

ヒルベルト - シュミット独立基準とランダム行列理論によるノイズ変数の推定

理学専攻・情報科学コース 川久保秀子 (指導教員: 吉田裕亮)

1 はじめに

本研究では、ヒルベルト-シュミット独立基準とランダム行列理論とを組み合わせることで Pure Free Meixner 分布の台を推定することにより、標本データに含まれるノイズ変数の集合を推定する方法を提案する。提案手法により、回帰問題において意味のある変数の最小部分集合を抽出することが可能となる。

2 ヒルベルト-シュミット独立基準 (HSIC)

2つの n 次元確率ベクトル V_j ($j = 1, \dots, p$) と Y の関係を調べる時、共分散や相関を用いても線形な関係しか考慮することができないが、再生核ヒルベルト空間 (RKHS) への特徴写像 $\Phi(V_j), \Phi(Y)$ は高次の情報を含むため、その共分散を考えることによって、もとの確率変数の高次の独立性を調べることができる。

このアイデアに基づき、特性的な正定値カーネル (Gaussian カーネル) による相互共分散作用素を用いて確率変数の独立性や依存性を調べる方法が、ヒルベルト-シュミット独立基準 (HSIC) [1] である。

経験的 HSIC は以下により求めることができる。

$$HSIC_j = \frac{1}{(n-1)^2} \text{Tr} [\bar{K}_{V_j} \bar{K}_Y] \quad (j = 1, \dots, p)$$

ここで、 \bar{K}_{V_j}, \bar{K}_Y は中心化グラム行列を表す。

3 ランダム行列理論 (RMT)

$N \times P$ 長方形行列 A の成分がそれぞれ独立に同一分布に従うとき、 $C = (1/n)AA^T$ を相関行列とよぶ。相関行列は $N \times N$ 対称ランダム行列である。

相関行列の漸近固有値分布は $Q = P/N$ の比を一定に保ちながら、 $N \rightarrow \infty, P \rightarrow \infty$ とした極限で Marcenko-Pastur 分布 (MP 分布)

$$p(t) = \begin{cases} \frac{\sqrt{(\lambda_+ - t)(t - \lambda_-)}}{2\pi Qt} & (\lambda_- < t < \lambda_+) \\ 0 & (\text{otherwise}) \end{cases}$$

に従う。ただし、 $\lambda_{\pm} = (1 \pm \sqrt{Q})^2$ とする。

また、相関行列 C の各行ベクトル c_i ($i = 1, \dots, N$) を Gaussian カーネルで RKHS に写像すると、 $K_{i,j} = \exp\{-\|c_i - c_j\|^2 / 2\sigma^2\}$ となり、その固有値分布は非常に大きな値を持つ固有値 1 つを除けば、ほぼ MP 分布の affine 変換となる。

MP 分布はランダム性と密接な関わりを持つため、ランダム行列理論 (RMT) をデータに対して用いた時、データのノイズ部は MP 分布にほぼ一致する。一方、MP 分布から外れる部分はランダム性に関わりが無いので、特に $\lambda_+ < t$ の部分はデータの構造部と考えることができる。MP 分布の台を推定できればノイズ部と構造部の境界を推定することができ、構造部のみを抽出することが可能となる。

4 提案手法のアイデア

いま、 V_j ($j = 1, \dots, p$) が独立に同一分布に従い、全変数が Y と関係を持たないノイズ変数であるとする。すると、 \bar{K}_{V_j}, \bar{K}_Y の固有値分布はそれぞれ MP 分布に近い分布に従うことが推測され、 $\{\text{Tr}[\bar{K}_{V_j} \bar{K}_Y]\}_{j=1}^p$ の分布もまた同様に MP 分布に近い分布に従うことが推測される。

検証の結果、HSIC によるノイズ変数の評価値 $\{HSIC_j\}_{j=1}^p$ は、図 1 のように MP 分布を含む Free Meixner 分布族に属する Pure Free Meixner 分布 (PFM 分布) [2] に従うことが推測された。ここで、平均 0、分散 1 の Free Meixner 分布族の確率密度関数を以下に示す。

$$q(t) = \begin{cases} \frac{\sqrt{4(1+b) - (t-a)^2}}{2\pi(bt^2 + at + 1)} & (\gamma_- < t < \gamma_+) \\ 0 & (\text{otherwise}) \end{cases}$$

また、Free Meixner 分布族の台は

$$(\gamma_-, \gamma_+) = (a - 2\sqrt{1+b}, a + 2\sqrt{1+b})$$

で表される。なお、 $b > 0$ かつ $a^2 < 4b$ のとき PFM 分布の確率密度関数を表し、 $b = 0$ かつ $a \neq 0$ のとき、MP 分布の確率密度関数を表す。

提案手法では RMT に MP 分布ではなく PFM 分布を採用し、ノイズ変数の推定を行う。

5 提案手法のアルゴリズム

- (1) $S^\ddagger = \emptyset$ とする。
- (2) $X = [V_1, \dots, V_j, \dots, V_p]$ と Y を標準化し、HSIC によって V_j と Y の独立性を調べる。
- (3) 得られた HSIC の評価値 $\{HSIC_j\}_{j=1}^p$ を降順に並べ、 s_p, \dots, s_1 とする。
- (4) $S_d = \{s_{p-d}, \dots, s_1\}$ ($d = 0, \dots, p-1$) を考え、各 S_d をそれぞれ標準化して 4 次標本モーメントまでを求める。
- (5) $s_{p-d} = \gamma_+^{(d)}$ と仮定し、3 次標本モーメント $m_3^{(d)} = \tilde{a}^{(d)}$ を用いて $\tilde{b}^{(d)}$ を求める。
- (6) $\tilde{a}^{(d)}$ と $\tilde{b}^{(d)}$ を用いて 4 次モーメントの推定値 $\tilde{m}_4^{(d)} = (\tilde{a}^{(d)})^2 + \tilde{b}^{(d)} + 2$ を求める。
- (7) (条件 1) $\tilde{b}^{(d)} = \theta_1^{(d)} > 0$, (条件 2) $4\tilde{b}^{(d)} - (\tilde{a}^{(d)})^2 = \theta_2^{(d)} > 0$, (条件 3) $|m_4^{(d)} - \tilde{m}_4^{(d)}| = \theta_3^{(d)} < 1$, (条件 4) $m_4^{(d)}/m_4^{(d+1)} = \theta_4^{(d)} \approx 1$ とし、条件 1 と条件 2 を満たし、かつ、条件 3 を満たす $\min d$ と条件 4 を満たす $\min d$ が一致するとき、その d を d^\ddagger とする。推定により d^\ddagger が定まらないときは、Gaussian カーネルのパラメータの値を変えて (2) から (7) を再度試行する。

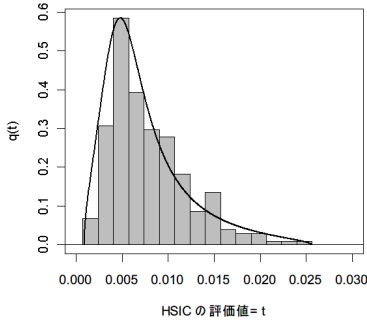


図 1: $\{HSIC_i\}_{i=1}^p$ の分布

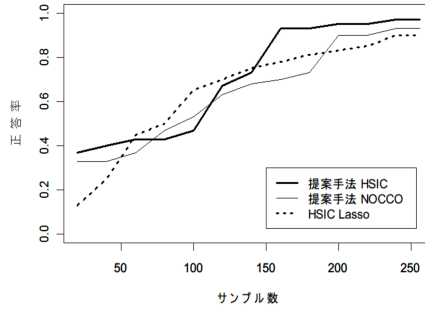


図 2: データ 1 の実験結果

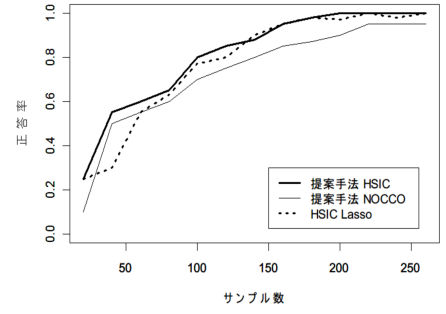


図 3: データ 2 の実験結果

- (8) $S^\ddagger \cup S_{dt}$ を新たに S^\ddagger とし, $S^\dagger = S_0 \setminus S_{dt}$ とおく .
- (9) $s_{max}^\dagger = \max \{s | s \in S^\dagger\}$ に対応する変数を Y' とおく . X の各変数と Y' の独立性を HSIC によって調べ, (3) から (7) を実行する . ここで $\dot{S} = S_0 \setminus S_{dt}$ とし, $S^\ddagger \cup (\dot{S} \cap S^\dagger \setminus \{s_{max}^\dagger\})$ を新たに S^\ddagger とする . また, $S^\dagger \setminus (\dot{S} \cap S^\dagger)$ を新たに S^\dagger とする .
- (10) $S^\dagger = \emptyset$ になるまで (3) から (9) (ただし (8) は飛ばす) を繰り返す .

以上により, S^\ddagger に対応する変数をノイズ変数, $S_0 \setminus S^\ddagger$ に対応する変数を Y と関係を持つ変数の最小部分集合として認識することができるようになる .

6 実験

独立な変数が同一分布に従うデータを用意して, 提案手法の有効性を検証する . 実験ではサンプル数を 20 ずつ増やしていき, 正答率を観察する . 目的変数と説明変数の独立性をより直接的に調べるためには, 正規化された相互共分散作用素である NOCCO を用いるのが適切であるが, NOCCO による評価値の分布もまた Pure Free Meixner 分布に従うことが推測されたため, 提案手法アルゴリズム (3) の HSIC による評価を NOCCO による評価に変更したのも実験に加えた . なお, 比較手法に HSICLasso を採用する .

データ 1 (加法モデル)

$$Y_i = v_{i1} \exp(-v_{i2}) + 2v_{i3} \log |3v_{i4}| + E,$$

$$Y = [y_1, \dots, y_i, \dots, y_n]^T, E \sim N(0, 1),$$

$$X = [X_1, \dots, X_i, \dots, X_n]^T = [V_1, \dots, V_j, \dots, V_p],$$

$$X_i = [v_{i1}, \dots, v_{ij}, \dots, v_{ip}], V_{i1} \sim N(0.5, 1.5) i.i.d.,$$

$$V_{i2} = \cos(V_{i1}), V_{i3} \sim N(0.3, 1.2) i.i.d.,$$

$$V_{i4} \sim N(0, 1) i.i.d., V_{i5} = 3V_{i4}, V_{i6} \sim N(0, 1) i.i.d.,$$

$$V_{ij} = 2V_{i6} + E \quad (j = 7, \dots, 10),$$

$$V_{ij} \sim N(0, 1) i.i.d. \quad (j = 11, \dots, 256).$$

データ 2 (非加法モデル)

$$Y_i = v_{i1} \exp\{2v_{i2}\} + v_{i3}^2 + E. \quad v_{i1} \sim N(0.5, 1.5) i.i.d.,$$

$$v_{i2} = \cos(v_{i1}), v_{i3} \sim N(0.3, 1.2) i.i.d.,$$

$$v_{i4} \sim N(0, 1) i.i.d., v_{i5} = v_{i4}^2, v_{i6} \sim N(0, 1) i.i.d.,$$

$$v_{ij} = 2v_{i6} + E \quad (j = 7, \dots, 10),$$

$$v_{ij} \sim N(0, 1) i.i.d. \quad (j = 11, \dots, 1000), E \sim N(0, 1)$$

データ 1 の実験結果は図 2, データ 2 の実験結果は図 3 のようになった . 図 2, 図 3 から, データ X の独立な変数が同一分布に従っていれば, 変数ごとに成分の単位が異なるケースや X に独立でない変数が含まれ

るケースにおいても, 提案手法が有効であることがわかる . また, 正答率の違いは回帰モデルによって多少あるものの, 提案手法と HSICLasso の精度は同程度であることがわかる . HSICLasso ではパラメータの設定や正答率の計算方法に正解を知っているという有利な前提を入れて実験を行なっているが, 提案手法は正解を知らない前提で精度を落とさず自動的に有用な変数の最小部分集合を推定することができるので, 付加価値を持つと考えることができる . また, 実験では正答率は NOCCO よりも HSIC の方が良いので, NOCCO の計算ではパラメータが増えることや逆行列を計算する必要のあることを考慮すると, 提案手法に用いる独立性の評価は HSIC で十分であることもわかった .

7 まとめ

本研究では, 回帰問題に対してヒルベルト - シュミット独立基準と Pure Free Meixner 分布に基づくランダム行列理論とを組み合わせることで, 目的変数と関係を持つ説明変数の最小部分集合を抽出する方法を提案した . 提案手法は HSIC のパラメータ最適化に関する 1 つの指標になっていると考えることもできる . 既存の変数選択法では選択する変数の数を人為的に決定してから変数選択を行うのが一般的であるが, 提案手法ではそのような人為的決定なしで, 既存手法の精度と同程度の精度を保ちつつ, 有用な変数の最小部分集合を自動的に抽出することができることが実験により確認された . また, 独立な変数が同一分布に従っていれば, 変数ごとに成分の単位が異なるケースや変数集合に独立でない変数が含まれるケースにおいても, 提案手法が有効であることが確認された .

今後の課題として, $\{HSIC_j\}_{j=1}^p$ の分布が PFM 分布に従うという予想に対して数学的証明を与えたいと考えている .

参考文献

- [1] Gretton, A. and Bousquet, O. and Smola, A. and Schölkopf, B. : *Measuring statistical dependence with Hilbert-Schmidt norms*, Algorithmic learning theory, pp.63-77, Springer (2005).
- [2] Anshelevich, M. : *Bochner-Pearson-type characterization of the free Meixner class*, Adv. in App. Math., vol.46 (2011), pp.24-45.