

## 非正則な共分散行列に対する客観的総合指数

数理情報学専攻 48216231

村上 湧哉

指導教員 清 智也 教授

## 1 はじめに

$p$  個の指標を持つ  $n$  個体のデータから一つの総合指数を算出し、その値をもとに各個体に対してランク付けを行うことがしばしば行われている。たとえば科目数を  $p$ 、生徒数を  $n$  としたときのある学校のあるテストの成績データ、 $p = 10$ 、選手数を  $n$  としたときの十種競技の成績データ、複数の項目からなる都道府県幸福度ランキングのデータ [4] などがこれに当てはまる。

自然な総合指数として、偏差値の平均やデータの第 1 主成分を用いることが考えられる。しかしこれらは、場合によってはある指標と負の相関を持つことがある。総合指数はすべての指標の傾向を捉えているべきであるという考えのもとでは、この性質は望ましくない。

[3] はこの総合指数の一つとして、任意の指標との相関が正の値となる客観的総合指数 (objective general index; OGI) を提案したが、OGI は指標数に対して個体数が少ない場合など、データの共分散行列が非正則となる場合には計算不可能となってしまうことがある。そこで本研究では、共分散行列が非正則な場合にも OGI に相当する値を算出することを目的として、共分散行列の正則化によるアルゴリズムの拡張を提案した。

## 2 行列の対角スケーリング

対角スケーリング問題とは、与えられた正方行列の左右から対角行列をかけて、行和、列和を 1 にする問題の総称である [1]。以下の対角スケーリング定理は、客観的総合指数の根拠となっている。

**定理 1** ([2]). 任意の正定値行列  $S \in \mathbb{R}^{p \times p}$  に対して、対角成分が正の対角行列  $D$  で、 $DSD$  の各行の和が 1、すなわち  $DSD\mathbf{1}_p = \mathbf{1}_p$  であるものが一意に存在する。

## 3 客観的総合指数

定理 1 より、以下の客観的総合指数が定義できる。

**定義 2** (OGI [3]).  $X = (\mathbf{x}_1, \dots, \mathbf{x}_p)$  は中心化されたデータ行列、 $S = \frac{1}{n}X^\top X$  とする。定理 1 より、 $S$  が正則ならば  $DSD\mathbf{1}_p = \mathbf{1}_p$  となる  $D$  が一意に存在し、 $D$  の対角成分を並べたベクトルを  $\mathbf{w} = \mathbf{w}(S) > \mathbf{0}$  と書

く。このとき、

$$\mathbf{g}_{\text{OGI}} = \sum_{i=1}^p w_i \mathbf{x}_i$$

を客観的総合指数 (objective general index; OGI) と呼ぶ。また、 $\mathbf{w}$  を客観的重み (objective weight) と呼ぶ。

定義 2 および定理 1 より、

$$\psi(\mathbf{w}) = \frac{1}{2} \sum_{i,j=1}^p S_{ij} w_i w_j - \sum_{i=1}^p \ln w_i$$

とおくと、 $\nabla \psi(\mathbf{w}) = \mathbf{0}$  のときに  $\mathbf{w}$  が客観的重みとなるから、OGI は狭義凸関数  $\psi$  の最小点として求められる。  $\psi$  は明らかに狭義凸であるから、OGI は凸計画問題の解として数値的に計算することが可能である。

## 4 提案手法

$S$  が非正則な場合、既存の方法では最適化問題が非有界となり解けないことがある。本研究ではこの問題に対して、以下のように  $S$  を修正することにより、正則化してから OGI を求めることを考える。

$$S \rightarrow S + \lambda D_S$$

ただし  $\lambda \in \mathbb{R}_{\geq 0}$  で、 $D_S \in \mathbb{R}^{p \times p}$  は  $S$  の非対角成分を 0 にしてできる対角行列とする。

## 4.1 正則化 OGI に関する理論的考察

ここではまず、 $\lambda \rightarrow 0$  の場合の解の挙動を理論的に調べた。具体的な低次の  $p$  に対する観察から、 $\lambda$  が小さい場合には  $w_i \propto \frac{1}{\sqrt{\lambda}}$  となることを予想し、 $w_i = \frac{a_i}{\sqrt{\lambda}}$  ( $a_i > 0$ ) とおいて目的関数を展開した。これによって  $\mathbf{a}$  について以下の最適化問題を得た。

$$\begin{aligned} \text{minimize} \quad & F_a(\mathbf{a}) = - \sum_i \log a_i + \frac{1}{2} \mathbf{a}^\top D_S \mathbf{a} \quad (1) \\ \text{subject to} \quad & \mathbf{a} > \mathbf{0} \\ & X\mathbf{a} = \mathbf{0} \end{aligned}$$

**定理 3.**  $\mathbf{a} = (a_1, \dots, a_p)^\top \in \mathbb{R}^p$  に関する最適化問題 (1) の解  $\mathbf{a}$  とラグランジュ乗数  $\boldsymbol{\nu}$  は、以下の関係を満たす。

$$\begin{cases} \|a_i \mathbf{x}_i + \frac{1}{2} \boldsymbol{\nu}\|^2 = 1 + \frac{1}{4} \|\boldsymbol{\nu}\|^2, & i = 1, \dots, p \\ X\mathbf{a} = \mathbf{0} \end{cases}$$

## 4.2 具体的な $X$ のもとでのシミュレーション

$n, p$  が小さい場合において, 具体的な  $X$  に対して提案手法を適用することにより, 定理 3 の正当性を確認した.  $n = 2, p = 5$  とし, データ  $X$  の各列ベクトルの配置を図 1 のようにした場合の提案手法の適用結果を図 2 に示した. この  $X$  について, 各列ベクトルは円周上に並び, かつ重心が原点となるように選ばれていることに注意する. グラフを見ると,  $\lambda \rightarrow 0$  では  $\mathbf{a} \approx k\mathbf{1}$  ( $k \in \mathbb{R}$ ) となっており, 確かに定理 3 を満たしていることがわかる.

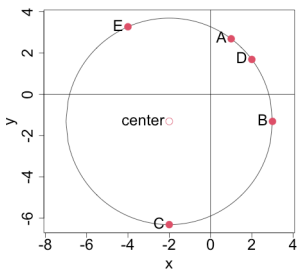


図 1:  $p = 5$  の例でのデータ  $X$  の配置.

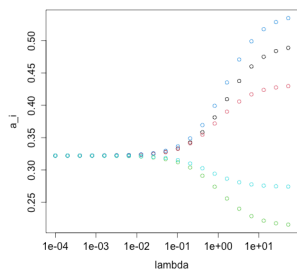


図 2:  $p = 5$  の例での  $\lambda$  ごとの  $a_i$  の様子.

## 5 数値実験

提案手法で得られる  $\mathbf{w}$  について,  $\lambda$  の取り方による変化を調べるための数値実験を行った. 既知の分布から発生させたデータ  $X \in \mathbb{R}^{n \times p}$  ( $n < p$ ) に提案手法を適用することで得られる客観的重みと, 真の共分散をもとに得られる客観的重みとの 2 乗誤差を用いて,  $\mathbf{w}$  の精度を評価した. 以下に具体的な数値実験の手順を示す.

1. 行列  $X_r = (x_{r,ij})_{ij} \in \mathbb{R}^{5 \times 10}$  を, それぞれ独立に標準正規分布に従う乱数として生成する.
2. 行列  $A$  を用意し,  $X = X_r A$  としてデータ行列  $X$  を生成する. このとき,  $S$  が非正則となるような  $X$  を得るため,  $n < p$  となるように設定した.
3.  $X$  に対して提案手法を適用し, 客観的重みを求める. このとき,  $\lambda \in \{10^{-4} \times 2^{i-1} \mid i = 1, \dots, 20\}$  の 20 通りを用いた.
4. 真の分布に対する共分散行列  $S_{\text{true}} = A^T A$  に対して求めておいた客観的重み  $\mathbf{w}_{\text{true}}$  を用いて, 2 乗誤差  $\text{err} = \|\mathbf{w}_{\text{true}} - \mathbf{w}\|^2$  を算出する.
5. 以上の操作を 1000 回繰り返す,  $\lambda$  の値ごとに  $\text{err}$  の平均をプロットする.

実験結果を抜粋して図 3~6 に示す. また, 各実験で用

いた  $A$  を以下に示す.

実験 1: 対角成分を 1 とし, 非対角成分は (1,2) 成分のみ 5, それ以外は 0 である行列.

実験 2: 対角成分を含む上三角成分をすべて 1 で埋めた行列.

実験 5: 単位行列を基本とし, 対角成分を除く上三角成分を標準正規分布から生成した行列.

実験 8: 実験 1 の  $A$  の (1,2) 成分を  $-1$  倍したもの.

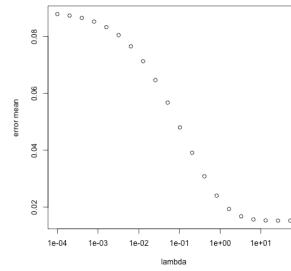


図 3: 実験 1 の結果.

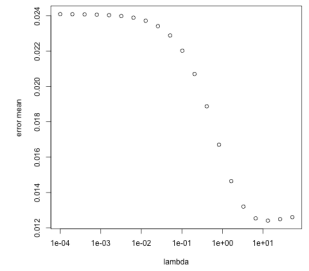


図 4: 実験 2 の結果.

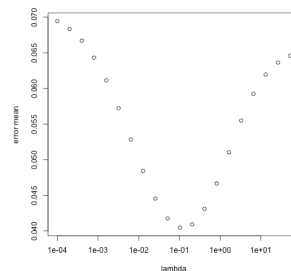


図 5: 実験 5 の結果.

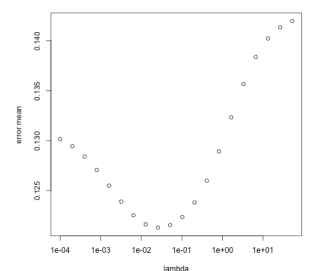


図 6: 実験 8 の結果.

実験の結果, 実験 1,2 を含むグループ 1, 実験 5,8 を含むグループ 2 の 2 グループで違った傾向が見られた. グループ 1 に関しては,  $\lambda$  が大きいほど誤差が小さくなる傾向があり, 一定値に収束しているように見えた. これに対してグループ 2 に関しては, 観測した範囲に誤差が最小となる点が見られた. これらの 2 つのグループ間の設定上の相違点としては, 調整に用いた行列  $A$  に負の成分が含まれるかどうかという点があった.

## 参考文献

- [1] M. Idel. A review of matrix scaling and sinkhorn's normal form for matrices and positive maps, 2016.
- [2] A.W. Marshall and I. Olkin. Scaling of matrices to achieve specified row and column sums. *Numerische Mathematik*, Vol. 12, pp. 83–90, 1968.
- [3] T. Sei. An objective general index for multivariate ordered data. *Journal of Multivariate Analysis*, Vol. 147, pp. 247–264, 2016.
- [4] 寺島実郎. 全 47 都道府県幸福度ランキング 2016 年版. 東洋経済新報社, 2016.