

# An Interpretable Kernel-based Conditional Independence Test with an Extension to Random Objects

RYUNOSUKE MIYAZAKI<sup>1</sup> and YOSHIMASA UEMATSU<sup>1</sup>

<sup>1</sup>*Department of Social Data Science, Hitotsubashi University*

## 1 背景

一般の距離空間上のデータに対しても適用可能なノンパラメトリックな条件付き独立性検定

$$H_0 : P_{XYZ} = P_{X|Z} \otimes P_{Y|Z} \otimes P_Z, \quad \text{vs.} \quad H_1 : P_{XYZ} \neq P_{X|Z} \otimes P_{Y|Z} \otimes P_Z$$

のための新たな手法を提案する。条件付き独立性は、統計学における多様な概念の基礎となる極めて重要な概念である。例えば、識別性、予測、データ選択といった多様な統計的問題 (Dawid 1979), 因果推論 (Pearl 2009)や変数選択 (Candès et al. 2018)などにも深く関わっている。

## 2 本研究の概要

我々は、カーネル法に基づく条件付き相互共分散作用素 (Fukumizu et al. 2007) と同値な作用素を新たに定義し、その推定量を構成することにより、新たな条件付き独立性検定手法を提案する。主な貢献は以下の通り。

1. 帰無仮説の分布の集合に関して一様にサイズを制御する一致性を持つ検定を提案。
2. カーネル選択に依存しない確率密度関数の距離として解釈可能な統計量を構成。
3. 収束レートに関する二重頑健性を示すことにより、バイアス項に関する緩やかな仮定の下で、漸近分布を導出。この理論はその他のカーネル法の手法にも適用可能。
4. ポーランド空間上の指数型カーネル関数が「特性的」であるための条件を導出。これにより、ワッサースタイン空間、球面、確率単体、実対称正定値行列など様々な空間に対して条件付き独立性検定だけでなく、二標本検定や独立性検定などが利用可能に。
5. シミュレーションにより、提案手法が既存の主要な手法と比較して、特に高次元データに対しても、検定のサイズ（第一種の過誤）を正確に制御しつつ、より高い検出力を示すことを確認。

## 参考文献

- Candès, E. J., Y. Fan, L. Janson, and J. Lv. 2018. “Panning for gold: ‘model-X’ knockoffs for high dimensional controlled variable selection.” *Journal of the Royal Statistical Society Series B* 80:551–577.
- Dawid, A. P. 1979. “Conditional Independence in Statistical Theory.” *Journal of the Royal Statistical Society. Series B (Methodological)* 41 (1): 1–31.
- Fukumizu, Kenji, Arthur Gretton, Xiaohai Sun, and Bernhard Schölkopf. 2007. “Kernel measures of conditional dependence.” In *Proceedings of the 20th International Conference on Neural Information Processing Systems*, 489–496.
- Pearl, Judea. 2009. “Causal Inference in Statistics: An Overview.” *Statistics Surveys* 3:96–146.