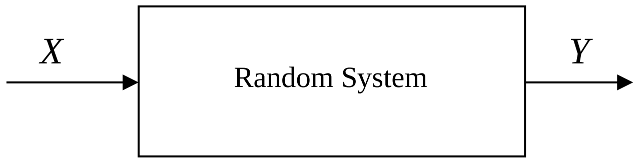
Class Exercise 5 – Estimation of scalar random variables

Definition of the problem:

Givan the random variable Y, we would like to estimate a hidden random variable X. The relation between X and Y is some statistical connection, for example,



where the system can be an addition of some type of random noise, a multiplication by another random variable, etc. Thus, there exists a statistical model which connects between the measurement Y and the parameter we would like to estimate X. Given the value of Y, we want to estimate the value of X. Denote the estimator, which is a function of Y, as $\hat{X} = g(Y)$.

The method of estimation depends on the criterion of quality of the estimation, namely, what is considered a good estimator. For example, an estimator which minimizes the error probability $\mathrm{P}(\hat{X} \neq X)$ can be found, as well as an estimator that minimizes the mean square error $\mathrm{E}[(X - \hat{X})^2]$. Overall, we would like to minimize:

$$D = \mathrm{E}[d(e)] \qquad e = X - \hat{X}$$

where $d(\cdot)$ is a non-negative distortion measure and e is the estimation error.

Criterion of Minimum Probability of Error:

Distortion measure appropriate for the problem:

$$d(e) = \begin{cases} 1, & e \neq 0 \\ 0, & e = 0 \end{cases}$$

The optimal estimator: the solution to the problem rises from the Maximum A-posteriori Probability (MAP) estimation, and its optimal estimator is given by:

$$\hat{X}_{MAP} = g_{MAP}(y) = \arg\max_x \mathrm{P}(X = x \mid Y = y)$$

Criterion of Minimum Mean Square Error:

Distortion measure appropriate for the problem:

$$d(e) = e^2$$

The solution is an estimator which brings to minimum the mean squared error (Minimum Mean Squared Error (MMSE) Estimator). Let us examine two cases: optimal estimator and optimal linear estimator.

The optimal estimator: the estimator which minimizes the mean squared error is

$$\hat{X}_{opt} = \mathrm{E}[X \mid Y]$$

Notice that $\mathrm{E}[X \mid Y]$ is **only** a function of Y !!!

- The optimal estimator gives zero expectation error, that is, $E[e] = 0$.
- The mean squared error in the case of the optimal estimator:
  $$Var(e) = \mathrm{E}[e^2] = \mathrm{E}[(X - \hat{X})^2] = \mathrm{E}\Big[\mathrm{E}[X^2 \mid Y] - \mathrm{E}^2[X \mid Y]\Big] = \mathrm{E}[Var(X \mid Y)]$$
  Namely, averaging of the conditional variance according to Y.
  Another expression for the mean squared error:
  $$Var(e) = \sigma_X^2 - \sigma_{\hat{X}_{opt}}^2$$
- Perpendicularity property – an estimator is an MMSE estimator iff the estimation error is orthogonal to all functions of the measurements:
  $$\forall g(Y) \qquad \mathrm{E}[e \cdot g(Y)] = 0$$

The optimal linear estimator: for all estimators of the form $aY + b$ (a, b constants), that which brings the mean squared error to a minimum is given by:

$$\hat{X}_{BLE} = \eta_X + \frac{\sigma_{XY}}{\sigma_Y^2}(Y - \eta_Y)$$

- The optimal linear estimator gives zero expectation error.
- The mean squared error in the case of the optimal linear estimator:
  $$Var(e) = \mathrm{E}[e^2] = \mathrm{E}[(X - \hat{X})^2] = \sigma_X^2 - \frac{\sigma_{XY}^2}{\sigma_Y^2} = \sigma_X^2(1 - \rho_{XY}^2)$$
- Perpendicularity property – an estimator is a BLE (Best Linear Estimator) iff the estimation error is orthogonal to all linear functions of the measurements:
  $$\forall a, b \qquad \mathrm{E}[e \cdot (aY + b)] = 0$$

Problem 1:

Let $X \sim N(0, \sigma^2)$, and denote $W = |X|$, $B = sign(X) = \begin{cases} 1 & X \geq 0 \\ -1 & X < 0 \end{cases}$.

1. Find the MMSE estimator of X given W (*Recall that W and B are independent*).

2. Calculate the mean squared error (MSE) of the estimator from section 2.

3. Calculate the MMSE estimator of X given B.

4. Calculate the MSE of the estimator from section 4.

5. Repeat sections 2-5 for the optimal linear estimator.

Solution:

1. Notice that $X = BW$. Since W and B are independent, we have:

$$\hat{X} = \mathrm{E}[X \mid W] = \mathrm{E}[BW \mid W] = W\,\mathrm{E}[B \mid W] = W\,\mathrm{E}[B] = 0$$

This result is intuitive since, given the absolute value of w, X can take either of the values $\pm w$ with equal probability, and the estimator gives the average.

2. Since the estimator is zero, we have $MSE = Var(X) = \sigma^2$.

3. Write

$$\hat{X} = \mathrm{E}[X \mid B] = \mathrm{E}[BW \mid B] = B\,\mathrm{E}[W \mid B] = B\,\mathrm{E}[W]$$

$$\mathrm{E}[W] = \mathrm{E}[|X|] = \int_{-\infty}^{\infty} |x|\, f_X(x)dx = 2\int_0^{\infty} x\, f_X(x)dx = \frac{2}{\sqrt{2\pi\sigma^2}}\int_0^{\infty} x e^{-\frac{x^2}{2\sigma^2}}dx = \frac{-2\sigma^2}{\sqrt{2\pi\sigma^2}}\, e^{-t}\Big|_0^{\infty} = \sqrt{\frac{2\sigma^2}{\pi}}$$

Substituting into the above, we have:

$$\hat{X} = B\,\mathrm{E}[W] = B\sqrt{\frac{2\sigma^2}{\pi}}$$

4. $MSE = Var(X) - Var(\hat{X}_{opt}) = Var(X) - \frac{2\sigma^2}{\pi}Var(B) = \sigma^2 - \frac{2\sigma^2}{\pi}\underbrace{Var(B)}_{1} = \sigma^2 - \frac{2\sigma^2}{\pi}$

5. We got in both sections that the optimal estimator is linear in the measurements hence it is also the optimal linear estimator.

Problem 2:

Given are random variables X_1, X_2, N, independent of one another, where $N \sim N(0,\sigma^2)$ and:

$$X_i = \begin{cases} 1 & w.p.\ 1/2 \\ -1 & w.p.\ 1/2 \end{cases} \qquad i = 1,2$$

Consider the measurements:

$$\begin{aligned} Y_1 &= X_1 + N \\ Y_2 &= X_2 + N \end{aligned}.$$

1. Find the MAP estimator of X_1 from Y_1.

A linear transformation is defined:

$$\begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} X_1 + X_2 + 2N \\ X_1 - X_2 \end{bmatrix}$$

2. The following estimator is suggested:

$$\hat{X}_1 = \begin{cases} sign(Z_1), & Z_2 = 0 \\ sign(Z_2), & Z_2 \neq 0 \end{cases}$$

Find the probability of error of the estimator, and compare it to the probability of error of the estimator from (1). Express your answer using the Q-function.

3. Find the estimator of X_2 from (Z_1, Z_2), which has probability of error equal to that of the estimator from section 2.

<u>Solution:</u>

1.

$$\hat{X}_1^{opt} = \hat{X}_1^{MAP} = \arg\max_{\alpha\in\{-1,1\}} P(X_1 = \alpha \mid Y_1 = y_1) = \arg\max_{\alpha\in\{-1,1\}} \frac{P(X_1 = \alpha) f_{Y_1|X_1}(y_1 \mid \alpha)}{f_{Y_1}(y_1)} =$$

$$= \arg\max_{\alpha\in\{-1,1\}} \frac{\frac{1}{2} f_{Y_1|X_1}(y_1 \mid \alpha)}{f_{Y_1}(y_1)} = \arg\max_{\alpha\in\{-1,1\}} f_{Y_1|X_1}(y_1 \mid \alpha) = \arg\max_{\alpha\in\{-1,1\}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y_1-\alpha)^2}{2\sigma^2}} =$$

$$= \arg\min_{\alpha\in\{-1,1\}} | y_1 - \alpha | = sign(y_1)$$

2.

$$\begin{aligned} P(X_1 \neq \hat{X}_1) &= P(Z_2 = 0)P(X_1 \neq \hat{X}_1 \mid Z_2 = 0) + P(Z_2 \neq 0)P(X_1 \neq \hat{X}_1 \mid Z_2 \neq 0) = \\ &= P(Z_2 = 0)P(X_1 \neq sign(Z_1) \mid Z_2 = 0) + P(Z_2 \neq 0)P(X_1 \neq sign(Z_2) \mid Z_2 \neq 0) = \\ &= P(X_1 = X_2)P(X_1 \neq sign(X_1 + X_2 + 2N) \mid X_1 = X_2) + \\ &\quad P(X_1 = -X_2)P(X_1 \neq sign(X_1 - X_2) \mid X_1 = -X_2) = \\ &= P(X_1 = X_2)P(X_1 \neq sign(2X_1 + 2N) \mid X_1 = X_2) + \\ &\quad P(X_1 = -X_2)\underbrace{P(X_1 \neq sign(2X_1) \mid X_1 = -X_2)}_{=0} = \\ &= P(X_1 = X_2)P(X_1 \neq sign(X_1 + N) \mid X_1 = X_2) \end{aligned}$$

Notice that, given $X_1 = 1$, the error event $\{X_1 \neq sign(X_1 + N) \mid X_1 = 1\}$ occurs iff $N < -1$, and, given $X_1 = -1$, the error event $\{X_1 \neq sign(X_1 + N) \mid X_1 = -1\}$ occurs iff $N > 1$. Furthermore, the random variable N is distributed symmetrically around zero and is independent of X_1, X_2. Therefore, we can get rid of the conditioning above:

$$P(X_1 \neq \hat{X}_1) = P(X_1 = X_2)P(X_1 \neq sign(X_1 + N)) = 0.5 \cdot P(X_1 \neq sign(Y_1))$$

We thus have:

$$P(X_1 \neq \hat{X}_1) = \frac{1}{2}P(X_1 \neq sign(Y_1)) = \frac{1}{2}P(X_1 = 1)P(Y_1 < 0 \mid X_1 = 1) + \frac{1}{2}P(X_1 = -1)P(Y_1 > 0 \mid X_1 = -1)$$

$$= \frac{1}{4}Q\left(\frac{1}{\sigma}\right) + \frac{1}{4}Q\left(\frac{1}{\sigma}\right) = \frac{1}{2}Q\left(\frac{1}{\sigma}\right)$$

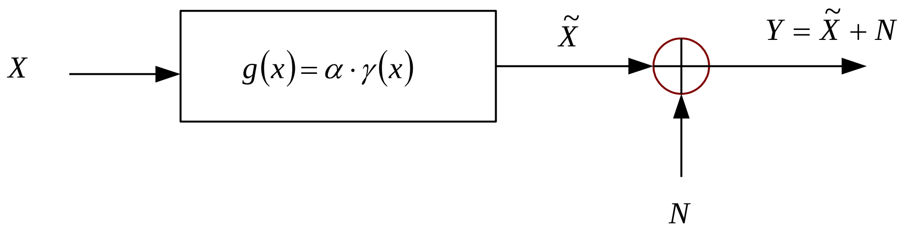
3. Based on symmetry, this is the estimator from section 2 with $-Z_2$ instead of Z_2:

$$\hat{X}_2 = \begin{cases} sign(Z_1), & Z_2 = 0 \\ sign(-Z_2), & Z_2 \neq 0 \end{cases}$$

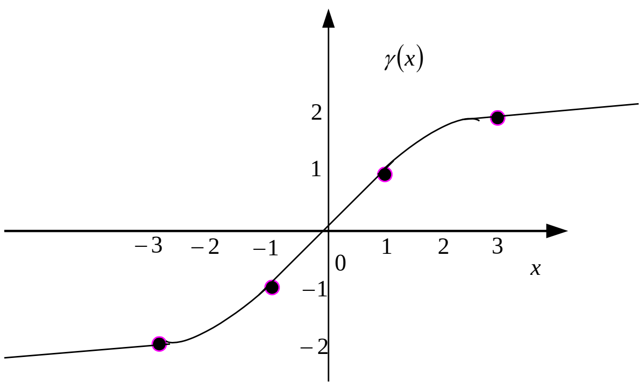
Extra Questions

Problem 3:

Consider a communication system, given by the following figure:



X represents the message at the entrance to the channel, and it is a discrete random variable that takes the values $\{-3,-1,1,3\}$ with equal probability. X passes through a non-linear amplifier with characteristics $g(x)=\alpha\cdot\gamma(x)$, where $\alpha>0$ is a constant and $\gamma(x)$ is the function shown in the following graph:



After passing through the amplifier, the random variable $\tilde{X}$ is obtained. Gaussian noise $N \sim N(0,\sigma^2)$, independent of X, is added to $\tilde{X}$, creating the random variable Y at the output of the system.

1. Calculate $f_Y(y)$.

Solution:

$$f_Y(y) = \sum_{x\in\{-3,-1,1,3\}} f_{Y|X}(y|x)\Pr\{X=x\} = \frac{1}{4}\sum_{x\in\{-3,-1,1,3\}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-g(x))^2}{2\sigma^2}} =$$

$$= \frac{1}{4}\frac{1}{\sqrt{2\pi\sigma^2}}\left(e^{-\frac{(y+2\alpha)^2}{2\sigma^2}} + e^{-\frac{(y+\alpha)^2}{2\sigma^2}} + e^{-\frac{(y-\alpha)^2}{2\sigma^2}} + e^{-\frac{(y-2\alpha)^2}{2\sigma^2}}\right)$$

2. Find the correlation coefficient between X and Y, $\rho_{XY} = \dfrac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$.

Solution:

$$\mathrm{E}[X] = \mathrm{E}[\tilde{X}] = \mathrm{E}[Y] = 0$$

$$Var(X) = \mathrm{E}[X^2] = \frac{1}{4}(2\cdot 1 + 2\cdot 9) = 5$$

$$Var(\tilde{X}) = \mathrm{E}[\tilde{X}^2] = \frac{1}{4}(2\cdot 1 + 2\cdot 4)\alpha^2 = 2.5\alpha^2$$

$$Var(Y) = Var(\tilde{X} + N) = Var(\tilde{X}) + Var(N) = 2.5\alpha^2 + \sigma^2$$

$$Cov(X,Y) = \mathrm{E}[X(\tilde{X}+N)] = \mathrm{E}[X\tilde{X}] + \mathrm{E}[XN] = \mathrm{E}[X\tilde{X}] + \underbrace{\mathrm{E}[X]\mathrm{E}[N]}_{=0} = \frac{1}{2}(1+6)\alpha = 3.5\alpha$$

$$\rho_{XY} = \frac{3.5\alpha}{\sqrt{5}\sqrt{2.5\alpha^2 + \sigma^2}} = \frac{7}{\sqrt{50 + 20\dfrac{\sigma^2}{\alpha^2}}}$$

3. Find the optimal linear estimator of X given Y, and find the estimation error.

Solution:

$$\hat{X}_{BLE} = \frac{Cov(X,Y)}{Var(Y)}\cdot Y = \frac{3.5\alpha}{2.5\alpha^2 + \sigma^2}Y = \frac{7\alpha}{5\alpha^2 + 2\sigma^2}Y$$

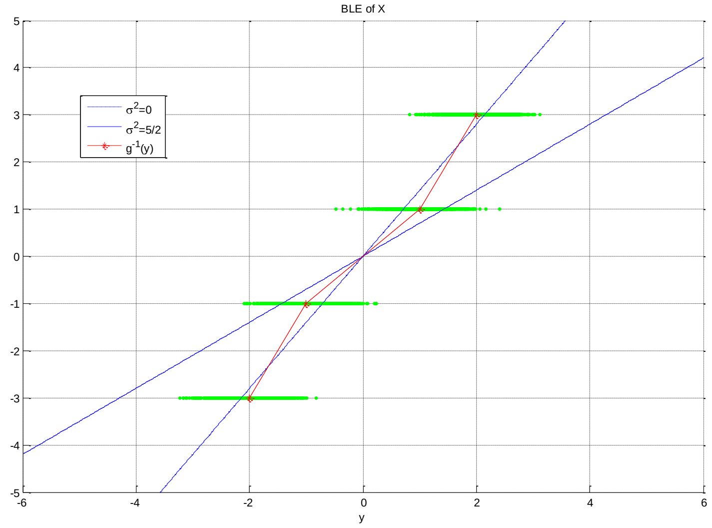
The estimation error:

$$MSE = Var(X)\left(1 - \rho_{XY}^2\right) = 5\left(1 - \frac{49}{50 + 20\dfrac{\sigma^2}{\alpha^2}}\right) = \frac{\alpha^2 + 20\sigma^2}{10\alpha^2 + 4\sigma^2}$$

From now on, assume $\alpha = 1$.

4. Draw the linear estimation function for $\sigma^2 = 0$ and for $\sigma^2 = \dfrac{5}{2}$.

Solution:



Notice that for measurements without noise ($\sigma^2 = 0$), the linear estimator is the line that best approximates the non-linear characteristics. The bigger the intensity of the noise, the less the estimator gives weight to the measurements and the more it approaches the expected value (zero).

5. Find the optimal estimator of X given Y, that minimizes the estimation error probability.

Solution:

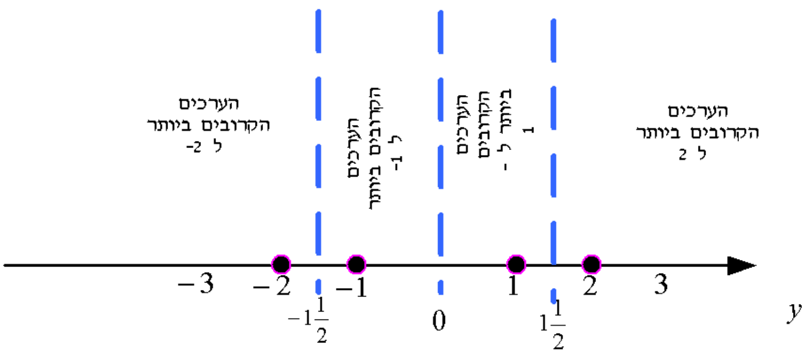
First, let us find the optimal estimator of $\tilde{X}$ given Y that brings the estimation error to a minimum. The estimation function:

$$h_{MAP,Y\to\tilde{X}}(y) = \arg\max_{\tilde{x}} \Pr\left\{\tilde{X} = \tilde{x} \mid Y = y\right\} = \arg\max_{\tilde{x}\in\{\pm1,\pm2\}} \frac{f_{Y|\tilde{X}}(y \mid \tilde{x})\Pr\left\{\tilde{X} = \tilde{x}\right\}}{f_Y(y)} =$$

$$= \arg\max_{\tilde{x}\in\{\pm1,\pm2\}} f_{Y|\tilde{X}}(y \mid \tilde{x})\Pr\left\{\tilde{X} = \tilde{x}\right\} = \arg\max_{\tilde{x}\in\{\pm1,\pm2\}} \frac{1}{4}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(y-\tilde{x})^2}{2\sigma^2}\right) =$$

$$= \arg\max_{\tilde{x}\in\{\pm1,\pm2\}} -(y-\tilde{x})^2 = \arg\min_{\tilde{x}\in\{\pm1,\pm2\}} (y-\tilde{x})^2 = \arg\min_{\tilde{x}\in\{\pm1,\pm2\}} |y-\tilde{x}|$$

Namely, for all $y \in \Re$, we are to find the numerical value $\tilde{x} \in \{\pm1,\pm2\}$ that is closest to it. It is easy to see that this requirement separates the number axis in the following way:



Therefore:

$$h_{MAP,Y\to\tilde{X}}(y) = \begin{cases} -2 & y < -\frac{3}{2} \\ -1 & \frac{-3}{2} \le y < 0 \\ 1 & 0 \le y < \frac{3}{2} \\ 2 & \frac{3}{2} \le y \end{cases} \qquad \hat{\tilde{X}}_{MAP} = h_{MAP,Y\to\tilde{X}}(Y)$$

Now, notice that since a one-to-one mapping connects X to $\tilde{X}$, in order to get the optimal estimator of X given Y that brings the estimation error to a minimum, it is enough to apply to the estimator $\hat{\tilde{X}}_{MAP}$ the inverse mapping from $\tilde{X}$ to X, which we will mark as $g^{-1}(\cdot)$. Let us demonstrate this:

$$\hat{X}_{MAP} = \arg\max_{x\in\{\pm1,\pm3\}} \Pr\left\{X = x \mid Y\right\} \underset{(1)}{=} \arg\max_{x\in\{\pm1,\pm3\}} \Pr\left\{\tilde{X} = g\left(x\right) \mid Y\right\} =$$

$$= g^{-1}\left(\arg\max_{g(x)\in\{\pm1,\pm2\}} \Pr\left\{\tilde{X} = g\left(x\right) \mid Y\right\}\right) = g^{-1}\left(\hat{\tilde{X}}_{MAP}\right)$$

(1) Since a one-to-one mapping exists from X to $\tilde{X}$ using the function $g(\cdot)$, it follows that the events $\{X = x\}$ and $\left\{\underbrace{\tilde{X}}_{g(X)} = g(x)\right\}$ are equivalent, so it may be substituted in the probability function. Specifically:

$$h_{MAP,Y\to X}\left(y\right) = g^{-1}\left(h_{MAP,Y\to\tilde{X}}\left(y\right)\right) = \begin{cases} -3 & y < -\frac{3}{2} \\ -1 & \frac{-3}{2} \le y < 0 \\ 1 & 0 \le y < \frac{3}{2} \\ 3 & \frac{3}{2} \le y \end{cases} \qquad \hat{X}_{MAP} = h_{MAP,Y\to X}\left(Y\right)$$