

# Data Mining & Machine Learning

## Project 3 - Sequential Estimation

Steinarr Hrafn Höskuldsson

September 2022

### Part 1.2

*Do you expect the batch estimate to be exactly  $(0, 1, -1)$  or  $(0, 1, 1)$ ? Which two parameters can be used to make this estimate more accurate?*

The batch estimate is expected to be around the mean but because of the randomness it will very unlikely be exactly the mean. Two parameters we can tune to decrease the difference in expected and observed means. The variance can be set to a smaller value. Causing the data points to be clustered closer to the mean. Another way is to increase the sample size  $N$ . Using more data points means the randomness will 'cancel out', so to speak, and bring the observed mean closer to the expected mean

### Independent

*What happens if the mean value changes (perhaps slowly) with time? What if  $\mu = (0, 1, -1)$  moves to  $\mu = (1, -1, 0)$  in 500 time ticks? How would we track the mean? Some sort of a forgetting could be added to the update equation. How would that be done?*

One way would be to only take the mean of the last  $M$  data points. This would be a moving average filter and would require the program to remember the last  $M$  data points which, depending on the program implementation, might be less than ideal. Another method better suited for our implementation might be to change the update equation to be

$$\mu_{ML}^N = \mu_{ML}^{N-1} + \frac{1}{\min(N, M)}(x_n - \mu_{ML}^{N-1})$$

Where  $M$  would be an appropriately chosen constant. This would prevent each successive update from having a smaller and smaller affect. Instead, every update would be able to have an appreciable affect.

In figure 1 we can see how such a system might behave when the mean is changing.  $M$  was selected as 75 by trial and error and, as can be seen, the graph follows the changing mean. The error of this sequential estimate hovers around 0.1. By choosing a smaller  $M$  the estimate follows the changing mean more closely but is more prone to fluctuate from random noise. Choosing a larger  $M$  results less noise in the estimation but the estimate will always lag in time.

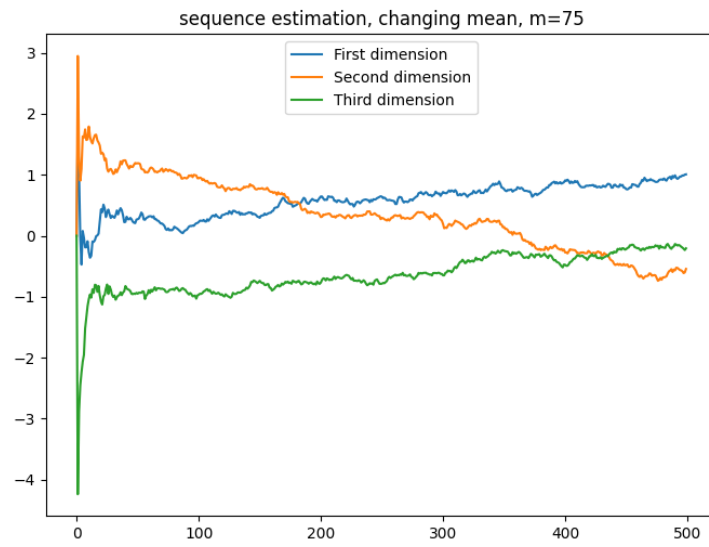


Figure 1: Sequential estimation of a dataset with a changing mean.

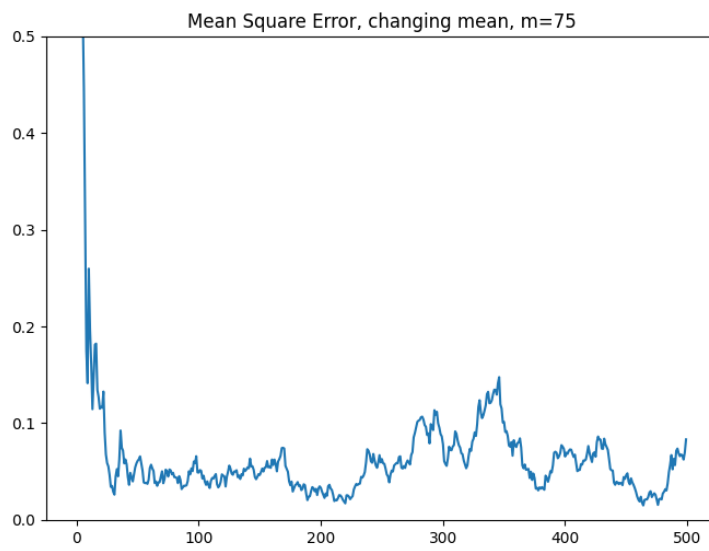


Figure 2: Error of the above estimation.

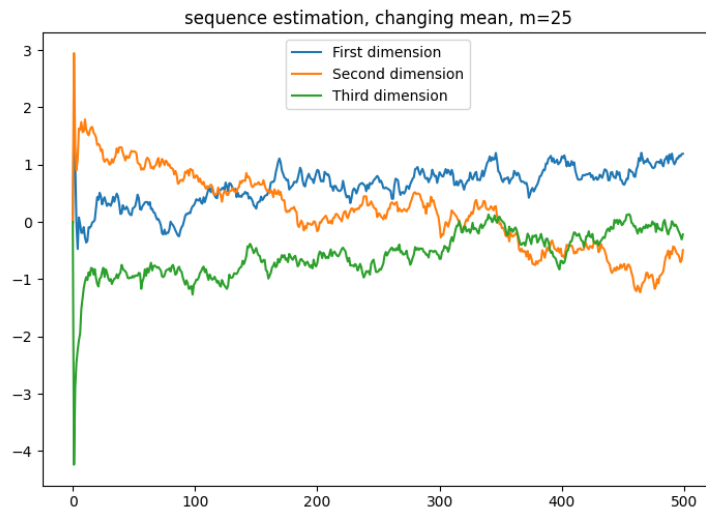


Figure 3: Smaller M

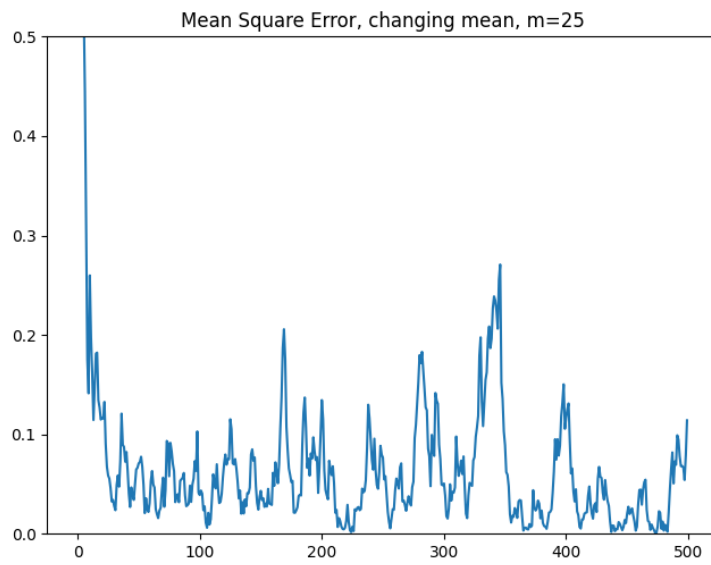


Figure 4: Error of the above estimation.

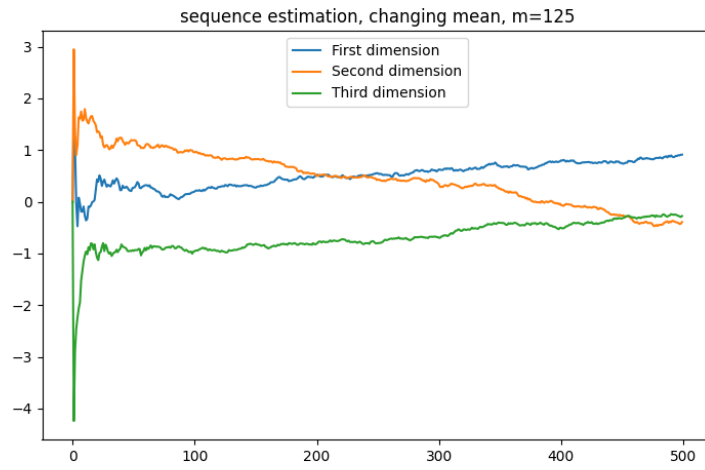


Figure 5: Larger M

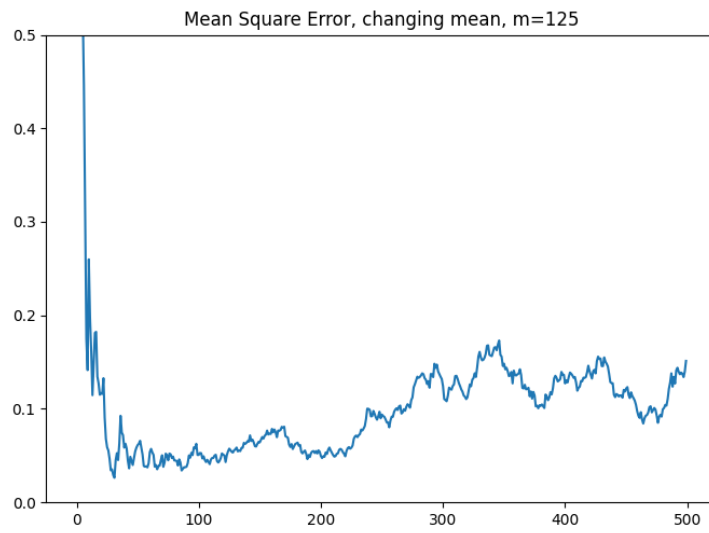


Figure 6: Error of the above estimation.