Homework 4

Due by
Wednesday 11 October 2017 17:00

Consider the paper "Testing linear hypotheses in high-dimensional regressions" (2013) by Bai, Jiang, Yao and Zheng that we studied in Lecture 9.

Question 1 [2 marks]

Let S_1 and S_2 be sample covariance matrices for p-dimensional observations of size n_1 and n_2, respectively. Let $V_n = S_1 S_2^{-1}$ where $n = (n_1, n_2)$ and assume $n_2 > p$. Make an appropriate choice of parameters, sample from V_n and produce a plot showing the histogram of eigenvalues of V_n compared to the LSD F_{y_1,y_2} given by equation (14) in the paper.

Hint: See pages 1210 and 1211 of the paper (above) and Workshop 2, Section 2.2. You can either simulate the data matrix **X** *using* `rnorm` *(and then construct the sample covariances) or draw the sample covariances directly using* `rWishart`.

Question 2 [3 marks]

In the paper, Theorem 3.1, it is proved that under the null hypothesis

$$T_n = \upsilon(f)^{-1/2}[-\log \Lambda_n - p\, F_{y_1,y_2}(f) - m(f)] \;\Rightarrow\; \mathcal{N}(0,1)$$

where $m(f)$, $\upsilon(f)$ and $F_{y_1,y_2}(f)$ are given in the paper in equations (26), (27), and (29), respectively.

Demonstrate numerically that this theorem works by making an appropriate choice of parameters, sampling a large number of T_n, and comparing the histogram of values of T_n against the density of a standard normal.

Hint: See page 1212 and notice that Λ_n is given in terms of **F** *and the quantity* **F** *is given in terms of the ratio of two Wishart matrices. Therefore, for this task, sample* **F** *by posing*

$$\mathbf{F} = \frac{n-q}{q_1} S_1^{-1} S_2, \qquad S_1 \sim W_p(\Sigma, n-q), \quad S_2 \sim W_p(\Sigma, q_1),$$

and using the `rWishart` *function in R. Now from* **F** *it should be straightforward to generate Λ_n. See Workshops 5, 6, and 7 where we have done similar CLT checks.*