Assessing Statistical Reliability of LiNGAM via Multiscale Bootstrap

長瀬　永遠

富山県立大学　情報基盤工学講座

April 5, 2023

はじめに

背景

定性分析等に基づいた事前知識なしでもデータ駆動的に因果関係が分析できる手法の一つとして LiNGAM がある．LiNGAM を用いることで因果関係の向きや有無が不明な複数の変数間に成り立つ因果グラフを一意に識別することができる．一方，LiNGAM を用いて完璧な因果グラフをもとめるためにはサンプリングデータが無限でないといけないという欠点がある．

目的

理論上完璧ではない LiNGAM による分析結果の信頼性を高めるための検証方法を紹介し，その手法を LiNGAM に適用する方法を提案する．

ブートストラップ法

母集団に対してサンプルデータが有限である場合に有効な検定．コンセプトはサンプルデータから復元抽出を行うことで疑似的なサンプルデータ（ブートストラップ）を量産し，それらに対する検定結果の平均をサンプルデータの検定結果とすること．

$$p^{BP} = \frac{1}{Q}\sum_{q=1}^{Q} f(X_q^*) \tag{1}$$

$X = (x_1, \cdots, x_n)$: サンプルデータ,
X_q^*: サンプルデータから復元抽出で作成したデータ,
$f(X)$: X について仮説が棄却された場合 0，それ以外の場合 1 となる関数,
Q: ブートストラップの数

マルチスケール・ブートストラップ法1

一般的なブートストラップ法ではブートストラップの数が有限個である（無限でない）ためバイアスや歪が生じる．そこで，ブートストラップのサイズを変えて様々なサイズのサンプルを作成することで漸近的に検定結果を補正することを考えたのがマルチスケール・ブートストラップ法である．

マルチスケール・ブートストラップ法の手順

1. 長さ $N' = r_k N$, $k = (1, \cdots, K)$ のブートストラップを作成する．ブートストラップ確立 $\tilde{p}(r_k)$ は以下のようになる．

$$\tilde{p}(r_k) = \frac{C(r_k)}{B_k} \tag{2}$$

B_k: ブートストラップの総数,
$C(r_k)$: ブートストラップのうち仮説が支持された回数

マルチスケール・ブートストラップ法2

2. $N' = r_k N$ のときのブートストラップ確立の理論値 $\pi(r; d, c)$ は以下の式で表されるため，手順 1 の結果をあてはめて回帰係数 d，c を推定する．

$$\pi(r; d, c) = 1 - \Phi(d\sqrt{r} + \frac{c}{\sqrt{r}}) \tag{3}$$

$\Phi(\cdot)$: 標準正規分布関数,
d: 符号付き距離,
c: 境界の曲率

3. 推定した d と c を用いて補正した確立値 $\tilde{p}$ を以下の式にて求める．

$$\tilde{p} = 1 - \Phi(d - c) \tag{4}$$

ブートストラップ法を用いたLiNGAMの検証1

帰無仮説

LiNGAM における二つの変数 x_i, x_j $(i \neq j)$ の間に以下の仮説を立てる．

1. H_{ij}^{+}: x_i が x_j に直接起因し，その接続強度が正である，すなわち $b_{ij} > 0$ であるという仮説
2. H_{ij}^{-}: x_i が x_j に直接起因し，その接続強度が負である，すなわち $b_{ij} < 0$ であるという仮説
3. H_{ji}^{+}: x_j が x_i に直接起因し，その接続強度が正である，すなわち $b_{ji} > 0$ であるという仮説
4. H_{ji}^{-}: x_j が x_i に直接起因し，その接続強度が負である，すなわち $b_{ji} < 0$ であるという仮説

ブートストラップ法を用いたLiNGAMの検証2

MB 法を用いた LiNGAM の検証の手順

LiNGAM における二つの変数 x_i, $x_j (i \neq j)$ の間に以下の仮説を立てる.

1. 長さ $N' = r_k N$ が整数となるように r_k を選択し, Q 個のブートストラップを作成する.
2. 各ブートストラップについてそれぞれ LiNGAM を実行し, $\tilde{p}(H_{ij}^{+})$ と $\tilde{p}(H_{ij}^{-})$ $(i \neq j)$ を求める.
3. 手順 2 の結果をもとに H_{ij}^{+} および H_{ij}^{-} $(i \neq j)$ が成り立つ確立を求める.

おわりに

まとめ

統計データのみからデータ駆動的に因果関係を求めることができるLiNGAM について，結果の信頼性を検証するためにマルチスケール・ブートストラップ法を用いた手法を提案した．

自分の研究について

中間発表の際に挙げていた課題のうち，因果関係の確からしさの検証に本論文の手法を用いることを考えている．