Advanced Statistics - Lab 05

1) Confirm by simulations the Johnson-Lindenstrauss Lemma. More precisely, generate N vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_N$, each with dimension d, where d is very large. Then, transform the vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_N$ to $\boldsymbol{y}_1, \boldsymbol{y}_2, ..., \boldsymbol{y}_N$, each with dimension n, where $n \ll d$, such that

$$\Pr\left\{(1-\delta)||\boldsymbol{x}_k - \boldsymbol{x}_j||_2 \leq ||\boldsymbol{y}_k - \boldsymbol{y}_j||_2 \leq (1+\delta)||\boldsymbol{x}_k - \boldsymbol{x}_j||_2\right\} \geq 1-\epsilon \tag{1}$$

holds for all vectors $k \neq j$. Repeat this process many times and thereby generate (1) by simulations and plot it for three different pairs of (δ, ϵ) as a function of n on a log-scale.

Upload the results on Moodle in a single PDF file or as the script itself that contains explanations, the code, and figures.

Important note: By failing to do the following, you will loose points:

- You must provide clear explanation of what your program is doing.
- Use a log-scale when piloting probabilities or tails.
- You must provide comments in your code in order for anyone to understand the code.
- You must not use in-bulid functions for obtaining the PDF, mean, variance, and probability.
- You must use different colors, lines, and markers in the plots, along with legends for each curve and suitable line-widths of the curves so that the figure is understandable.
- You muse clearly define what are the x and y axis in your figures.
- Finally, you must use caption that fully explains the figure.