

客観的総合指数の漸近的性質

東京大学・情報理工学系研究科 清 智也

Tomonari Sei

Graduate School of Information Science and Technology,
The University of Tokyo.

概要

客観的総合指数とは複数の量的変数を重み付けて得られる総合指数の一つであり、各変数との共分散が正となるように作られたものである。本稿では、客観的総合指数をデータから推定する場合の漸近的性質を考察する。特に、標準的な推定量を漸近的に優越する推定量を一つ与える。

1 はじめに

優劣を表すような確率変数 $X_1, \dots, X_p$ に対し、その重み付き和

$$G = \sum_{i=1}^p w_i X_i \quad (1)$$

を総合指数として用いる場面は応用上しばしば現れる。ここで $w_1, \dots, w_p$ は正の実数である。たとえば学生の成績評価において英語、数学、理科の成績から総合点を算出したい場合は $p = 3$ であり、 X_1, X_2, X_3 は各科目の点数、 w_1, w_2, w_3 は各科目に対応する重み係数ということになる。

著者が [4] で提案した OGI (客観的総合指数, Objective General Index) は、 $(X_1, \dots, X_p)$ の共分散行列 Σ に対して、方程式

$$\Sigma \mathbf{w} = \frac{1}{\mathbf{w}}, \quad \mathbf{w} \in \mathbb{R}_+^p, \quad (2)$$

の解 $\mathbf{w}$ を重み係数とするという手法である。ここで $\mathbb{R}_+$ は正の実数全体、 $\mathbf{1}$ は成分が 1 のみからなるベクトルを表し、 $\mathbf{1}/\mathbf{w}$ は成分ごとの商を表すものとする。Marshall and Olkin [3] の対角スケーリング定理により、行列 Σ が正

定値であれば、式 (2) の解は一意的に存在する¹。行列の対角スケーリングについては文献 [1] に詳しい解説がある。

本稿では、標本から w を推定することを考え、推定量の性質を議論する。基本的な推定量として、式 (2) における Σ を標本共分散行列 S に置き換えることが考えられる。こうして得られる推定量を標本 OGI と呼ぶ (2 節)。漸近理論の一般論から、標本 OGI は $n \rightarrow \infty$ のもとで一致性や漸近正規性を持つ。しかし、 n が十分に大きくない場合には標本 OGI に代わる推定量としてベイズ推定量や正則化推定量を用いることも考えられる。以下では主に p を固定した $n \rightarrow \infty$ のもとでの漸近理論、つまり大標本理論を展開するが、一致性については $p \rightarrow \infty$ の場合も考慮する。

構成は以下の通りである。2 節では問題設定を述べる。標本 OGI の一致性と漸近正規性はそれぞれ 3 節と 4 節で扱う。5 節では推定問題を統計的決定問題として定式化した後、標本 OGI を漸近的に優越する推定量を構成する。6 節では経験ベイズ推定量を提案する。7 節で各推定量のリスクを数値的に比較し、最後に 8 節で今後の課題についてまとめる。

2 問題設定

2.1 モデル

Σ を正定値行列とする。 $\mathbf{x}_{(1)}, \dots, \mathbf{x}_{(n)} \in \mathbb{R}^p$ は平均ゼロ、共分散行列 Σ の多変量正規分布 $N_p(\mathbf{0}, \Sigma)$ に従うランダム標本とする。このデータから、式 (2) によって定まる w を推定したい。

標本共分散行列を $S = (1/n) \sum_{t=1}^n \mathbf{x}_{(t)} \mathbf{x}_{(t)}'$ とおく。 w のもっとも自然な推定量は、式 (2) において Σ を S に置き換えて得られる推定量であり、

$$S\hat{v} = \frac{1}{\hat{v}}, \quad \hat{v} \in \mathbb{R}_+^p, \quad (3)$$

を満たす $\hat{v}$ である。これを標本 OGI と呼ぶ²。標本 OGI の漸近的な性質は 3 節と 4 節で明らかにする。

¹正定値でなくとも半正定値かつ狭義共正値であれば一意的に存在する。

²正確には標本 OGI の重み係数と呼ぶべきだが、煩雑になるので単に標本 OGI と呼ぶ。

本稿では、 w の推定量として S のみに依存するものを考える。正規性のもとで S は Σ の十分統計量であるからこの制約は不自然ではない。

Σ には w 以外の情報が含まれるが、これは局外母数として扱う。具体的には相関行列を局外母数として選ぶことができる。すなわち、 Σ に対応する相関行列を R とおくと、 w と R から Σ を復元することができる：

$$\Sigma = D_w^{-1} D_v R D_v D_w^{-1},$$

ここで v は $Rv = 1/v$ を満たす正のベクトルであり、 D_v は v を対角成分とする対角行列を表す。この Σ と (w, R) の対応は一つ一つである。

別の局外母数の選び方として、 $B = D_w \Sigma D_w$ がある。この行列は $B1 = 1$ を満たす。明らかに $\Sigma = D_w^{-1} B D_w^{-1}$ が成り立つ。

例 1. $p = 2$ とし、共分散行列を

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

と表す。このとき式 (2) を満たす w は

$$w = \left(\frac{1}{\sigma_1(1+\rho)}, \frac{1}{\sigma_2(1+\rho)} \right)'$$

であることが確かめられる。特に w と相関係数 ρ から Σ を復元できることが分かる。また

$$B = D_w \Sigma D_w = \begin{pmatrix} \frac{1}{1+\rho} & \frac{\rho}{1+\rho} \\ \frac{\rho}{1+\rho} & \frac{1}{1+\rho} \end{pmatrix}$$

を局外母数と思ってもよい。

2.2 推定量の共変性

ここで後の議論を見通し良くするため、推定量の共変性を定義しておく。正の対角行列 D に対し、 $X \sim N_p(0, \Sigma)$ ならば $DX \sim N_p(0, D\Sigma D)$ となり、これは群の作用になっている。また、この変換に対応して真の w は $D^{-1}w$ に変換される：

$$\Sigma w = \frac{1}{w} \quad \Rightarrow \quad (D\Sigma D)(D^{-1}w) = \frac{1}{D^{-1}w}.$$

そこで、推定量 $\hat{w} = \hat{w}(S)$ が $\hat{w}(DSD) = D^{-1}\hat{w}(S)$ を満たすとき、これを共変であるという。たとえば標本 OGI は共変性を満たす。共変推定量については、 $\Sigma 1 = 1$ を満たす Σ に対してのみ性質を調べておけば、他の場合も直ちに結果が得られる。先に述べた通り、 $\Sigma 1 = 1$ を満たす Σ は局外母数と考えることができる。

3 標本 OGI の一貫性

本節では 2 節で定義した標本 OGI の一貫性を調べる。標本 OGI を $\hat{v}$ と表記する。共変性により、 $\Sigma 1 = 1$ の場合だけ議論すれば十分である。このとき真の OGI は $w = 1$ であることに注意しよう。

推定量の一貫性は、 $\|\hat{w} - 1\|$ が 0 に確率収束することとして定義する。ただし $\|\cdot\|$ はユークリッドノルムを表す。以下、 Σ の成分を σ_{ij} と表し、 S の成分を S_{ij} と表す。 p は n に依存してよく、そのことを強調する場合は $p = p_n$ と書く。なお $p \rightarrow \infty$ の場合は、より弱い一貫性として $\hat{w}$ と 1 のなす角度の収束性が考えられる。これについては 8 節で補足する。

次の補題は基本的であり、以後しばしば用いられる。

補題 1. 正規性の仮定のもとで等式

$$\text{Cov}[S_{ij}, S_{kl}] = \frac{1}{n}(\sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk})$$

が成り立つ。

この補題をベースにして次の定理が導かれる。定理の証明は付録に記す。

定理 1. $\Sigma 1 = 1$ とし、

$$\lim_{n \rightarrow \infty} \frac{p_n}{n} \{\text{tr}(\Sigma) + 1\} = 0 \quad (4)$$

とする。このとき

$$\|\hat{v} - 1\| \xrightarrow{P} 0$$

が成り立つ。つまり標本 OGI は一貫性を持つ。

なお、条件 $\Sigma \mathbf{1} = \mathbf{1}$ より Σ は固有値 1 を持つので、式 (4) は

$$\lim_{n \rightarrow \infty} \frac{p_n}{n} \text{tr}(\Sigma) = 0$$

と書いても同じことである。

4 標本 OGI の漸近正規性と漸近展開

本節では古典的設定、つまり p は固定し $n \rightarrow \infty$ とした下で、標本 OGI の漸近正規性を確認し、形式的な漸近展開を導出する。この設定では一致性に関する定理 1 の十分条件は常に満たされる。本節の結果は次節で推定量のリスクの評価をする際に用いられる。これまでと同様、標本 OGI を $\hat{v}$ と記す。

4.1 漸近正規性

補題 2. $\Sigma \mathbf{1} = \mathbf{1}$ と仮定する。このとき

$$\sqrt{n}(\hat{v} - \mathbf{1}) \xrightarrow{\text{law}} N_p(\mathbf{0}, (\Sigma + \mathbf{I})^{-1}(p\Sigma + \mathbf{1}\mathbf{1}')(\Sigma + \mathbf{I})^{-1})$$

が成り立つ。ここで $\mathbf{I}$ は単位行列を表す。

Proof. デルタ法による。 $\hat{v} = \mathbf{1} + \tilde{v}/\sqrt{n}$ とおき方程式 (3) を展開すると

$$(\mathbf{I} + \mathbf{S})\tilde{v} = -\sqrt{n}(\mathbf{S}\mathbf{1} - \mathbf{1}) + o_p(1)$$

となる。中心極限定理と補題 1 から結果が従う。 $\square$

漸近共分散を上から評価してみよう。 $\mathbf{e} = \mathbf{1}/\sqrt{p}$ とおき、 $\mathbf{H} = \Sigma - \mathbf{e}\mathbf{e}'$ とおく。また $\mathbf{Q} = \mathbf{I} - \mathbf{e}\mathbf{e}'$ とおく。このとき仮定から $\mathbf{H}\mathbf{1} = \mathbf{Q}\mathbf{1} = \mathbf{0}$ が成り立つ。また、

$$(\Sigma + \mathbf{I})^{-1} = (2\mathbf{e}\mathbf{e}' + \mathbf{H} + \mathbf{Q})^{-1} = \frac{1}{2}\mathbf{e}\mathbf{e}' + (\mathbf{H} + \mathbf{Q})^{-},$$

となる。ただし $\mathbf{A}^-$ は $\mathbf{A}$ の Moore–Penrose 一般逆行列を表す。したがって

$$\begin{aligned} \mathbf{V} &:= (\Sigma + \mathbf{I})^{-1}(p\Sigma + \mathbf{1}\mathbf{1}')(\Sigma + \mathbf{I})^{-1} \\ &= \frac{p}{2}\mathbf{e}\mathbf{e}' + p(\mathbf{H} + \mathbf{Q})^{-}\mathbf{H}(\mathbf{H} + \mathbf{Q})^{-} \\ &\leq \frac{p}{2}\mathbf{e}\mathbf{e}' + \frac{p}{4}\mathbf{Q} \end{aligned}$$

となる。最後の不等式は、 $\mathbf{H}$ と $\mathbf{Q}$ が同時対角化可能であること、および任意の $h > 0$ に対して

$$\frac{h}{(1+h)^2} \leq \frac{1}{4}$$

が成り立つことによる。等号条件は $h = 1$ である。

この評価から、真の Σ に依存しない保守的な信頼領域を構成することが可能となる。たとえば $\mathbf{w}$ の第 i 成分 w_i の保守的な 95% 信頼区間は

$$\hat{v}_i \pm 1.96 \sqrt{\frac{p+1}{4n}} \hat{v}_i$$

となる。

4.2 漸近展開

ここでは形式的に標本 OGI を漸近展開する。

補題 3. $\Sigma \mathbf{1} = \mathbf{1}$ と仮定し、 $\tilde{\mathbf{S}} = \sqrt{n}(\mathbf{S} - \Sigma)$ とおく。このとき次の展開が成り立つ：

$$\hat{\mathbf{v}} = \mathbf{1} + \frac{\mathbf{r}}{\sqrt{n}} + \frac{\mathbf{u}}{n} + o_p(1/n). \quad (5)$$

ただし

$$\begin{aligned} \mathbf{r} &= -(\Sigma + \mathbf{I})^{-1} \tilde{\mathbf{S}} \mathbf{1}, \\ \mathbf{u} &= -(\Sigma + \mathbf{I})^{-1} (\tilde{\mathbf{S}} \mathbf{r} - \mathbf{r}^2) \end{aligned}$$

であり、 $\mathbf{r}^2 = (r_i^2)_{i=1}^p$ とする。

Proof. 方程式 $\mathbf{S} \hat{\mathbf{v}} = \mathbf{1}/\hat{v}$ に式 (5) と $\mathbf{S} = \Sigma + \tilde{\mathbf{S}}/\sqrt{n}$ を代入すると、

$$\begin{aligned} \mathbf{0} &= \mathbf{S} \hat{\mathbf{v}} - \frac{\mathbf{1}}{\hat{v}} \\ &= \frac{1}{\sqrt{n}} ((\Sigma + \mathbf{I}) \mathbf{r} + \tilde{\mathbf{S}} \mathbf{1}) + \frac{1}{n} ((\Sigma + \mathbf{I}) \mathbf{u} + \tilde{\mathbf{S}} \mathbf{r} - (\mathbf{r}^2)) + o_p(1/n) \end{aligned}$$

を得る。この両辺を比較すればよい。 □

補題 3 からバイアスの漸近展開が得られる。

補題 4. $\Sigma \mathbf{1} = \mathbf{1}$ とする。 $\hat{v}$ のバイアスは次のように展開される。

$$\begin{aligned} E[\hat{v}] - \mathbf{1} &= \frac{1}{n} E[\mathbf{u}] + o\left(\frac{1}{n}\right), \\ E[\mathbf{u}] &= \left(\frac{3}{8} + \frac{1}{2} \text{tr}(\Sigma(\Sigma + \mathbf{I})^{-1})\right) \mathbf{1} + p(\Sigma + \mathbf{I})^{-1} \text{diag}((\Sigma + \mathbf{I})^{-1} \Sigma (\Sigma + \mathbf{I})^{-1}). \end{aligned}$$

ここで $\text{diag}(\mathbf{A})$ は行列 $\mathbf{A}$ の対角要素からなるベクトルを表す。

Proof. $E[r_i] = 0$ は明らかである。あとは $E[\tilde{S}_{ij}\tilde{S}_{kl}] = \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}$ を利用して計算する。仮定 $\Sigma \mathbf{1} = \mathbf{1}$ より $(\Sigma + \mathbf{I})^{-1} \mathbf{1} = \frac{1}{2} \mathbf{1}$ となることに注意する。□

5 統計的決定問題

本節では OGI の推定問題を統計的決定問題として定式化し、標本 OGI を漸近的に優越する推定量を一つ構成する。共分散行列そのものの推定に関する決定理論については例えば [5] の 11 章を参照されたい。

5.1 損失関数

損失関数 $L(\Sigma, \hat{w})$ としては不変性

$$L(D\Sigma D, D^{-1}\hat{w}) = L(\Sigma, \hat{w})$$

を満たすものを考えるのが自然であろう。ここで D は任意の対角行列を表す。

例えば、次の損失関数は不変性を満たす：

$$L(\Sigma, \hat{w}) = - \sum_i \log \frac{\hat{w}_i}{w_i} + \frac{1}{2} \hat{w}' \Sigma \hat{w} - \frac{p}{2}. \quad (6)$$

ここで w は真の OGI である³。この損失関数は $\hat{w}$ について凸であり、かつ $\hat{w} = w$ で最小値ゼロを達成する。この損失関数を用いる利点の一つは、ベ

³実際には $\log w_i$ の項は推定量の比較には寄与しない。

イズ推定量が容易に求められる点である。すなわち、 Σ の事後期待値を $\bar{\Sigma}$ とおけば、ベイズ推定量は

$$\bar{\Sigma}\hat{w} = \frac{1}{\hat{w}} \quad (7)$$

の解となる。また式 (6) は $N_p(\mathbf{0}, \Sigma)$ から $N_p(\mathbf{0}, D_{\hat{w}}^{-1}A^{-1}D_{\hat{w}}^{-1})$ への Kullback-Leibler ダイバージェンスにおいて A を $\mathbf{1}\mathbf{1}'$ に近づけたときの極限として解釈することも可能である。

他にも、真値との比の二乗誤差

$$L_2(\Sigma, \hat{w}) = \left\| \frac{\hat{w}}{w} - 1 \right\|^2 \quad (8)$$

や、方向のみの誤差を測る損失関数

$$L_*(\Sigma, \hat{w}) = \min_{c>0} L(\Sigma, c\hat{w}) = - \sum_i \log \frac{\hat{w}_i}{w_i} + \frac{p}{2} \log \left(\frac{\hat{w}'\Sigma\hat{w}}{p} \right), \quad (9)$$

$$L_{2*}(\Sigma, \hat{w}) = \min_{c>0} L_2(\Sigma, c\hat{w}) = p - \frac{((\hat{w}/w)' \mathbf{1})^2}{\|\hat{w}/w\|^2} \quad (10)$$

が考えられる。いずれも不変性を満たす。

次の補題は $\sqrt{n}$ 一致推定量について、損失の最初のオーダーを求めたものである。導出は容易なので省略する。以下、損失関数を $L(\hat{w}) = L(\Sigma, \hat{w})$ などと略記する。

補題 5. $\Sigma \mathbf{1} = \mathbf{1}$ とする。 $\sqrt{n}$ 一致推定量 $\hat{w} = \mathbf{1} + \tilde{w}/\sqrt{n} + o_p(1/\sqrt{n})$ に対して損失関数 L, L_2, L_*, L_{2*} はそれぞれ

$$\begin{aligned} L(\hat{w}) &= \frac{1}{2n} \tilde{w}'(I + \Sigma)\tilde{w} + o_p(1/n), \\ L_2(\hat{w}) &= \frac{1}{n} \tilde{w}'\tilde{w} + o_p(1/n), \\ L_*(\hat{w}) &= \frac{1}{2n} \tilde{w}'(Q + H)\tilde{w} + o_p(1/n), \\ L_{2*}(\hat{w}) &= \frac{1}{n} \tilde{w}'Q\tilde{w} + o_p(1/n) \end{aligned}$$

と展開される。ただし $Q = I - \mathbf{1}\mathbf{1}'/p$, $H = \Sigma - \mathbf{1}\mathbf{1}'/p$ とおいた。

5.2 標本 OGI のリスク

標本 OGI のリスクを最初のオーダーまで求めておく。リスクは損失関数の期待値 $R(\hat{\mathbf{w}}) = E[L(\hat{\mathbf{w}})]$ として定義される。

補題 6. $\Sigma \mathbf{1} = \mathbf{1}$ とする。損失関数 L, L_2, L_*, L_{2*} に対する標本 OGI $\hat{\mathbf{v}}$ のリスクはそれぞれ

$$\begin{aligned} R(\hat{\mathbf{v}}) &= \frac{p}{2n} \left(\frac{1}{2} + \text{tr}(\Sigma(\mathbf{I} + \Sigma)^{-1}) \right) + o(1/n), \\ R_2(\hat{\mathbf{v}}) &= \frac{p}{n} \left(\frac{1}{4} + \text{tr}(\Sigma(\mathbf{I} + \Sigma)^{-2}) \right) + o(1/n), \\ R_*(\hat{\mathbf{v}}) &= \frac{p}{2n} \text{tr}(\mathbf{H}(\mathbf{Q} + \mathbf{H})^{-1}) + o(1/n), \\ R_{2*}(\hat{\mathbf{v}}) &= \frac{p}{n} \text{tr}(\mathbf{H}(\mathbf{Q} + \mathbf{H})^{-2}) + o(1/n) \end{aligned}$$

と展開される。ただし $\mathbf{Q} = \mathbf{I} - \mathbf{1}\mathbf{1}'/p$, $\mathbf{H} = \Sigma - \mathbf{1}\mathbf{1}'/p$ とし、逆行列は Moore–Penrose 一般逆行列と解釈するものとする。

Proof. 補題 5 の結果に対し、補題 2 の漸近分散を用いて形式的に期待値を計算する。その際 $(\mathbf{I} + \Sigma)^{-1}\mathbf{1} = \frac{1}{2}\mathbf{1}$ などに注意する。 $\square$

5.3 標本 OGI を優越する推定量

ここでは式 (6) の L を損失関数として採用し、標本 OGI を漸近的に優越する推定量を示そう。

標本 OGI に修正を加えた次の形の推定量を考える：

$$\hat{\mathbf{w}} = \hat{\mathbf{v}} + \frac{\hat{\mathbf{b}}}{n} + O_p(1/n^2). \quad (11)$$

ここで $\hat{\mathbf{b}} = \hat{\mathbf{b}}(\mathbf{S})$ は $\mathbf{S}$ に依存してよい。 $\hat{\mathbf{w}}$ が共変であるためには任意の対角行列 $\mathbf{D}$ に対して

$$\hat{\mathbf{b}}(\mathbf{DSD}) = \mathbf{D}\hat{\mathbf{b}}(\mathbf{S})$$

が成り立つ必要がある。

リスクの差の漸近形は次の定理で与えられる。証明は付録に示す。

定理 2. $\Sigma \mathbf{1} = \mathbf{1}$ とする。 $\hat{\mathbf{w}}$ と $\hat{\mathbf{v}}$ のリスクの差は次のように展開される：

$$R(\hat{\mathbf{w}}) - R(\hat{\mathbf{v}}) = \frac{1}{n^2} \left\{ \frac{1}{2} \mathbf{b}'(\mathbf{I} + \Sigma) \mathbf{b} + \mu \mathbf{1}' \mathbf{b} - \sum_{i,j,k} (\partial_{ij} b_k) A_{ijk} \right\} + o(1/n^2).$$

ここで $\mathbf{b} = \hat{\mathbf{b}}(\Sigma)$, $\partial_{ij} b_k = (\partial_{ij} \hat{b}_k)(\Sigma)$ であり,

$$\mu = \frac{1}{2} + \text{tr}(\Sigma(\mathbf{I} + \Sigma)^{-1}), \quad A_{ijk} = \sigma_{ik} 1_j + \sigma_{jk} 1_i$$

とする。また微分 $\partial_{ij} \hat{b}_k$ は、 $\mathbf{S}$ の摂動 $\dot{\mathbf{S}}$ に対する $\hat{b}_k$ の摂動が $\sum_{i,j} \dot{S}_{ij} (\partial_{ij} \hat{b}_k)$ となるように定める。

定理 2 の結果から、修正項 $\hat{\mathbf{b}}$ の良し悪しを論ずることができる。一般の場合は難しいので、関数 $\hat{\mathbf{b}}$ の形を限定して、そこに含まれるパラメータを調整することを考えよう。ここでは標本 OGI のスカラー倍として与えられる推定量のうち、漸近的に標本 OGI を優越する推定量を構築する。

まず、推定量の形を探るために次の補題を示す。

補題 7. $\hat{\mathbf{w}} = (1 + c/n)\hat{\mathbf{v}}$ とし、 $c \in \mathbb{R}$ は $\mathbf{S}$ によらない定数とする。このとき $R(\hat{\mathbf{w}})$ を漸近的な意味で最小にする c は $-\mu$ で与えられる。ただし μ は定理 2 で定義した量である。

Proof. 定理 2 において $\hat{\mathbf{b}}(\mathbf{S}) = c\hat{\mathbf{v}}$ とおき、 $c \in \mathbb{R}$ について最小化すればよい。標本 OGI の定義と陰関数定理から

$$\partial_{ij} v_k = -(\Sigma + \mathbf{I})_{ik}^{-1} 1_j$$

が成り立つ（添字 i, j は対称化すべきだが A_{ijk} と縮約を取るので問題ない）。したがって定理 2 から

$$R(\hat{\mathbf{w}}) - R(\hat{\mathbf{v}}) = \frac{1}{n^2} \{ pc^2 + p\mu c + p\mu c \} + o(1/n^2)$$

となり、 $c = -\mu$ のとき最小となる。 □

この補題で求めた $c = -\mu = -\frac{1}{2} - \text{tr}(\Sigma(\Sigma + \mathbf{I})^{-1})$ には真の Σ が含まれるため、 Σ に標本共分散 $\mathbf{S}$ をプラグインした推定量を考える：

$$\hat{\mathbf{w}} = \frac{\hat{\mathbf{v}}}{1 + \frac{1}{n}(\frac{1}{2} + \text{tr}(\mathbf{S}(\mathbf{S} + D_{\hat{\mathbf{v}}}^{-2})^{-1})}.$$

ただし推定量の正值性が保証されるよう、修正項は分母に施した。また共変性が成り立つように単位行列 $\mathbf{I}$ を $D_{\hat{\mathbf{v}}}^{-2}$ に置き換えている。この推定量をスケール修正推定量と呼ぶことにしよう。

次の定理に示すようにスケール修正推定量は標本 OGI を漸近的に優越する。証明は付録に示す。

定理 3. スケール修正推定量と標本 OGI のリスクの差は

$$\lim_{n \rightarrow \infty} n^2 \{R(\hat{\mathbf{w}}) - R(\hat{\mathbf{v}})\} = \frac{p}{2} - 2p \text{tr}(\Sigma(\Sigma + \mathbf{I})^{-2}) - p \left(\frac{1}{2} + \text{tr}(\Sigma(\Sigma + \mathbf{I})^{-1}) \right)^2 \leq -p$$

となる。最後の不等式の等号は $\Sigma = \mathbf{1}\mathbf{1}'/p$ のときに成立する。

なお、スケール修正推定量は式 (9), (10) のように方向を測る損失に対しては無力であり、さらなる優越推定量があるかどうかは理論的に興味深い問題である。

6 経験ベイズ推定量

ここでは共変性を持つような経験ベイズ推定量を一つ与え、さらにそれを近似することで簡便な推定量を導出する。

まず $\mathbf{S}$ の尤度は

$$p(\mathbf{S}|\Sigma) \propto \frac{|\mathbf{S}|^{(n-p-1)/2}}{|\Sigma|^{n/2}} \text{etr}\left(-\frac{n}{2}\Sigma^{-1}\mathbf{S}\right)$$

と書ける。これに対し、 Σ の事前分布のクラスとして逆 Wishart 分布

$$p(\Sigma|\mathbf{T}, \alpha) \propto \frac{|\mathbf{T}|^{\alpha/2}}{|\Sigma|^{(\alpha+p+1)/2}} \text{etr}\left(-\frac{\alpha}{2}\Sigma^{-1}\mathbf{T}\right)$$

を考える。ただし $\mathbf{T}$ は正の対角行列であり、 α は正の実数である。後で経験ベイズ推定量を定義する際、 α は固定し、 $\mathbf{T}$ のみデータから推定する。そのため事前分布で $\mathbf{T}$ に依存する因子は明示し、 α のみに依存する因子は省略した。

このとき事後分布は

$$p(\Sigma|\mathbf{S}, \mathbf{T}, \alpha) \propto \frac{|\mathbf{nS} + \alpha\mathbf{T}|^{(n+\alpha)/2}}{|\Sigma|^{(n+\alpha+p+1)/2}} \text{etr}\left(-\frac{1}{2}\Sigma^{-1}(\mathbf{nS} + \alpha\mathbf{T})\right)$$

となる。式 (7) のベイズ推定量の計算に必要な事後期待値は

$$\bar{\Sigma} = \frac{1}{n + \alpha - p - 1}(\mathbf{nS} + \alpha\mathbf{T}) \quad (12)$$

となる（例えば [5] の定理 4.7）。

周辺尤度は

$$\begin{aligned} p(\mathbf{S}|\mathbf{T}, \alpha) &= \int p(\mathbf{S}|\Sigma)p(\Sigma|\mathbf{T}, \alpha)d\Sigma \\ &\propto \int \frac{|\mathbf{S}|^{(n-p-1)/2}|\mathbf{T}|^{\alpha/2}}{|\Sigma|^{(n+\alpha+p+1)/2}} \text{etr}\left(-\frac{1}{2}\Sigma^{-1}(\mathbf{nS} + \alpha\mathbf{T})\right)d\Sigma \\ &\propto \frac{|\mathbf{S}|^{(n-p-1)/2}|\mathbf{T}|^{\alpha/2}}{|\mathbf{nS} + \alpha\mathbf{T}|^{(n+\alpha)/2}} \end{aligned}$$

となる。 $\mathbf{T}$ を対角行列として、 $p(\mathbf{S}|\mathbf{T}, \alpha)$ を $\mathbf{T}$ について最大化することにより、 $\mathbf{T}$ の推定量が得られ、結果として経験ベイズ推定量が得られる。この経験ベイズ推定量は 2 節で述べた意味で不変となる。

さて $\mathbf{T} = \text{diag}(t_1, \dots, t_d)$ に関する周辺尤度の停留条件は

$$\begin{aligned} &\frac{d}{dt_i} (\alpha \log |\mathbf{T}| - (n + \alpha) \log |\mathbf{nS} + \alpha\mathbf{T}|) \\ &= \alpha \left(\frac{1}{t_i} - (n + \alpha)(\mathbf{nS} + \alpha\mathbf{T})_{ii}^{-1} \right) = 0 \end{aligned} \quad (13)$$

となる。特に $\alpha \rightarrow 0$ のとき $t_i \rightarrow 1/(\mathbf{S}^{-1})_{ii}$ となるから、 $\mathbf{T} = (\text{diag}(\mathbf{S}^{-1}))^{-1}$ は (13) の近似解と見なせる。これを式 (12) に代入すれば近似的な経験ベイズ推定量が得られる。

7 数値実験

5 節で定義した 4 つの損失関数 L, L_2, L_*, L_{2*} に対し、標本 OGI (2 節)、スケール修正推定量 (5 節)、経験ベイズ推定量とその近似 (6 節) のリスクをモンテカルロ法により計算する。

真の共分散行列は

$$\Sigma = D \begin{pmatrix} 1 & -r & \cdots & \cdots & -r \\ & 1 & r & \cdots & r \\ & & \ddots & \ddots & \vdots \\ * & & & \ddots & r \\ & & & & 1 \end{pmatrix} D, \quad -\frac{1}{p-1} < r < 1, \quad (14)$$

とする。ここで対角行列 D は $\Sigma \mathbf{1} = \mathbf{1}$ を満たすようなスケーリング行列であり、陽に表すと $\kappa = (p-1)r$, $\lambda = 1 + (p-2)r$, $A = \lambda^2 - \kappa\lambda r$, $B = 2\lambda - \kappa r + r^2$ として

$$D = \text{diag}(\alpha, \beta, \dots, \beta), \quad \alpha = \frac{\kappa\beta + \sqrt{\kappa^2\beta^2 + 4}}{2}, \quad \beta = \left\{ \frac{B \pm \sqrt{B^2 - 4A}}{2A} \right\}^{1/2}$$

となる。ただし複号は $r > 0$ のとき $+$ とし、 $r < 0$ のとき $-$ とする。

図 1 は $n = 20$, $p = 10$ とし、各損失関数、各推定量のもとでリスクを計算した結果である。横軸は Σ に含まれるパラメータ r であり、区間 $[-1/p, 0.99]$ の等間隔な 30 個の点で計算を行った。また経験ベイズ推定量に含まれる定数 α は $\alpha = p$ とした。モンテカルロ法の実験回数は 10^4 回とした。

図 1 より、スケール修正推定量は損失関数 L と L_2 に対して標本 OGI を優越している。しかし L_* と L_{2*} に対しては効果がないことが分かる。また経験ベイズ推定量は L, L_2, L_* において標本 OGI を優越しており、近似経験ベイズ推定量も同等の性質を有している。

以上の計算に要した時間を比較すると表 1 のようになった。特に経験ベイズ推定量は周辺尤度の最大化に時間がかかっている。

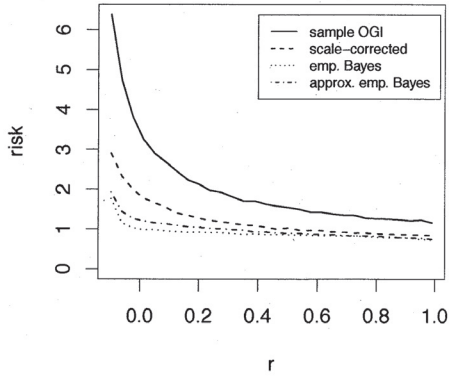
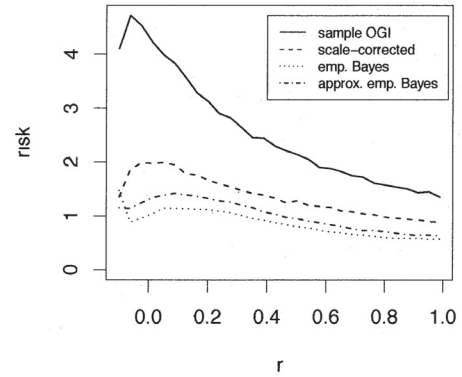
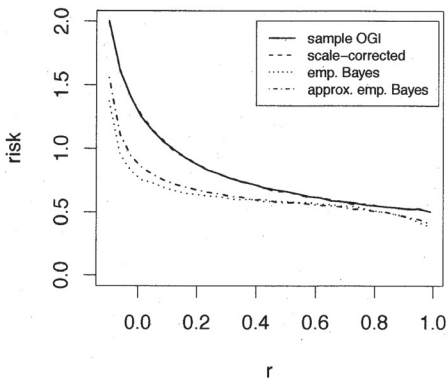
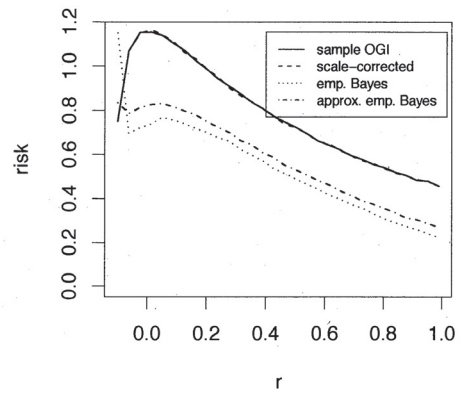
(a) 損失関数が L の場合。(b) 損失関数が L_2 の場合。(c) 損失関数が L_* の場合。(d) 損失関数が L_{2*} の場合。

図 1: 各損失関数における各推定量のリスク。真の共分散は式 (14) であり, $n = 20, p = 10$ とした。横軸は共分散に含まれるパラメータ r である。

表 1: 計算時間の比較 (単位は秒)。

推定量	標本 OGI	スケール修正	経験ベイズ	近似経験ベイズ
計算時間	118	139	648	133

8 今後の研究課題

3 節では変数の次元 p が標本サイズ n とともに大きくなる状況での一致性を考察した。しかしその条件は $p/n \rightarrow 0$ という条件を含んでおり、古典的な漸近論の範疇に収まってしまっている。これに対し、 $p/n \rightarrow c \in (0, 1)$ という設定で数値実験を行うと、 $\hat{w}/w$ と 1 のなす角度の意味でも不一致性が観察される。これは固有ベクトルの推定問題における不一致性（例えば [2] など）に類似しており、理論的な検討は興味深い研究課題である。

5 節では、スケール修正推定量が標本 OGI を漸近的に優越するという結果を定理 3 で示した。しかし総合指数という観点で言えば、スケールの変換はランキングを変えない変換であり、実用上の意味に乏しい。スケール修正以外の優越推定量を示すことも今後の課題である。

6 節で論じた経験ベイズ推定では、ハイパーパラメータ α を固定していた。実用上は α を交差検証法で選択すればよいと考えられるが、 α も含めて周辺尤度最大化を行った場合の性能も検討すべきである。また、 α を固定した場合を含め周辺最尤推定値が一意に存在するかどうか、さらに $p > n$ のときでも安定な推定量を構成できるか、などの問題が残されている。

7 節の数値実験では経験ベイズ推定量が比較的良好な性能を持つことが確認された。しかし限られた設定での実験しか行っていないため、一般的に成り立つかどうかは不明である。漸近リスクの比較が今後の課題である。

謝辞

本研究は科研費（課題番号：26108003, 17K00044）の助成を受けたものである。

A 各定理の証明

A.1 定理 1 の証明

補題 8. 関数 ϕ を

$$\phi(z) = -\log(1+z) + z$$

と定義する。このとき

$$\phi(z) \geq \begin{cases} z^2/8 & \text{if } -1 < z < 1, \\ z/8 & \text{if } z \geq 1 \end{cases}$$

が成り立つ。

Proof. テイラーの定理より, $|z_*| \leq |z|$ を満たす z_* が存在して

$$\phi(z) = \frac{z^2}{2(1+z_*)^2}$$

が成り立つ。よって $-1 < z < 1$ において $\phi(z) \geq z^2/8$ となる。また ϕ の凸性と $\phi(0) = 0$ から, $z \geq 1$ に対して

$$\phi(z) \geq \phi(1)z \geq \frac{z}{8}$$

が成り立つ。 □

$\hat{\mathbf{v}} = \mathbf{1} + \mathbf{r}$ とおき, $\|\mathbf{r}\|$ が 0 に確率収束することを示す。 $\boldsymbol{\eta} = \mathbf{S}\mathbf{1} - \mathbf{1}$ とおく。補題 1 と仮定より,

$$E[\|\boldsymbol{\eta}\|^2] = \text{tr}(\text{Cov}[\mathbf{S}\mathbf{1}, \mathbf{S}\mathbf{1}]) = \frac{p}{n}(\text{tr}\boldsymbol{\Sigma} + 1) \rightarrow 0$$

となる。特に任意の $\delta > 0$ に対して $P(\|\boldsymbol{\eta}\| \geq \delta) \rightarrow 0$ となる。 $\hat{\mathbf{v}}$ は

$$\psi(\mathbf{w}) := \sum_{i=1}^p (-\log w_i) + \frac{1}{2} \mathbf{w}' \mathbf{S} \mathbf{w}$$

の最小点であるから

$$\Delta := \psi(\mathbf{1} + \mathbf{r}) - \psi(\mathbf{1}) \leq 0$$

となる。この Δ を下から評価する。まず

$$\begin{aligned} \Delta &= \psi(\mathbf{1} + \mathbf{r}) - \psi(\mathbf{1}) \\ &= \sum_i (-\log(1 + r_i)) + \mathbf{1}' \mathbf{S} \mathbf{r} + \frac{1}{2} \mathbf{r}' \mathbf{S} \mathbf{r} \\ &\geq \sum_i (-\log(1 + r_i)) + \mathbf{1}' \mathbf{S} \mathbf{r} \\ &= \sum_i \{-\log(1 + r_i) + r_i(1 + \eta_i)\} \end{aligned}$$

と評価できる。ここで $I = \{i \mid r_i \geq 1\}$ とおけば、補題 8 より

$$\begin{aligned} \Delta &\geq \sum_{i \in I} \left(\frac{r_i}{8} + r_i \eta_i \right) + \sum_{i \notin I} \left(\frac{r_i^2}{8} + r_i \eta_i \right) \\ &\geq \sum_{i \in I} \left(\frac{r_i}{8} + r_i \eta_i \right) - \sum_{i \notin I} 2\eta_i^2 \end{aligned} \quad (15)$$

となる。さらに事象 $\Omega_1 = \{\|\boldsymbol{\eta}\| < 1/16\}$ に制限すれば、 $\eta_i > -1/16$ となるから

$$0 \geq \Delta \geq \sum_{i \in I} \frac{r_i}{16} - 2\|\boldsymbol{\eta}\|^2 \geq \frac{|I|}{16} - 2\|\boldsymbol{\eta}\|^2$$

となる。よって $|I| \leq 32\|\boldsymbol{\eta}\|^2 \leq 1/8 < 1$ となり、 $|I| = 0$ となる。よって式 (15) から

$$0 \geq \Delta \geq \sum_i \left(\frac{r_i^2}{8} + r_i \eta_i \right) \geq \frac{\|\mathbf{r}\|^2}{8} - \|\mathbf{r}\| \|\boldsymbol{\eta}\|$$

より $|\mathbf{r}| \leq 8\|\boldsymbol{\eta}\|$ が得られる。ゆえに任意の $\varepsilon > 0$ に対して

$$P(\|\mathbf{r}\| > \varepsilon) \leq P(\Omega_1^c) + P(8\|\boldsymbol{\eta}\| > \varepsilon) \rightarrow 0$$

となる。これで一致性が示された。

A.2 定理 2 の証明

$\hat{\mathbf{w}} = \hat{\mathbf{v}} + \boldsymbol{\delta}/n$ とおく。 $\boldsymbol{\delta} = \hat{\mathbf{b}} + O_p(1/n)$ である。まず損失関数の差は

$$\begin{aligned} L(\hat{\mathbf{w}}) - L(\hat{\mathbf{v}}) &= - \sum_i \log \left(\hat{v}_i + \frac{\delta_i}{n} \right) + \frac{1}{2} \left(\hat{\mathbf{v}} + \frac{\boldsymbol{\delta}}{n} \right)' \boldsymbol{\Sigma} \left(\hat{\mathbf{v}} + \frac{\boldsymbol{\delta}}{n} \right) + \sum_i \log \hat{v}_i - \frac{1}{2} \hat{\mathbf{v}}' \boldsymbol{\Sigma} \hat{\mathbf{v}} \\ &= - \sum_i \log \left(1 + \frac{\delta_i}{n\hat{v}_i} \right) + \frac{1}{n} \hat{\mathbf{v}}' \boldsymbol{\Sigma} \boldsymbol{\delta} + \frac{1}{2n^2} \boldsymbol{\delta}' \boldsymbol{\Sigma} \boldsymbol{\delta} \\ &= \frac{1}{n} \left(\boldsymbol{\Sigma} \hat{\mathbf{v}} - \frac{1}{\hat{\mathbf{v}}} \right)' \boldsymbol{\delta} + \frac{1}{2n^2} \mathbf{b}' (\mathbf{I} + \boldsymbol{\Sigma}) \mathbf{b} + o_p(1/n^2) \end{aligned}$$

となる。よってリスクの差は

$$R(\hat{\mathbf{w}}) - R(\hat{\mathbf{v}}) = \frac{1}{n} \mathbf{E} \left[\left(\boldsymbol{\Sigma} \hat{\mathbf{v}} - \frac{1}{\hat{\mathbf{v}}} \right)' \boldsymbol{\delta} \right] + \frac{1}{2n^2} \mathbf{b}' (\mathbf{I} + \boldsymbol{\Sigma}) \mathbf{b} + o(1/n^2)$$

となる。さて、補題 3 より $\hat{\mathbf{v}} = \mathbf{1} + \frac{\mathbf{r}}{\sqrt{n}} + \frac{\mathbf{u}}{n} + o(1/n)$, $E[\mathbf{r}] = 0$ とおけば

$$\begin{aligned}\Sigma\hat{\mathbf{v}} - \frac{1}{\hat{\mathbf{v}}} &= \left(\Sigma\mathbf{1} + \frac{\Sigma\mathbf{r}}{\sqrt{n}} + \frac{\Sigma\mathbf{u}}{n}\right) - \left(\mathbf{1} - \frac{\mathbf{r}}{\sqrt{n}} - \frac{\mathbf{u}}{n} + \frac{\mathbf{r}^2}{n}\right) + o_p(1/n) \\ &= \frac{1}{\sqrt{n}}(\mathbf{I} + \Sigma)\mathbf{r} + \frac{1}{n}((\mathbf{I} + \Sigma)\mathbf{u} - \mathbf{r}^2) + o_p(1/n)\end{aligned}$$

となり、また

$$\boldsymbol{\delta} = \mathbf{b} + \frac{1}{\sqrt{n}}(\tilde{S}_{jk}\partial_{jk}\mathbf{b}) + O_p(1/n)$$

となる。ただし $\mathbf{S} = \Sigma + \tilde{\mathbf{S}}/\sqrt{n}$ とおき、同じ添字が 2 回現れたら和を取る規則を用いた。したがって

$$\begin{aligned}E\left[\left(\Sigma\hat{\mathbf{v}} - \frac{1}{\hat{\mathbf{v}}}\right)'\boldsymbol{\delta}\right] &= \frac{1}{n}E[\mathbf{r}'(\mathbf{I} + \Sigma)(\partial_{ij}\mathbf{b})\tilde{S}_{ij}] + \frac{1}{n}((\mathbf{I} + \Sigma)E[\mathbf{u}] - E[\mathbf{r}^2])'\mathbf{b} + o(1/n) \\ &= \frac{1}{n}(-(\partial_{ij}b_k)E[\tilde{S}_{ij}\tilde{S}_{kl}1_l]) + \frac{1}{n}((\mathbf{I} + \Sigma)_{jk}^{-1}E[\tilde{S}_{ij}\tilde{S}_{kl}1_l]b_i) + o(1/n) \\ &= \frac{1}{n}(-(\partial_{ij}b_k)A_{ijk} + \mu 1_i b_i)\end{aligned}$$

となる。最後の等号は補題 1 による。

A.3 定理 3 の証明

以下、同じ添字が 2 回現れたら和を取る規則を用いる。まず

$$\hat{\mathbf{b}}(\mathbf{S}) = (-1/2 - \text{tr}(\mathbf{S}(\mathbf{S} + \mathbf{D}_{\hat{\mathbf{v}}}^{-2})^{-1}))\hat{\mathbf{v}}$$

として $\partial_{ij}b_k$ を求める。補題 7 の証明で述べた通り

$$\partial_{ij}v_k = -(\Sigma + \mathbf{I})_{ik}^{-1}1_j$$

が成り立つ (添字 i, j は対称化すべきだが A_{ijk} と縮約するので問題ない)。これを用いて

$$\begin{aligned}\partial_{ij}b_k &= \partial_{ij}\{(-1/2 - \text{tr}(\mathbf{S}(\mathbf{S} + \mathbf{D}_{\hat{\mathbf{v}}}^{-2})^{-1}))\hat{\mathbf{v}}_k\} \\ &= -(\Sigma + \mathbf{I})_{ij}^{-1}1_k + ((\Sigma + \mathbf{I})^{-1}\Sigma(\Sigma + \mathbf{I})^{-1})_{ij}1_k \\ &\quad + 2\sum_l((\Sigma + \mathbf{I})^{-1}\Sigma(\Sigma + \mathbf{I})^{-1})_{ul}(\Sigma + \mathbf{I})_{li}^{-1}1_j1_k \\ &\quad + (1/2 + \text{tr}(\Sigma(\Sigma + \mathbf{I})^{-1}))(\Sigma + \mathbf{I})_{ki}^{-1}1_j\end{aligned}$$

が得られ, したがって

$$\begin{aligned}
 (\partial_{ij}b_k)A_{ijk} &= (\partial_{ij}b_k)(\sigma_{ik}1_j + \sigma_{jk}1_i) \\
 &= -p + \frac{p}{2} + 2 \sum_l ((\Sigma + \mathbf{I})^{-1}\Sigma(\Sigma + \mathbf{I})^{-1})_{ll} p 1_l \\
 &\quad + p(1/2 + \text{tr}(\Sigma(\Sigma + \mathbf{I})^{-1}))^2 \\
 &= -\frac{p}{2} + 2p \text{tr}(\Sigma(\Sigma + \mathbf{I})^{-2}) + p(1/2 + \text{tr}(\Sigma(\Sigma + \mathbf{I})^{-1}))^2
 \end{aligned}$$

となる。また $\mathbf{b} = \mu \mathbf{1}$ より

$$\begin{aligned}
 \frac{1}{2} \mathbf{b}'(\mathbf{I} + \Sigma)\mathbf{b} - \mu \mathbf{1}'\mathbf{b} &= p\mu^2 - p\mu^2 \\
 &= 0
 \end{aligned}$$

となるので, 最終的に

$$\begin{aligned}
 &\lim_{n \rightarrow \infty} n^2(R(\hat{\mathbf{w}}) - R(\hat{\mathbf{v}})) \\
 &= \frac{1}{2} \mathbf{b}'(\mathbf{I} + \Sigma)\mathbf{b} - \mu \mathbf{1}'\mathbf{b} - (\partial_{ij}b_k)A_{ijk} \\
 &= \frac{p}{2} - 2p \text{tr}(\Sigma(\Sigma + \mathbf{I})^{-2}) - p(1/2 + \text{tr}(\Sigma(\Sigma + \mathbf{I})^{-1}))^2 \\
 &= \frac{p}{2} - 2p \left(\frac{1}{4} + \text{tr}(\mathbf{H}(\mathbf{H} + \mathbf{Q})^{-2}) \right) - p(1 + \text{tr}(\mathbf{H}(\mathbf{H} + \mathbf{Q})^{-1}))^2 \\
 &\leq -p
 \end{aligned}$$

が得られる。ただし $\mathbf{H} = \Sigma - \mathbf{1}\mathbf{1}'/p$, $\mathbf{Q} = \mathbf{I} - \mathbf{1}\mathbf{1}'/p$ とおいた。最後の不等式は $\mathbf{H} = \mathbf{0}$ のときに等号が成立する。

References

- [1] Idel, M. (2016). A review of matrix scaling and Sinkhorn's normal form for matrices and positive maps, Preprint arxiv:1609.06349.
- [2] Johnstone, I. M. and Lu, A. Y. (2009). On consistency and sparsity for principal components analysis in high dimensions, *J. Amer. Statist. Assoc.*, **104**, 682–693.

- [3] Marshall, A. W. and Olkin, I. (1968). Scaling of matrices to achieve specified row and column sums. *Numer. Math.*, 12, 83–90.
- [4] Sei, T. (2016). An objective general index for multivariate ordered data, *J. Multivariate Anal.*, 147, 247–264.
- [5] 竹村 彰通 (1991). 「多変量推測統計の基礎」, 共立出版.