

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

理学専攻 情報科学コース 伊藤里江 (指導教員：吉田裕亮)

1 はじめに

多変量観測データの分散共分散行列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考えられる。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 (support) 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なるはずである。

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を見ただ目で推測する難しい。そこで、本研究では 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、独立正規雑音がついたデータ行列に Gaussian カーネルを施した際に、その雑音部を特定することは可能であるかを検証する。

2 ランダム行列理論

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、

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

この確率密度関数を持つ確率分布が Marcenko-Pastur 分布である。

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ひとつである。カーネルには様々な種類のものがあるが、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事が知られている。

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。

ここで、 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、 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、 $Y = aX + b$ という関係にある。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。

3 自由キウムラント

自由キウムラント r_m をモーメント μ_m を用いて最初のいくつかを書き出すと、以下のようなになる。

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

Marcenko-Pastur 分布の場合、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ため $r_k = \lambda$ ($k \geq 1$) となり、すべての次数の自由キウムラントが等しくなる。

また、確率変数 X の k 次の自由キウムラントを $r_k(X)$ とし、 $Y = aX + b$ の k 次の自由キウムラントを $r_k(Y)$ 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。

ここで、Marcenko-Pastur 分布の自由キウムラントを $r_k = \lambda$ とすると、Gaussian カーネル行列のスペクトル分布の自由キウムラントは

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

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。

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、 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。

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。

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} 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

$$a = \frac{2p}{A} \exp\left(-\frac{2p\sigma^2}{A}\right)$$

$$b = \exp\left(\frac{-2p\sigma^2 + 2p}{A}\right) - \exp\left(-\frac{2p\sigma^2}{A}\right) - \frac{2p}{A} \exp\left(-\frac{2p\sigma^2}{A}\right)$$

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 の推定ができる。

5 実験

5.1 実験概要

Marcenko-Pastur 分布の台の最大値、最小値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。

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図 1 のようになる。

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。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。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図 2 のようになる。

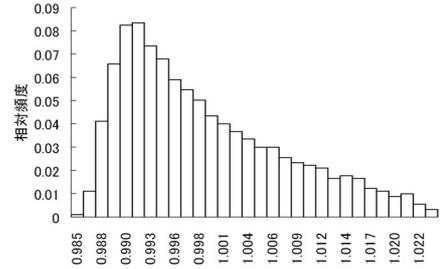


図 1: $N(0, 4)$ のスペクトル分布

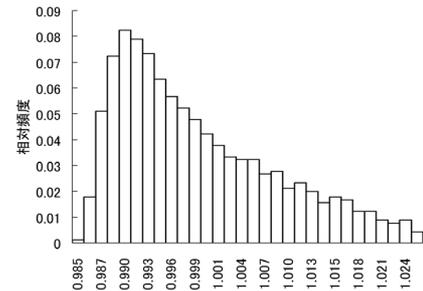


図 2: 構造データ + $N(0, 4)$ のスペクトル分布

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 1 と図 2 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測することができ、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。

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).