研究集会「生物情報を解明するための統計数学的基礎理論とその応用」
2009 年 10 月 1 日（木）– 3 日（土）・於: 日本女子大学

Wishart 分布のスケール行列の推定について[1]

今野　良彦[2]
日本女子大学理学部

0　概要

Wishart 分布のスケール行列の推定問題を統計的決定問題の立場から論じる．第 1 節では，統計的決定問題の用語を解説し，第 2 節では，Eaton (1986) を参考にして，統計的決定問題の立場から Wishart 分布ののスケール行列の推定問題における古典的な結果を概括し，第 3 節では，Andersson and Perlman (1993) による多変量正規 Lattice conditional independence model (LCI) モデルの基本的な結果を述べた後に，最も簡単な例である 3×3 の分散行分散行列の逆行列の (3, 2) 成分が 0 の条件付独立多変量正規モデルにおけるミニマックス推定量について述べる．

1　統計的決定問題

統計的決定問題を構成する要素はつぎの 5 つである：

(i) 標本空間 $(\mathscr{X}, \mathcal{B}_1)$; (ii) 母数空間 $(\Theta, \mathcal{B}_2)$; (iii) 行動空間 $(\mathscr{A}, \mathcal{B}_3)$;
(iv) 統計モデル $\{P_\theta : \theta \in \Theta\}$，すなわち，標本空間 $(\mathscr{X}, \mathcal{B}_1)$ 上の確率測度の族;
(v) 損失関数 $\mathbf{L} : \Theta \times \mathscr{A} \to [0, \infty)$，ただし，$\mathbf{L}$ は可測関数である．

標本 $x \in \mathscr{X}$ に対して，行動空間 $(\mathscr{A}, \mathcal{B}_3)$ 上の確率測度を対応させる関数 $\delta : x \mapsto \delta(\cdot\,|\,x)$ を統計的決定関数という．各 $x \in \mathscr{X}$ に対して，ある元 $a(x) \in \mathscr{X}$ が存在して，$\delta(\{a(x)\}|\,x) = 1$[3]のとき，決定関数 δ は非確率化決定関数という．与えられた決定関数 δ に対して，δ の危険関数 $\mathbf{R}(\theta, \delta)$ を $\mathbf{R}(\theta, \delta) = \int_{\mathscr{X}} \int_{\mathscr{A}} \mathbf{L}(\theta, a)\, \delta(da|\,x)\, P_\theta(dx)$ で定める．非確率化決定関数 δ[4]に対する危険関数は $\mathbf{R}(\theta, \delta) = \int_{\mathscr{X}} \mathbf{L}(\theta, a(x))\, P_\theta(dx)$ と簡単に書ける．

$\boxed{\text{決定関数のよさ}}$　決定関数全体の集合を $\mathscr{D}$ と書くことにする．$\delta_1, \delta_2 \in \mathscr{D}$ とする．δ_1 が δ_2 と同程度によいとは，すべての $\theta \in \Theta$ に対して，$\mathbf{R}(\theta, \delta_1) \le \mathbf{R}(\theta, \delta_2)$ が成り立つときにいう．δ_1 が δ_2 よりよいとは，δ_1 は δ_2 と同程度によく，さらに，すくなくとも一つの $\theta_0 \in \Theta$ が存在して，$\mathbf{R}(\theta_0, \delta_1) < \mathbf{R}(\theta_0, \delta_2)$ が成り立つときにいう．

$\boxed{\text{許容性}}$　δ_1 よりよい $\delta \in \mathscr{D}$ が存在しないとき，δ_1 は許容的であるという．

$\boxed{\text{完備類}}$　$D \subset \mathscr{D}$ を部分集合とする．任意の $\delta \in \mathscr{D} \setminus D$ に対して，ある $\delta' \in D$ が存在して，δ' が δ よりよい（同程度によい）とき，D は（本質的）完備類という．D が（本質的）完備類であって，D のいかなる真部分集合も（本質的）完備類にならないとき，D は最小（本質的）完備類という．

命題 1. 行動空間 $\mathscr{A}$ は $\mathbb{R}^k$ の凸集合とし，損失関数 $\mathbf{L}(\theta, a)$ は a の凸関数で，$\|a\| \to \infty$ のとき，$\mathbf{L}(\theta, a) \to \infty$ とする．このとき，非確率化決定関数の全体の集合は本質的完備類である．

命題 2. 統計量 t を十分とする．$\mathscr{X}$ または $\mathscr{A}$ は $\mathbb{R}^k$ の部分集合とする．t に基づく決定関数全体の集合は本質的完備類である．

---

[1]A copy of the Handout is available at http://mp-w3math.jwu.ac.jp/~konno/pdf/talk34.pdf
[2]email address: konno[at]fc[dot]jwu[dot]ac[dot]jp
[3]すなわち，$a(x)$ に退化.
[4]すなわち，$\delta(\{a(x)\}|\,x) = 1$ が成り立っている．

$\boxed{\text{ベイズ解}}$ ξ を $(\Theta, \mathcal{B}_2)$ 上の確率測度とする．これを事前確率とよぶことにする．δ_1 が事前確率 ξ に対するベイズ解であるとは，

$$\int_\Theta \mathbf{R}(\theta, \delta_1)\,\xi(d\theta) = \min_{\delta\in\mathscr{D}} \int_\Theta \mathbf{R}(\theta, \delta)\,\xi(d\theta)$$

をみたすときにいう．δ_1 が広義ベイズ解であるとは，$(\Theta, \mathcal{B}_2)$ 上の確率測度列 $\{\xi_n\}_{n=1}^\infty$ が存在して，

$$\int_\Theta \mathbf{R}(\theta, \delta_1)\,\xi_n(d\theta) - \min_{\delta\in\mathscr{D}} \int_\Theta \mathbf{R}(\theta, \delta)\,\xi_n(d\theta) \to 0, \qquad (n\to\infty)$$

がなりたつときにいう．

$\boxed{\text{ミニマックス決定関数}}$ δ_1 がミニマックス決定関数であるは，

$$\max_{\theta\in\Theta} \mathbf{R}(\theta, \delta_1) = \min_{\delta\in\mathscr{D}} \max_{\theta\in\Theta} \mathbf{R}(\theta, \delta)$$

が成り立つときにいう．

$\boxed{\text{不変性}}$ $(\mathscr{X}, G, \varphi)$, $(\Theta, \bar{G}, \bar{\varphi})$, $(\mathscr{A}, \tilde{G}, \tilde{\varphi})$ は変換群とする．以後，誤解のおそれがなければ，$G, \bar{G}, \tilde{G}$ および $\varphi, \bar{\varphi}, \tilde{\varphi}$ をすべて G, φ とかくことにする．決定問題が G 不変であるとは，つぎの条件をみたすときをいう．
(i) モデル $\{P_\theta : \theta\in\Theta\}$ は不変．すなわち，

$$gP_\theta = P_{g\theta}, \qquad (g\in G,\, \theta\in\Theta)$$

ただし，任意の $B\in\mathcal{B}_1$ に対して，$gP_\theta(B) = P_\theta(g^{-1}B)$ である．
(ii) 損失関数 $\mathbf{L}$ は不変．すわわち，

$$\mathbf{L}(g\theta,\, ga) = \mathbf{L}(\theta,\, a), \qquad (g\in\Theta,\, \theta\in\Theta,\, a\in\mathscr{A}).$$

決定関数 $\delta\in\mathscr{D}$ と $g\in G$ に対して，決定関数 $g\delta$ を

$$(g\delta)(B|\,x) = \delta(g^{-1}B|\,g^{-1}x), \qquad (B\in\mathcal{B}_3)$$

で定める．

決定関数 δ が G 不変であるとは，すべての $g\in G$ に対して，$g\delta = \delta$ が成り立つときにいう．δ が非確率化検定（$\delta(\{a(x)\}|\,x) = 1$ ）のとき，

$$\delta \quad \text{が不変} \quad \Longleftrightarrow a \quad \text{が不変} \quad \Longleftrightarrow a(gx) = ga(x), \qquad (g\in G)$$

となる．

命題 3. 決定問題が G 不変とする．任意の決定関数 δ に対して，

$$\mathbf{R}(g\theta,\, g\delta) = \mathbf{R}(\theta,\, \delta), \qquad (g\in G,\, \theta\in\Theta)$$

さらに，決定関数 δ が G 不変ならば，

$$\mathbf{R}(g\theta,\, \delta) = \mathbf{R}(\theta,\, \delta), \qquad (g\in G)$$

がわかる．

G が $\mathscr{X}$ に推移的に作用していれば，固定した元 $\theta_0\in G$ に対して，

$$\mathbf{R}(\theta,\, \delta) = \mathbf{R}(\theta_0,\, \delta_0)$$

となる．したがって，固定した θ_0 のもとで最良のものは G 不変のなかで最良となる．

最良 G 不変性とミニマックス性

命題 4. G は三角群とし，決定問題は G 不変とする．このとき，もし，ミニマックス決定関数が存在するならば，G 不変ミニマックス決定関数が存在する．もし，G 不変決定関数 δ_1 が

$$\max_{\theta\in\Theta} \mathbf{R}(\theta,\, \delta_1) = \min_{\delta\in\mathscr{D}^*} \max_{\theta\in\Theta} \mathbf{R}(\theta,\, \delta)$$

をみたすならば，δ_1 はミニマックスである．ただし，$\mathscr{D}^*$ は G 不変決定関数の全体である．

注意 この命題は Hunt-Stein の定理と呼ばれるものを簡単にしたものである．Hunt-Stein の定理が成立するための条件として，Amenability 条件（Lehmann and Casella (1998, page 422) を参照）が知られている．三角群はこの条件をみたしていることが知られている． □

2 Wishart 分布のスケール行列の推定問題

r 次元縦ベクトル確率変数列 $X_1,\, X_2,\, \ldots,\, X_n\ (n \geq r)$ は独立同一に多変量正規分布 $N_r(0,\, \Sigma)$ に従うと仮定する．ただし，$\Sigma \in \mathrm{Sym}^+(r,\, \mathbb{R})$[5] とする．このとき，$(X_1,\, X_2,\, \ldots,\, X_n)$ に基づき，Σ の推定問題を損失関数

$$\mathbf{L}(\Sigma,\, a) = \mathrm{Tr}\,(a\Sigma^{-1}) - \log \mathrm{Det}\,(a\Sigma^{-1}) - r \tag{1}$$

のもと[6]で考える．ただし，a は Σ の推定量で $S = \sum_{i=1}^n X_i X_i'$ を通してのみ $(X_1,\, X_2,\, \ldots,\, X_n)$ に依存する推定量とし，Tr , Det は行列のトレースと行列式，X_i' は縦ベクトル X_i の転置である．これは S は Σ の十分統計量であることと命題 2 から正当化される．この損失関数 $\mathbf{L}$ は James and Stein (1961) で提案された．この損失関数は非常に取り扱いが容易である．これは，つぎの性質をみたす：

(i) $\mathbf{L}(\Sigma,\, a) \geq 0$ で等号成立は $a = \Sigma$ のときのみ;

(ii) $\mathbf{L}(\Sigma,\, a)$ は第 2 の変数に関して凸関数;

(iii) 任意の $g \in \mathrm{GL}(r,\, \mathbb{R})$ [7] に対して，

$$\mathbf{L}(g\Sigma g',\, gag') = \mathbf{L}(\Sigma,\, a).$$

ただし，g' は行列 g の転置行列を示す.

損失関数の性質と命題 1 から非確率化推定量のみを考えれたよいことがわかる．損失関数 $\mathbf{L}$ に対応する危険関数 $\mathbf{R}$ を

$$\mathbf{R}(\Sigma,\, a) = \mathbb{E}[\mathbf{L}(\Sigma,\, a)]$$

で定義する．

2.1 定数リスクのミニマックス推定量

自然指数分布族の性質から，Σ の最尤推定量は $\widehat{\Sigma}_{\mathrm{mle}} = n^{-1}S$ となることが直ちにわかる．

$\mathscr{X} = \mathrm{Sym}^+(r,\, \mathbb{R})$，$G = \mathrm{GL}(r,\, \mathbb{R})$，$\varphi(g,\, x) = gxg'$ とすれば，$(\mathscr{X},\, G,\, \varphi)$ は変換群となる．群 $\mathrm{GL}(r,\, \mathbb{R})$ は $\mathscr{X}$ に推移的に作用するので，最良 $\mathrm{GL}(r,\, \mathbb{R})$ 不変推定量を求めることができる．すなわち，$\widehat{\Sigma}$ を G 不変とすれば，命題 3 から

$$\mathbf{R}(\Sigma,\, \widehat{\Sigma}) = \mathbf{R}(I_r,\, \widehat{\Sigma})$$

となる．さらに，任意の $\mathrm{GL}(r,\, \mathbb{R})$ 不変推定量は $\widehat{\Sigma} = cS$ と表現できることに注意する．ただし，c は正の定数である．なぜならば，$g = S^{-1/2}$ ととれば，

$$\widehat{\Sigma}(S) = S^{-1/2}\widehat{\Sigma}(I_r)S^{-1/2}$$

と表現できる．さらに，$g \in \mathcal{P}_r \subset \mathrm{GL}(r,\, \mathbb{R})$ を置換として，$\widehat{\Sigma}(I_r) = \widehat{\Sigma}(gI_rg') = g\widehat{\Sigma}(I_r)g'$ となることから，$\widehat{\Sigma}(I_r) = cI_r$ と書けることがわかる．

[5] $\mathrm{Sym}^+(r,\, \mathbb{R})$ は r 次正値実対称行列の全体を表わす.
[6] 上記損失関数以外には，$\mathrm{Tr}\,(a\Sigma^{-1} - I_r)^2$，および $\mathrm{Tr}\,(a - \Sigma)^2$ 等がある．前者は不変であるが，後者はそうでない.
[7] $\mathrm{GL}(r,\, \mathbb{R})$ は r 次正則行列の全体である.

命題 5. G 不変推定量の中で最小のものは $\widehat{\Sigma}_{\mathrm{mle}} = (1/n)S$ である．

Σ_{mle} は 最良 $\mathrm{GL}(r,\,\mathbb{R})$ 不変推定量 である．また，Σ_{mle} は $\mathrm{LT}^{+}(r,\,\mathbb{R})$ 不変[8]でもあることに注目する．$\mathrm{LT}^{+}(r,\,\mathbb{R})$ も $\mathrm{Sym}^{+}(r,\,\mathbb{R})$ に推移的に作用するので，命題 3 から，このクラスに属する推定量のリスクも $\Sigma = I_r$ のもとで計算すればよい．これから， $\mathrm{LT}^{+}(r,\,\mathbb{R})$ 不変で一番よいものを求めることができる．

命題 6 (James and Stein (1961))**.** (i) $S = TT'$ と分解する．ここで，$T \in \mathrm{LT}^{+}(r,\,\mathbb{R})$ である．$\mathrm{LT}^{+}(r,\,\mathbb{R})$ 不変推定量は

$$TAT',$$

の形で与えられる．ただし，$A \in \mathrm{Sym}(r,\,\mathbb{R})$ は定数である．
(ii) $\mathrm{LT}^{+}(r,\,\mathbb{R})$ 不変推定量のリスクの下限はつぎで与えられる：

$$\mathbf{R}(I_r,\,TAT') \geq \sum_{i=1}^{r}\left\{\log(n+r-2i+1) - \mathbb{E}[\log\chi^2_{n-i+1}]\right\}$$

(iii) 推定量 $\widehat{\Sigma}_{\min} = T\mathrm{diag}\,(1/(n+r-2i+1)|\,i=1,\,2,\,\ldots,\,r)T'$ は (ii) の下限に達する．したがって，命題 4 から，$\widehat{\Sigma}_{\min}$ はミニマックスである．

証明：$\mathrm{LT}^{+}(r,\,\mathbb{R})$ 不変性から $\widehat{\Sigma}(T^{-1}S(T')^{-1}) = T^{-1}\widehat{\Sigma}(I_r)(T')^{-1}$ となることより (i) はわかる．
つぎに，命題 4 と (i) から，$\mathrm{LT}^{+}(r,\,\mathbb{R})$ 不変推定量 $\widehat{\Sigma}$ に対して，

$$\mathbf{R}(\Sigma,\,\widehat{\Sigma}) \;=\; \mathbf{R}(I_r,\,\widehat{\Sigma}) = \mathrm{Tr}\,\{\mathbb{E}(T'T)A\} - \log\mathrm{Det}\,(A) - \sum_{i=1}^{r}\mathbb{E}[\log t_{ii}^2] - r$$

となる．しかし，Bartelett 分解から $t_{ii}^2 \sim \chi^2_{n-i+1},\, t_{ij} \sim N(0,\,1)\,(i \neq j)$, さらに，これらは互いに独立なので $\mathbb{E}[T'T] = \mathrm{diag}\,(n+r-2i+1|\,i=1,\,2,\,\ldots,\,r)$ となる．さらに，A の各成分に関して微分すれば，

$$\mathrm{diag}\,(n+r-2i+1|\,i=1,\,2,\,\ldots,\,r) - A^{-1} = 0$$

を得る．よって，損失関数の凸性から (ii) がわかる．(iii) は (ii) の証明をたどればよい． □

2.2 直交不変な推定量と SURE 法

Σ の直交不変な推定量を考える：

$$\widehat{\Sigma}(HSH') = H\widehat{\Sigma}(S)H', \qquad (H \in \mathrm{O}(r,\,\mathbb{R})).$$

ただし，$\mathrm{O}(r,\,\mathbb{R})$ は r 次の直交行列の全体である．このクラスの推定量は

$$\widehat{\Sigma} = H\mathrm{diag}\,(\phi_i|\,i=1,\,2,\,\ldots,\,r)H' \tag{2}$$

と書ける．ただし，$S = H\mathrm{diag}\,(\ell_1,\,\ldots,\,\ell_r)H'$ で $l = (\ell_1,\,\ldots,\,\ell_r)$ で $\phi_i(\ell)$ は微分可能とする．これは，$H = \mathrm{diag}\,(\pm 1,\,\ldots,\,\pm 1)$ とすればわかる．$\phi_i = (1/n)\ell_i$ とすれば，$\widehat{\Sigma}_{\mathrm{mle}}$ となることがわかる．

$\boxed{\text{動機}}$ $\widehat{\Sigma}_{\mathrm{mle}}$ の固有値は Σ の自然な推定量であるが，r が大きく，$\Sigma = I_r$ としたときにには，これらの推定量は偏りが生ずることが Johnsson (1982) で知られている結果から予想される：$0 < y < 1$ とする．$r/n \to y\,(n \to \infty)$ のとき，

$$\frac{1}{r}\sum_{i=1}^{r}\mathbb{I}\,\{\ell_i/r \leq x\} \xrightarrow{P} \int_a^x \frac{1}{u}\frac{\sqrt{(u-a)(b-u)}}{2\pi yu}\,du\,\mathbb{I}_{[a,\,b]}(x) + \mathbb{I}_{(b,\,\infty)}(x)$$

[8] $\mathrm{LT}^{+}(r,\,\mathbb{R})$ は r 次下三角行列で対角成分がすべて正の行列全体を表わす．

となる．ただし，各 x における $\xrightarrow{P}$ は確率収束を示し，

$$a = 1 + y - 2\sqrt{y} \qquad b = 1 + y + 2\sqrt{y}$$

である．また，集合 A に対して，$x \in A$ のとき，$\mathbb{I}_A(x) = 1$，その他のとき，$\mathbb{I}_A(x) = 0$ である．したがって，i/r と $1 - i/r$ が小さい $i \in \{1, 2, \ldots, r\}$ に対しては，ℓ_i/n は 1 に高い確率で離れていることがわかる．よって，n と r が近く，$\Sigma = I_r$ のとき，$1 - i/r$ が小さいもの（i が r に近い）に対しては，ℓ_i/n を縮小し，i/r がが小さいもの（i が 1 に近い）に対しては，ℓ_i/n を拡大するのがよいことが予想される．問題は，このような修正を行っても Σ に関して一様に $\widehat{\Sigma}_{\mathrm{mle}}$ よりよいものができるかである．実際にうまくできることを Stein (1973) で示している．

SURE 法：リスクの不偏推定量　直交群 $\mathrm{O}(r, \mathbb{R})$ の $\mathrm{Sym}^+(r, \mathbb{R})$ への作用は推移的でない．したがって，リスクは一般には定数でない．直交不変な推定量のリスクを評価するためには．リスクの直接的な評価が困難になる．これを克服する方法が Stein's Unbiased Risk Estimate 法（SURE 法）である．

まず，議論を簡単にするために，推定量のリスクを比較をするための本質的な部分を

$$\widetilde{\mathbf{R}}(\Sigma, \widehat{\Sigma}) = \mathbb{E}[\mathrm{Tr}\,(\widehat{\Sigma}\Sigma^{-1}) - \log \mathrm{Det}\,(\widehat{\Sigma})] \tag{3}$$

とおく．有限リスクを持つふたつの推定量 $\widehat{\Sigma}_1, \widehat{\Sigma}_2$ に対して，

$$\mathbf{R}(\Sigma, \widehat{\Sigma}_1) \geq \mathbf{R}(\Sigma, \widehat{\Sigma}_2) \Longleftrightarrow \widetilde{\mathbf{R}}(\Sigma, \widehat{\Sigma}_1) \geq \widetilde{\mathbf{R}}(\Sigma, \widehat{\Sigma}_2), \qquad (\forall \Sigma \in \mathrm{Sym}^+(r, \mathbb{R}))$$

となることに注意せよ．以後は，言葉の乱用となるが，$\widetilde{\mathbf{R}}$ のこともリスクと呼ぶことにする．

各推定量のリスクは Σ の関数となっている．リスクの一様な比較をするために，直交不変な推定量 $\widehat{\Sigma}$ に対するリスク（ Σ の関数）の不偏推定量を求めることを考える．形は $\widehat{\Sigma}$ に依存するが，ひとたび S を観測すれば，計算できる量，すなわち，推定量（これを $\widehat{\mathbf{R}}_{\widehat{\Sigma}}(S)$ と書くことにする）で

$$\widetilde{\mathbf{R}}(\Sigma, \widehat{\Sigma}) = \mathbb{E}[\widehat{\mathbf{R}}_{\widehat{\Sigma}}(S)]$$

をみたすものを構成することを目指す．有限のリスクを持つ任意の直交不変な推定量 $\widehat{\Sigma}$ に対して，幸運にもこのように，リスクの不偏推定量を見つけることができれば，期待値の基本的な性質から

$$\widehat{\mathbf{R}}_{\widehat{\Sigma}_1}(S) \geq \widehat{\mathbf{R}}_{\widehat{\Sigma}_2}(S) \Longrightarrow \widetilde{\mathbf{R}}(\Sigma, \widehat{\Sigma}_1) \geq \widetilde{\mathbf{R}}(\Sigma, \widehat{\Sigma}_2) \qquad ({}^\forall \Sigma \in \mathrm{Sym}^+(r, \mathbb{R}))$$

となる．さらに，ある Σ で

$$\mathbb{P}(\widehat{\mathbf{R}}_{\widehat{\Sigma}_1}(S) > \widehat{\mathbf{R}}_{\widehat{\Sigma}_2}(S)) > 0$$

ならば，

$$\widetilde{\mathbf{R}}(\Sigma, \widehat{\Sigma}_1) > \widetilde{\mathbf{R}}(\Sigma, \widehat{\Sigma}_2) \qquad (\exists \Sigma \in \mathrm{Sym}^+(r, \mathbb{R}))$$

がわかるのである．

方針をまとめるとつぎのふたつの段階に分かれる：

(1) 有限のリスクを持つ任意の直交不変な推定量 $\widehat{\Sigma}$ に対して，そのリスクの不偏推定量 $\widehat{\mathbf{R}}_{\widehat{\Sigma}}(S)$ を構成する．

(2) $\widehat{\mathbf{R}}_{\widehat{\Sigma}}(S)$ で推定量の比較を行う．

(1) の段階では，部分積分を用いる．(2) はある種の偏微分不等式をとくことになる．これを解く体系的な手法[9]は知らない．現段階では，いくつかのパターンを試してみて，解を見つけるにとどまっている．

[9]正確には，一様に改良する推定量を得るための体系的な手法は知られていない．しかし，一様に改良する推定量を得ることはできないが，母数のおおくの領域でリスクを大きく改善する推定量の構成法は知られている．だが，解析的な正当化はやはりむずかしいと思う．

2.3 部分積分の公式

$H\Sigma^{-1}H' = (a_{ij}(H))$ とおき，推定量 (2) を (3) に代入すると

$$\widetilde{\mathbf{R}}(\Sigma,\,\widehat{\Sigma}) = \mathbb{E}\left[\sum_{i=1}^{r}\phi_i(\ell)a_{ii}(H) - \log\phi_i(\ell)\right] \tag{4}$$

と書き直せる．したがって，上の式の右辺の括弧内の第一目のみが母数に依存している．それを部分積分の公式を用いて評価する．ただし，ここで，損失は ℓ と H にのみ依存しているので，ウイシャート行列の極分解に対する分布に関して部分積分の公式を使うと非常に見通しよく，不偏推定量を求めることができる．そのために，以下の記号を用意する：ℓ の同時分布は

$$K\exp\bigl\{-\frac{1}{2}\sum_{j=1}^{r}\ell_j a_{jj}(H)\bigr\}\bigl(\Pi_{j=1}^{r}\ell_j\bigr)^{(n-r-1)/2}\Pi_{j>i}(\ell_i-\ell_j)\,\mu_{\mathrm{O}(p,\,\mathbb{R})}\,\Pi_{j=1}^{r}d\ell_j.$$

これを踏まえてつぎの記号を用意する：

$$\begin{aligned}
F(\ell) &= K\bigl(\Pi_{j=1}^{r}\ell_j\bigr)^{(n-r-1)/2}\Pi_{j>i}(\ell_i-\ell_j)\\
G(a) &= \int_{\mathscr{K}}\exp\bigl\{-\frac{1}{2}\sum_{j=1}^{r}\ell_j a_{ii}(H)\bigr\}\,\mu_{\mathcal{O}(P)},\\
\mathcal{R}_j &= \{\ell^{(j-)} = (\ell_1,\ldots,\ell_{j-1},\,\ell_{j+1},\,\ldots,\,\ell_r)|\,\ell_1>\cdots>\ell_{j-1}>\ell_{j+1}>\cdots>\ell_r>0\},
\end{aligned}$$

とおく．ただし，$\ell_0=\infty$ と $\ell_{p+1}=0$ した．これらの記号を用いて，つぎの定理を示す．

補題 1 (Sheena (1995))**.**

$$\sum_{j=1}^{r}\mathbb{E}[\phi_j a_{jj}(H)] = \mathbb{E}\Biggl[\sum_{j=1}^{r}\Bigl\{2\frac{\partial\phi_j}{\partial\ell_j} + \bigl(n-r-1\bigr)\frac{\phi_j}{\ell_j} + 2\sum_{i\neq j}\frac{\phi_j}{\ell_j-\ell_i}\Bigr\}\Biggr]$$

証明：部分積分から

$$\begin{aligned}
\sum_{j=1}^{r}\mathbb{E}[\phi_j a_{jj}(H)] &= -2\sum_{j=1}^{r}\int_{\mathcal{R}_j}\Bigl(\int_{\ell_{j+1}}^{\ell_{j-1}}\varphi_j\,F(\ell)\frac{\partial}{\partial\ell_j}G(\ell)\,d\ell_j\Bigr)\,d\ell^{(j-)}\\
&= 2\sum_{j=1}^{r}\int_{\mathcal{R}_j}\int_{\ell_{j+1}}^{\ell_{j-1}}\frac{\partial}{\partial\ell_j}(\phi_j F(\ell))G(\ell)\,da_j\,d\ell^{(j-)}\\
&= 2\mathbb{E}\Biggl[\sum_{j=1}^{r}\frac{\partial}{\partial\ell_j}(\phi_j F(\ell))\frac{1}{F(\ell)}\Biggr]\\
&= \mathbb{E}\Biggl[\sum_{j=1}^{r}\Bigl\{2\frac{\partial\phi_j}{\partial\ell_j} + \bigl(n-r-1\bigr)\frac{\phi_j}{\ell_j} + 2\sum_{i\neq j}\frac{\phi_j}{\ell_j-\ell_i}\Bigr\}\Biggr],
\end{aligned}$$

となる． □

補題より，推定量 (2) に対して，

$$\widetilde{\mathbf{R}}(\Sigma,\,\widehat{\Sigma}) = \mathbb{E}\left[\sum_{j=1}^{r}\Bigl\{2\frac{\partial\phi_j}{\partial\ell_j} + \bigl(n-r-1\bigr)\frac{\phi_j}{\ell_j} + 2\sum_{i\neq j}\frac{\phi_j}{\ell_j-\ell_i} - \log\phi_j\Bigr\}\right] \tag{5}$$

を得る．

2.4 直交不変ミニマックス推定量

いきあたりばったりではあるが，推定量 (2) において

$$\phi_j^{(m)} = d_j \ell_j, \qquad d_j = \frac{1}{n+r-2j+1}, \qquad j = 1,\, 2,\, \ldots,\, r. \tag{6}$$

としたものを $\widehat{\Sigma}_{\mathrm{m}}$ とする．$d_1 < \cdots < d_r$ で $(1/r)\sum_{j=1}^{r} 1/d_j = n$ となっている．

定理 1 (Dey and Srinivasan (1985))**.** (2) と (6) で与えられる推定量 $\widehat{\Sigma}_{\mathrm{m}}$ に対して，

$$\mathbf{R}(\Sigma,\, \widehat{\Sigma}_{\mathrm{m}}) \le -\sum_{j=1}^{r} \log d_j - \sum_{j=1}^{r} \mathbb{E}[\log u_j^2],$$

が成立する．ただし，u_j ’は $\chi^2_{n-j+1}\,(j = 1,\, 2,\, \ldots,\, r)$ に従う．したがって，$\widehat{\Sigma}_{\mathrm{m}}$ はミニマックス．

証明：まず，

$$\mathbb{E}[\log \mathrm{Det}\, \widehat{\Sigma}_{\mathrm{m}}] \;=\; \sum_{j=1}^{r} \log d_j + \mathbb{E}[\log \mathrm{Det}\, S] = \sum_{j=1}^{p} \big\{ \log d_j + \log u_j^2 \big\} + \log \mathrm{Det}\, \Sigma$$

となることに注意する．この式と (5) より

$$\begin{aligned}
\mathbf{R}(\Sigma,\, \widehat{\Sigma}_{\mathrm{m}}) \;&=\; \mathbb{E}\Big[2\sum_{j=1}^{r}\sum_{i>j} \frac{d_j\ell_j - d_i\ell_i}{\ell_j - \ell_i} + 2\sum_{j=1}^{r} d_j + \big(n-r-1\big)\sum_{j=1}^{r} d_j \\
&\qquad - \sum_{j=1}^{r} \mathbb{E}\log u_j^2 - \sum_{j=1}^{r} \log d_j - p \Big] \\
&\le\; \sum_{j=1}^{r} \big(n+r-2j+1\big) d_j - \sum_{j=1}^{r} \mathbb{E}\log u_j^2 - \sum_{j=1}^{r} \log d_j - p \\
&=\; -\sum_{j=1}^{r} \mathbb{E}\log u_j^2 - \sum_{j=1}^{r} \log d_j.
\end{aligned}$$

となることからわかる． □

3 Lattice conditional indepedence モデル

集合 $I = \{1,\, 2,\, \ldots,\, r\}$ に対して，$\mathcal{K}$ を I の部分集合からなる分配環とし，集合の和（∪）と積（∩）について閉じて空集合を含むものとする．$X \sim N_r(0,\, \Sigma)$ とし、$K \in \mathcal{K}$ に対して，X_K を K の元に対応する座標の成分をもつ部分ベクトルとする．このとき、Lattice conditional independence (LCI) モデル $\boldsymbol{N}(\mathcal{K})$ を

すべての $L, M \in \mathcal{K}$ に対して，
$X_{L\cap M}$ を与えたとき，X_L と X_M は条件付き独立 (7)

であるような多変量正規分布 $N_r(0, \Sigma)$ の集合とする．また，$\boldsymbol{P}(\mathcal{K})$ をすべての $L,\, M \in \mathcal{K}$ に対して，(7) を満足するような非退化多変量正規分布の分散共分散行列[10]の集まりとする．条件付独立性による別のモデリングである Graphical model (Lauritzen (1996)) との関係については，Andersson *et al.* (1995) を参照のこと．

記号 $K \in \mathcal{K}$ に対して，$\langle K \rangle = \cup(\tilde{K} \in \mathcal{K} | \tilde{K} \subset K \quad$ かつ $\quad \tilde{K} \neq K)$, $[K] = K \backslash \langle K \rangle$ とする．したがって，$K = \langle K \rangle \cup [K]$ かつ $\langle K \rangle \cap [K]$ は空集合である．$\mathcal{K}$ は集合の和と積について閉じているので，$\langle K \rangle \in \mathcal{K}$ であるが，$[K] \in \mathcal{K}$ は必ずしも成立しない．更に，部分集合族を

$$\mathcal{J}(\mathcal{K}) = \{ K \in \mathcal{K} | [K] \text{ は空でない } \}$$

[10]したがって，正定値行列.

とおく．$J(\mathcal{K})$ に含まれる部分集合を $K_1, K_2, \ldots, K_q\,(q \le r)$ とする．順番は，$i < j$ ならば，$K_j \not\subset K_i$ となるように取ることにする．各 $K_i \in J(\mathcal{K})\,(i = 1, 2, \ldots, q))$ は $[K_i] \neq \emptyset$ で，$i \neq j$ ならば，$[K_i] \cap [K_j] = \emptyset$ であり，任意の $K \in \mathcal{K}$ に対して，

$$K = \cup_{K_i \subset K\,(i=1,\ldots,q)}[K_i]; \qquad I = [K_1] \cup [K_2] \cup \cdots \cup [K_q].$$

以後は，$[K_1], [K_2], \ldots, [K_q]$ の順に行列の成分は並んでいるとする．$L, M \subset I$ に対して，行列 A に対して，(i, j) 成分（$i \in L, j \in M$）を集めた部分行列を $A_{L,M}$ と記すことにする．つぎの集合を考える：

$$\boldsymbol{LT}(\mathcal{K}) = \{A \text{ は } I \text{ 次の正則行列}; \forall L, M \in J(\mathcal{K}) : L \neq M, \text{and}\, M \not\subset L \Longrightarrow A_{[L],[M]} = 0\}.$$

$\mathcal{K} = \{\emptyset, I\}$ でなければ，$K_1, K_2, \ldots, K_q$ は非減少（包含関係の意味）列であるので，$\boldsymbol{LT}(\mathcal{K})$ はブロック下三角行列になっている．$\Sigma \in \boldsymbol{P}(\mathcal{K})$ に対して，

$$\boldsymbol{LT}(\mathcal{K}) \times \boldsymbol{P}(\mathcal{K}) \ni (A, S) \mapsto ASA' \in \boldsymbol{P}(\mathcal{K})$$

となり，推移的な作用となる．ただし，A' は行列 A の転置行列とした．

$r = 3$ とし，LCI モデル $\boldsymbol{N}(\mathcal{K}_1)$ からのランダム標本 $X_1, X_2, \ldots, X_n (n \ge 3)$ に基づいて分散共分散行列 Σ の推定問題を損失関数 (1) のもとで考える．ここで，$\mathcal{K}_1 = \{\emptyset, 1, 12, 13, I\}$ [11]とする．(7) から $\mathcal{K}_1$ は

$$2 \perp\!\!\!\perp 3\,|\,1 \Longleftrightarrow (\Sigma^{-1})_{2,3} = (\Sigma^{-1})_{3,2} = 0 \Longleftrightarrow (\Sigma)_{3,2} = \frac{(\Sigma)_{3,1}(\Sigma)_{2,1}}{(\Sigma)_{1,1}}$$

となる制限になることに注意する．また，$\mathcal{J}(\mathcal{K}_1) = \{1, 12, 13\}$ であり，$K_1 = 1, K_2 = 12, K_3 = 13$ とおけば，K_1, K_2, K_3 は非減少列であり，$[K_1] = 1, [K_2] = 2, [K_3] = 3$ となる．変換

$$X_1 \mapsto \begin{pmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & 0 & a_{33} \end{pmatrix} X_1 =: AX_1, \qquad A \in \boldsymbol{LT}(\mathcal{K}_1) \tag{8}$$

のもとで，$\mathcal{K}_1$ により制限されたモデルは不変であることがわかる．すわわち，$\Sigma_A = \mathbb{COV}[AX_1]$ としたとき，

$$\frac{(\Sigma_A)_{3,1}(\Sigma_A)_{2,1}}{(\Sigma_A)_{1,1}} = (\Sigma_A)_{3,2}$$

をみたす．$a_{11}, a_{22}, a_{33} > 0$ とし，(8) により定まる変換群を $\boldsymbol{LT}^{+}(\mathcal{K}_1)$ と書くことにする．

LCI モデル $\boldsymbol{N}(\mathcal{K}_1)$ からのランダム標本 $X_1, X_2, \ldots, X_n$ に対して，$S = \sum_{i=1}^{n} X_i X_i' = (s_{ij})$ とおく．LCI モデルの性質を利用して，$\mathcal{K}_1$ で制約されたモデルの分散共分散行列 Σ の最尤推定量を求めると

$$n\hat{\Sigma}^{\text{mle}}_{\mathcal{K}_1} = \begin{bmatrix} s_{11} & s_{12} & s_{13} \\ s_{21} & s_{22} & s_{21}s_{11}^{-1}s_{13} \\ s_{31} & s_{31}s_{11}^{-1}s_{12} & s_{33} \end{bmatrix}$$

となる．これを分解して，$n\hat{\Sigma}^{\text{mle}}_{\mathcal{K}_1} = T_1 T_1'$ と書く．ただし，

$$T_1 = \begin{bmatrix} s_{11}^{1/2} & 0 & 0 \\ s_{21}s_{11}^{-1/2} & t_{22} & 0 \\ s_{31}s_{11}^{-1/2} & 0 & t_{33} \end{bmatrix} \in \boldsymbol{LT}^{+}(\mathcal{K}_1)$$

である．ここで，$t_{ii}^2 = s_{ii} - s_{i1}s_{11}^{-1}s_{1i}$ とおいたとき，$t_{ii} > 0 (i = 2, 3)$ である．ミニマックス推定量は $\hat{\Sigma}^{\text{m}}_{\mathcal{K}_1} = T_1 D_1 T_1'$ で与えられる．ただし，

$$D_1 = \ \text{diag}\ \left(\frac{1}{n+2}, \frac{1}{n-1}, \frac{1}{n-1}\right)$$

[11] 12 は集合 $\{1, 2, 3\}$ の部分集合 $\{1, 2\}$ の意味である．

である．更に，これを成分で書けば，

$$
\hat{\Sigma}^{\mathrm{m}}_{\mathcal{K}_1} = \begin{bmatrix} \frac{s_{11}}{n+2} & \frac{s_{12}}{n+2} & \frac{s_{13}}{n+2} \\ \frac{s_{21}}{n+2} & \frac{1}{n-1}\left(s_{22} - \frac{3}{n+2}s_{21}s_{11}^{-1}s_{12}\right) & \frac{s_{21}s_{11}^{-1}s_{13}}{n+2} \\ \frac{s_{31}}{n+2} & \frac{s_{31}s_{11}^{-1}s_{12}}{n+2} & \frac{1}{n-1}\left(s_{33} - \frac{3}{n+2}s_{31}s_{11}^{-1}s_{13}\right) \end{bmatrix}
$$

となることがわかる．

4 むすび

前節の議論において得られた推定量 (9) は，許容的かどうか不明である．すくなくとも 3 節の手法を用いて，推定量 (9) を改良する推定量を求めることはできないように思われる．

Andersson and Perlman (1993) で得られた結果を用いると 前節の結果を一般の Lattice conditional independence model に拡張することが可能である．Konno (2001) を参照．Andersson and Madsen (1998) は LCI 構造に対称性も加えてモデルを提案している．Massam and Neher (1998) は，Andersson and Perlman (1993) の議論を Jordan algebra を用いて拡張している．これらの設定における推定の議論も興味のあるところである．

一方，正値対称行列の空間上の確率分布である古典的な Wishart 分布とその性質を等質錐（すなわち，homogeneous cones，自己同型群の部分群が推移的に作用する開凸錐）上の分布に一般化を行う研究がいくつかある．たとえば，Andersson and Wojnar (2004), Ishi (2000)．さらに，等質性をみたさないものも含んだモデルとしては，Andersson and Klein (2008), Letac and Massam (2007), Robert (2000) がある．これらの研究で一般化された Wishart 分布が統計推測理論でそのような役割を果たす可能性があるかは興味のあるところである．

参考文献

Andersson, S.A. and Klein, T.: On Rietz and Wishart distributions associated with decomposable undirected graphs., Preprint Nr. 06/2008. Feb. 2008.

Andersson, S.A., Madigan, D., Perlman, M.D., and Triggs, C.M.: On the relation between conditional independence models determined by finite distributive lattices and by directed acyclic graphs. *J. Statist. Plann. Inference* **48** (1995) 25–46.

Andersson, S.A., Madigan, D., Perlman, M.D., and Triggs, C.M.: A graphical characterization of lattice conditional independedece models, *Ann. Math. Artifical Intelligence* **21** (1997) 27–50.

Andersson, S.A. and Madsen, J.: Symmetry and lattice conditional independence in a multivariate normal distribution. *Ann. Statist.* **26** (1998) 525–572.

Andersson, S.A. and Perlman, M.D.: Lattice models for conditional independence in a multivariate normal models, *Ann. Statist.* **21** (1993), 525–572.

Andersson, S.A. and Wojnar, G.G.: Wishart distributions on homogeneous cones, *J. Theoret. Probab.* **17** (2004) 781–818.

Bonder, J. and Milnes, P.: Amenability: A survey for statistical applications of Hunt-Stein and related conditions on groups, *Z. Wahrsch. verw. Gebiete* **57** (1981) 103–128.

Dawid, A. and Lauritzen, S.L.: Hyper-Markov laws in the statistical analysis of decomposable graphical models, *Ann. Statist.* **21** (1993) 1272–1317.

Dey, D.K. and Srinivasan, C.: Estimation in a covariance matrix under Stein's loss, *Ann. Statist.* **13** (1985) 1581–1591.

Eaton, M.L.: Group Invariance Application in Statistics. Regional conference series in Probability and Statistics Vol. 1. Institute of Mathematical Statistics (1986).

Ishi, H.: Positive Rietz distributions on homogeneous cones, *J. Math. Soc. Japan* **52** (2000) 161–181.

James, W. and Stein, C.: Estimation with quadratic loss. In: Proc. Fourth Berkeley Symp. Math. Statist. Prob. (1961) **1** 361-380, Univ. California Press.

Konno, Y.: Inadmissibility of the maximum likelihood estimator of normal covariance matrices with the lattice conditional independence. *J. Multivariate Anal.* **79** (2001) 33-51.

Lauritzen, S.L.: Graphical Models, Oxford Univ. Press (1996).

Lehmann, E.L. and Casella, G.: Theory of Point Estimation 2nd ed., Springer (1998).

Letac, G.G. and Massam, H.: Wishart distributions for decomposable graphs, *Ann. Statist.* **35** (2007) 1278–1323.

Massam, H. and Neher, E.: Estimation and testing for lattice conditional independence models on Euclidean Jordan algebras, *Ann. Statist.* **26** (1998) 1051–1082.

Neher, E.: Transformations groups of the Andersson-Perlman cone, *J. Lie Theory* **9** (1999) 203–313.

Roverat, A.: Cholesky decomposition of hyper inverse Wishart matrix, *Biometrika* **87** (2000) 99–112.

Sheena, Y.: Unbiased estimator of risk for an orthogonally invariant estimator of a covariance matrix. *J. Japan Statist. Soc.* (1995) **25** 35-48.

Stein, C.: Inadmissibility of the usual estimator for the mean of a multivariate distribution. *Proc. Third Berkeley Symp. Math. Statist. Prob.* **1**, University of California Press (1956), 197–206.

Stein, C.: Estimation of mean of a multivariate normal distribution, *in Proceedings of Prague Symposium on Asymptotic Statistics* (1973), pp. 345–381.