

ランダム行列理論を用いた Gaussian カーネルにおける雑音の推定

伊藤 里江

お茶の水女子大学 大学院人間文化創成科学研究科

本研究では, Gaussian カーネル行列のスペクトル分布から Marcenko-Pastur 分布に対応する部分を推定し, 特徴空間に写像されたデータにおける雑音部の推定を試みる. ランダム行列理論を用いて雑音推定を行うことにより, データの本質的な構造を抽出することが可能になると考えられる. そこで, 本研究では独立正規雑音がついたデータ行列に Gaussian カーネルを施し, その雑音部を特定することは可能であるかを検証する.

A method of an estimation of noise in Gaussian kernel matrices by random matrices

Rie Itoh

Graduate School of Humanities and Sciences, Ochanomizu University

In this study, we investigate an estimation of the part of spectra corresponding to the support of Marcenko-Pastur distribution from the spectral distribution of Gaussian kernel matrix by the data mapped into the characteristic space. It would enable us to extract an essential structure of given data. Especially, we shall see some numerical results on the estimation for the part of noise in Gaussian kernel matrices.

1 はじめに

ランダム行列は 20 世紀はじめ, 数理統計学の研究で Wishart らによって考案され, 1950 年代 Wigner が原子核物理学にランダム行列理論が応用されるようになった. このような歴史をもつランダム行列は, 数理物理学, 統計学など多くの分野から様々な観点で興味を持たれ研究されている. 情報学や工学の分野で不確実性, 不確定性を定量的に取り扱う必要のある局面でも重要な役割を果たし, 現在ランダム行列の応用範囲は多岐に及んでいる.

代表的なランダム行列として, ガウス直交ランダム行列のスペクトル密度分布が Wigner 分布 (半円分布) に, また Wishart 行列のスペクトル密度分布が Marcenko-Pastur 分布に従うことは知られている.

したがって, 多変量観測データから Wishart 行列に対応する部分, すなわちデータのスペクトル密度

分布から Marcenko-Pastur 分布に対応する部分を推定することにより, データのより本質的な部分を取り出すことが可能となると思われる. Marcenko-Pastur 分布は正の有限な区間でのみ 0 でない値をとる. 一般にこのような領域は台 (support) と呼ばれている.

観測されたデータのスペクトル分布から Marcenko-Pastur 分布部分の台の最小値および最大値を推定できたならば, 最大値の側がデータの本質的な構造部と雑音部の切り分けを行う境界値と考えられるであろう.

しかし, 実際のデータに適用した際, どの固有値までが Marcenko-Pastur 分布の台に含まれるのか, すなわち, どこまでが雑音部を指しているのかを見ただ目で推測する難しい. 本研究では Marcenko-Pastur 分布と照らし合わせることで, すなわち Marcenko-Pastur 分布の affine 変換により合理的に導き出す手法について考える.

2 ランダム行列理論

一般に要素が確率変数であるような行列はランダム行列と呼ばれる。適当な対称性のみを課したランダム行列の固有値や固有ベクトルの統計的な性質を考えることがランダム行列理論である。

2.1 Wishart 行列

その性質が良く調べられているランダム行列のひとつとして Wishart 行列がある。

各成分が独立に標準正規分布 $N(0, 1)$ に従う確率変数からなる $n \times p$ の長方形行列 C とする。このランダム行列 C から

$$S = \frac{1}{n} C^t C$$

として得られる $n \times n$ 対称ランダム行列 S は Wishart 行列と呼ばれている。 $p/n = \lambda \geq 1$ を保ったまま $n \rightarrow \infty, p \rightarrow \infty$ の極限をとると、 S の固有値の経験分布は以下の確率密度関数に収束することが知られている。

$$\rho(t) = \frac{1}{2\pi} \frac{\sqrt{-(t - \lambda_{\max})(t - \lambda_{\min})}}{\lambda t}$$

ここで、

$$\lambda_{\min}^{\max} = (1 \pm \sqrt{\lambda})^2 \quad (1)$$

この密度関数の台は、もちろん $t \in [\lambda_{\min}, \lambda_{\max}]$ である。すなわち、 $\rho(t)$ は上のような極限で与えられる無限次元行列の作用素のスペクトル密度関数である。

また、この確率密度関数を持つ確率分布は Marcenko-Pastur 分布と呼ばれている。

2.2 Gaussian カーネル

カーネルはパターン認識において使われる手法の一つで、線形分離不可能なものを、カーネルを用いて高次元な特徴空間に写像することにより、線形分離可能な状態にすることができる。一般に、特徴空間における内積をデータから直接計算する手段を与えるものとして利用されるが、その中で Gaussian カーネル行列は相関行列と同じような振る舞いをすることが知られている。

実は後述のように、ランダムベクトルの Gaussian カーネルを用いて写像したデータに対しても、その固有値の分布は Wishart 行列の場合と同等 (ほとんど Wishart 行列の場合の Affine 変換) になることが近年示されている。

すなわち、上の行列 C の各行ベクトル $x_i; (i = 1, 2, \dots, n)$ を Gaussian カーネルで写像すると、特徴空間における $n \times n$ 行列 (M_{ij}) は

$$M_{ij} = K(i, j) = \exp\left(-\frac{\|x_i - x_j\|^2}{A}\right) \quad (2)$$

となる。

ただし、 $\|\cdot\|^2$ は通常のユークリッド 2 乗距離で、 $A \in \mathbb{R}$ は適当なパラメータである。

以下に、実際に、このような Gaussian カーネル行列 (M_{ij}) 固有値の分布がどのように振る舞うかを乱数による数値実験の例により挙げておく。

$n \times p = 200 \times 600$ 行列 C の各成分に、独立に標準正規分布 $N(0, 1)$ に従う乱数を発生させて、上の式で行列 S を構成し、その 600 個の固有値のヒストグラムを描いたものが図 1 である。なお、図 1 の実線による曲線は、パラメータ $\lambda = 3$ の Marcenko-Pastur 分布の密度関数である。

また、先に正規乱数を発生して構成した行列 C の各行ベクトルを用いて (2) 式に従って $A = 400$ とし Gaussian カーネル行列を構成し、その固有値のヒストグラムを描いたものが図 2 である。

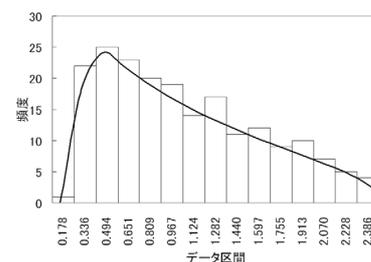


図 1: Wishart 行列の固有値の数値実験

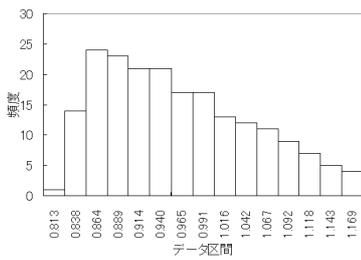


図 2: Gaussian カーネル行列の固有値の数値実験

図 1 と 図 2 のヒストグラムを見比べてみると形状が非常に酷似していることが見てとれるであろう。実際、固有値を共に昇順に並べプロットしたものが下の図 3 である。ただし、Gaussian カーネル行列には非常に大きな固有値がひとつ存在するので、それを除いた 599 組に関するプロットである。

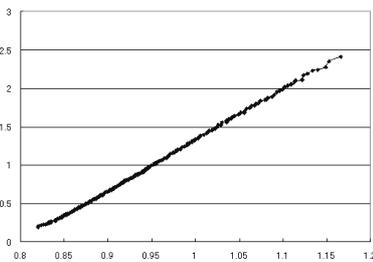


図 3: 上の 2 つの固有値集合の関連性

図 3 のプロットが直線であることより、2 つの固有値の密度関数の関連性は、先にも述べたように X を Wishart 行列のスペクトル密度関数を確率密度関数に持つ確率変数とし、 Y を Gaussian カーネル行列のスペクトル分布を確率密度関数に持つ確率変数とすると、 $Y = aX + b$ という関係があることが見てとれる。

したがって、推定したい Gaussian カーネル行列のスペクトル密度関数の台の最大値、最小値は (1) 式より

$$\tilde{\lambda}_{\min}^{\max} = a \left(1 \pm \sqrt{\frac{p}{n}} \right)^2 + b \quad (3)$$

となり、 n は与えられるので、この式中の a, b, p が推定できれば、最適な密度関数の台の最大値、最小値を求められることになる。もちろん、 $\lambda = p/n$ であつたので、Marcenko-Pastur 分布のパラメータ λ

と shift パラメータ b , dilation パラメータ a が推定できればよいことになる。これは、与えられたデータ (固有値) に最適な分布パラメータの推定問題で、本研究では Marcenko-Pastur 分布の性質が良く反映されるようなモーメント法と本質的に同じ方法で推定することを考える。

3 自由キウムラント

Marcenko-Pastur 分布はランダム行列理論では Wishart 行列のスペクトル密度関数として知られていたが、近年の急速に発展している自由独立性に基づく自由確率論においてはポアソン分布の役割を果たす分布である。^[2] 通常確率論では、 n 次キウムラントは n 次以下のモーメントを用いてあらわすことができる。^[3] 自由キウムラント (自由独立性に基づく合成積を線形化する半不変量) も同様に n 次以下のモーメントを用いてあらわされる。

実際、 m 次のモーメントを μ_m とし、モーメント母関数 (指数型ではない) を

$$M(z) = \sum_{m=0}^{\infty} \mu_m z^m$$

とおく。また m 次の自由キウムラントを r_m とし、その母関数を

$$C(z) = \sum_{m=0}^{\infty} r_m z^m$$

とする。ただし、ここで便宜的に $r_0 = 1$ とする。このとき $M(z)$ と $C(z)$ の間には

$$M\left(\frac{z}{C(z)}\right) = C(z)$$

の関係があることが知られている。この関係より、自由キウムラント r_m は m 次以下のモーメントを用いて、表されることが分かる。最初のいくつかの r_m を書き出すと、以下のようなになる。

$$\begin{aligned} r_1 &= \mu_1, \\ r_2 &= \mu_2 - \mu_1^2, \\ r_3 &= \mu_3 - 3\mu_2\mu_1 + 2\mu_1^3, \\ r_4 &= \mu_4 - 4\mu_3\mu_1 - 2\mu_2^2 + 10\mu_2\mu_1^2 - 5\mu_1^4, \\ &\vdots \end{aligned}$$

3 次までの自由キュムラントの式は通常確率論での関係と同じであるが、4 次以上の高次の自由キュムラントの式は異なるものである。

Marcenko-Pastur 分布の場合、 $r_k = \lambda$ ($k \geq 1$) となり、すべての次数の自由キュムラントが等しくなる。これは、通常確率論での Poisson 分布の特徴付けである「すべてのキュムラントが一定」ということに対応している。

また、確率変数 X の k 次自由キュムラントを $r_k(X)$ とし、 $Y = aX + b$ の k 次自由キュムラントを $r_k(Y)$ とあらわすと、

$$\begin{aligned} r_1(Y) &= ar_1(X) + b, \\ r_k(Y) &= a^k r_k(X) \quad k \geq 2 \end{aligned}$$

であることも分かる。この確率変数の shift と dilation に関する自由キュムラントの関係は通常キュムラントと全く同じである。

また、Gaussian カーネル行列のスペクトル分布は標準 Gaussian 直交ランダム行列のスペクトル分布の affine 変換であることは既に述べた通りである。したがって、標準 Gaussian 直交ランダム行列のスペクトル分布に対応する Marcenko-Pastur 分布の自由キュムラントを $r_k = \lambda$ とすると、Gaussian カーネル行列のスペクトル分布の自由キュムラントは

$$\begin{aligned} r_1 &= a\lambda + b \\ r_2 &= a^2\lambda \\ r_3 &= a^3\lambda \\ r_4 &= a^4\lambda \\ &\vdots \end{aligned}$$

となる。スペクトルのデータから 4 次までのモーメントを求め、上のモーメント-キュムラント公式で 4 次までの自由キュムラントを計算し、これらの式から、パラメータである a, b, λ の値を推定することが可能である。これは通常のいわゆるモーメント法と呼ばれる確率分布のパラメータ推定手法と同等である。

4 分散 σ^2 の推定

$\{X_i\}_{i=1}^n$ を各成分が独立に標準正規分布 $N(0, 1)$ に従う p 次元ランダムベクトルとする。N. El Karoui は適当な微分可能性の要件を満たす関数 f を用いて (i, j) 成分が

$$M_{ij} = f\left(\frac{X_i^t X_j}{p}\right)$$

で与えられるような $n \times n$ カーネルランダム行列のスペクトル分布の $n \rightarrow \infty$ の漸近挙動を調べた。[1]

その結果 $M = (M_{ij})$ のスペクトル分布は、以下の式で与えられる \widetilde{M} のスペクトル分布に漸近的に等しくなることが示された。

$$\widetilde{M} = f(0)\mathbf{1}\mathbf{1} + f'(0)\frac{X^t X}{p} + v_p \text{Id}_n \quad (4)$$

ただし、 X は第 i 行が X_i で与えられる $n \times p$ 行列で

$$v_p = f\left(\frac{\text{Trace}(\Sigma)}{p}\right) - f(0) - f'(0)\frac{\text{Trace}(\Sigma)}{p},$$

である。さらに、 Σ は $\{X_i\}_{i=1}^n$ の共分散行列である。また、 Id_n は n 次元単位行列を表す。

ここで $\mathbf{1}\mathbf{1}$ はすべての成分が 1 の行列である 1 つの固有値に対応するが、ランク落ちを防ぐためのものであり、漸近的なスペクトル分布には影響しない。これが先に述べた 1 つの非常に大きな固有値に対応する。

式 (4) を見ると $X^t X$ の項から dilation パラメータ a が、単位行列の項から shift パラメータ b が導かれることが分かる。

この公式を用いて、ランダムベクトルの各成分が iid in $N(0, 1)$ ではなく、iid in $N(0, \sigma^2)$ の場合の、この分散の値 σ^2 の推定も可能となることが分かる。

実際、 $\{X_i\}_{i=1}^n$ を各成分が独立に $N(0, \sigma^2)$ に従う p 次元ランダムベクトルとし、 (i, j) 成分が

$$M_{ij} = \exp\left(-\frac{\|X_i - X_j\|^2}{A}\right)$$

で与えられるような $n \times n$ Gaussian カーネル行列の場合 (4) 式より a, b の値は

$$\begin{aligned} a &= \frac{2p}{A} \exp\left(-\frac{2p\sigma^2}{A}\right) \\ b &= \exp\left(\frac{-2p\sigma^2 + 2p}{A}\right) - \exp\left(-\frac{2p\sigma^2}{A}\right) \\ &\quad - \frac{2p}{A} \exp\left(-\frac{2p\sigma^2}{A}\right) \end{aligned}$$

と表すことができる。この式により、先に推定された λ すなわち p と a, b から σ^2 推定できる。

5 実験

5.1 実験概要

Gaussian カーネルによって写像したデータのスペクトルから, Marcenko-Pastur 分布の台の最大値, 最小値を導き出し, どこまでが雑音なのかを推定する。その推定にあたっては, 上位のスペクトルを取り除きながら最適なスペクトル密度関数の台の最大値, 最小値を見つけることにし, その評価基準として, 平均対数尤度

$$LL = \frac{1}{N} \sum_{i=1}^N \log \rho(x_i)$$

を用いる。ただし, N は $n - (\text{取り除いた固有値の数})$ 。この平均対数尤度を大きくするものが最適な台の範囲と推定する。

また, 正規雑音 $N(0, \sigma^2)$ の標準偏差 σ の値は 4 節の手法で推定する。

5.1.1 $N(0, 4)$ の数値実験

各成分に平均 0, 標準偏差 2 の正規乱数を発生させた 900×1200 行列を生成。式 (2) 中の A の値は $A = 1500$ として, その行列を Gaussian カーネルで写像し, 実験を行った。第 1 最大固有値を取り除いた 899 個の固有値のみでは $LL = -5.96$, そこから固有値を 1 個取り除くと $LL = -5.99$, 2 個取り除くと $LL = -6.01$ となった。よって第 1 最大ベクトルを取り除いたスペクトル分布が Marcenko-Pastur 分布に対応する部分であると考えられ, その際の標準偏差については $\sigma = 1.87$ と推定される。また, このときのスペクトル分布は次のようになる。

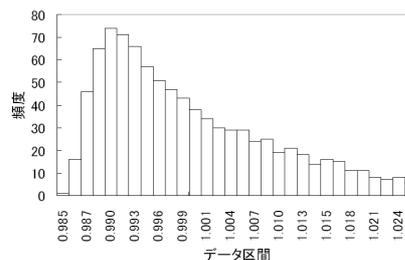


図 4: $N(0, 4)$ のスペクトル分布

5.1.2 構造データに正規ノイズを加えた数値実験

1 ~ 300 行が $(10, 0, 0, 0, \dots, 0)$, 301 ~ 600 行が $(0, 10, 0, 0, \dots, 0)$, 601 ~ 900 行が $(0, 0, 10, 0, \dots, 0)$ であるような構造データを 900×1200 行列として用意した。そこに $N(0, 4)$ の正規雑音を加え, Gaussian カーネルを用いて写像を行った。このときのスペクトル分布は以下の通り。

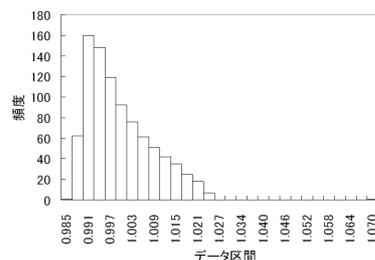


図 5: 構造データ + $N(0, 4)$ のスペクトル分布

先の実験同様, 上位の固有値を 3 個取り除くと $LL = -6.28$, 4 個取り除くと $LL = -6.34$ となり, Marcenko-Pastur 分布に対応するのは 3 個取り除いた際の分布であることが分かる。(1 個, 2 個取り除いた際は, $\lambda < 1$ となることから明らかに不適切であると判断される。) また, このときの標準偏差は $\sigma = 1.86$ と推定できる。スペクトルを 3 個取り除いた際の分布は次のようになる。

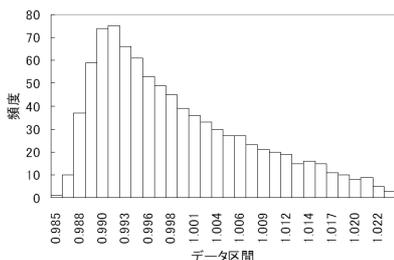


図 6: 構造データ $+N(0, 4)$ のスペクトル分布 (3 個のスペクトルを取り除いた後)

6 まとめ

ランダム行列理論を Gaussian カーネルで写像したデータに応用することで、そのデータのノイズに対応するスペクトル分布、すなわち Marcenko-Pastur 分布を推定した。さらに、構造データにノイズを加えた場合でも同様に応用することは可能か、という検証も行ったが、前節図 4 と図 6 のヒストグラムの形状は非常に似ており、構造データに加えられたノイズ部を取り出すことができたと思われる。また、いずれの場合でもノイズの標準偏差を推定することができ、ランダム行列理論を応用することは有効であったといえる。

7 今後の課題

今回、Gaussian カーネルのパラメータ A の値は自ら勝手に定めてしまったが、本来はそれぞれのデータに応じて最適な値が存在しているはずである。その最適な A の値を用いることにより、スペクトル密度分布は変わってくる考えられ、その際にはより良い σ^2 推定もできると思われる。そこで、今後の課題として、その最適な A を求め、標準偏差等を再度推定し直したいと思う。

参考文献

[1] El Karoui, N. *The spectrum of kernel random matrices*, Technical Report 748, Department

of Statistics, UC Berkeley. To appear in *The Annals of Statistics*, (2007).

[2] D. V. Voiculescu, K. Dykema, and A. Nica, *Free random variables*, CRM Monograph Series, 1 American Mathematical Society, Providence, RI, 1992.

[3] A. Stuart and J. K. Ord, *Kendall's advanced theory of statistics*. Vol. 1, Distribution theory, Sixth ed., Edward Arnold, London; copublished in the Americas by Halsted Press, New York, 1994.

[4] 麻生英樹, 津田宏治, 村田昇, *パターン認識と学習の統計学*, 岩波書店, 東京, 2005