Soham Chatterjee
Email: soham.chatterjee@tifr.res.in
Course: Probability Theory

Assignment - 5
Dept: STCS
Date: November 10, 2024

Problem 1

Let X, Y_1, Y_2 be three random variables with joint density f_{X,Y_1,Y_2}. For a fixed y_1, consider two random variables $\tilde{X}$, $\tilde{Y}_2$ with joint distribution $g_{\tilde{X},\tilde{Y}_2}$ defined as $g_{\tilde{X},\tilde{Y}_2}(x, y_2) = f_{X,Y_2|Y_1}(x, y_2 \mid y_1)$. Show that

$$\mathbb{E}[\tilde{X} \mid \tilde{Y} = y_2] = \mathbb{E}[X \mid Y_1 = y_1, Y_2 = y_2]$$

Solution: We have

$$g_{\tilde{X}|\tilde{Y}_2}(x \mid y_2) = \frac{g_{\tilde{X},\tilde{Y}_2}(x, y_2)}{g_{\tilde{Y}_2}(y_2)} = \frac{f_{X,Y_2|Y_1}(x, y_2 \mid y_1)}{f_{Y_2|Y_1}(y_2 \mid y_1)} = f_{X|Y_1,Y_2}(x \mid y_1, y_2)$$

Therefore $\mathbb{E}[\tilde{X} \mid \tilde{Y} = y_2] = \mathbb{E}[X \mid Y_1 = y_1, Y_2 = y_2]$. ■

Problem 2

Consider the Kalman filtering problem for the scalar system:

$$\begin{aligned} X_k &= \alpha X_{k-1} + W_k \\ Y_k &= hX_k + Z_k \end{aligned}$$

as described in class (i.e., $W_k \sim N\left(0, \sigma_W^2\right)$ i.i.d, $Z_k \sim N\left(0, \sigma_Z^2\right)$, and X_1 are independent). The initial condition is $X_1 \sim N\left(0, \sigma_{X_1}^2\right)$. For the numerical exercises below you can assume $\sigma_{X_1}^2 = \sigma_Z^2 = \sigma_W^2 = h = 1$.

(a) Plot sample paths of the process $\{X_k\}$ for different values of α. Pick a representative set of values of α to show the effect of α on how the sample paths look like. Can you explain qualitatively the effect?

(b) Let $\hat{X}_k = E\left[X_k \mid Y_1, \dots, Y_k\right]$. For those sample paths of $\{X_k\}$ plotted in part (a), plot in the same figure the sample paths of the estimates $\left\{\hat{X}_k\right\}$. What is the qualitative effect of α on the estimation errors?

(c) Let $\tilde{X}_k = E\left[X_k \mid Y_k\right]$. This is the state estimate based only on the current observation. For the sample paths in (a) and (b), plot the sample paths of $\left\{\tilde{X}_k\right\}$ in the same figure as well. How does the difference in the accuracy of the estimators $\hat{X}_k$ and $\tilde{X}_k$ depend on the value of α ? Explain qualitatively.

(d) Let f_k be the conditional distribution of X_k give the observations up to time k. For your favorite value of α, plot f_k for several values of k to get a feel of how the distribution evolves in time. Do these distributions depend on the random outcome of the experiment? How?

(e) What happens to the distribution of X_k as $k \to \infty$? Give a quantitative answer. Does your answer depend on α ? Does your answer depend on $\sigma_{X_1}^2$?

(f) What happens to the MMSE estimation error σ_k^2 of $\hat{X}_k$ as $k \to \infty$? Does it converge to zero, a finite non-zero value or infinity? How does your answer depend on α ? An answer supported by numerical evidence together with some analysis would be fine; it doesn't have to be totally rigorous.

Solution:

(a) Here we have taken the values of α to be $\{-1, 0.8, 1, 1.2\}$ in Figure 1(a). Here we can see that when the value of α is 1.2 then the sample value increases. And when the value of α is -1 it oscillates around 0. But for $\alpha = 0.8$ the sample values remains close to 0. Therefore the sample values converges when $|\alpha| < 1$ and otherwise diverges.

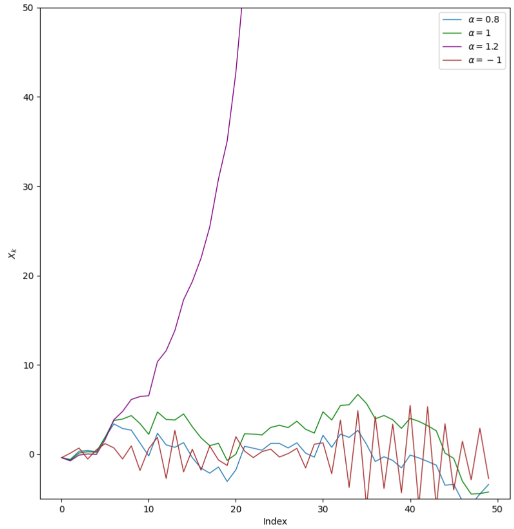
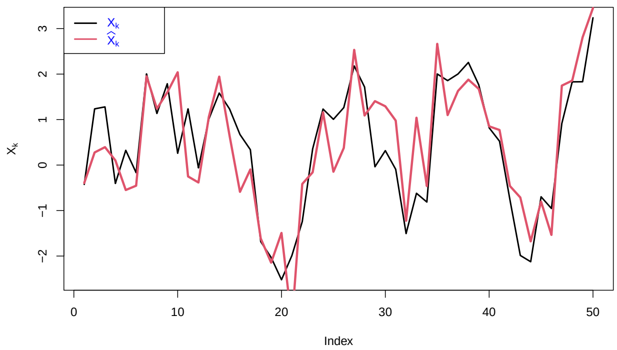
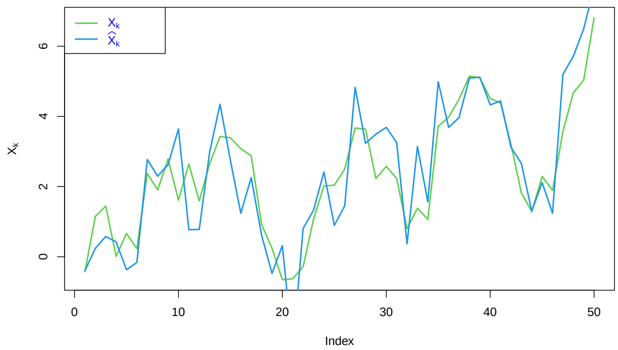


Figure 1: Plot of X_k for different $\alpha \in \{-1, 0.8, 1, 1.2\}$

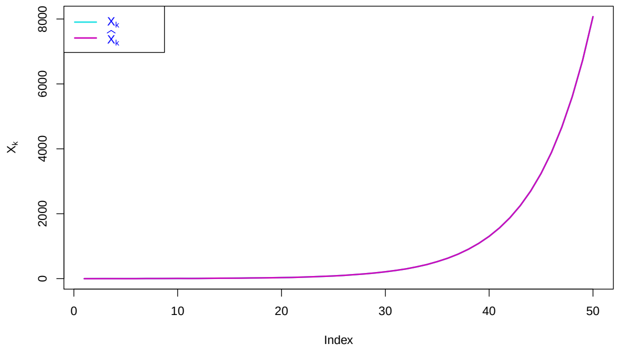
(b) In the following plot we can see that the predicted values $\mathbb{E}[X \mid Y_1, \ldots, Y_k]$ matches almost correctly with the sample values X_k. From the plots we conclude that as $|\alpha$ becomes larger it has lesser effect on the estimation which we also showed in part (f) where we showed if $|\alpha|$ becomes larger then the MMSE estimation is independent of α.



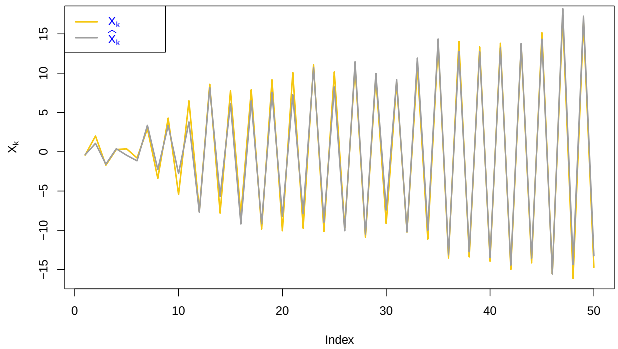
(a) Plot of X_k vs $\hat{X}_k$ for $\alpha = 0.8$



(b) Plot of X_k vs $\hat{X}_k$ for $\alpha = 1$



(c) Plot of X_k vs $\hat{X}_k$ for $\alpha = 1.2$



(d) Plot of X_k vs $\hat{X}_k$ for $\alpha = -1$

Figure 2: Compared X_k and predicted $\hat{X}_k$

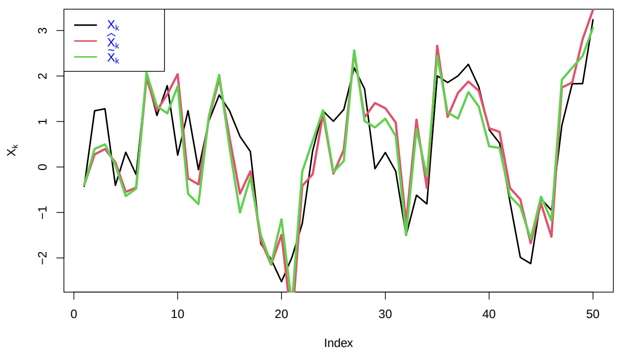
(c) Here we compare X_k, $\hat{S}_k$ and $\tilde{X}_k = \mathbb{E}[X_k \mid Y_k]$ for all values of α. Now we have $\mathrm{Cov}(X_k, Y_k) = h\rho_k^2$ and $\mathrm{Var}[Y_k] = h^2\rho_k^2 + \sigma_Z^2$. Hence we have

$$\mathbb{E}[X_k \mid Y_k] = \frac{h\rho_k^2 Y_k}{h^2\rho_k^2 + \sigma_Z^2}$$

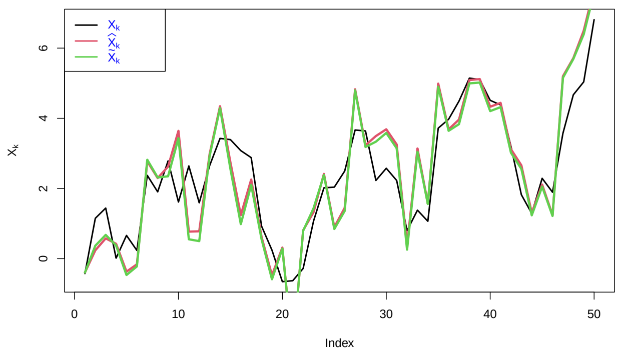
So we have

$$\begin{aligned}
\mathbb{E}[X_k - \tilde{X}_k]^2 &= \mathbb{E}\left[X_k - \frac{h\rho_k^2 Y_k}{h^2\rho_k^2 + \sigma_Z^2}\right]^2 \\
&= \frac{1}{\left(h^2\rho_k^2 + \sigma_Z^2\right)^2}\mathbb{E}\left[(h^2\rho_k^2 + \sigma_Z^2)X_k - h\rho_k^2(hX_k + Z_k)\right]^2 \\
&= \frac{1}{\left(h^2\rho_k^2 + \sigma_Z^2\right)^2}\mathbb{E}[\sigma_Z^2 X_k - h\rho_k^2 Z_k]^2 \\
&= \frac{\sigma_Z^4\rho_k^2 + h^2\rho_k^4\sigma_Z^2}{\left(h^2\rho_k^2 + \sigma_Z^2\right)^2} = \frac{\sigma_Z^2\rho_k^2}{h^2\rho_k^2 + \sigma_Z^2}
\end{aligned}$$

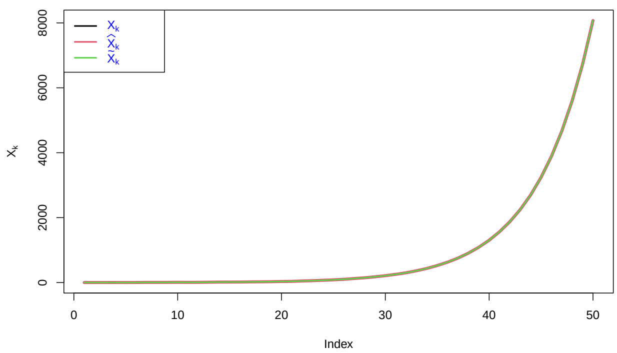
Now this is the MMSE estimation of σ_k^2 of X_k which comparing with part (f) we can see that we obtained the same estimation value. Therefore both $\hat{X}_k$ and $\tilde{X}_k$ are equally good estimating sample values.



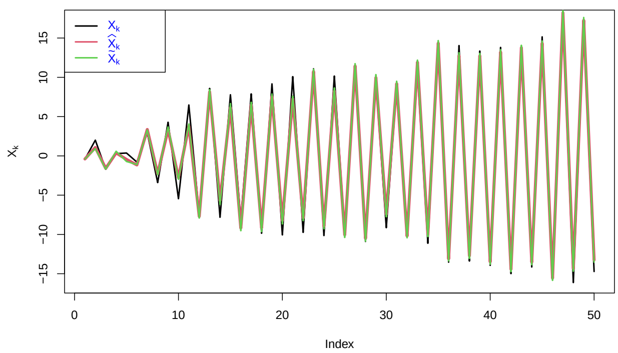
(a) Plot of X_k vs $\hat{X}_k$ vs $\tilde{X}_k$ for $\alpha = 0.8$



(b) Plot of X_k vs $\hat{X}_k$ vs $\tilde{X}_k$ for $\alpha = 1$



(c) Plot of X_k vs $\hat{X}_k$ vs $\tilde{X}_k$ for $\alpha = 1.2$



(d) Plot of X_k vs $\hat{X}_k$ vs $\tilde{X}_k$ for $\alpha = -1$

Figure 3: Compared X_k and predicted $\hat{X}_k$ and $\tilde{X}_k$

(d) Here we plot f_k for values $k \in [10]$ with $\alpha = 0.8$. Now the conditional distribution $X \mid Y_1, \ldots, Y_k$ approaches the distribution $N\left(0, \frac{\sigma 2}{1-\alpha^2}\right)$ for large k and also we notice from the plot that this doesn't depend on the Y_k.

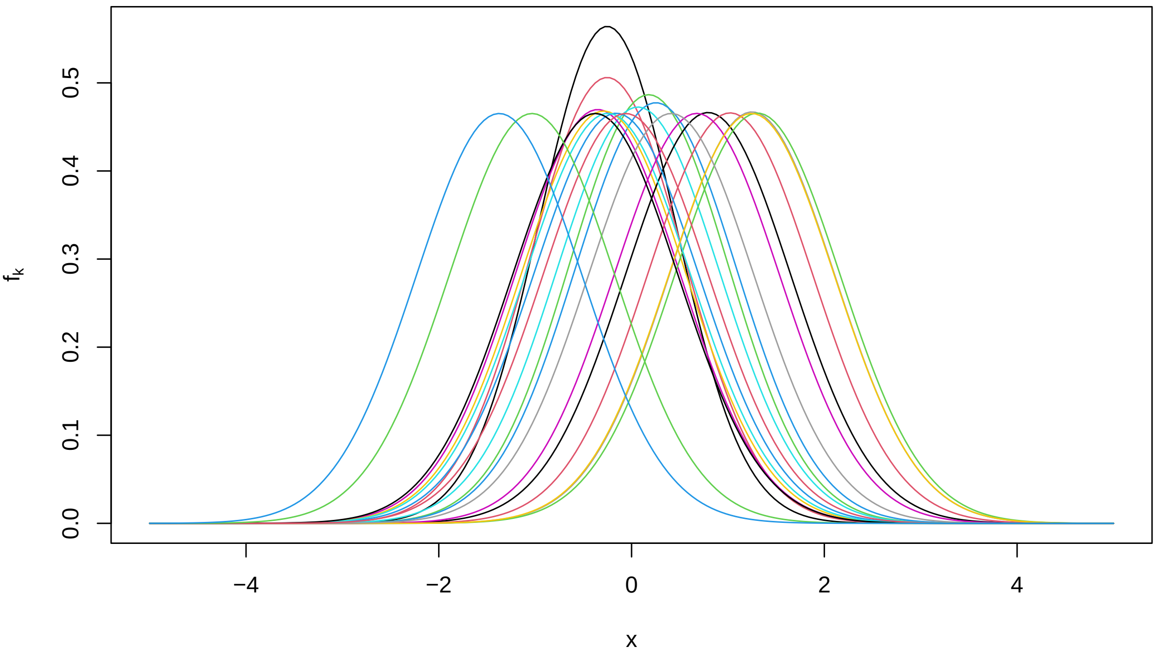


Figure 4: Plot of density f_k of $X_k \mid Y_1, \ldots, Y_k$ for $\alpha = 0.8$, $k \in [10]$

(e) We will induct on k. Since X_1, W_2 are independent and we have $X_2 = \alpha X_1 + W_2$ hence $X_2 \sim N\left(0, \alpha^2 \sigma_{X_1}^2 + \sigma_W^2\right)$. Now X_{k-1} and W_k are independent and we have $X_k = \alpha X_{k-1} + W_k$. By inductive hypothesis X_{k-1} follows Gaussian Distribution. Hence X_k also follows Gaussian Distribution. Hence $\mathbb{E}[X_k] = 0$. Now we have to calculate $\operatorname{Var}[X_k]$.

$$\operatorname{Var}[X_k] = \alpha^2 \operatorname{Var}[X_{k-1}] + \sigma_W^2 = \alpha^2(\alpha^2 \operatorname{Var}[X_{k-2}] + \sigma_W^2) + \sigma_W^2 = \cdots = \alpha^{2k-2}\sigma_{X_1}^2 + \sigma_W^2 \sum_{i=0}^{k-1} \alpha^{2i}$$

Therefore $X_k \sim N\left(0, \alpha^{2k-2}\sigma_{X_1}^2 + \sigma_W^2 \sum_{i=0}^{k-1} \alpha^{2i}\right)$. Hence if $|\alpha| < 1$, $\lim_{k\to\infty} \operatorname{Var}[X_k] = \frac{\sigma_W^2}{1-\alpha^2}$. Hence as $k \to \infty$, $X_k \to N\left(0, \frac{\sigma_W^2}{1-\alpha^2}\right)$. Now if $|\alpha| \geq 1$, then as $k \to \infty$, α^{2k-2} diverges. Therefore $\operatorname{Var}[X_k]$ diverges to $+\infty$.

If $|\alpha| < 1$ then $\operatorname{Var}[X_k] = \frac{\sigma_W^2}{1-\alpha^2}$. Hence it doesn't depend on $\sigma_{X_1}^2$.

(f) Let ρ_k denote the variance of X_k. Then we have the formula

$$\rho_n = \alpha^2 \rho_{n-1}^2 + \sigma_W^2 \quad \text{for } n \geq 2$$

Hence we know the behavior of the conditional variance σ_k^2 of X_k. Hence we know the MMSE of $X_k \mid Y_1, \ldots, Y_k$ as $k \to \infty$. Now we have

$$\sigma_k^2 = \frac{\rho_k^2 \sigma_Z^2}{h^2 \rho_k^2 + \sigma_Z^2} = \frac{\sigma_Z^2}{h^2 + \frac{\sigma_Z^2}{\rho_k^2}}$$

From the previous part if $|\alpha| < 1$ then $\lim_{k\to\infty} \rho_k = \frac{\sigma_W^2}{1-\alpha^2}$ and if $|\alpha| \geq 1$ then as $k \to \infty$, ρ_k diverges to $+\infty$. Therefore when $|\alpha| < 1$, $\lim_{k\to\infty} \sigma_k^2 = \frac{\sigma_Z^2 \sigma_W^2}{h^2 \sigma_W^2 + (1-\alpha^2)\sigma_Z^2}$ and when $|\alpha| \geq 1$ we have $\lim_{k\to\infty} \sigma_k^2 = \frac{\sigma_Z^2}{h^2}$.

■

Problem 3

For the system in Problem 2 derive a recursive algorithm for computing the one-step ahead estimator: $\mathbb{E}[X_k \mid Y_1, Y_2, \ldots, Y_k, Y_{k+1}]$. This means we can look ahead one step to estimate the state.

Solution:

■