EE 503: Problem Set #11 : Solutions

I. LEON-GARCIA PROBLEM 7.22

Solution:

a) We have $X_i \sim Bern(1/2)$ so $m = 1/2$ and $\sigma^2 = 1/4$.

$$\begin{aligned} P\left[40 \leq \sum_{i=1}^{100} X_i \leq 60\right] &= P\left[-10 \leq \sum_{i=1}^{100}(X_i - 1/2) \leq 10\right] \\ &= P\left[\frac{-10}{\sqrt{100\sigma^2}} \leq G_{100} \leq \frac{10}{\sqrt{100\sigma^2}}\right] \\ &= P\left[-2 \leq G_{100} \leq 2\right] \\ &= 1 - 2Q(2) \\ &= 0.954499736103642 \end{aligned}$$

Also

$$\begin{aligned} P\left[50 \leq \sum_{i=1}^{100} X_i \leq 55\right] &= P\left[0 \leq \sum_{i=1}^{100}(X_i - 1/2) \leq 5\right] \\ &= P\left[0 \leq G_{100} \leq \frac{5}{\sqrt{100\sigma^2}}\right] \\ &= P\left[0 \leq G_{100} \leq 1\right] \\ &\approx 1/2 - Q(1) \\ &= 0.341344746068543 \end{aligned}$$

Note that we could use a continuity correction here with $[39.5, 60.5]$ and $[49.5, 55.5]$ to get slightly more accurate results, but we do not bother to do so (students are not expected to do the continuity correction on any of these CLT problems).

b) We have

$$\begin{aligned} P\left[400 \leq \sum_{i=1}^{1000} X_i \leq 600\right] &= P\left[-100 \leq \sum_{i=1}^{1000}(X_i - 1/2) \leq 100\right] \\ &= P\left[\frac{-100}{\sigma\sqrt{1000}} \leq G_{1000} \leq \frac{100}{\sigma\sqrt{1000}}\right] \\ &= P\left[\frac{-200}{\sqrt{1000}} \leq G_{1000} \leq \frac{200}{\sqrt{1000}}\right] \\ &\approx 1 - 2Q(200/\sqrt{1000}) \\ &= 0.999999999746037 \end{aligned}$$

Also

$$\begin{aligned} P\left[500 \leq \sum_{i=1}^{1000} X_i \leq 550\right] &= P\left[0 \leq \sum_{i=1}^{1000}(X_i - 1/2) \leq 50\right] \\ &= P\left[0 \leq G_{1000} \leq \frac{50}{\sqrt{1000\sigma^2}}\right] \\ &= P\left[0 \leq G_{1000} \leq \frac{100}{\sqrt{1000}}\right] \\ &\approx 1/2 - Q(100/\sqrt{1000}) \\ &= 0.499217298870999 \end{aligned}$$

II. Leon-Garcia problem 7.25

Solution:
We have $\{X_i\}$ i.i.d. $Exp(\lambda)$ with $\lambda = 1/36$ and so $m = 36$ and $\sigma^2 = 36^2$. So

$$\begin{aligned} P\left[\sum_{i=1}^{16} X_i < 600\right] &= P\left[\sum_{i=1}^{16}(X_i - 36) < 24\right] \\ &= P[G_{16} < \frac{24}{\sigma\sqrt{16}}] \\ &= P[G_{16} < 1/6] \\ &\approx 1 - Q(1/6) \\ &= 0.566183832610904 \end{aligned}$$

III. Leon-Garcia problem 7.15

Solution:
For $\epsilon > 0$ we have by the Chebyshev inequality:

$$\begin{aligned} P[|N(t)/t - \lambda| \geq \epsilon] &= P[|N(t) - \lambda t| \geq \epsilon t] \\ &\leq \frac{Var(N(t))}{\epsilon^2 t^2} \\ &= \frac{\lambda t}{\epsilon^2 t^2} \\ &= \frac{\lambda}{\epsilon^2 t} \end{aligned}$$

IV. Leon-Garcia problem 7.18

Solution:
We have $\{X_i\}$ are i.i.d. $N(0,1)$. For $\epsilon > 0$ the Chebyshev inequality gives

$$P[|M_n - \mu| < \epsilon] \geq 1 - \frac{\sigma^2}{n\epsilon^2}$$

In our case we have $\sigma^2 = 1$, $\mu = 0$, $M_n \sim N(0, 1/n)$, so letting $G = \sqrt{n}M_n \sim N(0,1)$ gives

$$P[|M_n - \mu| < \epsilon] = P[|M_n| < \epsilon] = P[|G| < \epsilon\sqrt{n}] = 1 - 2Q(\epsilon\sqrt{n})$$

and

$$1 - \frac{\sigma^2}{n\epsilon^2} = 1 - \frac{1}{n\epsilon^2}$$

- Case $n = 16$: We see the bound is upheld in all cases considered in the table below. For $\epsilon < 0.25$ the Chebyshev-inspired bound is useless.

ϵ	$1 - 2Q(\epsilon\sqrt{n})$	$1 - \frac{1}{n\epsilon^2}$
0.2	0.576289202833207	-0.562500000000000
0.25	0.682689492137086	0
0.3	0.769860659556583	0.305555555555556
0.4	0.890401416600884	0.609375000000000
0.5	0.954499736103642	0.750000000000000
1	0.999936657516334	0.937500000000000
1.5	0.999999998026825	0.972222222222222

- Case $n = 81$: We see the bound is upheld in all cases considered in the table below. For $\epsilon < 1/9$ the Chebyshev-inspired bound is useless.

ϵ	$1 - 2Q(\epsilon\sqrt{n})$	$1 - \frac{1}{n\epsilon^2}$
0.11	0.677825880978338	-0.020304050607081
0.2	0.928139361774148	0.691358024691358
0.3	0.993066052393919	0.862825788751715
0.4	0.999681782819685	0.922839506172839
0.5	0.999993204653751	0.950617283950617
1	1	0.987654320987654
1.5	1	0.994513031550069

V. Covariance matrices

Let X= $[X_1\ X_2\ X_3]^T$ be a random vector with μ= [5 -5 6]^T. And the covariance given by:

$$\begin{bmatrix} 5 & 2 & -1 \\ 2 & 5 & 0 \\ -1 & 0 & 4 \end{bmatrix}$$

Calculate the mean and variance of $Y = A^T X + B$, where A=(2 -1 2)^T and B=5.
Solution:

$$\begin{aligned} M_Y &= A^T * Mx + B \\ &= 27 + 5 \\ &= 32 \end{aligned}$$

$$\begin{aligned} K_Y &= A^T * Kx * A \\ &= (2\ -1\ 2) * (6\ -1\ 6)^T \\ &= 25 \end{aligned}$$

VI. MAP and ML estimator

X and Y are discrete random variables with joint probability mass function given by the table

		Y		
		-1	0	1
X	-1	K	$2K$	$3K$
	0	0	K	$2K$
	1	0	0	K

Where K is a non-negative constant.
(a) Find the value of K
(b) Find the best MAP estimator for X given Y.
(c) Find the best ML estimator for X given Y.
Solution:(a)

$$3K + 2 * 2K + 3K = 1$$
$$K = \frac{1}{10}$$

(b)

$$\hat{X}_{MAP}(y) = argmax_x P(X = x|Y = y)$$

For $y = 1$, choose $x = -1$
For $y = 0$, choose $x = -1$
For $y = -1$, choose $x = -1$
So, $\hat{X}_{MAP}(y) = -1 \quad \forall y \in \{-1, 0, 1\}$.
(c)

$$\hat{X}_{ML}(y) = argmax_x P(Y = y|X = x)$$

For $y = 1$, choose $x = 1$
For $y = 0$, choose $x = 0$ or $x = -1$ [this is a tie]
For $y = -1$, choose $x = -1$

$$\hat{X}_{ML}(y) = \begin{cases} 1 & y = 1 \\ 0 \ or \ -1 & y = 0 \\ -1 & y = -1 \end{cases}$$

VII. Minimum MSE estimator

Given the following pdf

$$f_{X,Y}(x,y) = \begin{cases} 2 & x + y < 1, x \geq 0, y \geq 0 \\ 0 & elsewhere \end{cases}$$

(a) Find the best constant MSE estimator $\hat{X}_C = C$ for X and its mean-square error.
(b) Find the best homogeneous linear MSE estimator $\hat{X}_{HL} = aY$ for X (given Y) and its mean-square error.
(c) Find the best non-homogeneous linear MSE estimator $\hat{X}_{HL} = aY + b$ for X (given Y) and its mean-square error.
Solution:

$$f(x) = \begin{cases} 2(1-x) & 0 \leq x \leq 1 \\ 0 & elsewhere \end{cases}$$

(a) Constant mean square error estimator $\hat{X}_C = C$:

$$error = E[(X - C)^2] = E[X^2] - 2CE[X] + C^2$$

Let the derivative of error equals to 0:

$$\begin{aligned} -2E[X] + 2C &= 0 \\ C &= E[X] \end{aligned}$$

$$\begin{aligned} E[X] &= \int_0^1 x f(x) dx = \frac{1}{3} \\ E[X^2] &= \int_0^1 x^2 f(x) dx = \frac{1}{6} \end{aligned}$$

Therefore:

$$\begin{aligned} C &= \frac{1}{3} \\ error &= \frac{1}{18} \end{aligned}$$

(b) Homogeneous linear MSE estimator $\hat{X}_{HL} = aY$:

$$error = E[(X - aY)^2] = E[X^2] - 2aE[XY] + E[a^2Y^2]$$

Let the derivative of error equals to 0:

$$\begin{aligned} -2E[XY] + 2aE[Y^2] &= 0 \\ a &= \frac{E[XY]}{E[Y^2]} \end{aligned}$$

$$\begin{aligned} E[Y^2] &= E[X^2] = \frac{1}{6} \\ E[XY] &= \int_0^1 \int_0^{1-y} 2xy dx dy \\ &= \int_0^1 y(1-y)^2 dy \\ &= \frac{1}{12} \end{aligned}$$

Therefore:

$$a = \frac{1}{2}$$
$$error = \frac{1}{6} - \frac{1}{12} + \frac{1}{24}$$
$$= \frac{1}{8}$$

(c) Non-homogeneous linear MSE estimator $\hat{X}_{HL} = aY + b$:

$$error = E[(X - aY - b)^2] = E[X^2] - 2E[X(aY + b)] + E[(aY + b)^2]$$

Let the derivative of error equals to 0 (for a and b respectively):
Derivative for b:

$$-2E[X] + 2aE[Y] + 2b = 0$$
$$b = E[X] - aE[Y]$$

Derivative for a:

$$-2E[XY] + 2aE[Y^2] + 2bE[Y] = 0$$
$$-2E[XY] + 2aE[Y^2] + 2E[X]E[Y] - 2aE[Y]^2 = 0$$
$$a = \frac{E[XY] - E[X]E[Y]}{E[Y^2] - E[Y]^2}$$

Therefore:

$$a = \frac{1/12 - 1/9}{1/6 - 1/9}$$
$$= -\frac{1}{2}$$
$$b = \frac{1}{3} + \frac{1}{6}$$
$$= \frac{1}{2}$$
$$error = \frac{1}{24}$$

VIII. Sensor measurement and estimation 1

A device parameter X is a continuous random variable that takes values over some interval $[a, b]$. We take a measurement Y. Let $\hat{X}(Y)$ be some estimator of X based on Y. Define $\tilde{X}(Y) = [\hat{X}(Y)]_a^b$, which projects $\hat{X}(Y)$ onto the interval $[a, b]$. Show that the MSE for $\tilde{X}(Y)$ is less than or equal to the MSE for $\hat{X}(Y)$. Hint: Use the fact that $W \leq Z \implies \mathbb{E}[W] \leq \mathbb{E}[Z]$.

Solution:

Claim: $(\tilde{X}(Y) - X)^2 \leq (\hat{X}(Y) - X)^2$.

Proof: Consider three cases:

1) Case 1: Suppose $\hat{X}(Y) \in [a, b]$. In this case $\tilde{X}(Y) = \hat{X}(Y)$ and so the desired result holds with equality.
2) Case 2: Suppose $\hat{X}(Y) > b$. Then $\tilde{X}(Y) = b$ and

$$0 \leq b - X = \tilde{X}(Y) - X \leq \hat{X}(Y) - X$$

and so

$$(\tilde{X}(Y) - X)^2 \leq (\hat{X}(Y) - X)^2$$

3) Case 3: Suppose $\hat{X}(Y) < a$. This is similar to case 2.

We use the claim to finish up:

$$(\tilde{X}(Y) - X)^2 \leq (\hat{X}(Y) - X)^2 \implies \mathbb{E}\left[(\tilde{X}(Y) - X)^2\right] \leq \mathbb{E}\left[(\hat{X}(Y) - X)^2\right]$$

IX. Sensor measurement and estimation 2

A device parameter X is a continuous random variable that takes values over $[-2, 2]$ and is uniform over that interval. We take a measurement $Y = \rho X + N$ where $\rho \neq 0$ is a given parameter and $N \sim N(0, \sigma^2)$ is independent noise. Define $\hat{X}_1(Y) = aY$ as the linear MMSE estimator (for an optimized a). Define $\hat{X}_2(Y) = [\hat{X}_1(Y)]_{-2}^{2}$. Define $\hat{X}_3(Y) = \mathbb{E}[X|Y]$. Assume $\rho = 0.7$.

a) Compute the optimal a and the corresponding MSE for estimator $\hat{X}_1(Y)$.

b) Compute $\hat{X}_3(y) = \mathbb{E}[X|Y = y]$ for all $y \in \mathbb{R}$. It will be an integral divided by another integral (if you want, you can compute the numerator in closed form, and you can compute the denominator in terms of the $Q()$ function if you want).

c) Write a program that, for values $\sigma^2 \in [0.01, 20]$ does the following: Given σ^2, run $n = 10000$ samples to generate i.i.d. (X_i, Y_i) values. Take the empirical values $MSE1_{empirical} = \frac{1}{n}\sum_{i=1}^{n}(\hat{X}_1(Y_i) - X_i)^2$. Plot as a function of σ^2 and compare with the theoretical value for MSE in part (a), also compare with the constant $Var(X)$ which is the MSE obtained by the best constant estimator.

d) In the same graph, compare $MSE2_{empirical} = \frac{1}{n}\sum_{i=1}^{n}(\hat{X}_2(Y_i) - X_i)^2$. Explain your observations.[1]

Solution:

a) Here X and Y both have zero mean so the optimal linear estimator $\hat{X}(Y) = aY + b$ has $b^* = 0$, and $Var(X) = \mathbb{E}\left[X^2\right]$, $Var(Y) = \mathbb{E}\left[Y^2\right]$ and $Cov(X, Y) = \mathbb{E}[XY]$ and so

$$a^* = \frac{\mathbb{E}[XY]}{\mathbb{E}[Y^2]}$$
$$MSE^* = \mathbb{E}\left[X^2\right] - \frac{\mathbb{E}[XY]^2}{\mathbb{E}[Y^2]}$$

Now

$$\begin{aligned}
\mathbb{E}[XY] &= \mathbb{E}[X(\rho X + N)] \\
&= \rho\mathbb{E}\left[X^2\right] + \mathbb{E}[X]\,\mathbb{E}[N] \\
&= \rho\frac{4}{3}
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}\left[Y^2\right] &= \mathbb{E}\left[(\rho X + N)^2\right] \\
&= \rho^2\mathbb{E}\left[X^2\right] + 2\rho\mathbb{E}[X]\,\mathbb{E}[N] + \mathbb{E}\left[N^2\right] \\
&= \rho^2\frac{4}{3} + \sigma^2
\end{aligned}$$

So

$$\begin{aligned}
a^* &= \frac{\rho(4/3)}{\rho^2(4/3) + \sigma^2} = \frac{\rho}{\rho^2 + (3/4)\sigma^2} \\
MSE^* &= (4/3) - \frac{\rho^2(4/3)^2}{\rho^2(4/3) + \sigma^2} \\
&= \frac{(4/3)\sigma^2}{\rho^2(4/3) + \sigma^2} \\
&= \frac{\sigma^2}{\rho^2 + (3/4)\sigma^2}
\end{aligned}$$

[1] Neely constructed this problem because he expected there to be a significant difference between the empirical MSE for $\hat{X}_1, \hat{X}_2, \hat{X}_3$, with MSE getting better for better estimators. He observed that, in this particular case, all MSE values were roughly the same. Can you confirm these results? You can also play around with different parameter choices in this and the next problem.

b) We have

$$\begin{aligned}\hat{X}_3(y) &= \mathbb{E}\left[X|Y=y\right]\\ &= \int_{-2}^{2} x f_{X|Y}(x|y)dx\\ &= \int_{-2}^{2} x \frac{f_{Y|X}(y|x)f_X(x)}{f_Y(y)}dx\\ &= \frac{\int_{-2}^{2} x f_{Y|X}(y|x)f_X(x)dx}{\int_{-2}^{2} f_{Y|X}(y|x)f_X(x)dx}\\ &= \frac{\int_{-2}^{2} x f_{Y|X}(y|x)(1/4)dx}{\int_{-2}^{2} f_{Y|X}(y|x)(1/4)dx}\\ &= \frac{\int_{-2}^{2} x f_{Y|X}(y|x)dx}{\int_{-2}^{2} f_{Y|X}(y|x)dx}\end{aligned}$$

Now we have

$$\begin{aligned}F_{Y|X}(y|x) &= P[Y \leq y|X=x]\\ &= P[\rho X + N \leq y|X=x]\\ &= P[N \leq y-\rho x|X=x]\\ &= P[N \leq y-\rho x] \quad \text{[by independence of } X \text{ and } N]\\ &= F_N(y-\rho x)\end{aligned}$$

Thus

$$f_{Y|X}(y|x) = \frac{d}{dy}F_N(y-\rho x) = f_N(y-\rho x) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(y-\rho x)^2}{2\sigma^2}}$$

Substituting this gives

$$\boxed{\hat{X}(y) = \mathbb{E}\left[X|Y=y\right] = \frac{\int_{-2}^{2} x e^{-\frac{(y-\rho x)^2}{2\sigma^2}}dx}{\int_{-2}^{2} e^{-\frac{(y-\rho x)^2}{2\sigma^2}}dx}}$$

c/d) The data is shown in Fig. 1. As σ^2 varies over the interval $[0.01, 20]$ the simulated value of MSE1 closely matches the theoretical computed value of $\frac{\sigma^2}{\rho^2+(3/4)\sigma^2}$. When noise goes to infinity, it approaches the horizontal asymptote of $\mathbb{E}\left[X^2\right]$ which is the MSE associated with the best constant estimator that does not use the measurement Y. Surprisingly, there is no noticeable difference between MSE1, MSE2, MSE3, even though we would assume:

$$MSE1 > MSE2 > MSE3$$

In particular, it seems that $MSE1 \approx MSE3$ (and so MSE2 is sandwiched in the middle so is also close to MSE1). If X had a Gaussian distribution then it can be shown (X, Y) is jointly Gaussian, in which case the best nonlinear estimator is the same as the best linear estimator. While X is not Gaussian here, the noise N is Gaussian so perhaps this is "close enough" to the jointly Gaussian case for the best linear estimator to be almost as good as the best nonlinear estimator.

X. Sensor measurement and estimation 3

A device parameter X is a continuous random variable that takes values over $[-1, 1]$ and is uniform over that interval. We take a measurement $Y = \rho X^3 + N$ where $\rho \neq 0$ is a given parameter and $N \sim N(0, \sigma^2)$ is independent noise. Define $\hat{X}_1(Y) = aY$ as the linear MMSE estimator (for an optimized a). Define $\hat{X}_3(Y) = \mathbb{E}\left[X|Y\right]$. Assume $\rho = 0.7$ and observe that Y now depends on X^3 rather than X.

a) Compute the optimal a and the corresponding MSE for estimator $\hat{X}_1(Y) = aY$, the best linear estimator.

b) Compute $\hat{X}_3(y) = \mathbb{E}\left[X|Y=y\right]$ in terms of an integral. The numerator of the integral cannot be solved in closed form.

c) Write a program that, for values $\sigma^2 \in [0.0001, 0.1]$ does the following: Given σ^2, run $n = 1000$ samples to generate i.i.d. (X_i, Y_i) values. Take the empirical values $MSE1_{empirical} = \frac{1}{n}\sum_{i=1}^{n}(\hat{X}_1(Y_i) - X_i)^2$. Plot as a function of σ^2 and compare with the theoretical value for MSE in part (a), also compare with the constant $Var(X)$ which is the MSE obtained by the best constant estimator, and with $MSE3_{empirical} = \frac{1}{n}\sum_{i=1}^{n}(\hat{X}_3(Y_i) - X_i)^2$. You can use the matlab integral command if you like (a sample will be given on Piazza). Your MSE3 should always be at least as good as MSE1, but should be a little better when σ^2 is small.

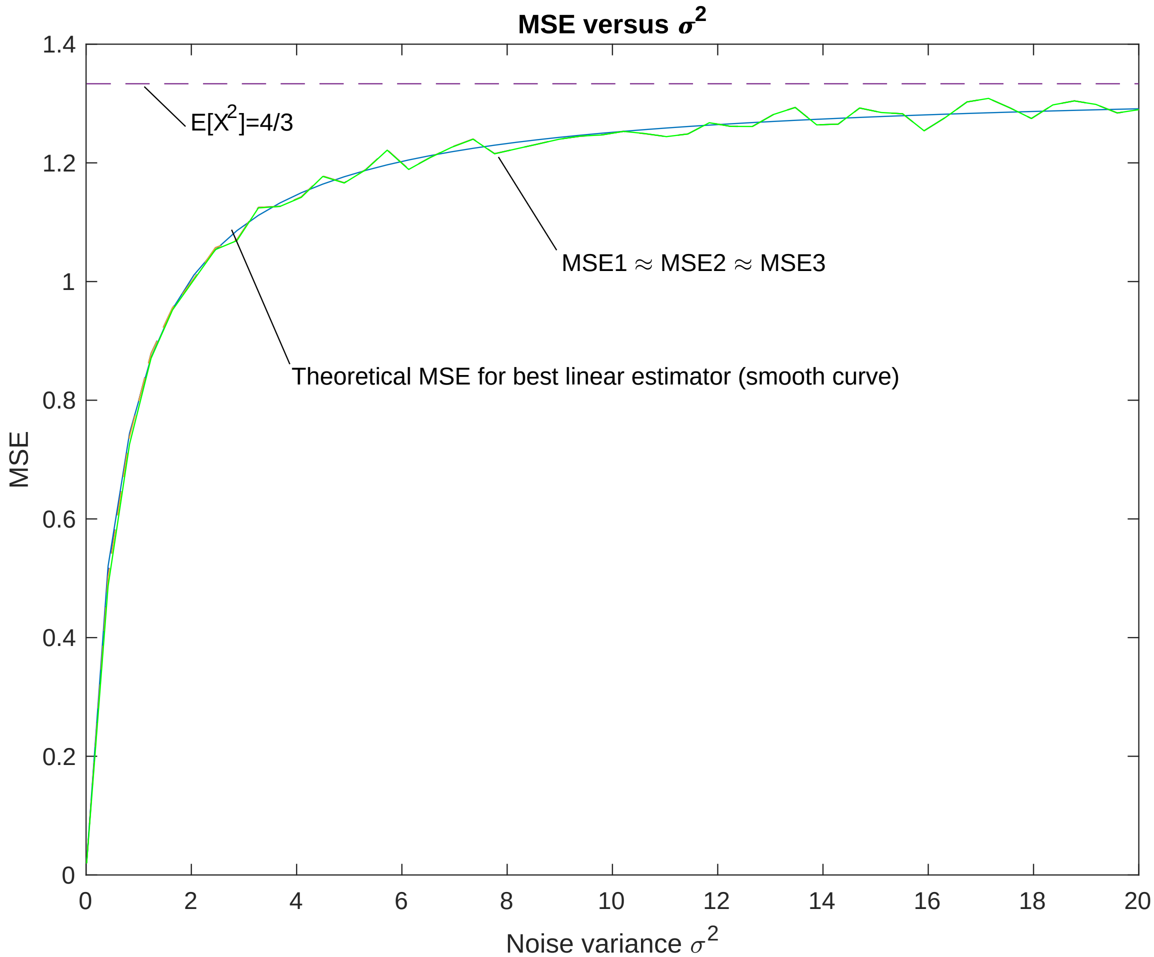

Fig. 1. Simulation for 10000 samples in problem IX with $Y = \rho X + N$.

Solution:

a) Here we also have $\mathbb{E}[X] = 0, \mathbb{E}[Y] = 0$ so the best linear estimator has $\hat{X}(Y) = aY + b$ with $b^* = 0$, and $Var(X) = \mathbb{E}\left[X^2\right]$, $Var(Y) = \mathbb{E}\left[Y^2\right]$ and $Cov(X,Y) = \mathbb{E}[XY]$ and so

$$a^* = \frac{\mathbb{E}[XY]}{\mathbb{E}[Y^2]}$$
$$MSE^* = \mathbb{E}\left[X^2\right] - \frac{\mathbb{E}[XY]^2}{\mathbb{E}[Y^2]}$$

Now

$$\begin{aligned}\mathbb{E}[XY] &= \mathbb{E}\left[X(\rho X^3 + N)\right]\\ &= \rho\mathbb{E}\left[X^4\right] + \mathbb{E}[X]\,\mathbb{E}[N]\\ &= \rho(1/5)\end{aligned}$$

$$\begin{aligned}\mathbb{E}\left[Y^2\right] &= \mathbb{E}\left[(\rho X^3 + N)^2\right]\\ &= \mathbb{E}\left[\rho^2 X^6 + 2\rho X^3 N + N^2\right]\\ &= \rho^2\mathbb{E}\left[X^6\right] + 2\rho\mathbb{E}\left[X^3\right]\mathbb{E}[N] + \mathbb{E}\left[N^2\right]\\ &= \rho^2(1/7) + \sigma^2\end{aligned}$$

So

$$a^* = \frac{\rho(1/5)}{\rho^2(1/7) + \sigma^2}$$
$$MSE^* = (1/3) - \frac{\rho^2(1/5)^2}{\rho^2(1/7) + \sigma^2}$$

b) As before, we have

$$\begin{aligned}\hat{X}(y) &= \mathbb{E}\left[X|Y = y\right] \\ &= \int_{-1}^{1} x f_{X|Y}(x|y)dx \\ &= \int_{-1}^{1} x \frac{f_{Y|X}(y|x)f_X(x)}{f_Y(y)}dx \\ &= \frac{\int_{-1}^{1} x f_{Y|X}(y|x)f_X(x)dx}{\int_{-1}^{1} f_{Y|X}(y|x)f_X(x)dx} \\ &= \frac{\int_{-1}^{1} x f_{Y|X}(y|x)dx}{\int_{-1}^{1} f_{Y|X}(y|x)dx}\end{aligned}$$

Now we have

$$\begin{aligned}F_{Y|X}(y|x) &= P[Y \leq y|X = x] \\ &= P[\rho X^3 + N \leq y|X = x] \\ &= P[N \leq y - \rho x^3|X = x] \\ &= P[N \leq y - \rho x^3] \quad \text{[by independence of } X \text{ and } N\text{]} \\ &= F_N(y - \rho x^3)\end{aligned}$$

Thus

$$f_{Y|X}(y|x) = \frac{d}{dy}F_N(y - \rho x^3) = f_N(y - \rho x^3) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(y-\rho x^3)^2}{2\sigma^2}}$$

Substituting this gives

$$\boxed{\hat{X}(y) = \mathbb{E}\left[X|Y = y\right] = \frac{\int_{-1}^{1} x e^{-\frac{(y-\rho x^3)^2}{2\sigma^2}}dx}{\int_{-1}^{1} e^{-\frac{(y-\rho x^3)^2}{2\sigma^2}}dx}}$$

c) Data is plotted in Fig. 2. Note that the simulated data for MSE1 closely matches its theoretical value $(1/3) - \frac{\rho^2(1/5)^2}{\rho^2(1/7)+\sigma^2}$ for all σ^2 values tested. Note that

$$\lim_{\sigma^2 \to 0} MSE1(\sigma^2) = (1/3) - (7/25) = 4/75 = 0.05333$$

So the linear estimator error does not go to zero when the noise is zero. We see this in Fig. 2 for $\sigma^2 \approx 0$. On the other hand, the nonlinear estimator has MSE that goes to 0 when $\sigma^2 \to 0$. So we see that the nonlinear estimator is strictly better than the linear estimator in the low noise regime. However, for medium and high levels of noise, both estimators have similar MSE. While not shown in the figure, as $\sigma^2 \to \infty$ both estimators have MSE that approaches $1/3$, which is the MSE associated with the best constant estimator that does not use any measurement Y.

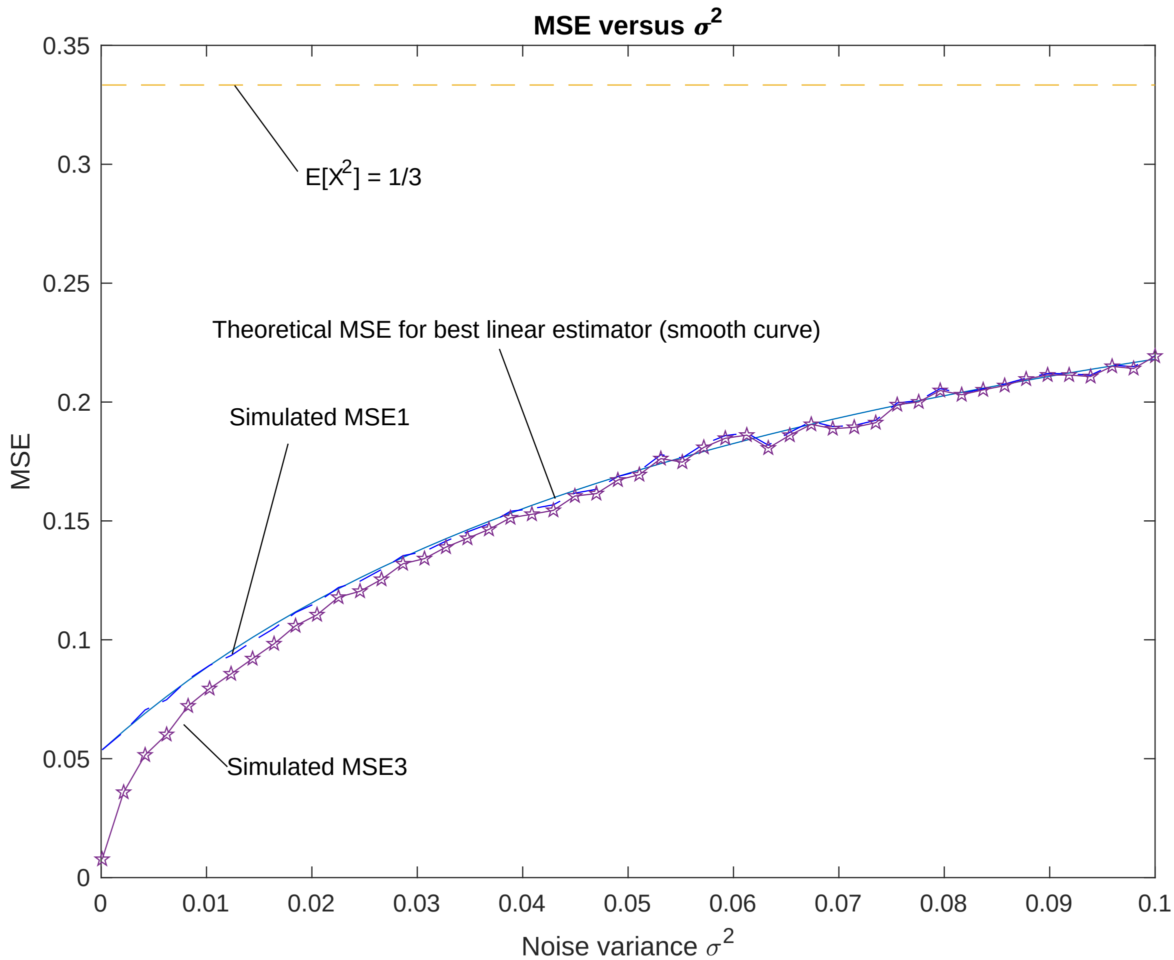

Fig. 2. Simulation data for 10000 samples in Problem X with $Y = \rho X^3 + N$.