重み付きの縮小推定量

湯浅 良太 統計思考院 助教

1 導入

都道府県ごとに様々なものに対する物価の指数を考えたり、複数の野球チームで打率や長打率、ホームラン数などの複数のデータが得られたり、といった状況ではデータが自然に行列構造を持つことが想定される.

こうした行列型のデータで平均の推定を行う場合、行列構造を無視してベクトルの形にして、ベクトル型の縮小推定量であるJames-Stein推定量を用いることと、行列型の縮小推定量であるEfron-Morris推定量を用いることで推定精度を向上させることが期待できる. どちらの推定量を用いる方が良いだろうか. データに基づいてどちらの推定量を用いるべきかを決めたい.

本ポスターは久保川達也教授(東京大学)との共同研究[3]に基づく.

2 提案する縮小推定量の形

行列データが$\boldsymbol{X} \sim N_{m\times p}(\boldsymbol{\Theta}, \boldsymbol{I_m} \otimes \boldsymbol{I_p})$に従って得られているとき、平均$\boldsymbol{\Theta}$を推定したい. このとき2つのよく知られた縮小推定量として

・James-Stein推定量

$$\widehat{\boldsymbol{\Theta}}^{JS} = \boldsymbol{X}\left(\boldsymbol{I_p} - \frac{mp-2}{\mathrm{tr}(\boldsymbol{X}'\boldsymbol{X})}\boldsymbol{I_p}\right)$$

・Efron-Morris推定量

$$\widehat{\boldsymbol{\Theta}}^{EM} = \boldsymbol{X}\left(\boldsymbol{I_p} - (m-p-1)(\boldsymbol{X}'\boldsymbol{X})^{-1}\right)$$

がある.

重み付きの縮小推定量

$$\widehat{\boldsymbol{\Theta}}^{WSE} = \widehat{w}\widehat{\boldsymbol{\Theta}}^{EM} + (1-\widehat{w})\widehat{\boldsymbol{\Theta}}^{JS}$$

を提案する.

3 重みの決定

重み$\widehat{w}$をどのように決めるかが重要となる.

・決め方1：SURE(Steinの不偏なリスク関数の推定量)を最小化

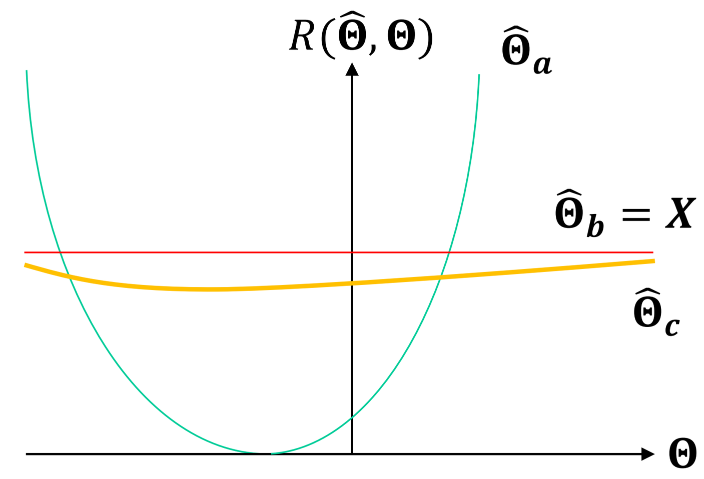
・決め方2：事前分布に対して行列正規分布とベクトル正規分布の混合を仮定した、経験ベイズ推定量として導出

詳細は[3]を参照. 以降詳細を省略している部分は基本的に[3]を参照.

4 推定精度に関する理論

推定量を$\widehat{\boldsymbol{\Theta}}$とすると、その推定精度の良し悪しを損失関数と呼ばれるもので測る. 例えば$L(\widehat{\boldsymbol{\Theta}}, \boldsymbol{\Theta}) = ||\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}||^2 = \sum_{i,j}(\widehat{\Theta}_{ij} - \Theta_{ij})^2$という形のように、真の平均から離れた値をとる程大きな値をとり、真の平均を上手く推定できている時程小さな値をとるようなものを考える. このときリスク関数を$\boldsymbol{X}$に関する期待値で、$R(\boldsymbol{\Theta}, \widehat{\boldsymbol{\Theta}}) = E[L(\widehat{\boldsymbol{\Theta}}, \boldsymbol{\Theta})]$と定める.

右の図は3つの推定量に対して、リスクの値を表したイメージ図である. どの推定量が良いか.



ここでは、リスクが最も悪くなる場合を比較するミニマックス性という基準を考える. つまり、$\sup_{\Theta} R(\boldsymbol{\Theta}, \widehat{\boldsymbol{\Theta}})$を最小化する推定量をミニマックスであるという. $\widehat{\boldsymbol{\Theta}}_{\boldsymbol{a}}$はミニマックスではない. 正規分布では$\widehat{\boldsymbol{\Theta}}_{\boldsymbol{b}} = \boldsymbol{X}$はミニマックスである事が知られている. $\widehat{\boldsymbol{\Theta}}_{\boldsymbol{c}}$は$\widehat{\boldsymbol{\Theta}}_{\boldsymbol{b}}$よりグラフが下にあり、ミニマックスである.

重みを決め方1,2のどちらのケースでもミニマックス性の成立を証明することができる.

5 分散が未知の場合

分散が未知の場合、$\boldsymbol{X} \sim N_{m\times p}(\boldsymbol{\Theta}, \boldsymbol{I_m} \otimes \boldsymbol{\Sigma}), \boldsymbol{S} \sim W_p(\boldsymbol{n}, \boldsymbol{\Sigma})$を考える. とくに$n < p$の場合も含めてミニマックス性を証明することができる. $n < p$の場合には一般化逆行列を用いる.

6 Positive-part 推定量による改良

$\boldsymbol{XX}'$や$\boldsymbol{XS}^{+}\boldsymbol{X}'$の固有値を用いて縮小推定量を書き直す事ができる. 一般的に固有値の縮小を行う推定量では、固有値を縮小しすぎて負になってしまう場合には0で打ち切るPositive-part推定量を用いることで改良することができる. 分散が未知の場合には[1],[2]に証明がある. 分散が既知の場合にも分散が未知の場合と同様の方法で証明を行う事が出来る.

7 推定量の特徴

$\boldsymbol{XX}'$や$\boldsymbol{XS}^{+}\boldsymbol{X}'$の固有値が全て近い値をとる場合は、すべての固有値を同じように縮小することが望ましく、James-Stein推定量がよい. 固有値がばらばらの値をとる場合には、それぞれの固有値の大きさに対応した縮小が望ましく、Efron-Morris推定量が良い. 提案した重みはちょうどそのようになっており、数値的にもそれぞれのよい部分を活かしていることが分かる.

参考文献

[1] H. Tsukuma, Shrinkage minimax estimation and positive-part rule for a mean matrix in an elliptically contoured distribution, Statist. Probab. Lett. 80 (3–4) (2010) 215–220.

[2] H. Tsukuma, T. Kubokawa, A unified approach to estimating a normal mean matrix in high and low dimensions, J. Multivariate Anal. 139 (2015) 312–328.

[3] R. Yuasa, T. Kubokawa, Weighted shrinkage estimators of normal mean matrices and dominance properties, J. Multivariate Anal. 194 (2023)