Indian Institute of Technology Bombay

EP219 Data analysis and interpretation
Assignment 3
Dated : 25 - 9 - 2017

— — — — — — — — — — — — — — — —

Prove that for a sample of N variable pairs (X_i, Y_i), $i = 1..N$, the expectation value of the following quantity is exactly the covariance of X, Y.

$$C_{XY}^{\text{est}} = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \overline{X})(Y_i - \overline{Y})$$

Here $\overline{X}$ is the sample average of X_i (similarly $\overline{Y}$ is the sample average of Y_i). You may assume that successive trials are independent from each other. Write up the proof in your report.

Your Assignment

1. Find the full estimated covariance matrix of (pairwise) the number of facebook friends, number of posts and number of likes from the previous assignment.
2. Find the estimated correlation coefficient for each pair. Explain what the sign and magnitude of the correlation coefficient tells you.

Notes: (as before try to follow these guidelines)

- Make sure python 2.7/2.8 is installed. We will prefer this to python 3 for this course.
- Make sure to label all your plots, axes etc. Install latex so that you can use latex symbols in the plot legends.
- Try to experiment with histogram bins, axes range, colors, linestyles, plot markers, displaying multiple plots on the same image, saving plots to pdfs etc.
- Comment your code with detailed comments! Uncommented code will receive no credit.
- Try to follow best programming practices in python. https://gist.github.com/sloria/7001839