Vol.2012-MPS-91 No.39
Vol.2012-BIO-32 No.39
2012/12/7

ヒルベルト-シュミット独立基準に基づくノイズ変数の除去

川久保秀子[1,a)]　吉田裕亮[1,b)]

概要：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て自由 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，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．

キーワード：ヒルベルト-シュミット独立基準，ランダム行列理論，自由 Meixner 分布，変数選択．

Elimination of redundant features via Hilbert Schmidt Independence Criterion

Hideko Kawakubo[1,a)]　Hiroaki Yoshida[1,b)]

Abstract: Combining Hilbert Schmidt Independence Criterion and Random Matrix Theory, we propose a method of estimating a set of redundant features. The minimum subset of non-redundant features can be extracted by estimating the support of the pure free Meixner distribution.

Keywords: Hilbert Schmidt Independence Criterion, Random Matrix Theory, the pure free Meixner distribution，feature selection.

1. はじめに

意味のある変数の最小部分集合を求めることができれば, 複雑なデータ構造を簡略化して把握することが可能となる．また，高次元データにおける学習の高速化や，バイオインフォマティクスにおける遺伝子選択にも役立つ．

ヒルベルト-シュミット独立基準（HSIC）は 2 つの確率ベクトルの独立性を測るためのノンパラメトリックな手法であり, 実装が容易で, 回帰問題の変数選択に用いることができると考えられる．HSIC は，全ての説明変数同士が独立であるという仮定の下では有用で，独立性を高い精度で数値化して評価するが, HSIC のみを用いて意味のある変数の最小部分集合を選択することは困難である．そこで, 本研究では HSIC とランダム行列理論とを組み合わせて，標本データに含まれるノイズ変数を推定する方法を提案する．

[1] お茶の水女子大学 大学院人間文化創成科学研究科
Graduate School of Humanities and Sciences, Ochanomizu University
[a)] kawakubo.hideko@is.ocha.ac.jp
[b)] yoshida@is.ocha.ac.jp

2. ヒルベルト-シュミット独立基準（HSIC）

近年，平均や分散などの統計量を再生核ヒルベルト空間（RKHS）で考えることによって古典的な統計概念を扱えることが明らかとなり，それに基づいたノンパラメトリックな統計的推論手法が開発されている [1].

2 つの n 次元確率ベクトル $\boldsymbol{v}_i (i = 1, \cdots, p)$ と $\boldsymbol{Y}$ の関係を調べるとき，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, $\boldsymbol{v}_i$ と $\boldsymbol{Y}$ を RHKS に写像すると，それらの特徴写像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．このアイディアに基づき，特性的な正定値カーネル（ガウシアンカーネル）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）[2] である．

経験的 HSIC は以下により求めることができる．

$$HSIC_i = \frac{1}{(n-1)^2} Tr\left[\overline{\boldsymbol{K}}_{\boldsymbol{v}_i} \overline{\boldsymbol{K}}_{\boldsymbol{Y}}\right] \quad (i = 1, \ldots, p) \quad (1)$$

ここで，$\overline{\boldsymbol{K}}_{\boldsymbol{v}_i}, \overline{\boldsymbol{K}}_{\boldsymbol{Y}}$ は中心化グラム行列を表す．

Vol.2012-MPS-91 No.39
Vol.2012-BIO-32 No.39
2012/12/7

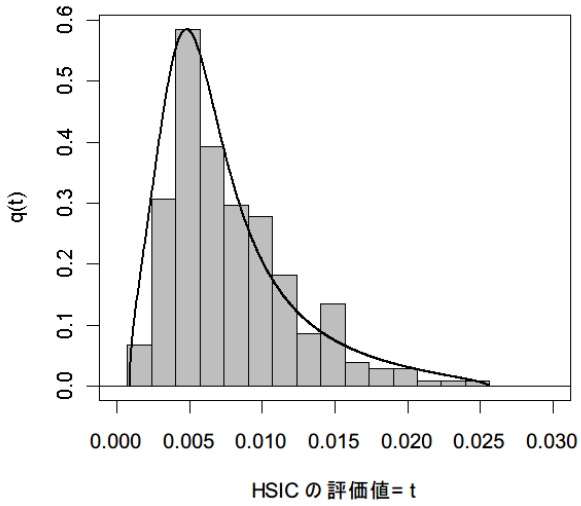

図 1 HSIC の評価値 $\{HSIC_i\}_{i=1}^{p}$ の分布

3. ランダム行列理論（RMT）

$N \times P$ 長方行列の成分がそれぞれ独立同一分布に従うとき，$\boldsymbol{C} = \boldsymbol{A}\boldsymbol{A}^T$ 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，$N \to \infty, P \to \infty$ とした極限で Marcenko-Pastur 分布（MP 分布）

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

に従う．ただし，$\lambda_\pm = \left(1 \pm \sqrt{Q}\right)^2$ とする．

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と関わりが無いので，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 分布の台を推定することができればノイズ部と構造部の境界を知ることができ，構造部のみを抽出することが可能となる [3].

4. 提案手法

相関行列をガウシアンカーネルによって RKHS に写像して得られた行列の固有値分布は，非常に大きな値を持つ固有値 1 つを除けば，ほぼ MP 分布のアフィン変換になることが知られている．

いま，$\boldsymbol{v}_i (i = 1, \ldots, p)$ と $\boldsymbol{Y}$ が独立同一分布に従っているとすると，$\boldsymbol{v}_i \boldsymbol{v}_i{}^T, \boldsymbol{Y}\boldsymbol{Y}^T$ は相関行列の一種であると考えられる．よって，それらをガウシアンカーネルで写像した $\overline{\boldsymbol{K}}_{\boldsymbol{v}_i}, \overline{\boldsymbol{K}}_{\boldsymbol{Y}}$ もそれぞれ MP 分布に近い分布に従うことが考えられ，相関行列の固有値分布と HSIC の評価値 $\{HSIC_i\}_{i=1}^{p}$ の分布の類似性が推測される．

また，RMT では MP 分布の確率変数として相関行列の固有値が対応しているのに対し，HSIC の評価値の分布では分布の確率変数として $\overline{\boldsymbol{K}}_{\boldsymbol{v}_i}\overline{\boldsymbol{K}}_{\boldsymbol{Y}}$ の固有値の和の定数倍が対応していることから，ここでも相関行列の固有値分布と HSIC の評価値の分布の類似性が推測される．

以上の類似性を考慮して検証を行った結果，$\boldsymbol{Y}$ と独立な関係を持つ HSIC の評価値 $\{HSIC_i\}_{i=1}^{p}$ は，図 1 のように MP 分布を含む自由 Meixner 分布族に従っていることが推測された．ここで，平均 0，分散 1 に標準化した自由 Meixner 分布の確率密度関数を以下に示す [4].

$$q(t) = \begin{cases} \dfrac{\sqrt{4(1+b) - (t-a)^2}}{2\pi(bt^2 + at + 1)} & (\gamma_- < t < \gamma_+) \\ 0 & (\text{otherwise}) \end{cases} \quad (3)$$

ただし，$b > 0$ かつ $a^2 < 4b$ とする．また，自由 Meixner 分布の台は

$$(\gamma_-, \gamma_+) = \left(a - 2\sqrt{(1+b)}, a + 2\sqrt{(1+b)}\right) \quad (4)$$

で表される．なお，MP 分布の確率密度関数は式 (3) の $b = 0$ かつ $a \neq 0$ の場合に相当する．

提案手法ではモーメント法を用いて自由 Meixner 分布の台を推定することにより，ノイズ変数の推定を行う．自由 Meixner 分布のモーメント母関数は以下のように表され，

$$M(t) = \frac{1 + 2b + at - \sqrt{(1-at)^2 - 4(1+b)t^2}}{2(t^2 + at + b)} \quad (5)$$

3 次モーメント，4 次モーメントはそれぞれ，

$$\left.\frac{d^3 M(t)}{dt^3}\right|_{t=0} = a \quad (6)$$

$$\left.\frac{d^4 M(t)}{dt^4}\right|_{t=0} = a^2 + b + 2 \quad (7)$$

となることを利用する．まず，式 (6) より 3 次標本モーメントから推定した $\tilde{a}$ を用いて式 (4) により $\tilde{b}$ を求め，次に，$\tilde{a}$ と $\tilde{b}$ から式 (7) によって求めた 4 次モーメントの推定値 $\tilde{m}_4$ と 4 次標本モーメントの一致性を調べて，自由 Meixner 分布の台の推定を行う．

4.1 提案手法のアルゴリズム

1. $\boldsymbol{X} = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_j, \ldots, \boldsymbol{v}_p]$ と $\boldsymbol{Y}$ を標準化し，HSIC によって $\boldsymbol{v}_j$ と $\boldsymbol{Y}$ の独立性を調べる．
2. 得られた HSIC の評価値 $\{HSIC_j\}_{j=1}^{p}$ を降順に並べ，$s_p, \cdots, s_1$ とする．
   このとき，s_l の l の値が小さい程，s_l に対応する変数 $\boldsymbol{v}_j$ は $\boldsymbol{Y}$ との独立性が高いことを意味する．
3. $S_d = \{s_{p-d}, \ldots, s_1\}$ $(d = 0, \ldots, p-1)$ を考え，各 s_j を標準化した値を新たに s_j と表す．ここで S_d の k 次標本モーメントを求める．

$$m_k^{(d)} = \frac{1}{p-d}\sum_{j=1}^{p-d} s_j^k \quad (8)$$

Vol.2012-MPS-91 No.39
Vol.2012-BIO-32 No.39
2012/12/7

表 1 d と s_l，s_l と $\boldsymbol{v}_j$ の対応関係

d	$s_l = s_{p-d}$	s_l の値	対応する変数 $\boldsymbol{v}_j$
0	s_{256}	0.1621	$\boldsymbol{v}_4$
1	s_{255}	0.0933	$\boldsymbol{v}_2$
2	s_{254}	0.0534	$\boldsymbol{v}_1$
3	s_{253}	0.0408	$\boldsymbol{v}_3$
4	s_{252}	0.0226	$\boldsymbol{v}_{220}$
5	s_{251}	0.0212	$\boldsymbol{v}_{188}$
6	s_{250}	0.0210	$\boldsymbol{v}_{100}$

表 2 パラメータ $\tilde{b}$ と 4 次モーメント

d	$\tilde{b}$	$\widetilde{m}_4^{(d)}$	$m_4^{(d)}$
0	1.652	90.223	107.151
1	3.798	53.726	70.782
2	4.715	21.124	28.380
3	5.436	12.806	14.414
4	0.746	4.044	4.404
5	0.560	3.671	4.088
6	0.703	3.679	3.874

表 3 $\hat{d}$ の推定

d	$\lvert m_4^{(d)} - \widetilde{m}_4^{(d)} \rvert$	$m_4^{(d-1)}/m_4^{(d)}$
0	16.928	−
1	17.055	1.513
2	7.256	2.494
3	1.608	1.968
4	0.359	3.272
5	0.417	1.077
6	0.194	1.055

4. $s_{p-d} = \gamma_+$ と仮定して 3 次標本モーメント $m_3^{(d)} = \tilde{a}^{(d)}$ から式 (4) により $\tilde{b}^{(d)}$ を求める．
5. 式 (7) により $\tilde{a}^{(d)}$ と $\tilde{b}^{(d)}$ を用いて 4 次モーメントの推定値 $\widetilde{m}_4^{(d)}$ を求める．
6. $\lvert m_4^{(d)} - \widetilde{m}_4^{(d)} \rvert < r$ かつ $\arg\max_d \dfrac{m_4^{(d-1)}}{m_4^{(d)}}$ を満たす d を $\hat{d}$ とする．
7. 上記 6. の 2 つの条件を満たす $\hat{d}$ が定まらないときは，1. に戻ってガウシアンカーネルのパラメータの値を変えて再度提案手法を試行する．

$S_{\hat{d}}$ に対応する変数の集合 $\boldsymbol{v}_j$ は $\boldsymbol{Y}$ と独立なノイズ変数の集合を表す．以上により γ_+ が推定され，$s_l \le \gamma_+$ に対応する変数をノイズ変数，$\gamma_+ < s_l$ に対応する変数をデータ $\boldsymbol{X}$ の構造部として認識することができる．

5. 実験

サンプル数 $n = 256$，次元数 $m = 256$ として，次のようなデータを用意する．

$y_i = -2\sin(2v_{i1}) + v_{i2}^2 + v_{i3} + \exp(-v_{i4}) + E.$

$\boldsymbol{Y} = [y_1, \ldots, y_i, \ldots, y_n]^T$,$E \sim N(0,1)$

$\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_i, \ldots, \boldsymbol{x}_n]^T = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_j, \ldots, \boldsymbol{v}_m]$,

$\boldsymbol{x}_i = [v_{i1}, \ldots, v_{ij}, \ldots, v_{im}]$, $\{v_{ij}\} \sim i.i.d.N(0,1)$.

$\boldsymbol{Y}$ と関係を持つ $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4$ の 4 変数を $\boldsymbol{X}$ の構造部として抽出し，それ以外の変数をノイズ変数として認識できるかどうかの実験を行う．

表 1 はガウシアンカーネルのパラメータを 1 としたときの提案手法 2. までの結果の一部で，s_l に対応する変数 $\boldsymbol{v}_j$ を示す．また，表 2 は更に提案手法 5. までを実行して得られた結果を示している．この結果をもとに提案手法 6. を行うと表 3 のようになる．

このとき $r = 1$ とすると，提案手法 6. の 2 つの条件を満たすのは $d = 4$ のときであることがわかる．よって，$\hat{d} = 4$ と推定することができ，これは S_4 に属する $\{s_{252}, \ldots, s_1\}$ のみを用いて推定を行ったとき自由 Meixner 分布への一致性が高くなることを意味する．S_4 は $\{\boldsymbol{v}_4, \boldsymbol{v}_2, \boldsymbol{v}_1, \boldsymbol{v}_3\}$ に対応する $\{s_{256}, s_{255}, s_{254}, s_{253}\}$ を $\{s_l\}_{l=1}^{256}$ から抜いた集合であるから，$\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4\}$ がデータの構造部であり，S_4 に属する $\{s_{252}, \ldots, s_1\}$ はノイズ部であることを認識できたことがわかる．

ところで，$4 < d$ では $\boldsymbol{Y}$ と独立な変数のみの集合となっているので，どの s_l を境界としても自由 Meixner 分布に従うことが考えられる．そのため，表 3 を見ると明らかなように，$\lvert m_4^{(d)} - \widetilde{m}_4^{(d)} \rvert < r$ の評価のみで $\hat{d}$ を推定することはできない．$\arg\max_d \dfrac{m_4^{(d-1)}}{m_4^{(d)}}$ の評価も併せて考慮し，自由 Meixner 分布の性質が最初に現れる $\hat{d}$ の推定を行う必要があることがわかる．

6. まとめ

HSIC と RMT とを組み合わせて自由 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し，$\boldsymbol{Y}$ と関係を持つ変数の最小部分集合を抽出できることが確認された．今後の課題として，変数同士が独立でない場合でも $\boldsymbol{Y}$ と関係を持つ変数の最小部分集合を抽出可能な方法を考察していきたい．

参考文献

[1] 福水健次: カーネル法入門, 朝倉書店 (2010).
[2] 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).
[3] 相馬亘, 藤原義久, 尹熙元：経済物理とランダム行列–株式市場にある本質的な構造の抽出 (ランダム行列の広がり–その多彩な応用), 数理科学, vol.45, No.2, pp.44–49, サイエンス社 (2007).
[4] Anshelevich, M.：*Bochner-Pearson-type characterization of the free Meixner class*, Adv. in App. Math., vol.46 (2011), pp.24–45.