Some tests for covariance matrices with large dimension

M. Rauf Ahmad[1], Martin Ohlson[1] and D. von Rosen[1,2]

[1] *Linköping University, Sweden, email: {muahm, mohl}@mai.liu.se*
[2] *Swedish University of Agricultural Sciences, Uppsala, Sweden, email: Dietrich.von.rosen@et.slu.se*

Keywords: covariance testing, high-dimensionality, sphericity.

Let $\mathbf{X}_k = (X_{k1}, \ldots, X_{kp})'$, $k = 1, \ldots, n$, be n independent and identically distributed random vectors where $\mathbf{X}_k \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. We present test statistics for

$$H_0 : \boldsymbol{\Sigma} = \mathbf{I} \quad \text{and} \quad H_0 : \boldsymbol{\Sigma} = \kappa \mathbf{I},$$

when p may be large, and may even exceed n, where $\kappa > 0$ is any constant. The test statistics are constructed using unbiased and consistent estimators composed of quadratic and bilinear forms of the random vectors $\mathbf{X}_k$. Under very general settings, the proposed test statistics are shown to follow an approximate normal distribution, for large n and p, inclusive of the case when $p > n$, or even $p \gg n$. The statistics are based on minimal conditions avoiding the usually adopted stringent assumptions found in the literature for similar high-dimensional inferences, for example assumptions on the traces of powers of the covariance matrix $\boldsymbol{\Sigma}$, or assumptions on the relations between p and n (see, for example, [2], [1]: Chs. 5&8). The performance of the test statistics is shown through simulations. It is demonstrated that the test statistics are accurate for both, size control and power for moderate n and any p, where p can be much large than n. The real life application of the statistics is also illustrated using practical data sets.

References

[1] Fujikoshi, Y., Ulyanov, V. U., Shimizu, R. (2010). *Multivariate Statistics: High-Dimensional and Large-Sample Approximations.* Wiley, New York.

[2] Ledoit, O., Wolf, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *The Annals of Statistics*, **30**(4), 1081-1102.