ASSIGNMENT #2

Comparison of the exponential and running mean for random walk model

Group 2

Ekaterina Karmanova

Timur Chikichev

Iaroslav Okunevich

Nikita Mikhailovskiy

Part 1. Determination of optimal smoothing constant in exponential mean

1. To conduct a simulation experiment, we have generated a true trajectory and it measurements by using random walk model:

$$X_i = X_{i-1} + w_i$$

Where: X – trajectory point,

w – normally distributed random noise

Size of trajectory was equal to 3000 and 300 points;

Initial condition was $X_1 = 10$;

Parameters of noise:

(Variance) $\sigma_w^2 = 9$;
(Mean) M = 0

To generate normally distributed noise in Python we were using function:

```
np.random.normal(mean, standart_deviation, size),
```

and definite iteration FOR to make value for each point of trajectory.

By using the same method, we defined measurements:

$$z_i=X_i+\eta_i$$

With noise, that has variance:

$\sigma_n^2 = 12$;

2. To statistically identify σ_n^2 and σ_w^2, we solved the system of equations (7,8) and used equations (5,6) from Topic 2 “Quasi-optimal approximation under uncertainty” p. 55:

$$\sigma_w^2 = E[\rho^2] - E[\nu^2],$$

$$\sigma_n^2 = E[\nu^2] - \frac{E[\rho^2]}{2},$$

Where

$$E[\rho^2] = \frac{1}{N-2}\sum_{k=3}^{N}(\omega_k + \omega_{k-1} + \eta_k - \eta_{k-2})^2,$$

$$E[\nu^2] = \frac{1}{N-1}\sum_{k=2}^{N}(\omega_k + \eta_k - \eta_{k-1})^2.$$

Result of calculations:

For 3000 points:

$$\sigma_w^2 = 9.997 \; and \; \sigma_n^2 = 11.946$$

For 300 points:

$$\sigma_w^2 = 11.737 \; and \; \sigma_n^2 = 11.633$$

Relative errors of results:

For 3000 points:

$$\varepsilon(\sigma_w^2) = 11\% \; and \; \varepsilon(\sigma_n^2) = 0.45\%$$

For 300 points:

$$\varepsilon(\sigma_w^2) = 30.41\% \; and \; \varepsilon(\sigma_n^2) = 3.06\%$$

Based on the result, we can say that with a greater number of points, the accuracy of variance will increase.

3. We determined the optimal smoothing coefficient in exponential smoothing:

$\chi = 0.75$,

$\alpha = 0.56873$.

4. Plot results of exponential smoothing:

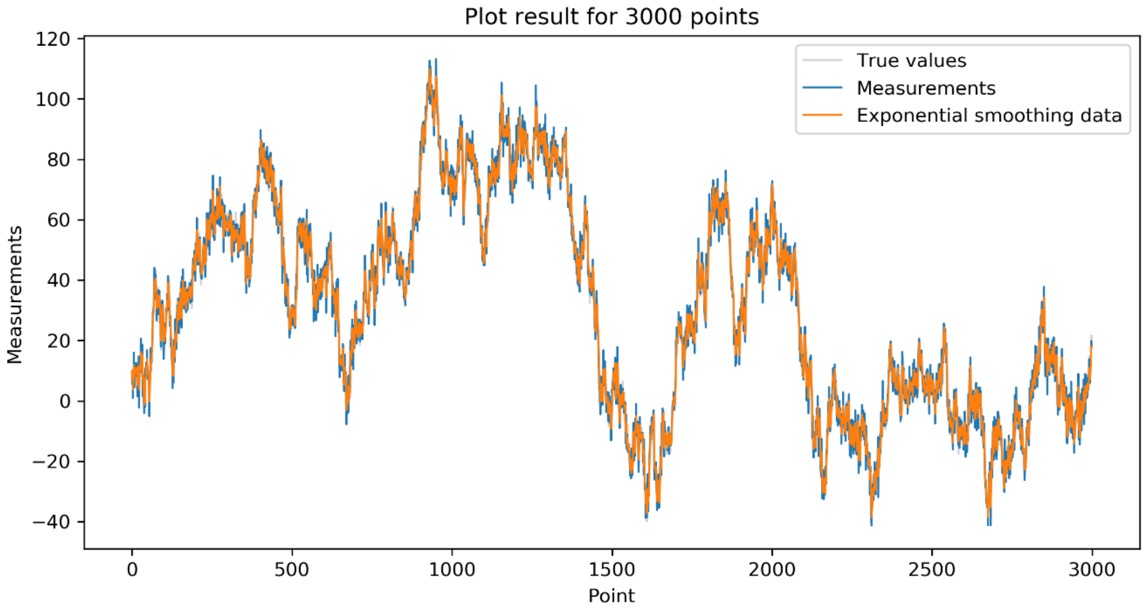

Figure 1 – Plot of result for 3000 points

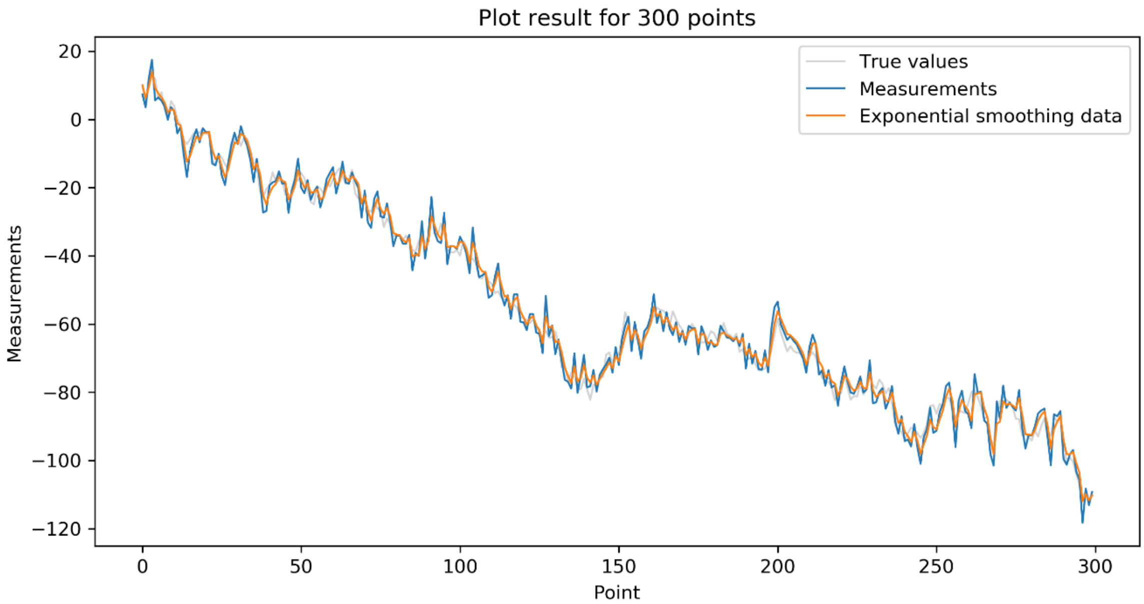

Figure 2 – Plot of result for 300 points

It is quite difficult to analyze original plots. On fig. 3, a zoomed plot of result for 300 points is presented. By analyzing this figure, we can say that the exponential smoothed curve repeats the shape of measurements, but the amplitude of smoothed lower. Exponential smoothing produces forecasts that little lag behind the actual trend.

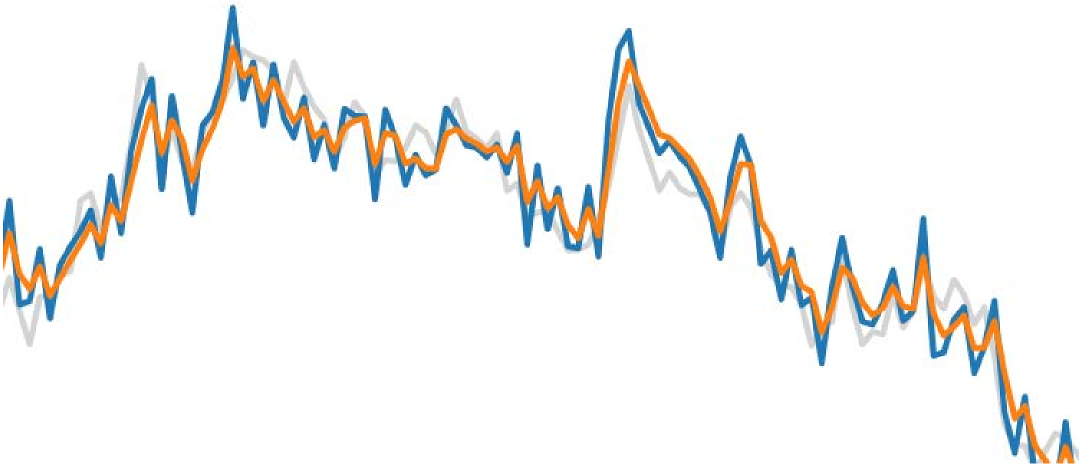

Figure 3 – Zoomed plot of result for 300 points,

Blue – Measurements;

Orange – Exponential smoothing;

Grey – true values.

Part 2 Comparison of methodical errors of exponential and running mean

1. We generated a true trajectory and measurements of it the same way as in part 1 with initial parameters:

Size of trajectory is 300 points;

Initial condition was $X_1 = 10$;

Parameters of noise:

(Variance) $\sigma_w^2 = 28^2$;

(Mean) $M = 0$

2. Parameters of measurements:

$\sigma_n^2 = 97^2$

3. Optimal smoothing coefficient for exponential mean is equal:

$\alpha = 0.25$

4. Window size for running mean smoothing calculated with this formula:

$$M = \frac{2-\alpha}{\alpha} = \frac{1.75}{0.25} = 7$$

5. We applied a running mean and an exponential mean for the same measurements to compare them. The comparison plot is in fig. 4.

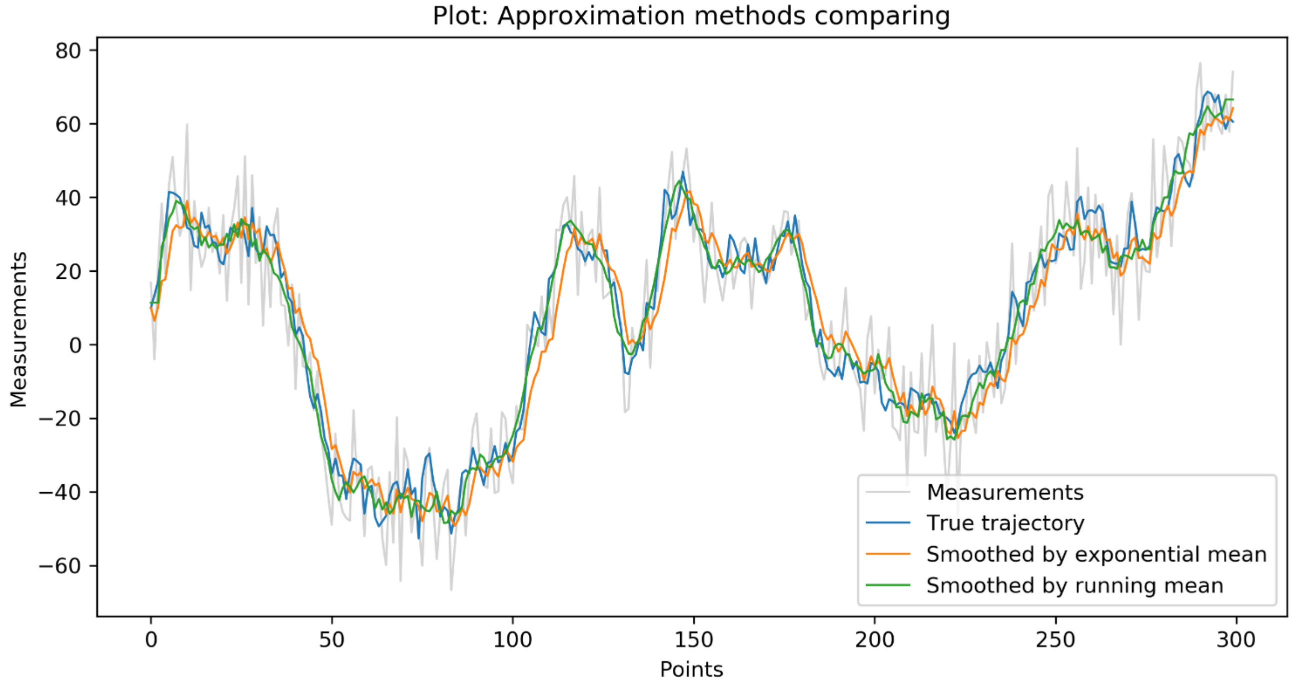

Figure 4 – Approximation methods comparing

6. The curve approximated with running mean loses data on fast-changing sections. Curve smoothed with an exponential mean has a shift to the right side. Those disadvantages give an error while analyzing and somewhere produce inversed changes to true trajectory. On the other side, these parameters of smoothing are optimal for current measurements, and changing of them does not give us any improvement.

Conclusions

During this assignment, we learn about the main advantages and disadvantages of two different quasi-optimal smoothing methods.

The exponential mean is easy to implement. Only three pieces of data are required for this method. Firstly, it needs the forecast for the most recent period. Secondly, it needs the actual value for that period. Finally, it needs the value of the smoothing constant, a weighting factor that reflects the weight given to the most recent data values. This method reacts quicker to the reverse in trend and shows the upward trend faster than the running mean method.

Smoothed curve by the running mean method is closer to the true trajectory and has a much smaller shift with the same variance of smoothing as in the exponential method. In the other side, it is less sensitive in insignificant sudden changes of true trajectory.

In addition, we found out that determination of window size for running mean depends on exponential smoothing coefficient only if the trajectory is a random walk model. Otherwise, if the function of the initial trajectory is known we cannot implement this method.

Moreover, we learned how to identify statistically variance of noise.