Homework 2

Due by
Tuesday 22 August 2017 17:00

Question 1 [5 Points]

Let Y be a random vector with multivariate Gaussian distribution $N_p(0, \Gamma)$. Show that if $\text{rank}(\Gamma) = p$ then

$$Y'\Gamma^{-1}Y \sim \chi^2(p),$$

where $\chi^2(p)$ denotes the chi-squared distribution with p degrees of freedom.

Question 2 [5 Points]

Calculate the following integral

$$I = \frac{1}{2\pi i} \oint_{|z|=1} z^2 e^{1/z} \, \mathrm{d}z$$

Question 3 [5 Points]

Calculate the following integral

$$I = \oint_{|z|=2} \frac{3+z}{z+z^2} \, \mathrm{d}z$$

Question 4 [5 Points]

Consider the situation where we may want to explain each response variable $Y \in \mathbf{R}$ by a p-dimensional variable $\mathbf{X} \sim \text{Unif}([0,1]^p)$.

Suppose our data consists of n i.i.d. observations $(Y_i, \mathbf{X}_i)_{\{i=1,\ldots,n\}}$ of the variables Y and $\mathbf{X}$. We could then model them with the classic regression equation

$$Y_i = f(\mathbf{X}_i) + \varepsilon_i, \quad i = 1, \ldots, n$$

with $f : [0,1]^p \to \mathbf{R}$ and $\varepsilon_1, \ldots, \varepsilon_n$ are independent and centered random variables.

It is typical to assume that the function f is smooth and we can estimate $f(x)$ by some averaging of the Y_i associated to the $\mathbf{X}_i$ in the vicinity of x. The simplest version of this idea is the k-nearest neighbor estimator where $f(x)$ is estimated by the mean of the Y_i associated with the k points $\mathbf{X}_i$ that are nearest to x. This works well in a low-dimensional setting as it is easy to make sense of what "nearest points" means.

(a) Show that the notion of nearest points vanishes as the dimensionality p increases by plotting the histogram of the distribution of pairwise-distances

$$\{\|\mathbf{X}_i - \mathbf{X}_j\| : 1 \leq i < j \leq n\}$$

for $n = 100$ and dimensions $p = 2$, 10, 100 and 1000. [3 Points]

(b) What do you observe? [2 Point]

This homework is to be submitted through Wattle in digital form only as per ANU policy. The R code for Question 4 must be supplied. If you use any references (note: this will never count against you), please clearly indicate which ones.