゼロとなるブートストラップ確率は 41.9% と比較的高く、一定の統計的信頼度が出ていることや、図 2 の構造も鑑みると、引き続き何らかの交絡因子の存在も視野に入れつつ、説明可能なロジックを模索する価値はある。

^{*6}ただし、ブートストラップの再標本化はランダムであり、毎回同じ結果が得られるわけではないことには注意が必要である。

^{*7} その他の因果関係についての考察についても、別途発表予定。

2.4 本研究に関する課題

ここまで DirectLiNGAM を用いた計算の結果と新たな因果関係の存在可能性、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 9 つに対して、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。
- 研究時間割合や基盤的経費の現実の配分やこれらによる因果関係は、分野ごとに異なる可能性。
- 実はこの分析では、その年度中に影響が生じることを暗に仮定していたが、実際にはある要因の変化による影響は年単位で遅れて現れる可能性。

今後、博士課程進学率に関してより正確に因果関係を突き詰めていく上では、本研究の結果をもってそのまま因果関係を断定することはせず、あくまで示唆とし、例えば治部らによる調査 [治部 21] の個票データを用いたサンプルサイズの問題の克服や、分野に応じた因果関係の相違点の抽出、遅延効果込みでの因果探索等が期待される。

3 年齢別の議論に拡張した場合の LiNGAM の試行的応用

前節で述べた研究は、特に博士課程進学率のみに着目して変数を絞り、試行的に分析したものである。一方、研究力強化・若手研究者支援という大きな構想の実現に向けては、大学院生から PI 級研究者までの各フェーズについて、各種政策要因間の因果関係を正しく整理し、一体的な議論を構築する必要がある。そのためには例えば、

- 大学・研究機関の PI 級研究者、ポスドクといった様々な属性を考慮した議論
- 各フェーズを特徴づける年齢に応じた議論
- 各種競争的資金等を始めとする変数の追加

を加味した統計的因果探索が期待される場所である。

これらのアプローチに共通するのは、いずれも変数を増やす必要があることである。しかしながら現実には、統計情報を利用した分析においては、前述の通りデータ点数が限られてしまいやすい。(変数の数 > データ点数) となった場合、少なくとも独立成分分析による LiNGAM (ICA-LiNGAM) [Shimizu06] では計算できな

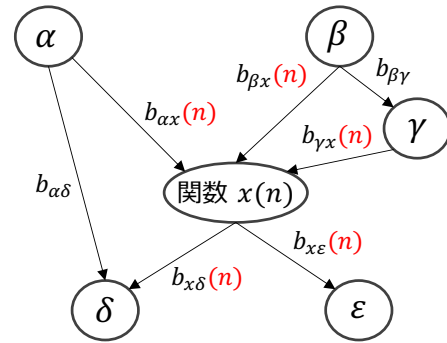


図 3: 関数を含めた因果探索で求めたい因果グラフのイメージ。

い。DirectLiNGAM は、そのアルゴリズムの性質上計算結果自体は返すことはあるものの、(変数の数 > データ点数) の関係が著しいほど結果的信頼性は低下する。

3.1 LiNGAM の拡張による非線形な"関数の"因果探索

上述の通り、例えば博士課程進学率について、もし年齢に応じた議論を行うとすると、厳密には 1 歳刻みの値を全て変数としてデータセットに組み込む必要がある。例えば 26 歳から 50 歳までのみを考慮するとしても、変数の数はこれだけの追加で 25 個増えてしまう。この場合、データの点数によっては DirectLiNGAM をかけても信頼性が著しく低下してしまう。

そもそもこの問題は、イメージとして図 3 に示す通り、因果探索に含めない変数である年齢 n に依存する博士進学率 $x(n)$ に対し因果探索を行い、変数 λ との間の直接因果効果を示す係数 $b_{x\lambda}(n)$ (あるいは $b_{\lambda x}(n)$) を n 依存性込みで求める問題に他ならない。そこで例えば、 $x(n)$ が 2 次関数で $x(n) = a_0 + a_1n + a_2n^2$ のように表される場合に $b_{x\lambda}(n)$ を見積もることを考える。 n が 1 から 30 までの全ての整数値をとるとき、これらの関数値全てに因果探索をかけようとする変数が 30 個増えることになる。しかし、以下の条件を仮定し、 $a_0 \sim a_2$ の 3 つのみを変数に加えて LiNGAM で分析することにより、信頼度を極端に低下させることなく、かつ概ね同等の n 依存性を示す $b_{x\lambda}(n)$ を、 $a_i (i = 0, 1, 2)$ に対する λ の直接因果効果を示す係数 $b_{a_i\lambda}$ によって式 (3) で見積もれる*8。

- a_0, a_1, a_2 の間に相互に直接的な因果関係がない
- 各 n に対し、 $x(n)$ 同士について相互に直接的な因果関係がない

$$b_{x\lambda}(n) = \sum_{i=0}^2 \frac{\partial x(n)}{\partial a_i} b_{a_i\lambda} \quad (3)$$

*8 この計算の数学的な詳細説明と妥当性検証のためにモデル計算した結果は別途発表予定。

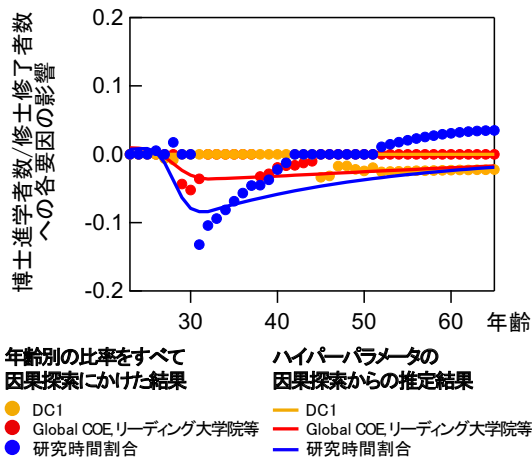


図 4: 関数に対する因果探索を実際のデータで行った結果。

このように、もし仮に因果探索の対象としない変数の非線形関数を含めて、因果探索を行う必要が生じた場合でも、関数形を特定し回帰することで特徴的なパラメータ（上の例では $a_0 \sim a_2$ ）を求め、それぞれを変数としてデータセットに加えて LiNGAM で分析することで、非線形な関数に関する因果推論の可能性が開かれる。特に、関数形がよくわかっていない場合には、別講演の予稿 [高山 21b] でも述べたように、

- 数理モデルを構築して回帰することにより特徴的なパラメータを抽出
- 関数の因果探索を実行し、関数とその周りの因果関係を調べるとともに、特徴量と各変数の因果関係から特徴量の性質を調査
- 領域知識と新たに抽出した特徴量の性質をもとに、数理モデルの再検討

というサイクルで、対象となる現象の数理的解明と統計的因果推論を両輪で行っていくことが望ましい。

最後に例として、図 4 ではこのサイクルに基づき、（社会人を含む⁹⁾ 博士課程進学者数と修士課程修了者数の比率について、年齢依存性込みで LiNGAM で試行的に解析した結果を示す¹⁰⁾。23～65 歳の年齢ごとの比率（全 43）を変数として取り入れて DirectLiNGAM で分析した結果（丸で示したデータ点）に対し、比率を適当な関数形を仮定してフィットし、少数の特徴的なパラメータ

⁹⁾ 学校基本調査では、社会人博士を除いた修士課程からの直接進学者に関する年齢別の人数が公開されていないため、試行的ではあるものの、社会人を含む形での解析とした。厳密な年齢別の博士課程進学者率を変数にするには、学校基本調査で公開されていないデータ利用が必要となる。

¹⁰⁾ この解析についても詳細は別途発表予定。

を抽出して上述の考え方と同様に変数に組み込み因果探索し、式 (3) と同様に解析的に年齢別の因果係数を求めた結果（実線）は、定性的な振る舞いが一致することがわかる。

4 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、統計的因果探索手法である LiNGAM を用い博士課程進学者率に関する因果関係の推定を行うとともに、その結果について簡単に考察した。また、若手研究者支援という幅広い議論に向けた解析の高度化の取組の一例を紹介した。

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、また、年齢依存性等を含めた議論の拡大等が期待される。

参考文献

- [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>
- [高山 21a] 高山正行, 星野利彦.: 博士人材の年齢別人材流動モデルと試行的な将来予測. *NISTEP Discussion Paper*, No.193, Feb 2021. <https://doi.org/10.15108/dp193>
- [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.: 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.
- [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.