Tutorial - Week 7

Please read the related material and attempt these questions before attending your allocated tutorial. Solutions are released on Friday 4pm.

We continue with the ideas covered in the week 6 tutorial and consider the paper **[B]**.

Question 1

We consider the problem of a one-sample covariance hypothesis test. Suppose $\mathbb{x}_1, \ldots, \mathbb{x}_n$ is a p-dimensional sample where we assume $p < n$. Let Σ be the (population) covariance matrix, we want to test

$$H_0 : \Sigma = I_p \quad \text{vs.} \quad H_1 : \Sigma \neq I_p,$$

where I_p is the $p \times p$ identity matrix. Note that testing $\Sigma = A$ with an arbitrary (positive definite) covariance matrix A can always be reduced to the null hypothesis above by applying the transformation $x \mapsto A^{-1/2}x$ to the data. Given the sample covariance matrix $\mathbb{S}_n$ of the data and setting

$$L^* = \operatorname{tr} \mathbb{S}_n - \log |\mathbb{S}_n| - p,$$

the likelihood ratio test (LRT) statistic is

$$T_n = n \cdot L^*.$$

Keeping p fixed while letting $n \to \infty$, then the classic theory states that T_n converges to the χ^2_ν distribution with $\nu = \frac{1}{2}p(p+1)$ under H_0.

In the paper **[B]**, they show under the assumption $n \to \infty$ and $p \to \infty$ such that $p/n \to y \in (0, 1)$ that[1], under H_0 their corrected likelihood ratio test (CLRT) given by

$$\widetilde{T}_n = \upsilon(g)^{-1/2}\big[L^* - p \cdot F^{y_n}(g) - m(g)\big]$$

converges to a standard Normal distribution $N(0, 1)$. Here, $F^{y_n}(g) = 1 - (y_n - 1)/y_n \cdot \log(1 - y_n)$, $m(g) = -\log(1 - y_n)/2$, and $\upsilon(g) = -2\log(1 - y_n) - 2y_n$.

(a) Sampling $\mathbb{x}_1, \ldots, \mathbb{x}_n$ from $N_p(0, I_p)$, perform a simulation to check that the LRT is approximated by a χ^2 distribution for $p = 3$ and $n = 500$ when H_0 is true. Is it a good fit?

(b) Sampling $\mathbb{x}_1, \ldots, \mathbb{x}_n$ from $N_p(0, I_p)$, perform a simulation to check that the CLRT is approximated by a standard Normal distribution for $p = 3, 50, 250$ and $n = 500$ when H_0 is true. What do you observe as p varies?

(c) Calculate the size of the classic LRT test with $\alpha = 0.05$. For (p, n) taking the values $(5, 500)$, $(10, 500)$, $(50, 500)$, $(100, 500)$, $(300, 500)$. What do you notice about the size of the test as p increases with respect to n? It could be useful to plot the size with respect to p.

[1] Theorem 3.1

(d) Calculate the power of the classic LRT test with $\alpha = 0.05$ under the alternative where the data is sampled from $N_p(0, \Sigma)$ where Σ has a *compound symmetric structure* $\Sigma = (1-\rho)I_p + \rho\mathbf{1}\mathbf{1}'$ where $\mathbf{1}$ is a p-dimensional vector of 1's, I_p is a $p \times p$ identity matrix, and $\rho \in (0,1)$ is a constant. Fixing $n = 500$, $\rho = 0.10$, plot the power of the test for $p = 1, 2, 3, 4, 5, 10, 25, 50, 100$. What do you notice about the power of the test as p increase with respect to n? It could be useful to plot the power with respect to p.

(e) Now repeat (c) but for the CLRT.

(f) Now repeat (d) but for the CLRT.

References

[A] Anderson (2003). An introduction to Multivariate Statistical Analysis. Wiley.

[B] Bai, Jiang, Yao, Zheng (2009). Corrections to LRT on large-dimensional covariance matrix by RMT. Annals of Statistics Vol 37, No. 6B, 3822–3840.