The 2nd ims-APRM2012 Meeting (July 2-4, 2012 at Tsukuba)

Shrinkage estimators for large covariance matrices in multivariate normal distribution

Yoshihiko KONNO, Japan Women's University

June 29, 2012

Outline

In this talk we consider the problem of estimating large covariance matrices in a decision-theoretic manner when the dimension of variables, p, is larger than the number of observations, n. Population distributions include not only real multivariate distributions but also complex multivariate distributions. See Konon (JMVA, 2009)

Outline

(1) Notation and ettup of the problem

(2) Class of estimators and Stein's unbiased risk estimate(SURE) for this problem

(3) Improved estimators

(4) Concluding remarks

Notation and settup

⋆ Let $X_1,\, X_2,\, \dots,\, X_n \sim N_p(0,\, \mathbf{\Sigma})$, independently. Each $X_i\,(i = 1,\, 2, \dots,\, n)$ is a p-variate column vector. Here $\mathbf{\Sigma}$ is unknown $p \times p$ positive definite matrix.

⋆ n equals "sample size minus 1" and p is dimension of the variables.

⋆ Defien Wishart matrix($p\times p$ matrix) as $\boldsymbol{W} := \sum_{k=1}^{n} X_i X_i'$. Here "$'$" denotes the transpose of the matrix and the vector.

⋆ Consider the estimation problem of covariance matrix $\mathbf{\Sigma}$ under a loss function

$$\text{loss: } L(\widehat{\mathbf{\Sigma}},\, \mathbf{\Sigma}) = \mathrm{Tr}\,(\widehat{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1} - I_p)^2; \qquad \text{risk: } R(\widehat{\mathbf{\Sigma}},\, \mathbf{\Sigma}) := \mathbb{E}[L(\widehat{\mathbf{\Sigma}},\, \mathbf{\Sigma})].$$

Here $\widehat{\mathbf{\Sigma}}$ is an estimator for $\mathbf{\Sigma}$, and $\boldsymbol{I}_p$ denotes $p \times p$ identity matrix.

⋆ Our goal is to compare the risk among the estimators uniformly in $\mathbf{\Sigma}$.

Some remarks

⋆ Without loss of generality, we can assume that the mean is zero.

⋆ The Wishart matrix Wishart $\boldsymbol{W}$ is positive-deffinite $\Longleftrightarrow n \geq p$.

⋆ When $p > n$, the distribution of $\boldsymbol{W}$ exists, but its density does not exists in usual sense.

⋆ Examples of loss functions which are invariant under the group of the transformations $\widehat{\boldsymbol{\Sigma}} \mapsto \boldsymbol{A}\widehat{\boldsymbol{\Sigma}}\boldsymbol{A}'$; $\boldsymbol{\Sigma} \mapsto \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}'$ ($\boldsymbol{A}$ is $p \times p$ nonsingular)

$$L(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) = \mathrm{Tr}\,(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1} - I_p)^2;\ L_S(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) = \mathrm{Tr}\,(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}) - \log \mathrm{Det}(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}) - p.$$

Here Det is the determinants of the matrix. But we can not use L_S if $p > n$ as we can not evaluate the expected value of L_S of $n^{-1}\boldsymbol{W}$.

Drawback of $n^{-1}\boldsymbol{W}$

$\star$ $\mathbb{E}[n^{-1}\boldsymbol{W}] = \boldsymbol{\Sigma}$. But the eigenvalues of $n^{-1}\boldsymbol{W}$ are more spread-out than those of $\boldsymbol{\Sigma}$.

$\star$ When $n \geq p$, $\boldsymbol{\Sigma}$ is p.d. But when $p > n$, $n^{-1}\boldsymbol{W}$ semi-p.d (not invertable!). This makes everything difficult!

$\star$ Under the loss L_S, the eigenvalues of $n^{-1}\boldsymbol{W}$ are shrunk and expanded to obtain improved estimators. See Stein (Prague Symp, 1977), Dey and Srivastava (AOS, 1985), Haff (AOS, 1991) *et al.*

$\star$ To evaluate risk, SURE(Stein's unbiased risk estimate) and eigenvalue-caluculus were worked well. $\rightarrow$ Do these senario work for the case when $p > n$ and the loss L?

$\star$ When $p > n$, Ledoit and Wolf (JMVA, 2004) considered a loss $\mathrm{Tr}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma})^2$ (not invariant!) and obtained asymptotically efficient estimators(roughly speaking) among the linear combinations of $n^{-1}\boldsymbol{W}$ and $\boldsymbol{I}_p$.

Class of estimators

⋆ Decompose $\boldsymbol{W} = \sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i'$; $\ell_1 \geq \cdots \geq \ell_n$ are the eigenvalues of $\boldsymbol{W}$, and

$$\boldsymbol{W} = \boldsymbol{O}_1 \boldsymbol{L} \boldsymbol{O}_1', \qquad \boldsymbol{L} = \mathrm{Diag}(\ell_1, \ldots, l_n);$$

$\boldsymbol{O}_1$ is a $p \times n$ semi-orthogonal s.t. $\boldsymbol{O}_1' \boldsymbol{O}_1 = \boldsymbol{I}_n$.

Our estimators
$$\widehat{\boldsymbol{\Sigma}} = \boldsymbol{O}_1 \boldsymbol{\Psi}(\boldsymbol{L}) \boldsymbol{O}_1', \tag{1}$$

Here $\boldsymbol{\Psi} := \boldsymbol{\Psi}(\boldsymbol{L}) = \mathrm{Diag}(\psi_1, \psi_2, \ldots, \psi_n)$ and $\psi_k := \psi_k(\boldsymbol{L}) (k = 1, 2, \ldots, n)$ are differentiable functions from $\mathbb{R}^n_{\geq}$ to $\mathbb{R}$.

Our goal Evaluate the risk $\mathbb{E}[\mathrm{Tr}\,(\widehat{\boldsymbol{\Sigma}} \boldsymbol{\Sigma}^{-1} - \boldsymbol{I}_p)^2]$, which depends on $\boldsymbol{\Sigma}$ generally.

SURE(Stein's Unbiased risk estimate), part 1

$\star$ Derive unbiased estimator $\widehat{\boldsymbol{R}}(\widehat{\boldsymbol{\Sigma}})$ (which depends only on $\psi_1, \ldots, \psi_n\ \ell_1, \ldots, \ell_n$) for the risk $\mathbb{E}[\mathrm{Tr}\,(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1} - \boldsymbol{I}_p)^2]$, i.e.,

$$\mathbb{E}[\mathrm{Tr}\,(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1} - \boldsymbol{I}_p)^2] = \mathbb{E}[\widehat{\boldsymbol{R}}(\widehat{\boldsymbol{\Sigma}})]$$

$\star$ As $\mathbb{E}[\mathrm{Tr}\,(n^{-1}\boldsymbol{W}\boldsymbol{\Sigma}^{-1} - \boldsymbol{I}_p)^2]$ is constant,

$$\widehat{\boldsymbol{R}}(\widehat{\boldsymbol{\Sigma}}) \leq \mathbb{E}[\mathrm{Tr}\,(n^{-1}\boldsymbol{W}\boldsymbol{\Sigma}^{-1} - \boldsymbol{I}_p)^2] = \text{constant}$$

implies that

$$\mathbb{E}[\mathrm{Tr}\,(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1} - \boldsymbol{I}_p)^2] \leq \mathbb{E}[\mathrm{Tr}\,(n^{-1}\boldsymbol{W}\boldsymbol{\Sigma}^{-1} - \boldsymbol{I}_p)^2] = \text{constant.}$$

For the class (1), we derive unbiased risk estimate $\widehat{\boldsymbol{R}}(\widehat{\boldsymbol{\Sigma}})$.

SURE, part 2

Result Recall that $\boldsymbol{W} = \boldsymbol{O}_1 \boldsymbol{L} \boldsymbol{O}_1'$, $\boldsymbol{L} = \mathrm{Diag}(\ell_1, \ldots, l_n)$. For the class of estimators $\widehat{\boldsymbol{\Sigma}} = \boldsymbol{O}_1 \mathrm{Diag}(\psi_1(\boldsymbol{L}), \ldots, \psi_n(\boldsymbol{L}))\boldsymbol{O}_1'$, we have

$$
\begin{aligned}
R(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) &= \mathbb{E}[\mathrm{Tr}\,(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1} - \boldsymbol{I}_p)^2] = \mathbb{E}[\widehat{\boldsymbol{R}}(\widehat{\boldsymbol{\Sigma}})] \\
\widehat{\boldsymbol{R}}(\widehat{\boldsymbol{\Sigma}}) &= \sum_{k=1}^{n} \Big\{ (p-n-1)\big(\frac{\widetilde{\psi}_k^{(1)}}{\ell_k} - 2\frac{\psi_k}{\ell_k}\big) + 2\big(\frac{\partial \widetilde{\psi}_k^{(1)}}{\partial \ell_k} - 2\frac{\partial \psi_k}{\partial \ell_k}\big) \\
&\qquad + \sum_{b \neq k}^{n} \frac{(\widetilde{\psi}_k^{(1)} - 2\psi_k) - (\widetilde{\psi}_b^{(1)} - 2\psi_b)}{\ell_k - \ell_b} \Big\} + p.
\end{aligned}
$$

Here, for $k = 1, 2, \ldots, n$,

$$
\widetilde{\psi}_k^{(1)} = \frac{(p-n-1)\psi_k^2}{\ell_k} + 4\psi_k \frac{\partial \psi_k}{\partial \ell_k} + 2\psi_k \sum_{b \neq k}^{n} \frac{\psi_k - \psi_b}{\ell_k - \ell_b}.
$$

Improved estimators

Estimators Assume that $p > n$ and consider estimators of the form

$$\widehat{\mathbf{\Sigma}}_t = \frac{1}{p+n+1}\Big(\boldsymbol{W} + \frac{t}{\mathrm{Tr}\,\boldsymbol{W}^+}\boldsymbol{O}_1\boldsymbol{O}_1'\Big).$$

Here $\boldsymbol{O}_1$ is $p \times n$ semi-orthogonal matrix consisting of column eigenvectors coresponding to the positive eigenvalues of $\boldsymbol{W}$ and $\boldsymbol{W}^+$ is the Moore-Penrose inverse of $\boldsymbol{W}$, and t is positve constant.

Results We use SURE and evaluate the risk of $\widehat{\mathbf{\Sigma}}_t$ to get the following: If

$$0 < t < 2(n-1)(p-n-1)/\{(p-n+1)(p-n+2)\},$$

then

$$R(\widehat{\mathbf{\Sigma}}_t, \mathbf{\Sigma}) \le R(n^{-1}\boldsymbol{W}, \mathbf{\Sigma})$$

for all $\mathbf{\Sigma}$.

Concluding remarks: modification of improved estimators

⋆ $\widehat{\boldsymbol{\Sigma}}_t$ improves upon the usual estimator $n^{-1}\boldsymbol{W}$. But $\widehat{\boldsymbol{\Sigma}}_t$ is not p.d.

⋆ Modify $\widehat{\boldsymbol{\Sigma}}_t = \frac{1}{p+n+1}\big(\boldsymbol{W} + \frac{t}{\mathrm{Tr}\,\boldsymbol{W}^{+}}\boldsymbol{O}_1\boldsymbol{O}_1'\big)$ to get:

$$\widetilde{\boldsymbol{\Sigma}}_{\mathrm{HF}} = \frac{1}{p+n+1}\{\boldsymbol{W} + \frac{t_0}{\mathrm{Tr}\,\boldsymbol{W}}\boldsymbol{I}_p\}, \qquad t_0 = \frac{(n-1)(p+n+1)}{(p-n+1)(p-n+3).}.$$

⋆ Unfortunately, we can not evaluate the risk of the estimators $\widetilde{\boldsymbol{\Sigma}}_{HF}$ by using SURE! But we can see that this estimator is better than the usual estimator via the simulation results we did.

Thank you!