

情報幾何のピタゴラス関係に基づく指標を用いた因果推論

数理情報学専攻 48206231 細井 一成

指導教員 駒木 文保 教授

1 はじめに

統計的因果推論は、観測されたデータに基づき、変数間の因果関係を推論する問題である。経済学や、神経科学、医学などの様々な学問に応用されている分野であり、多くの種類の因果関係に対して、正しい推論結果を出すより一般的な手法が存在しないか、盛んに研究が行われている [2].

そのような手法のうち、Information-Geometric Causal Inference(IGCI) は情報幾何学的な観点に着目した手法 [4] であるが、変数間の非確率的な関係性を仮定していた。本研究ではこの情報幾何学的観点に基づいた上で、より一般的な変数に対し用いることができる因果推論の手法を提案し、数値実験を通じて IGCI との比較を行った。

2 既存研究

因果推論の代表的なモデルの一つに、以下のような加算ノイズモデル (Addictive Noise Model, ANM) が存在する [3, 5].

$$Y = f(X) + E$$

ここで、 X が原因変数、 Y が結果変数、 E がノイズである。この加算ノイズモデルの $X = x$ で条件付けた Y の確率分布の性質を考えると、その確率分布 $P_{Y|x}$ のエントロピー $S_{Y|x}$ について、以下の等式が成り立つ。

$$S_{Y|x} = S_N \text{ (const.)}$$

この性質に着目し、[4] では、より一般的なモデルに対し適応されうる、以下のような因果関係 $X \rightarrow Y$ を定義している。

仮定 1. $X \rightarrow Y$ は以下と同値。

$h(x)$ が $X = x$ における $P_{Y|x}$ の状態を表している”構造関数”とする。この時、参照確率分布 U_X を用いて以下が成り立つ：

$$\int h(x)P_X(x)dx \approx \int h(x)U_X(x)dx.$$

$h(x)$ や U_X の定義により、この因果関係 $X \rightarrow Y$ は様々な性質をとりうるが、[4] では U_X に一様分布やガ

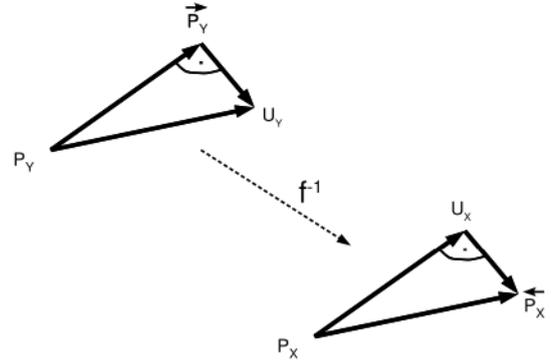


図 1. (1) と (2) の対応関係 ([4] からの引用.) 双方とも情報幾何的ピタゴラスの関係であり、 f^{-1} による変数変換で移り変わる関係性がある。

ウス分布族を用い、 $h(x)$ として以下のような KL ダイバージェンス D を用いた構造関数に注目した。ここで、 $\vec{P}_Y(y) := \int P(y|x)U_Y(x)dx$ とおく。

$$\begin{aligned} h_3(x) &:= \int \log \frac{\vec{P}_Y(y)}{U(y)} P(y|x) dy \\ &= D(P_{Y|x} \| U_Y) - D(P_{Y|x} \| \vec{P}_Y) \end{aligned}$$

このとき、以下の性質が成り立つ。

定理 1. (Janzing et al. (2012)) $\int h_3(x)P_X(x)dx = \int h_3(x)U_X(x)dx$ は、以下と同値：

$$D(P_Y \| U_Y) = D(P_Y \| \vec{P}_Y) + D(\vec{P}_Y \| U_Y).$$

さらに、単調増加関数 f に関して、 $Y = f(X)$ という非確率的な関係性を仮定すれば、以下のような情報幾何学的な関係性を導き出せる。ただし、 $\overleftarrow{P}_X(x) := \int P(x|y)U_X(y)dy$ とおく。

$$D(P_X \| \overleftarrow{P}_X) = D(P_X \| U_X) + D(U_X \| \overleftarrow{P}_X), \quad (1)$$

$$D(P_Y \| U_Y) = D(P_X \| U_X) + D(\vec{P}_Y \| U_Y), \quad (2)$$

$$D(P_X \| \overleftarrow{P}_X) = D(P_Y \| \vec{P}_Y) + D(\vec{P}_Y \| U_Y). \quad (3)$$

式 (3) に着目すると、 $X \rightarrow Y$ を仮定すると $D(P_X \| \overleftarrow{P}_X) > D(P_Y \| \vec{P}_Y)$ であり、同様に $Y \rightarrow X$ を仮定すると $D(P_X \| \overleftarrow{P}_X) < D(P_Y \| \vec{P}_Y)$ であるとわかる。この点に着目し、[4] では以下のような因果指標 Information-Geometric Causal Inference (IGCI) を提案している。

$$C_{X \rightarrow Y} := D(P_X \| U_X) - D(P_Y \| U_Y) \quad (4)$$

$C_{X \rightarrow Y} < 0$ ならば $X \rightarrow Y$, $C_{X \rightarrow Y} > 0$ ならば $Y \rightarrow X$ と推論する.

3 提案手法

IGCI には $Y = f(X)$ という仮定が存在した. さらに一般的な確率分布に対して用いることができる情報幾何学的因果指標を考えるため, 以下の点に着目した.

定理 2. $Y = f(X)$ である時, $\int h_3(x)P_X(x)dx = \int h_3(x)U_X(x)dx$ は以下と同値:

$$D(P_{X,Y} \| U_Y P_{X|Y}) = D(P_{X,Y} \| U_X P_{Y|X}) + D(U_X P_{Y|X} \| U_Y P_{X|Y}).$$

以下のような新しい構造関数 $h_4(x)$ を提案する.

$$h_4(x) = \int \log \frac{U_X P_{Y|X}}{U_Y P_{X|Y}} P_{Y|X}(y) dy$$

この h_4 には, 以下のような情報幾何学的関係性が備わっている.

定理 3. $\int h_4(x)P_X(x)dx = \int h_4(x)U_X(x)dx$ であることと, 以下が同値:

$$D(P_{X,Y} \| U_Y P_{X|Y}) = D(P_{X,Y} \| U_X P_{Y|X}) + D(U_X P_{Y|X} \| U_Y P_{X|Y})$$

h_4 による因果関係は, $Y = f(X)$ を仮定した, h_3 による因果関係の拡張となっている.

定理 3 に基づき, 以下のような因果指標 $C_{X \rightarrow Y}$ を提案する.

$$C_{X \rightarrow Y} := \log(|D(P_{X,Y} \| U_Y P_{X|Y}) - D(P_{X,Y} \| U_X P_{Y|X}) - D(U_X P_{Y|X} \| U_Y P_{X|Y})|) - \log(|D(P_{X,Y} \| U_X P_{Y|X}) - D(P_{X,Y} \| U_Y P_{X|Y}) - D(U_Y P_{X|Y} \| U_X P_{Y|X})|)$$

4 数値実験

提案手法の性質を確認するため, データセット X_1, \dots, X_{70} を乱数, および $X_j = X_i^2$ などの関係式を用いて生成し, X_i, X_j の間に $X_j = f(X_i)$ の生成式による因果関係 $X_i \rightarrow X_j$ があるかどうかを推定する課題を行った. IGCI, 提案指標の推定値 $\hat{C}_{X \rightarrow Y}, \hat{C}_{Y \rightarrow X}$ を用いて, 100 サンプルを用いた課題を, 100 回行い正解率 (Accuracy) を集計した. $\hat{C}_{X \rightarrow Y}, \hat{C}_{Y \rightarrow X} < -z$ ならば $X \rightarrow Y$, $\hat{C}_{X \rightarrow Y}, \hat{C}_{Y \rightarrow X} > z$ ならば $Y \rightarrow X$ と推測したとき, z を調整して $X \rightarrow Y$ または $Y \rightarrow X$ と因果を判定したデータの率 (Decision rate) を調整した時

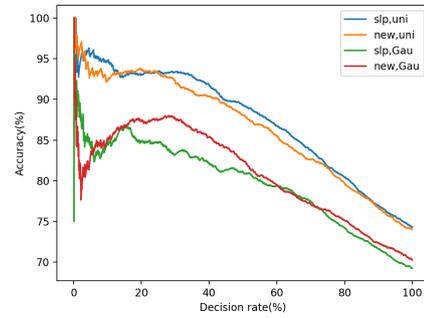


図 2. ノイズのない実験: 青, 緑が既存手法 [4] で橙, 赤が提案手法.

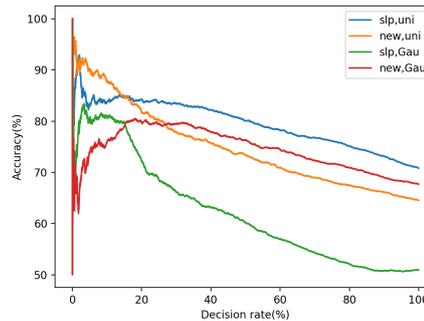


図 3. ノイズのある実験: 青, 緑が既存手法 [4] で橙, 赤が提案手法.

に, その正解率 (Accuracy) がどう変化するかを示した図が図 2 である. 青, 橙では参照確率分布 U_X, U_Y に一様分布, 赤, 緑にはガウス分布族を用いている. 提案手法と IGCI が同様の精度を持つことがわかる. 一方, 加算ノイズ E を追加し, $X_i = f(X_j) + E$ の関係性でデータを生成した図 3 では, IGCI では z が小さい時に Accuracy が低下した一方, 提案手法では比較的正確に因果推論ができていることがわかる.

参考文献

- [1] S. Amari: Information geometry on hierarchy of probability distributions. *IEEE Transactions on Information Theory*, Vol. 47, pp. 1701–1711, 2001.
- [2] A. Dawid: Beware of the DAG!. *Causality: objectives and assessment*, PMLR, pp. 59–86, 2010.
- [3] P. Hoyer, D. Janzing, J. Mooij, J. Peters, and B. Schölkopf: Nonlinear causal discovery with additive noise models. *NIPS*, Vol. 21, 2008.
- [4] D. Janzing, J. Mooij, K. Zhang, J. Lemeire, J. Zscheischler, P. Daniušis, B. Steudel, and B. Schölkopf: Information-geometric approach to inferring causal directions. *Artificial Intelligence*, Vol. 182, pp. 1–31, 2012.
- [5] J. Peters, D. Janzing, and B. Schölkopf: *Elements of causal inference: foundations and learning algorithms*, The MIT Press, 2017.