行列正規分布に関する一般化ベイズ縮小推定量の優越性

湯浅 良太 統計思考院 助教

1 概要

多変量正規分布の平均を推定する際に、最尤推定量である標本平均を用いる事が自然であるが、3次元以上の場合はデータをある方向に縮小する縮小推定と呼ばれる推定量を用いて推定精度を改善する事ができる. 本ポスターでは、久保川達也教授(東京大学)との共同研究[5]に基づき、データが行列として得られるような場合の縮小推定について考える. 特に、行列正規分布に関するベイズ縮小推定量を提案し、標本平均や標本分散共分散行列を用いるに比べより良い推定精度を持つという優越性と呼ばれる性質について議論する.

2 縮小推定について

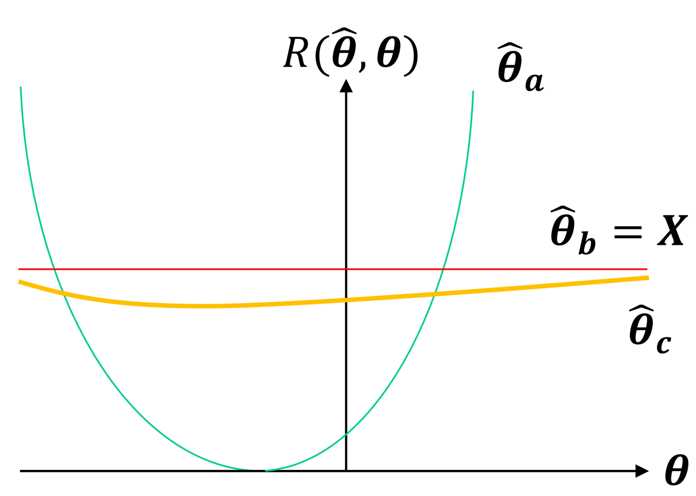
縮小推定量は、平均に事前分布を仮定するベイズ推定から自然に生じる. データが$\boldsymbol{X} \sim N_m(\boldsymbol{\theta}, \boldsymbol{I_m})$と多変量正規分布に従うとする. $\boldsymbol{X} = (X_1, \dots, X_m)'$, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)'$と表し、事前分布を$\theta_i \sim N(\mu, A)$と仮定する. これは、$\boldsymbol{X}$がシーズン初め頃の野球選手の打率で、シーズン終了時の打率を推定する為に$\boldsymbol{\theta}$を推定する場合だと、プロ野球選手では能力がある程度共通の分布に従っているだろうという状況を想定する事に対応する. 他には、地価を推定する際に似たような傾向を持つだろう地域のデータを用いる事を想定する事などにも対応しうる.

事前分布を課した結果として、ベイズ推定量として事後分布による平均$\hat{\theta}_i^{Bay} = \mu + \frac{A}{1+A}(X_i - \mu) = \frac{A}{1+A}X_i + \frac{1}{1+A}\mu$が得られる. データと事前分布の平均の重み付き和となっている. このμと$\frac{1}{1+A}$を最尤推定量で推定すると$\hat{\theta}_i^{S} = \left(1 - \frac{m-3}{\sum_{i=1}^m (X_i - \bar{X})^2}\right) X_i + \frac{m-3}{\sum_{i=1}^m (X_i - \bar{X})^2}\bar{X}$という形の推定量になる. μを0とすると、James-Stein推定量 : $\hat{\theta}_i^{JS} = \left(1 - \frac{m-2}{\sum_{i=1}^m X_i^2}\right) X_i$が得られる.

3 推定量の比較基準

統計的決定理論と呼ばれる数理統計の枠組みでの推定量の比較基準について簡単に説明する. 引き続き$\boldsymbol{X} \sim N_m(\boldsymbol{\theta}, \boldsymbol{I_m})$の場合を考える. 推定量を$\hat{\boldsymbol{\theta}}$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$L(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) = ||\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}||^2 = \sum_{i=1}^m (\hat{\theta}_i - \theta_i)^2$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Xに関する期待値で、$R(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) = E[L(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta})]$と定める.

右の図は3つの推定量に対して、リスクの値をグラフで表している. どの推定量が良い推定量であるか.



グラフで比べると、常に下にある方が良い推定量である. これを優越性と呼ぶ. つまり、$R(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_1) \le R(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_2)$が全ての$\boldsymbol{\theta}$について成立し、ある$\boldsymbol{\theta_0}$で$R(\boldsymbol{\theta_0}, \hat{\boldsymbol{\theta}}_1) < R(\boldsymbol{\theta_0}, \hat{\boldsymbol{\theta}}_2)$が成立するとき、$\hat{\boldsymbol{\theta}}_1$は$\hat{\boldsymbol{\theta}}_2$を優越すると言う. 優越される事がない事を許容的であるという. $\hat{\boldsymbol{\theta}}_b$は$\hat{\boldsymbol{\theta}}_c$に優越されるので許容的ではない. $\hat{\boldsymbol{\theta}}_c$を用いる方が良い. しかし、$\hat{\boldsymbol{\theta}}_a$も許容的になってしまう. 別の基準として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、$\sup_{\boldsymbol{\theta}} R(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$を最小化する推定量をミニマックスであるという. $\hat{\boldsymbol{\theta}}_a$はミニマックスではない. 正規分布では$\hat{\boldsymbol{\theta}}_b = \boldsymbol{X}$はミニマックスであり、$\hat{\boldsymbol{\theta}}_c$もミニマックスである.

4 ベクトル正規分布に対するベイズ縮小推定

許容的でかつミニマックスである事は、良い推定量である1つの保証となる. 許容的であるためには、事前分布がimproperであることも許した場合のベイズ推定量である一般化ベイズ推定量でなければならないことが知られている. そこで、本研究ではミニマックスな一般化ベイズ推定量を導く事を目的とする. 一般に一般化ベイズ推定量は複雑な積分を含む形になり扱いづらい. それに関連してベクトルの場合には、[1]により積分を含まないシンプルな形でのミニマックスな一般化ベイズ推定量が得られている. さらにこの推定量に関して、[2][3]によって近年許容性に関する研究も行われている. そこでは次のような事前分布が考えられている.

- $\boldsymbol{\theta} | \lambda, \sigma^2 \sim N_m(\boldsymbol{0}, \sigma^2(\lambda^{-1} - 1)\boldsymbol{I}_m)$
- $\lambda \sim \lambda^a (1-\lambda)^b 1_{[0,1)}$
- $\pi(\sigma^{-2}) \propto (\sigma^{-2})^e$

これを行列の場合に拡張したものを扱う.

5 本研究での事前分布と一般化ベイズ推定量

[1]によるベクトルの場合の事前分布を行列の場合に拡張する. $\boldsymbol{X} \sim N_{m\times p}(\boldsymbol{\Theta}, \boldsymbol{I_m} \otimes \boldsymbol{\Sigma})$, $\boldsymbol{S} \sim W_p(n, \boldsymbol{\Sigma})$とする. 事前分布は次の通り.

- $\boldsymbol{\Theta} | \boldsymbol{\Sigma}, \boldsymbol{\Omega} \sim N_{m\times p}(\boldsymbol{0}, \boldsymbol{\Omega} \otimes \boldsymbol{\Sigma})$
- $\pi(\boldsymbol{\Omega}) \propto |\boldsymbol{I_m} + \boldsymbol{\Omega}|^{-\frac{a}{2}-m} |\boldsymbol{\Omega}|^{\frac{c}{2}}$
- $\pi(\boldsymbol{\Sigma}^{-1}) \propto |\boldsymbol{\Sigma}^{-1}|^{\frac{b-1}{2}}$

これは[4]のconcluding remarksであげられている事前分布であるが、特にその推定量の導出や推定量に関する議論の性質などはなされていない. $b = a - n + p$とすると、次用ようなシンプルな形で推定量が得られる.

- $\boldsymbol{X} - k_0 \boldsymbol{X} \{\boldsymbol{I_p} + (1 - k_0)\boldsymbol{S}^{-1}\boldsymbol{X}'\boldsymbol{X}\}^{-1}$が平均行列に関するミニマックスな一般化ベイズ推定量である.
- $\widetilde{\boldsymbol{\Sigma}} - k_0 \widetilde{\boldsymbol{\Sigma}} \{\boldsymbol{I_p} + (1 - k_0)\boldsymbol{S}^{-1}\boldsymbol{X}'\boldsymbol{X}\}^{-1}$が分散共分散行列に関する一般化ベイズ推定量である.
    - $\widetilde{\boldsymbol{\Sigma}} = \frac{1}{m+c+1}\boldsymbol{S}$, k_0はm, pと事前分布のパラメータによる.

ミニマックス性や$\boldsymbol{S}/n$に対する優越性のためにはm, n, pや事前分布のパラメータに関する条件が必要となる. とくに$\boldsymbol{S}/n$に対する優越性については[5]での議論では$p > m$である必要がある. 詳しい条件や議論については[5]を参照.

許容性に関しては本研究や[2][3]をもとに今後考えたい.

参考文献

[1] Maruyama, Y. and Strawderman, W. E. (2005). A new class of generalized Bayes minimax ridge regression estimators. Ann. Statist., 33, 1753-1770.

[2] Maruyama, Y. and Strawderman, W. E. (2020). Admissible Bayes equivariant estimation of location vectors for spherically symmetric distributions with unknown scale. Ann. Statist. 48, 1052–71.

[3] Maruyama, Y., and Strawderman, W. E. (2021). Admissible estimators of a multivariate normal mean vector when the scale is unknown. Biometrika, 108(4), 997-1003.

[4] Tsukuma, H. (2009). Generalized Bayes minimax estimation of the normal mean matrix with unknown covariance matrix. J. Multivariate Anal., 100, 2296–2304.

[5] Yuasa, R and Kubokawa, T (2021). Generalized Bayes Estimators with Closed forms for the Normal Mean and Covariance Matrices. arXiv (https://arxiv.org/abs/2108.06041)