Testing Covariance Structure in High Dimensional Setting

Tõnu Kollo, Dietrich von Rosen, Marju Valge

University of Tartu, Estonia
Swedish University of Agricultural Sciences, Sweden
Linköping University, Sweden

The International Conference on Trends and Perspectives in Linear Statistical Inference (LinStat2014)
Linköping, August 24-28, Sweden
Simple hypothesis: \( H_0: \Sigma = \Sigma_0 \), where \( \Sigma_0 \) is known

- Covariance matrix is an identity matrix
- All parameters of \( \Sigma \) are fixed

Composite hypothesis:

- \( \Sigma \) is block-diagonal
- \( \Sigma \) has Kronecker product structure
- Not all parameters of \( \Sigma \) are fixed
Denote:

- $X \sim P(\theta)$, where $\theta$ is $r$-vector
- $L(X)$ - sample likelihood
- Hypothesis $H_0$: $\theta = \theta_0$

Likelihood Ratio Test (LRT) statistic:

- $LRT = -2\ln[maxL(X|\theta_0)/maxL(X|\theta)]$
- Under $H_0$ LRT is asymptotically $\chi^2(r)$
It has been shown that when the dimension of the covariance matrix increases the likelihood ratio test will stop working correctly and will almost always reject the null hypothesis.

In order to deal with this problem, corrections to the test have been made so that it could be used in high-dimensional case as well (see Bai, Jiang, Yao and Zheng (2009) for an example).
RAO’S SCORE TEST

Due to Rao (1948)

Rao’s Score Test (RST) statistic

- $RST = S(\theta_0)'I^{-1}(\theta_0)S(\theta_0)$ for simple hypothesis $H_0: \theta = \theta_0$

- $RST = S(\hat{\theta})'I^{-1}(\hat{\theta})S(\hat{\theta})$ for composite hypothesis $H_0: H(\theta) = C$, where $\hat{\theta}$ is the MLE of $\theta$, $H'(\theta) = (h_1(\theta), ..., h_t(\theta))$, $C' = (c_1, ..., c_t)$ and $t < r$.

- $S()$ - sample score function, $I()$ - Fisher’s information matrix

- Under $H_0$ RST is asymptotically $\chi^2(r)$

Note: RST does not require $\theta$ to be estimated for simple hypothesis testing and does not require any estimation under alternative hypothesis.
Rao’s Score Test works when $\text{rank}(\Sigma)$ is large and Likelihood Ratio Test does not work.

Simulation experiment was carried out in order to empirically test the hypothesis.
The simulation experiment was carried out using R software and the following procedure:

1. Generate $X_1, \ldots, X_n$ from multivariate $N_p$ distribution with mean vector $\mu$ and covariance matrix $\Sigma_0$.

2. Use LRT and RST to test $H_0$: $\Sigma = \Sigma_0$.

3. Repeat $N$ times, count the number of $H_0$ rejected by both LRT and RST.

4. Run for different $p = \text{rank}(\Sigma)$. 
Figure 1. Simulation results: $H_0: \Sigma = I$
Figure 1. Simulation results: $H_0: \Sigma = \Psi \otimes \Sigma_0$
Rao’s Score Test should be preferred to Likelihood Ratio Test in high dimensional setting.

Future questions of interest:

- Comparison with Wald’s Score Test,

- Simulations for different hypothesis types and sample sizes.