Multivariate Data Analysis Final Assignment

Marios Marios (M.Marinos@student.tudelft.nl)

November 5, 2020

1. Calculate and plot the position estimation error

On the X-axis we can see the N that goes from 0 to 1000 and on y axis it's the estimation error $\epsilon(n+1)$ with highest value of approximately 1.5.

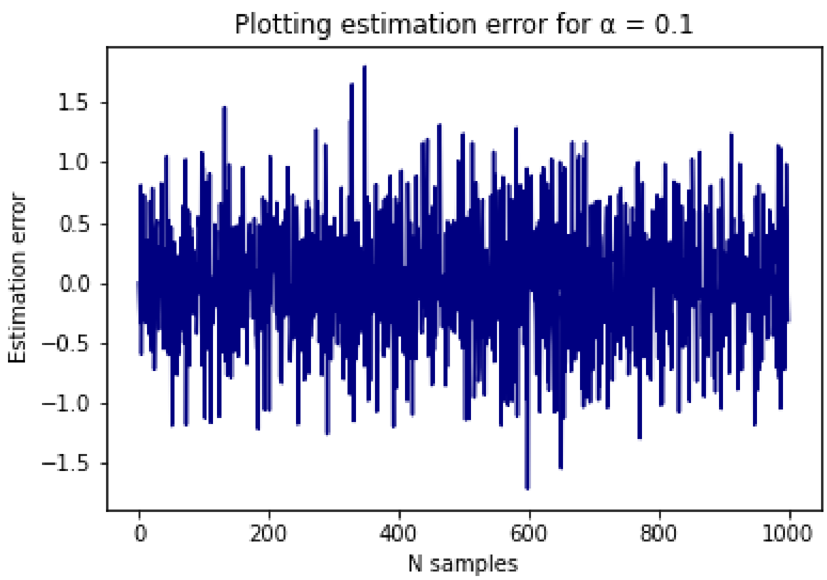

Figure 1: Plotting the estimation error for $\alpha = 0.1$

2. Estimate the variance of position estimation error for different alpha values.

Continuing the question 1. I created a new column on the dataframe that is the square of estimation error, and then got the mean of it. That was done for each different alpha value from -1 to 1. On the figure below we can see that the most suitable value of alpha is near $\alpha =$ -0.12 to -0.1 which is the minimum of our curve. Finally, we can observe as the alpha increases the variance.

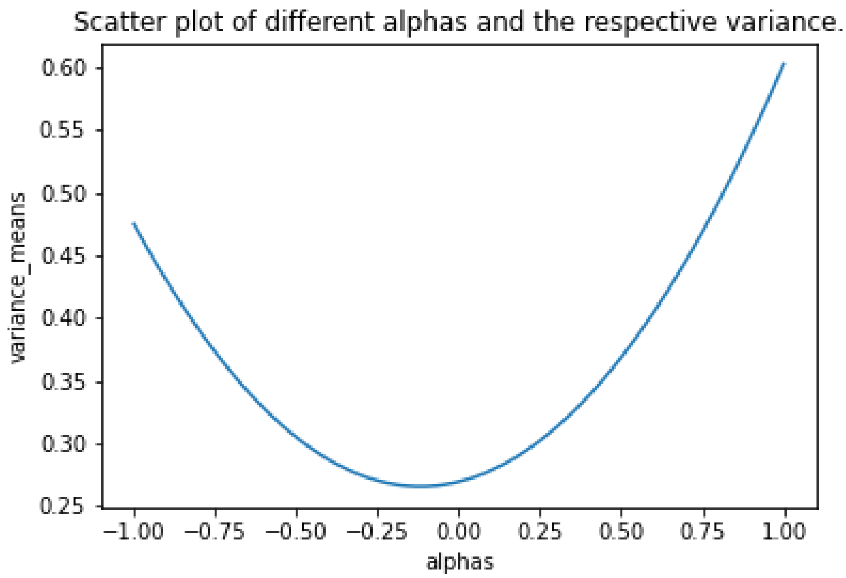

Figure 2: Plotting the variance mean for α values ranging from -1 to 1

3. Show the variance of the estimation error can also be expressed in terms of the auto-correlation function of the process X(n).

Starting by this equation $\sigma_\epsilon^2 = E[\epsilon(n+1)^2] = E[(X(n+1) - \hat{X}(n+1))^2]$ we have by expanding using the formula $(a-b)^2$ the equation (2).

$$\hat{X}(n+1) = X(n) + \alpha(X(n) - X(n-1)) \tag{1}$$

$$\sigma_\epsilon^2 = E[X^2(n+1) - 2X(n+1)\hat{X}(n+1) + \hat{X}^2(n+1)] \tag{2}$$

Using in the equation 2 the equation 1 we get :

$$\begin{gathered}
\sigma_\epsilon^2 = E[X^2(n+1) - 2X(n+1)(X(n) + \alpha(X(n) - X(n-1))) + (X(n) + \alpha(X(n) - X(n-1)))^2] \\
\Rightarrow E[X^2(n+1) - 2X(n+1)(X(n) + \alpha X(n) - \alpha X(n-1)) + (X(n) + \alpha X(n) - \alpha X(n-1))^2] \\
By\ applying\ the\ formula\ (a+b-c)^2 = +a^2 + b^2 + c^2 + 2ab - 2ac - 2bc\ wehave\ : \\
\Rightarrow E[X^2(n+1) - 2X(n)X(n+1) - 2\alpha X(n)X(n+1) + 2\alpha X(n+1)X(n-1) \\
+X^2(n) + \alpha^2 X^2(n) + \alpha^2 X^2(n-1) + 2\alpha X^2(n) - 2\alpha X(n)X(n-1) - 2a^2 X(n)X(n-1)] \\
\Rightarrow E[(2\alpha^2 + 2\alpha + 2)X^2(n) + (-2\alpha^2 - 4\alpha - 2)X(n)X(n+1) + 2\alpha X(n)X(n+2)] \\
\Rightarrow (2 + 2\alpha + 2\alpha^2)R_X(0) + (-2 - 4\alpha - 2\alpha^2)R_X(1) + 2\alpha R_X(2)
\end{gathered}$$

$$\Rightarrow Therefore,\ \sigma_\epsilon^2 = (2 + 2\alpha + 2\alpha^2)R_X(0) + (-2 - 4\alpha - 2\alpha^2)R_X(1) + 2\alpha R_X(2) \tag{3}$$

4. Minimize the variance with respect to a.

In order to find the minimum of equation (3) we need to take the derivative with respect to a and find where it's equal to 0.

$$\begin{aligned}
\frac{\partial \sigma_\epsilon^2}{\partial \alpha} &= (2+2\alpha+2\alpha^2)' R_X(0) + (-2-4\alpha-2\alpha^2)' R_X(1) + (2\alpha)' R_X(2) \\
&\Rightarrow (2+4\alpha)R_X(0) + (-4-4\alpha)R_X(1) + 2R_X(2) = 0 \qquad (4) \\
&\Rightarrow 2R_X(0) + 4\alpha R_X(0) - 4R_X(1) - 4\alpha R_X(1) + 2R_X(2) = 0 \\
&\Rightarrow 4\alpha R_X(0) - 4\alpha R_X(1) = -2R_X(0) + 4R_X(1) - 2R_X(2) \\
&\Rightarrow \alpha(4R_X(0) - 4R_X(1)) = 4R_X(1) - 2R_X(2) - 2R_X(0) \\
&\Rightarrow \alpha = \frac{4R_X(1) - 2R_X(2) - 2R_X(0)}{4R_X(0) - 4R_X(1)} \\
&\Rightarrow \alpha = \frac{2R_X(1) - R_X(2) - R_X(0)}{2(R_X(0) - R_X(1))} \qquad (5)
\end{aligned}$$

If we also take the second derivative we get $4R_X(0) - 4R_X(1)$ which is positive if we consider the fact that $R_X(0) \geq |R_X(k)| \forall k$. Hence, the α found is minimum.

5. Calculate and plot optimal value of α as a function of ρ

If we plug in the function given $R_X(k) = \sigma_X^2 \rho^{|k|}$ into equation (5) we get $\alpha = \frac{2\sigma_X^2\rho - \sigma_X^2\rho^2 - \sigma_X^2}{2(\sigma_X^2 - \sigma_X^2\rho)}$ .If we get out the factor σ^2 in both the numerator and the denominator and thus they cancel.

$$\alpha = \frac{2\rho - \rho^2 - 1}{2(1-\rho)} \Rightarrow \frac{-(\rho-1)^2}{2(1-\rho)} \Rightarrow \frac{\rho - 1}{2} \qquad (6)$$

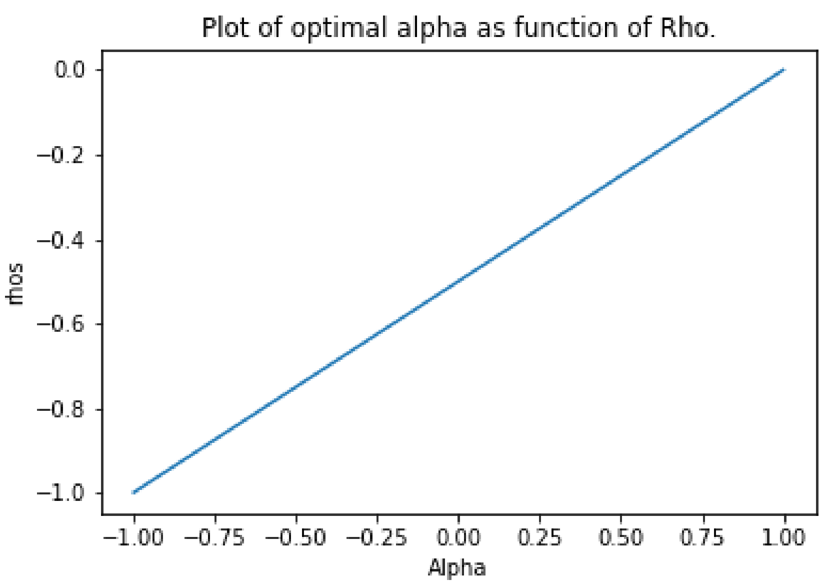

Figure 3: Plotting the variance mean for α values ranging from -1 to 1

6. Estimate the value of from the data X(n), and find the optimal value of α

Now, if we calculate the

$$R_X(1) = E[(n)X(n+1)] = \sigma_X^2 \rho$$
$$R_X(2) = E[(n)X(n+2)] = \sigma_X^2 \rho^2$$

So we have the case of k = 1, k = 2 and we have to estimate the $E[X(n)X(n+1)]$ and $E[X(n)X(n+2)]$ respectively to find out the estimation of ρ.

$$E[(n)X(n+1)] = \frac{1}{N}\sum_{i=1}^{N-1} X(i)X(i+1) \Rightarrow \rho = 0.8686$$
$$E[(n)X(n+2)] = \frac{1}{N}\sum_{i=1}^{N-2} X(i)X(i+2) \Rightarrow \rho^2 = 0.76797\cdots \Rightarrow \rho = \pm\sqrt{0.76797}$$

Given the values of ρ, we can calculate the optimal α based on the equation (6). For k = 1 we have $\alpha = -0.0657$. If we now consider the k = 2 and do the same we have for the 2 different $\rho' s$ we get $\alpha = -0.06183\ldots$. So we can conclude that the optimal value of α is around $\alpha \approx -0.063\ldots$

7. Estimate the autocorrelation function based on the input file

To estimate autocorrelation function using the following equation :

$$R_X(k) = \frac{1}{N}\sum_{i=1}^{N-k} X(i) * X(i+k) \tag{7}$$

We have to consider that we can **only sum** until $N - k$ rather than N as we have finite amount of data. Using equation 6. we get the below plot :

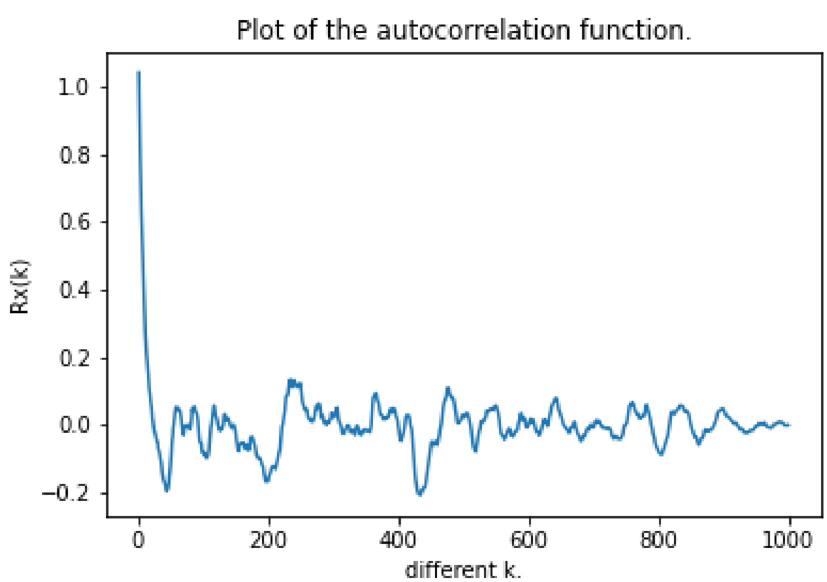

Figure 4: Plotting the variance mean for α values ranging from -1 to 1

8. Use the result under part (g.) and Equation (4) to plot the variance of position estimation error (n + 1) as a function of α

Using the estimate of the Autocorrelation function and equation (3) we plot the variance of position estimation error as a function of α. $\alpha \in [-1, 1]$ increasing it by 0.01.

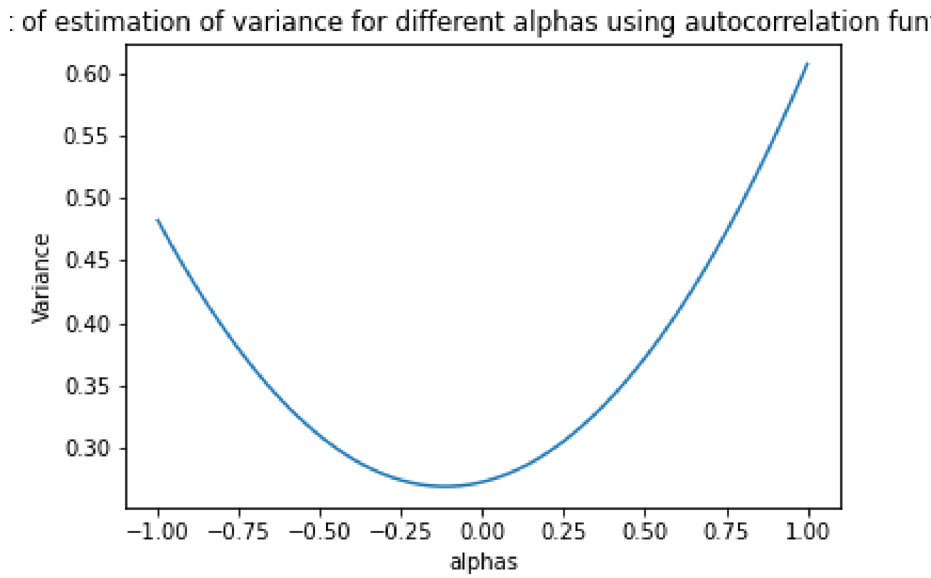

Figure 5: Plotting the variance of position estimation error as a function of α.

By observing Figure 5 a suitable value for a would be around -0.11 which line up with the previous results and it looks very similar in what we got on question (2).

9. Use the result under parts (d.) and (g.) to find the optimal value of α

If we fill in the $R_X(0), R_X(1), R_X(2)$ from question 7. in equation (4), we can calculate the actual optimal minimum α :

$$(4\alpha + 2) * 1.0418186912944625 + (-4\alpha - 4) * 0.9055397698880611 + 2 * 0.8005967705204181 = 0 \tag{8}$$

By solving equation (8), the optimal α is :

a	-0.114969804997615

10. . Explain the differences (if they exist) between the calculated optimal/suitable values of α in parts (b.), (f.), and (i.).

The values obtained from question 2, 9 seems very similar, if not exactly the same, thus there is no difference in calculated values of α . If we consider now the optimal α that was obtained from question 6 is a bit difference than the α obtained from question 2 and question 9. That makes quite sense, because calculating the α on question 6 we only consider one estimated value of the autocorrelation function whereas on the two others we get into account all the data provided.

Question 2	-0.117...
Question 6	-0.063...
Question 9	-0.115...

Table 1: Table of different optimal alphas.