Statistical Inference On the High-dimensional Gaussian Covariance Matrix

Xiaoqian SUN, Colin Gallagher, Thomas Fisher

Department of Mathematical Sciences, Clemson University

June 6, 2011

Outline

- Introduction
- Hypothesis Testing on the Covariance Matrix
- Estimation of the Covariance Matrix
- Conclusions and Future Work.

Problem Setup

Consider $X_1, X_2, \ldots, X_N \sim N_p(\mu, \Sigma)$:

- $\mu \in R^p$ and $\Sigma > 0$
- Both μ and Σ are unknown.
- $(\bar{X}, S)$ is a sufficient statistic.
- Σ is the parameter of interest.

Statistical Inference

Statistical Inference

- Classical inference
  - Based on the likelihood approach
  - Assume $N = n + 1 > p$ and $N \to \infty$ with p fixed
  - Results appeared on most multivariate analysis textbooks

Statistical Inference

- Classical inference
  - Based on the likelihood approach
  - Assume $N = n + 1 > p$ and $N \to \infty$ with p fixed
  - Results appeared on most multivariate analysis textbooks
- High-dimensional Inference
  - Assume both $(n, p) \to \infty$
  - No general approach
  - Fujikoshi, Ulyanov and Shimizu (2010) "*Multivariate Statistics : High-Dimensional and Large-Sample Approximation*", Wiley

High-Dimensional Data Sets

Examples:

1. Microarray gene data in genetics
2. Financial data in stock markets
3. Curve data in engineering
4. Image data in computer science

.....

High-Dimensional Data Sets

Examples:

1. Microarray gene data in genetics
2. Financial data in stock markets
3. Curve data in engineering
4. Image data in computer science

.....

Comments:

- The dimensionality exceeds the sample size, i.e. $p > N$.
- Collecting additional data may be expensive or infeasible.
- Few data analysis before 1970
- Fast computers $\Rightarrow$ New methods needed

Hypothesis Testing on the Sphericity

Consider

$$H_0 : \Sigma = \sigma^2 I \quad \text{vs.} \quad H_1 : \Sigma \neq \sigma^2 I.$$

Hypothesis Testing on the Sphericity

Consider

$$H_0 : \Sigma = \sigma^2 I \quad \text{vs.} \quad H_1 : \Sigma \neq \sigma^2 I.$$

The likelihood ratio test (LRT) for this hypothesis is,

$$\Lambda(\mathbf{x}) = \left(\frac{\prod_{i=1}^{p} l_i^{1/p}}{\sum_{i=1}^{p} l_i/p} \right)^{\frac{1}{2}pN}$$

where $l_1, l_2, \ldots, l_p \geq 0$ are the eigenvalues of the MLE for Σ.

Hypothesis Testing on the Sphericity

Consider

$$H_0 : \Sigma = \sigma^2 I \quad \text{vs.} \quad H_1 : \Sigma \neq \sigma^2 I.$$

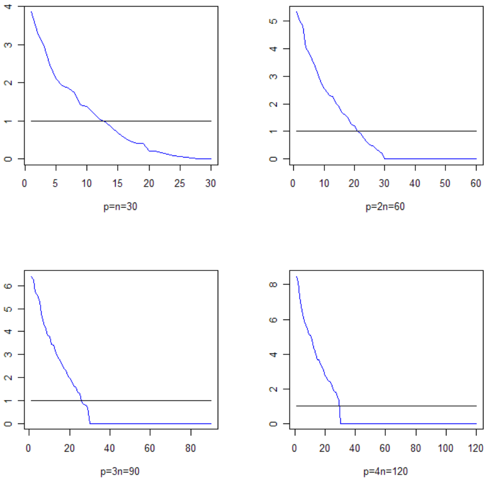
The likelihood ratio test (LRT) for this hypothesis is,

$$\Lambda(\mathbf{x}) = \left(\frac{\prod_{i=1}^{p} l_i^{1/p}}{\sum_{i=1}^{p} l_i / p} \right)^{\frac{1}{2} pN}$$

where $l_1, l_2, \ldots, l_p \geq 0$ are the eigenvalues of the MLE for Σ.

- When $p > n$, $\hat{\Sigma}$ will be singular, and hence have 0-eigenvalues.
- Even when $p \leq n$, the eigenvalues of S disperse from the true ones

Sample Eigenvalue Dispersion ($\Sigma = I$)



Effects on LRT under High-Dimensionality

- If $p > N$, the LRT is degenerate
- If $N > p$, but $p \to N$, the LRT will become computational degenerate/unreliable
- The LRT cannot be used in a high-dimensional situation.

Previous Work on High-Dimensional Sphericity Test

- John (1971) U test statistic,

$$U = \frac{1}{p}\mathrm{tr}\left[\left(\frac{S}{(1/p)\mathrm{tr}(S)} - I\right)^2\right].$$

- Its based on the 1st and 2nd arithmetic means.
- Ledoit and Wolf (2002) show its (n, p)-asymptotic null distribution is $N(1, 4)$.
- Its (n, p)-asymptotic distribution under the alternative is unknown.

Open Question about tests based on r^{th} Mean

The r^{th} mean of p nonnegative reals, $\{\lambda_1, \ldots, \lambda_p\}$ is given by

$$M(r) = \begin{cases} \left(\dfrac{1}{p} \displaystyle\sum_{i=1}^{p} \lambda_i^r \right)^{1/r} & \text{if } r \neq 0 \\ \displaystyle\prod_{i=1}^{p} \lambda_i^{1/p} & \text{if } r = 0 \end{cases}$$

Open Question about tests based on r^{th} Mean

The r^{th} mean of p nonnegative reals, $\{\lambda_1, \ldots, \lambda_p\}$ is given by

$$M(r) = \begin{cases} \left(\frac{1}{p} \sum_{i=1}^{p} \lambda_i^r \right)^{1/r} & \text{if } r \neq 0 \\ \prod_{i=1}^{p} \lambda_i^{1/p} & \text{if } r = 0 \end{cases}$$

- The LRT is based on the geometric, $M(0)$, and the first arithmetic, $M(1)$, means.
- John's U statistic is based on $M(1)$ and $M(2)$.
- Open question: Construct a test based on $M(r)$ and $M(t)$ for $r, t > 0$?

Srivastava Test for Sphericity

Srivastava (2005) constructs a test based on $M(1)$ and $M(2)$ using a parametric function of Σ. Consider the Cauchy-Schwarz inequality,

$$\left(\sum_{i=1}^{p} \lambda_i^r \times 1^r\right)^2 \leq p\left(\sum_{i=1}^{p} \lambda_i^{2r}\right).$$

Thus the ratio

$$\psi_r = \frac{\left(\sum\limits_{i=1}^{p} \lambda_i^{2r}/p\right)}{\left(\sum\limits_{i=1}^{p} \lambda_i^{r}/p\right)^2} \geq 1$$

with equality holding if and only if $\lambda_i = \lambda$, some constant λ, for all $i = 1, \ldots, p$.

Tests based on Cauchy-Schwarz Inequality

- $$H_0 : \Sigma = \sigma^2 I \quad \text{vs} \quad H_A : \Sigma \neq \sigma^2 I$$

  $$\Leftrightarrow H_0 : \psi_r = 1 \quad \text{vs} \quad H_A : \psi_r > 1.$$

Tests based on Cauchy-Schwarz Inequality

- $$H_0 : \Sigma = \sigma^2 I \quad \text{vs} \quad H_A : \Sigma \neq \sigma^2 I$$
  $$\Leftrightarrow H_0 : \psi_r = 1 \quad \text{vs} \quad H_A : \psi_r > 1.$$
- Srivastava (2005) finds unbiased and consistent estimators for the numerator and denominator of ψ_r when $r = 1$.

Tests based on Cauchy-Schwarz Inequality

- $$H_0 : \Sigma = \sigma^2 I \quad \text{vs} \quad H_A : \Sigma \neq \sigma^2 I$$
  $$\Leftrightarrow H_0 : \psi_r = 1 \quad \text{vs} \quad H_A : \psi_r > 1.$$
- Srivastava (2005) finds unbiased and consistent estimators for the numerator and denominator of ψ_r when $r = 1$.
  - The distributions under both the null and alternative hypotheses, as $(n, p) \to \infty$.
  - The test procedure is consistent as $(n, p) \to \infty$.

Tests based on Cauchy-Schwarz Inequality

- $$H_0 : \Sigma = \sigma^2 I \quad \text{vs} \quad H_A : \Sigma \neq \sigma^2 I$$
  $$\Leftrightarrow H_0 : \psi_r = 1 \quad \text{vs} \quad H_A : \psi_r > 1.$$
- Srivastava (2005) finds unbiased and consistent estimators for the numerator and denominator of ψ_r when $r = 1$.
  - The distributions under both the null and alternative hypotheses, as $(n, p) \to \infty$.
  - The test procedure is consistent as $(n, p) \to \infty$.
- We explore the case of $r = 2$.

Some Assumptions for the New Testing Procedure

Suppose $\mathbf{X}_1, \ldots, \mathbf{X}_N \sim N_p(\boldsymbol{\mu}, \Sigma)$, $N = n + 1$.
Make the following assumptions

$$\text{(A)} : \text{As } p \to \infty,\ a_i \to a_i^0,\ 0 < a_i^0 < \infty, i = 1, \ldots, 16,$$
$$\text{(B)} : \text{As } (n, p) \to \infty, \frac{p}{n} \to c,\ \text{where } 0 < c < \infty,$$

where

$$a_i = \frac{1}{p}\mathrm{tr}\Sigma^i = \frac{1}{p}\sum_{j=1}^{p} \lambda_j^i$$

and the λ_js are the eigenvalues of the covariance matrix, i.e. a_i is the i^{th} arithmetic mean of the eigenvalues of the covariance matrix.

An Unbiased and Consistent Estimator for a_4

Theorem

An unbiased and (n, p)-consistent estimator of $a_4 = \sum_{i=1}^{p} \lambda_i^4 / p$ is given by

$$\hat{a}_4 = \frac{\tau}{p}\left[trS^4 + b \cdot trS^3 trS + c^* \cdot (trS^2)^2 + d \cdot trS^2 (trS)^2 + e \cdot (trS)^4\right],$$

where

$$b = -\frac{4}{n},\ c^* = -\frac{2n^2 + 3n - 6}{n(n^2 + n + 2)}, d = \frac{2(5n + 6)}{n(n^2 + n + 2)},$$

$$e = -\frac{5n + 6}{n^2(n^2 + n + 2)}, \tau = \frac{n^5(n^2 + n + 2)}{(n + 1)(n + 2)(n + 4)(n + 6)(n - 1)(n -}$$

Consistent Estimators for a_2 and ψ_2

- Srivastava (2005) provides an unbiased and consistent estimator for a_2 which is

$$\hat{a}_2 = \frac{n^2}{(n-1)(n+2)} \frac{1}{p} \left[\mathrm{tr} S^2 - \frac{1}{n} (\mathrm{tr} S)^2 \right].$$

Consistent Estimators for a_2 and ψ_2

- Srivastava (2005) provides an unbiased and consistent estimator for a_2 which is

$$\hat{a}_2 = \frac{n^2}{(n-1)(n+2)} \frac{1}{p} \left[\mathrm{tr} S^2 - \frac{1}{n} (\mathrm{tr} S)^2 \right].$$

- Thus an (n, p)-consistent estimator for ψ_2 is provided by

$$\hat{\psi}_2 = \frac{\hat{a}_4}{\hat{a}_2^2}.$$

Asymptotic Result

Theorem

Under assumptions (A) and (B), as $(n, p) \to \infty$

$$\frac{n}{\sqrt{8(8+12c+c^2)}}\left(\frac{\hat{a}_4}{\hat{a}_2^2} - \psi_2\right) \xrightarrow{D} N(0, \xi_2^2),$$

where

$$\xi_2^2 = \frac{1}{(8+12c+c^2)a_2^6}\Big(\frac{4}{c}a_4^3 - \frac{8}{c}a_4 a_2 a_6 - 4a_4 a_2 a_3^2 + \frac{4}{c}a_2^2 a_8 + 4a_6 a_2^3 + 8a_2^2 a_5 a_3 + 4ca_4 a_2^4 + 8ca_3^2 a_2^3 + c^2 a_2^6\Big).$$

Test Statistic under H_0

Corollary

Under H_0, $\psi_2 = 1$, *as* $(n, p) \to \infty$,

$$T = \frac{n}{\sqrt{8(8 + 12c + c^2)}} \left(\frac{\hat{a}_4}{\hat{a}_2^2} - 1 \right) \xrightarrow{D} N(0, 1).$$

Under H_0, $\xi_2^2 = 1$ since each $\lambda_i = \lambda$, for $i = 1, \ldots, p$ and some constant λ.

Power Function under General Asymptotics

Theorem

Under assumptions (A) and (B), as $(n, p) \to \infty$ *the above testing procedure based on* T *is* (n, p)*-consistent.*

Power Function under General Asymptotics

Theorem

Under assumptions (A) and (B), as $(n, p) \to \infty$ the above testing procedure based on T is (n, p)-consistent.

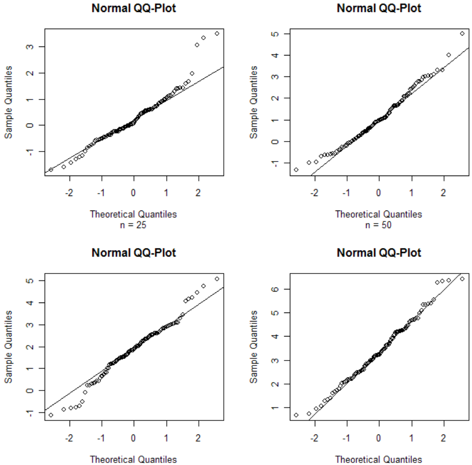
For large n and p, the power function of T is

$$Power_\alpha(T) \simeq \Phi\left(\frac{n\left(\frac{\hat{a}_4}{\hat{a}_2^2} - 1\right)}{\xi_2\sqrt{8(8+12c+c^2)}} - \frac{z_\alpha}{\xi_2}\right).$$

Under assumptions (A) and (B), we know ξ_2^2 is constant. From the properties of $\Phi(\cdot)$, it is clear that $Power_\alpha(T) \to 1$ as $(n, p) \to \infty$.

QQ-Plots for increasing (n, p) under H_A

500 observed values of T, with $p/n = 2$ under H_A with $\Sigma = \text{diag}(\lambda_1, \ldots, \lambda_p)$ with $\lambda_i \sim Unif(0.5, 1.5)$.



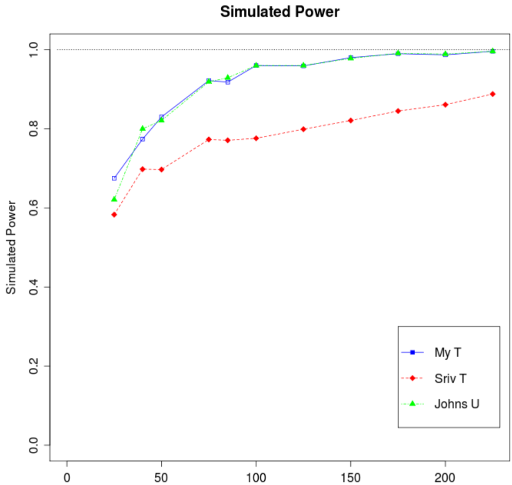
Power Study

- Simulate 1000 observed values of T under H_0:$\Sigma = I$ and find T_α such that
$$P(T > T_\alpha) = \alpha.$$
T_α is the estimated critical point at significance level α.
- Simulate from a p-dimensional normal distribution with zero mean vector and a *near* spherical covariance matrix. Define *near* spherical matrices to be of the form,

$$\Sigma = \sigma^2 \begin{pmatrix} \phi & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}, \phi \neq 1.$$

Simulation Power Functions

Simulated Power for each test, $c = 1$ with $\phi = 3.5$



Data Analysis

Gene Expression levels of 72 patients either suffering from acute lymphoblastic leukemia or acute myeloid leukemia were measured on Affymetric oligonucleotite microarrays.

- 47 and 25 patients of each respective leukemia type.
- Use a pooled covariance with only $n = 70$ degrees of freedom.
- Data is comprised of $p = 3571$ genes.

Data Analysis

Gene Expression levels of 72 patients either suffering from acute lymphoblastic leukemia or acute myeloid leukemia were measured on Affymetric oligonucleotite microarrays.

- 47 and 25 patients of each respective leukemia type.
- Use a pooled covariance with only $n = 70$ degrees of freedom.
- Data is comprised of $p = 3571$ genes.

- $T = 242.4386$, $T_{Sri} = 2294.9184$, and $U_J = 2326.7520$.
- p-value ≈ 0 for all three tests and thus H_0 is rejected.

Estimation of the Covariance Matrix

Estimation of the Covariance Matrix is typically achieved with the sample covariance matrix, i.e.

$$
\begin{aligned}
S &= \frac{1}{N-1}\sum_{i=1}^{N}(\mathbf{x_i}-\bar{\mathbf{x}})(\mathbf{x_i}-\bar{\mathbf{x}})' \\
&= \frac{1}{n}(\mathbf{X}-\bar{\mathbf{X}})(\mathbf{X}-\bar{\mathbf{X}})'
\end{aligned}
$$

where $\bar{\mathbf{x}}$ is the sample mean vector and $\bar{\mathbf{X}}$ is a matrix, with the columns composed of repeating $\bar{\mathbf{x}}$.

Properties of Sample Covariance Matrix

Pros

- S is an unbiased and N-consistent estimator for Σ.
- S is based on the MLE of Σ.
- S^{-1} can be used to estimate the precision matrix Σ^{-1}.
- Works well when $N > p$.

Cons

- When $p > N$, S is singular, and hence an estimate for the precision matrix is not possible.
- S becomes ill-conditioned as $p \to N$.
- As $p \to N$ or $p > N$, the eigenvalues of S diverge from the eigenvalues of Σ.

Need Good Estimators for Σ

A good estimate for Σ is needed in many statistical applications:

- Hotelling's T^2 statistic requires an estimate of the precision matrix
- Factor Analysis
- Principal Components
- Discrimination and Classification
- Time-Series Analysis

Stein-type Shrinkage Estimation for Σ

Consider a convex combination of the empirical sample covariance matrix with that of a target matrix,

$$S^* = \lambda M + (1 - \lambda)S,$$

where $\lambda \in [0, 1]$ is known as the shrinkage *intensity* and M is a shrinkage *target* matrix. M is chosen such that:

- It is well-structured, Positive Definite and well-conditioned.
- It will be *biased*, but will have *less variance*.

Stein-type Shrinkage Estimation for Σ

Consider a convex combination of the empirical sample covariance matrix with that of a target matrix,

$$S^* = \lambda M + (1 - \lambda)S,$$

where $\lambda \in [0, 1]$ is known as the shrinkage *intensity* and M is a shrinkage *target* matrix. M is chosen such that:

- It is well-structured, Positive Definite and well-conditioned.
- It will be *biased*, but will have *less variance*.

How to find a suitable λ?

Historical Approach and Optimal Intensity

Historical approaches

- Maximizing Cross-Validation.
- Bootstrap methods, Bayesian approach.
- MCMC Methods.

Historical Approach and Optimal Intensity

Historical approaches

- Maximizing Cross-Validation.
- Bootstrap methods, Bayesian approach.
- MCMC Methods.

Ledoit and Wolf (2003) show with respect to the squared loss $\|\Sigma^* - \Sigma\|^2$, or quadratic risk, an optimal λ will always exist.

Ledoit and Wolf (2004) Main Results

Consider the target matrix, $M = a_1 I$ where $a_1 = \mathrm{tr}\Sigma/p$.

Ledoit and Wolf (2004) Main Results

Consider the target matrix, $M = a_1 I$ where $a_1 = \mathrm{tr}\Sigma/p$. Define:

$$\begin{aligned} \alpha^2 &= \|\Sigma - a_1 I\|^2, \\ \beta^2 &= E[\|S - \Sigma\|^2], \\ \delta^2 &= E[\|S - a_1 I\|^2], \end{aligned}$$

and $\delta^2 = \alpha^2 + \beta^2$.

A calculus-based minimization of the objective function $E[\|\Sigma^* - \Sigma\|^2]$ provides the result

$$\lambda = \beta^2/(\alpha^2 + \beta^2) = \beta^2/\delta^2, \quad 1 - \lambda = \alpha^2/\delta^2.$$

Ledoit and Wolf (2004) Main Results

Consider the target matrix, $M = a_1 I$ where $a_1 = \mathrm{tr}\Sigma/p$. Define:

$$\begin{aligned} \alpha^2 &= \|\Sigma - a_1 I\|^2, \\ \beta^2 &= E[\|S - \Sigma\|^2], \\ \delta^2 &= E[\|S - a_1 I\|^2], \end{aligned}$$

and $\delta^2 = \alpha^2 + \beta^2$.

A calculus-based minimization of the objective function $E[\|\Sigma^* - \Sigma\|^2]$ provides the result

$$\lambda = \beta^2/(\alpha^2 + \beta^2) = \beta^2/\delta^2, \quad 1 - \lambda = \alpha^2/\delta^2.$$

Unfortunately, $\Sigma^* = \frac{\beta^2}{\delta^2} a_1 I + \frac{\alpha^2}{\delta^2} S$ is not a *bona fide* estimator since it depends on knowledge of the covariance matrix Σ.

Estimators of the Optimal Intensity

Recent approaches at estimating the optimal λ

- Ledoit and Wolf (2004) provide *n*-consistent estimators of α^2, β^2 and δ^2.
- Schäfer and Strimmer (2005) provide an unbiased estimator for λ.
- Under the assumption of Normality of the data, Chen, Wiesel and Hero (2009) provide an unbiased estimator for λ by utilizing the Rao-Blackwell theorem.

Estimators of the Optimal Intensity

Recent approaches at estimating the optimal λ

- Ledoit and Wolf (2004) provide *n*-consistent estimators of α^2, β^2 and δ^2.
- Schäfer and Strimmer (2005) provide an unbiased estimator for λ.
- Under the assumption of Normality of the data, Chen, Wiesel and Hero (2009) provide an unbiased estimator for λ by utilizing the Rao-Blackwell theorem.

Each performs well as *n* grows large.

Our Approach

Assume that

$$E[\mathrm{tr}(S)] = \mathrm{tr}(\Sigma)$$

and

$$E[\mathrm{tr}(S^2)] = \frac{n+1}{n}\mathrm{tr}\Sigma^2 + \frac{1}{n}(\mathrm{tr}\Sigma)^2.$$

Both hold in many cases, specifically when data comes from a multivariate normal distribution.

Explicit Calculation of λ

Hence we can explicitly calculate

$$\begin{aligned}\delta^2 = E[\|S - a_1 I\|^2] &= E[\|S\|^2] - 2a_1 E[\langle S, I\rangle] + a_1^2\|I\|^2 \\ &= \frac{n+1}{n}a_2 + \frac{p-n}{n}a_1^2.\end{aligned}$$

Likewise, we expand the term α^2 as follows

$$\alpha^2 = \|\Sigma - a_1 I\|^2 = a_2 - a_1^2.$$

where $a_i = \mathrm{tr}\Sigma^i/p$.

A similar result holds for β^2 but is not needed.

Reduced Problem

Under the normality assumption and

$$\text{(A)} : \text{As } p \to \infty, a_i \to a_i^0, 0 < a_i^0 < \infty, i = 1, \ldots, 4,$$
$$\text{(B)} : n = O\left(p^{\delta}\right), 0 \leq \delta \leq 1,$$

Srivastava (2005) finds unbiased and (n, p)-consistent estimators for a_1 and a_2:

$$\hat{a}_1 = \mathrm{tr}S/p$$

and

$$\hat{a}_2 = \frac{n^2}{(n-1)(n+2)} \frac{1}{p} \left[\mathrm{tr}S^2 - \frac{1}{n}(\mathrm{tr}S)^2\right].$$

Reduced Problem

Under the normality assumption and

$$\text{(A)} : \text{As } p \to \infty, a_i \to a_i^0, 0 < a_i^0 < \infty, i = 1, \ldots, 4,$$
$$\text{(B)} : n = O\left(p^\delta\right), 0 \leq \delta \leq 1,$$

Srivastava (2005) finds unbiased and (n, p)-consistent estimators for a_1 and a_2:

$$\hat{a}_1 = \mathrm{tr}S/p$$

and

$$\hat{a}_2 = \frac{n^2}{(n-1)(n+2)} \frac{1}{p} \left[\mathrm{tr}S^2 - \frac{1}{n}(\mathrm{tr}S)^2\right].$$

From Assumption (B), the estimators for a_1 and a_2 should be quite accurate in large p, small n situations.

Other Target Matrices

Analogous results hold for the targets $M = I$, and $M = D$ where D is the diagonal matrix comprised of the diagonal elements of S.

- Ledoit and Wolf (2004) only provide an estimator for the $M = a_1 I$ case, but its easily adapted to $M = I$.
- Chen, Wiesel and Hero (2009) only provide a result for $M = a_1 I$.
- Schäfer and Strimmer (2005) provide unbiased estimators for several targets (including some not discussed here) including $M = I$ and $M = D$.
- We can explicitly calculate the optimal shrinkage intensity, λ, in terms of a_1 and a_2.

Simulation Setup

A simulation study justifies our proposed estimator.

- Sample $n + 1$ observations from a p-dimensional multivariate normal distribution with zero mean vector and covariance matrix Σ.
- Σ is a random positive definite matrix with eigenvalues uniformly distributed over $(0.5, 10.5)$.
- The $n + 1$ samples of p dimension are used to compute the various shrinkage estimators.
- The process is repeated $m = 1000$ times with the same covariance matrix Σ.

First we explore the estimation of λ.

Simulation of Optimal λ for $M = a_1 I$, $n = 40$, $p = 20$

	λ_{new}	λ_{LW}	λ_{RBLW}	λ_{Schaf}
Simulated Mean	0.6595865	0.6265542	0.6424440	0.6407515
Standard Error	0.0000602	0.0000616	0.0000588	0.0000608

Table: λ estimation for $n = 40, p = 20$, $M = a_1 I$

Since the true covariance matrix is known in the simulation, the optimal intensity can be calculated exactly, it is 0.6503192.

Simulation of Optimal λ for $M = a_1 I$, $n = 5$, $p = 100$

	λ_{new}	λ_{LW}	λ_{RBLW}	λ_{Schaf}
Simulated Mean	0.9887804	0.6634387	0.7909521	0.7950775
Standard Error	0.0000218	0.0000532	0.0000176	0.0000194

Table: λ estimation for $n = 5, p = 100$, $M = a_1 I$

With the optimal intensity at $\lambda = 0.9868715$.

Improvement over Sample Covariance Matrix

How do the Optimal Stein-type shrinkage estimators improve over the sample covariance matrix? We look at the simulated risk

$$Risk(S^*) = E[\|S^* - \Sigma\|^2]$$

and the percentage relative improvement in average loss (PRIAL)

$$PRIAL(S^*) = \frac{E[\|S - \Sigma\|^2] - E[\|S^* - \Sigma\|^2]}{E[\|S - \Sigma\|^2]} \times 100.$$

Simulation of Stein-type Shrinkage Estimators, $M = a_1 I$

Same setup as before, the *true* Σ is a random positive definite matrix with eigenvalues uniformly distributed between (0.5, 10.5).

Estimator	S	S^*_{LW}	S^*_{RBLW}	S^*_{Schaf}	S^*_{new}
Risk	529.332	68.525	29.446	28.781	8.800
SE on Risk	2.536	0.762	0.193	0.206	0.020
PRIAL	0	87.054	94.437	94.563	98.338
Cond. Num.	∞	15.946	8.602	8.455	1.702

Table: Shrinkage estimation for $n = 5, p = 100,\ M = a_1 I$

Data Example on E.coli Data

Schmidt-Heck et al (2004) identified 102 genes, of 4,289 protein coding genes, as differentially expressed in one or more samples after induction of a recombinant protein on the microorganism Escherichia coli. The data monitored all 4,289 protein coding genes at 8 different times after the induction of the protein.

Target	$M = a_1 I$	$M = I$	$M = \text{diag}(S)$
New Estimators	156.73	155.95	468.37
LW-Type	384.89	382.97	NA
RBLW-Type	212.23	NA	NA
Schäfer-Strimmer	288.79	287.35	715.25

Table: Condition Numbers for estimators and common targets on E.coli data, $p = 102$, $N = 8$

Conclusion remarks and possible future work

- A new testing procedure for the sphericity
- Stein-type shrinkage estimators
- Good performances by simulation studies

Conclusion remarks and possible future work

- A new testing procedure for the sphericity
- Stein-type shrinkage estimators
- Good performances by simulation studies
- Possible to release the condition $p/n \to c \in (0, \infty)$?
- Dropping the normality assumption?
- Other tests based on $M(r)$ and $M(t)$?
- Other loss functions (Stein-type shrinkage estimators)?
- Possible Bayesian approaches?

References

1. Fisher, T., Sun, X. and Gallagher, C. (2010) "A New Test for Sphericity of the Covariance Matrix for High Dimensional Data", *J. of Multivariate Anal.*, V. 101, pp.2554-2570
2. Fisher, T. and Sun, X. (2011), "Improved Stein-Type Shrinkage Estimators for the Multivariate Normal Covariance Matrix", *Computat. Statist. and Data Anal.*, 55, 1909-1918
3. Srivastava, M. (2005), "Some tests concerning the covariance matrix in high dimensional data", *J. Japan Statist. Soc.*, Vol.35, pp251-272
4. Ledoit, O. and Wolf, M. (2002), "Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size", *Ann. Statist.*, Vol. 30, pp1081-1102
5. Ledoit, O. and Wolf, M. (2004), "A well-conditioned estimator for large-dimensional covariance matrices", *J. Multivariate Anal.*, V. 88, pp.365-411

Thank you! ☺