Covariance matrix estimation of an e. s. d. in high dimensional setting

D. Fourdrinier [1] M.A. Haddouche [2] F. Mezoued [3]

21 December 2018

INSA de Rouen (LITIS)

[1] Université de Normandie, UNIROUEN, UNIHAVRE, INSA Rouen, LITIS
France

[2] École Nationale Supérieure de Statistique et de l'Économie Appliquée, LAMOPS
Algeria
Université de Normandie, UNIROUEN, UNIHAVRE, INSA Rouen, LITIS
France

[3] École Nationale Supérieure de Statistique et de l'Économie Appliquée, LAMOPS
Algeria

Outline

Introduction

Why we care about the Covariance matrix?

- Statistical Network Analysis Network of Neurons: Which part of brain communicates during a given task?
- Climate Data Analysis The climate correlations among geographical regions.
- Financial Data Analysis Portfolio management in Finance.

The considered model

Consider the additive model

$$\boldsymbol{Y} = \boldsymbol{M} + \mathcal{E}, \qquad \mathcal{E} \sim E\left(\mathbf{0}_{\boldsymbol{np}}, I_n \otimes \Sigma\right) \tag{1.1}$$

where

- $\boldsymbol{Y}$ is an observed $n \times p$ matrix such that $p > n$

- $\boldsymbol{M}$ is the unknown matrix parameter.

- $\mathcal{E}$ is an e. d. noise with covariance matrix proportional to Σ.

Canonical form

As $rank(\boldsymbol{M}) = q < n < p$, there exists a semi-orthogonal $n \times (n - q)$ matrix Q_2 such that

$$Q_2^\top \boldsymbol{M} = \mathbf{0}\,.$$

Complete Q_2 with Q_1 to form an $n \times n$ orthogonal matrix $Q = (Q_1 Q_2)$:

$$Q^\top \boldsymbol{Y} = \begin{pmatrix} Q_1^\top \\ Q_2^\top \end{pmatrix} \boldsymbol{Y} = \begin{pmatrix} \boldsymbol{Z} \\ \boldsymbol{U} \end{pmatrix} = \begin{pmatrix} Q_1^\top \\ Q_2^\top \end{pmatrix} \boldsymbol{M} + Q^\top \mathcal{E} = \begin{pmatrix} \theta \\ \mathbf{0} \end{pmatrix} + Q^\top \mathcal{E},$$

- Q_1 and Q_2 are resp. $n \times q$, $n \times m$, with $m = n - q$.

- $\boldsymbol{Z}$ and $\boldsymbol{U}$ are resp. $q \times p$, $m \times p$.

Density

Assume that $\boldsymbol{\mathcal{E}} = \boldsymbol{Y} - \boldsymbol{M}$ has a density with respect to the Lebesgue measure (i.e. Σ is invertible) of the form

$$\boldsymbol{\mathcal{E}} \mapsto |\Sigma|^{-n/2} f\{\mathrm{tr}\left(\boldsymbol{\mathcal{E}}\,\Sigma^{-1}\,\boldsymbol{\mathcal{E}}^{\top}\right)\}.$$

The function f is called the density generator function.

Ex: The normal multivariate dist.

$$f(\cdot) \propto exp\{-\frac{1}{2}(\cdot)\}.$$

$\rightsquigarrow$ The density of $Q^\top \mathcal{E}$ is the same as that of $\mathcal{E}$.

$\rightsquigarrow$ It is follows that $(\boldsymbol{Z}^\top \boldsymbol{U}^\top)^\top = Q^\top \boldsymbol{Y}$ has an e. s. d. around the matrix $(\boldsymbol{\theta}^\top \mathbf{0}^\top)^\top$.

The joint density of $\boldsymbol{Z}$ and $\boldsymbol{U}$ is

$$(\boldsymbol{z}, \boldsymbol{u}) \mapsto |\Sigma|^{-n/2} f\left[\operatorname{tr}\{(\boldsymbol{z}-\theta)\Sigma^{-1}(\boldsymbol{z}-\theta)^\top\} + \operatorname{tr}\{\boldsymbol{u}\Sigma^{-1}\boldsymbol{u}^\top\}\right]. \quad (1.2)$$

Bellow $E_{\theta,\Sigma}$ will be the expectation w.r.t the density (1.2).

Related expectations

Let

$$F^*(t) = \frac{1}{2}\int_t^\infty f(u)\,du \quad \text{and} \quad F^{**}(t) = \frac{1}{2}\int_t^\infty F^*(u)\,du.$$

Bellow $E^*_{\theta,\Sigma}$ will be the expectation w.r.t. the density

$$(\boldsymbol{z},\boldsymbol{u}) \mapsto \frac{1}{K^*}|\Sigma|^{-n/2}\,F^*\big[\,\mathrm{tr}\{\,(\boldsymbol{z}-\theta)\Sigma^{-1}(\boldsymbol{z}-\theta)^\top\} + \mathrm{tr}\{\Sigma^{-1}\,s\}\,\big]$$

and $E^{**}_{\theta,\Sigma}$ the expectation w.r.t. the density

$$(\boldsymbol{z},\boldsymbol{u}) \mapsto \frac{1}{K^{**}}|\Sigma|^{-n/2}\,F^{**}\big[\,\mathrm{tr}\{\,(\boldsymbol{z}-\theta)\Sigma^{-1}(\boldsymbol{z}-\theta)^\top\} + \mathrm{tr}\{\Sigma^{-1}\,s\}\,\big]$$

where K^* and K^{**} are normalizing constant.

Densities subclass

We reduce the class of densities in (1.2) to the subclass such that

$$c \leq \frac{F^*(t)}{f(t)} \leq b\,,$$

Examples:

- The logistic type distribution

$$f(t) \propto \frac{\exp\left(-\beta t - \gamma\right)}{\left(1 + \exp\left(-\beta\, t - \gamma\right)\right)^2}$$

where $\beta > 0$ and $\gamma > 0$. In this case

$$c = \frac{1}{2\beta} \qquad b = \frac{(1 + e^{-\gamma})}{2\beta}.$$

- The m. n. d, $c = b = 1$ since $F^* = f$. See [2] for more examples.

D. Fourdrinier, F. Mezoued, and W. E. Strawderman. Bayes minimax estimators of a location vector for densities in the Berger class. Electronic Journal of Statistics. *6:783–809, 2012.*

Usual estimators

The usual estimators are of the form

$$\hat{\Sigma}_a = a\,S$$

with $S = U^\top U$ and a is a positive constant.

We consider the quadratic risk

$$R(\hat{\Sigma}, \Sigma) = E_{\theta,\Sigma}[\mathrm{tr}(\hat{\Sigma}\Sigma^{-1} - I_p)^2], \tag{1.3}$$

where $\hat{\Sigma}$ estimates Σ. For $\hat{\Sigma} = \hat{\Sigma}_a$, the best constant a is

$$a_o = \frac{1}{K^{**}(m+p+1)}\,. \tag{1.4}$$

$\hat{\Sigma}_{a_o}$ is unbiased.

What is wrong with $\hat{\Sigma}_{a_o}$

High Dimensional Setting: p is large compared to n (S is not invertible).

$\hat{\Sigma}_{a_o} = a_o S$ is not a good estimator of Σ. In fact,

- larger eigenvalues are overestimated; smaller eigenvalues are underestimated,
- $\hat{\Sigma}_{a_o}$ is inadmissible.

General estimators

Competitive estimators

We consider competitive estimators of the form

$$\hat{\Sigma}_G = a_o \left(S + SS^{+} G(\boldsymbol{Z}, S) \right),$$

where S^{+} is the Moore-Penrose inverse of S and $SS^{+} G(\boldsymbol{Z}, S)$ is a $p \times p$ symmetric matrix function.

$\rightsquigarrow$ $\hat{\Sigma}_G$ dominates $\hat{\Sigma}_{a_o}$ if the risk difference

$$\Delta(G) = R(\hat{\Sigma}_G, \Sigma) - R(\hat{\Sigma}_{a_o}, \Sigma) \leq 0,$$

for all Σ, with strict inequality for some Σ.

Risk difference

The following proposition insures the finiteness of the risk difference $\Delta(G)$.

Proposition 2.1

Assume that $E_{\theta,\Sigma}\left[\operatorname{tr}\left(\Sigma^{-1}S\right)^2\right] < \infty$ *and* $E_{\theta,\Sigma}\left[\operatorname{tr}(\Sigma^{-1}SS^+G)^2\right] < \infty$*. Then*

$$\begin{aligned} \Delta(G) = a^2 E_{\theta,\Sigma}\left[\operatorname{tr}\left(\Sigma^{-1}SS^+(2S+G)\Sigma^{-1}SS^+G\right)\right] \\ -2aE_{\theta,\Sigma}\left[\operatorname{tr}\left(\Sigma^{-1}SS^+G\right)\right] < \infty. \end{aligned} \tag{2.1}$$

$\rightsquigarrow$ Under which conditions on G, $\hat{\Sigma}_G$ dominates $\hat{\Sigma}_{a_0} = a_0 S$?

Stein-Haff type identity

The following Lemma is useful to deal with the dependence of the unknown parameter Σ^{-1} in the risk difference (2.1).

Lemma 2.2

Let $V(z,s)$ be a $p \times p$ matrix function such that, for any fixed z, $V(z,s)$ is weakly differentiable with respect to s and assume $E_{\theta,\Sigma}\left[\left|\operatorname{tr}\left(\Sigma^{-1} S S^{+} V\right)\right|\right] < \infty$. Then we have

$$E_{\theta,\Sigma}\left[\operatorname{tr}\left(\Sigma^{-1} S S^{+} V\right)\right] = K^{*} E^{*}_{\theta,\Sigma}\Big[\operatorname{tr}\Big((m - p \wedge m - 1)\, S^{+} V \\ + 2\, S S^{+}\, \mathcal{D}_s\{S S^{+} V\}^{\top}\Big)\Big].$$

Here $\mathcal{D}_S$ denotes the Haff-operator whose generic term is

$$d_{ij} = \frac{1}{2}(1 + \delta_{ij})\frac{\partial}{\partial S_{ij}}$$

with $\delta_{ij} = 1$ when $i = j$ and $\delta_{ij} = 0$ when $i \neq j$.

Dominance result

Theorem 2.3

Consider a density as in (1.2) such that $c \leq F^*(t)/f(t) \leq b$. If

$$\operatorname{tr}\left[2\, S^+ S\, \mathcal{D}_S\{SS^+ G\} - S^+ G\right] \geq 0\,,$$

the estimator $\hat{\Sigma}_G$ improves over $\hat{\Sigma}_{a_o}$ as soon as

$$\operatorname{tr}\Big[2S^+ S\, \mathcal{D}_S\{SS^+ T^*\}^\top - S^+ T^* \\ - 2(p+m+1)\frac{c^2}{b^2}\left(2S^+ S\, \mathcal{D}_S\{SS^+ G\} - S^+ S\right)\Big] \leq 0\,.$$

where

$$T^* = 4(S+G)\,\mathcal{D}_s\{S^+ SG\} + G\left(2mI_p - (p-m+1)S^+ G\right).$$

Orthogonally invariant estimators

Konno's estimator

Decompose S and S^+ as

$$S = H_1\, L\, H_1^\top \quad \text{and} \quad S^+ = H_1\, L^{-1}\, H_1^\top$$

where

H_1 is a $p \times m$ semi-orthogonal matrix ($H_1^\top H_1 = I_m$).

$H = (H_1 \; H_2)$ is $p \times p$ orthogonal matrix.

H_2 is a $p \times (p - m)$ matrix .

and

$$L = \mathrm{diag}(l_1, \ldots, l_m) \quad \text{and} \quad l_1 > l_2 > \cdots > l_m > 0\,.$$

Set

$$G(X,S) = \frac{t}{\mathrm{tr} S^+} I_p$$

where

$$t \text{ is a positive constante}.$$

The resulting estimator $\hat{\Sigma}_G = a_o \, (\, S + SS^+ G(X,S))$ is

$$\hat{\Sigma}_G = a_o \, H_1 \, \mathrm{diag}\big(l_1 + \frac{t}{\mathrm{tr} S^+}, l_2 + \frac{t}{\mathrm{tr} S^+}, \ldots, l_m + \frac{t}{\mathrm{tr} S^+}\big) H_1^\top .$$

Applying Theorem (2.3), it can be seen that $\hat{\Sigma}_G$ improves over $\hat{\Sigma}_{a_o}$ as soon as

$$0 < t < \frac{2(p-m-1)}{(p-m+3)(p-m+1)} \left((p+m+1)\frac{c^2}{b^2} - (p+2) \right) .$$

Generalizing the Konno's estimator

Set

$$G(X,S) = \frac{1}{\mathrm{tr} S^{+}} t\left(\frac{1}{\mathrm{tr} S^{+}}\right) I_p$$

and

$$t\left(\frac{1}{\mathrm{tr} S^{+}}\right) = \frac{\alpha}{\beta + 1/\mathrm{tr} S^{+}},$$

where $\alpha \geq 0$ and $\beta > 0$.

Applying Theorem (2.3), it can be seen that $\hat{\Sigma}_G$ improves over $\hat{\Sigma}_{a_o}$ since

$$\frac{\alpha}{\beta} \leq \min(t_1, t_2)$$

where

$$t_1 = \frac{2\,(p-m-1)}{(p-m+1)\,(p-m+3)}\left((p+m+1)\frac{c^2}{b^2} - p - 2\right).$$

$$t_2 = \frac{1}{2}\left[2\,p - 3\,m + 5 - (p+m+1)\frac{c^2}{b^2} - 4\,m^2\right].$$

Perspectives

Σ non-invertible

Let Σ be a $p \times p$ matrix with rank = $r < p$.

While $\mathcal{E}$ has no density w.r.t the Lebesgue measure, it has a density w.r.t the Hausdorff measure given by

$$\prod_{i=1}^{r} \lambda_i^{-n/2} f\left(\operatorname{tr}\left\{\mathcal{E}^\top \Sigma^+ \mathcal{E}\right\}\right)$$

where λ_i are the non zeros eigen-values of Σ. See [3] and [4]

J.A Díaz-García, V. Leiva-Sánchez and M. Galea. Singular Elliptical Distribution: Density and Applications. Communications in Statistics - Theory and Methods. *5:665–681, 2002.*

J.A Díaz-García, G. González-Farías. Singular random matrix decompositions. Journal of Multivariate Analysis. *1:109 – 122, 2005.*

Canonical form of the additive model

$$Q^\top \boldsymbol{Y} = \begin{pmatrix} Q_1^\top \\ Q_2^\top \end{pmatrix} \boldsymbol{Y} = \begin{pmatrix} \boldsymbol{Z} \\ \boldsymbol{U} \end{pmatrix} = \begin{pmatrix} Q_1^\top \\ Q_2^\top \end{pmatrix} \boldsymbol{M} + Q^\top \mathcal{E} = \begin{pmatrix} \theta \\ \boldsymbol{0} \end{pmatrix} + Q^\top \mathcal{E},$$

The density of $Q^\top \mathcal{E}$ is it the same as that of $\mathcal{E}$?

Estimation problems in the singular elliptical setting

Estimation of the cov. matrix Σ under the invariant squared loss

$$L(\hat{\Sigma}, \Sigma) = \mathrm{tr}\left(\hat{\Sigma}\Sigma^{+} - \Sigma\Sigma^{+}\right)^{2}.$$

Estimation of the precision matrix Σ^{+} under the Frobenius loss

$$L(\hat{\Sigma}^{+}, \Sigma^{+}) = \parallel \hat{\Sigma}^{+} - \Sigma^{+} \parallel_F^2 .$$

Estimation of the discriminant coef. $\eta = \Sigma^{+}\boldsymbol{M}$ under the squared loss

$$L(\hat{\eta}, \eta) = \parallel \hat{\eta} - \eta \parallel_2^2 .$$

The singular gaussian case

Orthogonally invariant setting, see [5]

D. Chételat, M. T. Wells. Improved second order estimation in the singular multivariate normal model. Journal of Multivariate Analysis , *147:11 – 19, 2016.*

Stein-Haff type identity of the form

$$E\left[\operatorname{tr}\left(\Sigma^{+}H_1\Phi(L)H_1^{\top}\right)\right] = E\left[\sum_{i=1}^{n\wedge r}\left\{(|n-r|-1)\frac{\phi_i}{l_i} + 2\frac{\partial\phi_i}{\partial l_i} + 2\sum_{j>i}\frac{\phi_i-\phi_j}{l_i-l_j}\right\}\right]$$

where $rank(S) = n \wedge r$.

Generalization to the singular elliptical setting

- How to derive a Stein-Haff identity to the elliptical singular dist. in the orthogonally invariant setting?
- How to derive an identity of the form

$$E\left[\mathrm{tr}\left(\Sigma^{+}SS^{+}G(S)\right)\right]?$$

Thank you for your attention !

For references and other details, I can be reached at

mohamed-anis.haddouche@etu.univ-rouen.fr

D. Fourdrinier, F. Mezoued, and M.T. Wells.
Estimation of the inverse scatter matrix of an elliptically symmetric distribution.
Journal of Multivariate Analysis, 143:32–55, 2016.

D. Fourdrinier, F. Mezoued, and W. Strawderman.
Bayes minimax estimators of a location vector for densities in the Berger class
Electronic Journal of Statistics, 6:783–809, 2012.

J.A Díaz-García, V. Leiva-Sánchez and M. Galea
Singular Elliptical Distribution: Density and Applications
Communications in Statistics - Theory and Methods , 5:665-681, 2002.

J.A Díaz-García, G. González-Farías
Singular random matrix decompositions
Journal of Multivariate Analysis , 1:109 – 122, 2005.

D. Chételat, M. T. Wells
Improved second order estimation in the singular multivariate normal model
Journal of Multivariate Analysis , 147:11 – 19, 2016.