Assessment 3

Due by Tuesday 19 September 2023 09:00

The Fisher limiting spectral distribution (LSD), denoted $F_{s,t}$, has the density function

$$f_{s,t}(x) := \frac{1-t}{2\pi x(s+tx)}\sqrt{(b-x)(x-a)}, \quad a \le x \le b,$$

where

$$a := a(s,t) := \frac{(1-h)^2}{(1-t)^2}, \quad b := b(s,t) := \frac{(1+h)^2}{(1-t)^2}, \quad h := h(s,t) := (s+t-st)^{1/2}.$$

Question 1 [12 marks]

Suppose we had two independent p-dimensional vector samples $\mathbb{X} := \{\mathbb{x}_1, \ldots, \mathbb{x}_{n_1}\}$ and $\mathbb{Y} := \{\mathbb{y}_1, \ldots, \mathbb{y}_{n_2}\}$ where $p \le n_2$. We assume that each sample comes from a (possibly different) population distribution with i.i.d. components and finite second moment.

[2] (a) How is the Fisher LSD related to the two vector samples $\mathbb{X}$ and $\mathbb{Y}$? What is the relationship between the two parameters (s,t) of $F_{s,t}$ and the three values (p, n_1, n_2) describing the dimensionality and sizes of $\mathbb{X}$ and $\mathbb{Y}$?

[2] (b) Now that you explained the relationship between $\mathbb{X}$, $\mathbb{Y}$, the values (p, n_1, n_2) and the parameters (s,t) of $F_{s,t}$ in part (a), what would you expect the empirical density of eigenvalues be for the following three choices of triplets (p, n_1, n_2):

$$(50, 100, 100), \qquad (75, 100, 200), \qquad (25, 100, 200).$$

Plot the three densities on the same figure with an appropriate legend.

[2] (c) Perform a simulation study for the same choices of the triplets (p, n_1, n_2) that are given in part (b). Setup your experiment correctly to demonstrate a histogram of eigenvalues and compare them to the appropriately parametrised densities from part (b). For each triplet (p, n_1, n_2), plot the histogram of eigenvalues, overlay the appropriate density, and ensure the plot is appropriately titled. Show the code for your simulation study.

[2] (d) Derive a formula for the first moment of the Fisher LSD in terms of its parameters s and t. You can assume that $h > t$ as this always holds by the definition of h.

[2] (e) Perform a numerical experiment to confirm the formula you derived in part (d). That is, choose 5 values of (s,t) then use your simulation study code (from part c) to generate empirical eigenvalues for those values and calculate their sample mean. Compare the sample means to your formula.

[2] (f) Describe what happens to the first moment when $t \to 0$ and $t \to 1$.

Question 2 [2 marks]

Show that the second moment

$$\int x^2 \, p_{s,t}(x) \, dx = \frac{h^2 + 1 - t}{(1-t)^3}.$$

Question 3 [2 marks]

Show that the variance equals $h^2/(1-t)^3$.