Vol.2011-MPS-86 No.21
Vol.2011-BIO-27 No.21
2011/12/2

ランダム行列理論のデータ距離行列への応用

小　林　由　香[†1]

本研究では, 二点間の距離が推定されているデータにおいて, 距離行列からランダム行列の理論を用いて雑音量の推定を行い, MDS に応用する手法について考察する.

An application of random matrx theory to data-distance matrices

Yuka Kobayashi[†1]

In this study, given the distance matrix, which is constituent from the estimated data between the two points with noise, using the we shale propose the method which enables us to estimate the power of the noise in the distanse and apply to the *MDS*.

1. はじめに

本研究では, ノイズを含む二点間の距離が推定されたデータにおいて, ランダム行列理論を用いてデータ距離行列 (三角行列) に含まれる雑音量の推定を行う. 特に MDS(Multi Dimensional Scaling) 法の手法に用いられる内積行列から, 雑音量の推定を行う手法を提案する.

2. 半円分布

ランダム行列とは, 一般に確率変数を要素にもつ行列である. 特に成分が, $a_{ii}=0$ かつ $a_{ij}=a_{ji}$ が成り立ち, 確立変数の族 $a_{ij}(1\le i<j\le N)$ が独立に $N(0,\sigma^2)$ に従うランダム行列は, ガウス型直交ランダム行列とも呼ばれ, その固有値経験分布は行列のサイズ N を $N\longrightarrow\infty$ とするとき, 以下の中心 0 の半円分布に収束されることが知られている.

一般に, 中心が $m\in$, 半径 $r>0$ の半円分布 $\omega_{m,r}$ とは, 密度関数 $P(t)$ が

$$P(t)=\frac{2}{\pi r^2}\sqrt{r^2-(t-m)^2} \tag{1}$$

で与えられる分布である.

半円分布 $\omega_{m,r}$ の k 次モーメント m_k は

$$m_k=\frac{2}{\pi r^2}\int_{-r}^{r}t^k\sqrt{r^2-t^2}dt=\begin{cases}0, & (k=2m+1),\\ \frac{(2m)!}{m!(m+1)!}(\frac{r^2}{4})^m, & (k=2m),\end{cases} \tag{2}$$

で与えられる. また, $C_m=\frac{(2m)!}{m!(m+1)!}$ は m 次 Catalan 数と呼ばれる組合せ論でよく現れる数でもある.

3. MDS(Multi Dimensional Scaling) 法

MDS 法とは幾つかの対象があり, 任意の 2 つの対象間の距離が Euclid 距離として推定されている場合, Young-Householder の定理をもとに対象を Euclid 空間の点として布置（位置付ける）方法である. n 個の対称 $O_1,O_2,\ldots,O_n$ のうち, O_n を原点とし, 残りの $n-1$ 個の対称を終点とする (列) ベクトルを $x_1,x_2,\ldots,x_{n-1}$ とする. さらに, 対象 O_i の第 j 軸の座標値を x_{ij} とし, 行列 $X=x_{ij}$ を定義する. 行列 X は, $n-1$ 個の対象の埋め込まれる空間の次元数分の座標値を対象ごと各行に並べたものである. この時, この空間の次元数 r は,

$$B=b_{ij}=XX^t \tag{3}$$

と置けば, $r=\mathrm{rank}X=\mathrm{rank}B$ となる. 行列 B は, $n-1$ 個の対象間の内積を要素とする行列で, 内積行列である.

$$B=(b_{ij})=\begin{pmatrix}(\overrightarrow{x_1}\mid\overrightarrow{x_1}) & (\overrightarrow{x_1}\mid\overrightarrow{x_2}) & \ldots & (\overrightarrow{x_1}\mid\overrightarrow{x_{n-1}})\\ (\overrightarrow{x_2}\mid\overrightarrow{x_1}) & (\overrightarrow{x_2}\mid\overrightarrow{x_2}) & \ldots & (\overrightarrow{x_2}\mid\overrightarrow{x_{n-1}})\\ \vdots & \vdots & \ddots & \vdots\\ (\overrightarrow{x_{n-1}}\mid\overrightarrow{x_1}) & (\overrightarrow{x_{n-1}}\mid\overrightarrow{x_2}) & \ldots & (\overrightarrow{x_{n-1}}\mid\overrightarrow{x_{n-1}})\end{pmatrix} \tag{4}$$

とする. つぎに, ベクトル $v_{ij}=x_j-x_i$ を定義すると, 対象 O_i と O_j 間のユークッリド距離 d_{ij} を Euclid ノルムを用いて表現すれば

$$\| v_{ij}\|^2=d_{ij}^2=v_{ij}^t v_{ij}=(x_j-x_i)^t(x_j-x_i)=\| x_i\|^2+\| x_j\|^2-2b_{ij} \tag{5}$$

†1 お茶の水女子大学大学院 人間文化創成科学科
Graduate school of Humanities and Sciences, Ochanomizu University

ⓒ 2011 Information Processing Society of Japan

Vol.2011-MPS-86 No.21
Vol.2011-BIO-27 No.21
2011/12/2

と書ける. $\| x_i \|^2 = d_{ni}^2$ より

$$b_{ij} = \frac{1}{2}(d_{ni}^2 + d_{nj}^2 - d_{ij}^2) \tag{6}$$

である. つまり, 対象間の距離情報（距離の二乗）から内積行列が構成されることになる. 行列 B は対象 O_n を原点とした場合の内積行列であり, どの対象を原点に取るかで行列 B は一般にことなる. 誤差を含む場合には原点の取り方には工夫が必要であり, どの対象 (点) も対等に扱うことなので, n 点の重心を原点に選ぶことが多い. n 個の対象行列の重心を原点に選んだときの各点の位置ベクトルの内積行列 B_c を求めるには, 対象間 2 乗距離を要素とする $n \times n$ の行列 $T = \{\hat{d}_{ij}^2\}$ を作る. 次に, 中心化行列

$$J_n = I - \frac{1}{n}\mathbf{1}_n\ {}^t\mathbf{1}_n \tag{7}$$

を作り, ただし, I_n は n 次単位行列 ${}^t\mathbf{1}_n = (1, 1, \ldots, 1)$, J_n は

$$J_n = \begin{pmatrix} 1-\frac{1}{n} & -\frac{1}{n} & \ldots & -\frac{1}{n} \\ -\frac{1}{n} & 1-\frac{1}{n} & \ldots & -\frac{1}{n} \\ \vdots & \vdots & \ddots & -\frac{1}{n} \\ -\frac{1}{n} & -\frac{1}{n} & \ldots & 1-\frac{1}{n} \end{pmatrix} \tag{8}$$

となる. これを用いて, 次のような計算により, 重心を原点とする位置ベクトルから構成される. 内積行列 B_c は

$$B_c = -\frac{1}{2} J_n T J_n^t \tag{9}$$

と求められる. これは Young-Householder 変換と呼ばれている. MDS では, B_c スペクトル分解を行うことにより, 布置座標を求めることになる.

本研究では, この内積行列の点の配置構造以外のスペクトル (固有値) より雑音部を推定する手法を提案する. すなわち, 2 点間の距離が推定された r 次元に埋め込まれたデータより内積行列 B_c を求める. r 個の優固有値を除いたスペクトルからモーメント法を用いて, 半円分布と比較することにより雑音量 (すなわち, 密度関数の半径) の推定を行う.

4. 実 験 例

まず, ノイズを含み 2 点間の距離が推定されている, 2 次元の構造データを 200 点用意する. これらの距離行列から, 先に述べた MDS 法により内積行列 B_c を求める. B_c のスペクトルから, 2 個の優固有値を除き, 各次数のモーメントを計算する (モーメントの各値は 表 1 のようになる). この場合は, $r = 21.2$ と推定され, 推定された半径を元に半円分布と B_c のヒストグラムの密度と重ね合わせると, 図 1 のようになり, モーメント法により半径, すなわち雑音量, の推定ができたと考えられる.

	2 乗	4 乗	6 乗	8 乗	10 乗
理論値	1	2	5	14	42
モーメント	1.02028	2.14757	5.65716	16.6713	52.5366

表 1 モーメント
Table 1 moment

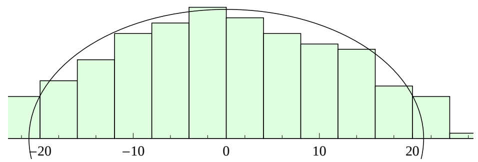

図 1 B_n のスペクトルと Wigner 分布
Fig. 1 nonliner structure and Wigner distribution

5. ま と め

データ距離行列とランダム行列理論の Wigner 半円分布を用いて, 雑音部の推定を行った. 雑音部の推定は目視ではなくモーメント法を用いて, 定量的に行う手法を提案した. 今後の課題として, 距離行列から生成される様々なタイプの行列とランダム行列との関連によりノイズ推定の精度を上げることも可能と考えられる.

参 考 文 献

1) 広尾太郎,「ランダム行列の広がり」, 数理科学, 第 45 巻 2 号, サイエンス社, (2007 年)

ⓒ 2011 Information Processing Society of Japan