SEMにおける因果同定の問題

○狩野 裕
大阪大学 大学院基礎工学研究科

1. 結論　本シンポジウムテーマは「因果は本当に証明できるのか?」である．もう少し正確に書くなら,「観察 (相関) データに基づいた統計的手法によって」という副詞句がつく[1]．そして，筆者にはSEM(Structural Equation Modeling) に関わる議論をせよということである．

このテーマに関する筆者の見解はシンプルである:

> **(現在のところ) 因果を証明することはできない．しかし，SEM による因果分析は継続すべきである**

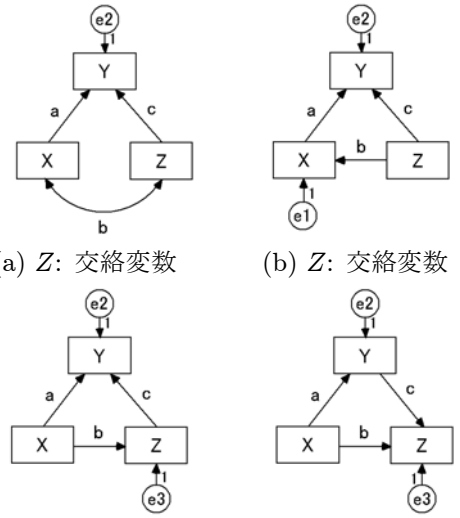
上記の見解は今更強調すべきことではなく，衆目の一致するところと言ってよいと思う．実際，少なくとも社会科学においてはSEMによる因果分析はスタンダードな手法になっているし，また，論文の結論には「 ... という関係が証明された」ではなく,「 ... という関係が示唆された」と表現されている．多くの研究者は適切にSEMを適用していると言えよう．

本発表では，継続がなぜ重要かということについて二つの観点から論じてみたい．

2. 反例の増強　一つ目は，なぜ「(現在のところ) 因果は証明できない」のか，ということに関係する．観察研究では，個体間の変動 (集団における変動) を個体内の変化に直接リンクさせることが相関と因果を結びつけることになる．しかし，個体間の変動が個体内の変化に直接対応しないことがあり，その代表例が第三変数の存在である．

原因変数と結果変数と目されるXとYとの因果関係の立証において，両者と関係するXとY以外の変数を第三変数と言う．ここではそれをZで表す．Zの存在に気づかず，XとYだけで因果分析を行うと誤った結論を導くことがある．第三変数を完全に特定することが困難であるという事実が，観察研究では因果を証明できないという主張の根拠になっている．

図1に示すように，第三変数ZのXとYへの関わり方にはいくつかのパターンがある．因果立証に立ちはだかるZの役割に「交絡変数」と「サンプルセレクション」があり，それぞれ，(a)(b) と (d) に対応する．



(a) Z: 交絡変数　(b) Z: 交絡変数

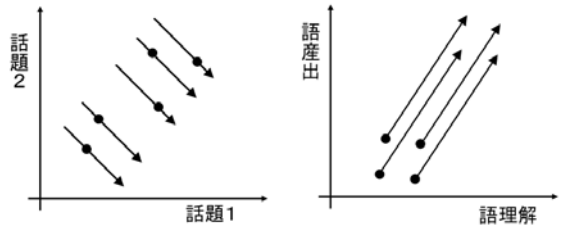
(c) Z: 中間変数　(d) Z: 合流点

図 1: 第三変数 Z の役割と重回帰モデル

ここで議論しておきたいことは，個体間の変動が個体内の変化に直接対応しない状況は交絡変数とサンプルセレクションだけかということである．

南風原-小松 (1999) は発達心理学の分野において興味ある二つの例を検討している (図 2)．第一は，２種類の話題についての親子間の会話量X,Yである．発達的には，Xが増えるにつれYが減少することが自然であるが (個体内変化)，会話の総量の個体差のために，Xが多い親子ほどYも多い (個体間変動) という正の相関が観測される (例１)．第二は，ダウン症児の語理解Xと語産出Yの関係である．個体を固定して時間に沿って追いかけるとXとYの間には正相関がみられるが，時点を固定して個体間の相関を計算すると０に近くなるという現象が報告されている (例２)．例１の方は，親子間の会話総量という第三変数を導入し図3のモデルにより適切に分析することができる．親子間の会話総量で調整することで，会話１と会話２との間の負の相

[1]準実験とよばれることもある (繁桝 2002).



例１ 親子間の会話　　例２ 語理解と語産出

図 2: 個体間変動と個体内変化 (南風原-小松 (1999) を単純化したもの)

関を同定することができる．図 3 では誤差共分散を導入しているが，会話１と会話２との間に矢印を引くことも可能である．ただ，どちらを原因変数とするかが難しい．

例２の状況はより複雑である．時刻を固定した状態，特に初期状態でのばらつきの原因について適切な情報が必要である．経時データが利用可能であれば，切片項をランダム変数とした潜在曲線モデル（混合モデル）での分析が有効である．個体によって傾きが変化するならば傾きもランダム効果とすることができる．しかし一時点の横断データと適当な第三変数のみから語理解と語産出間の正相関が同定できるかどうかは不明である．

交絡変数とサンプルセレクションは相関研究から因果関係を同定することができない反例である．しかし，これら以外にも反例はありえるだろう．(観察データの) 因果分析を継続することで，相関と因果のギャップを埋めていく必要がある．また，未観測の交絡変数については，そのような変数の存在を警鐘する統計量が開発できれば理想的ではある．

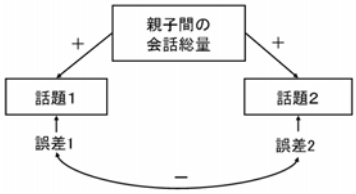


図 3: 例１と交絡変数

3. 因果分析は積み重ねる　Mulaik-James (1995) は自然科学 (物理学) と比して社会科学を次のように述べている．

> He argues that unlike in the physical sciences, there is in field X ”no reliable methodology in place for identifying the crucial variables or discovering the functional form of their relationships”, where X may be econometrics, sociology, psychology, or biology.

無作為割り付けができない状況においては因果分析は観察データに基づかざるを得ない．観察データにおける因果分析の定番が SEM である．SEM による因果分析が因果を立証するためにはさまざまな仮定が必要である．前節で見たように，交絡変数が存在しないこと因果関係を崩すようなサンプルセレクションが行われていないこと，また，母集団分布に関する諸仮定，因果の線形性 (線形モデルを用いるならば) など，細かいものも含めれば無数の仮定がある．そして，その多くは仮定の妥当性を吟味することが難しい．このような状況における科学的研究として，SEM を用いた因果分析に意味があるのだろうか．Mulaik-James (1995) は次のように述べている．

> The presumption of such criticisms is that ... we should be able to tell in any application of such models whether all these assumptions are satisfied and, if we cannot, we should not use the model. But that is not the way we reason with experience. We approach experience dialectically, that is, with imperfect knowledge about our assumptions, and gather evidence for or against our models gradually over time, while debating among ourselves as we go along as to whether we have or have not properly met our assumptions.

参考文献

[1] Mulaik and James (1995). Objectivity and reasoning in science and structural equation modeling. In *Structural Equation Modeling: Concept, Issue, and Applications*. Sage Publications.

[2] 南風原-小松 (1999). 発達研究の観点から見た統計 — 個の発達と集団統計量との関係を中心に — 「児童心理学の進歩」pp.214-238. 金子書房.

[3] 狩野裕 (2002).「構造方程式モデリング，因果推論，そして非正規性」 竹内啓 (編著) 多変量解析の展開 – 隠れた構造と因果を推理する – pp.65-129 (Part II). 岩波書店.

[4] 繁桝算男 (2002). 心理学における因果関係の分析．理論心理学研究.