ECE541: Stochastic Signals and Systems	Fall 2018
Chapter 6 : Estimation	
Dr. Salim El Rouayheb	*Scribe: Serge Kas Hanna, Lu Liu*

1 Estimation Based On Single Observation

Suppose we wish to estimate the a values of a RV X by observing the values of another random variable Y. This estimate is represented by $\hat{X}$ and the estimation error is given by $\epsilon = X - \hat{X}$. A popular approach for determining $\hat{X}$ the estimate of X given Y, is by minimizing the conditional MSE (mean squared error):

$$MSE = E[\epsilon^2|Y] = E[(X - \hat{X})^2|Y].$$

Theorem 1. *The MMSE (minimum mean squared estimate) of X given Y is $\hat{X}_{MMSE} = E[X|Y]$.*

In other words, the theorem states that $\hat{X}_{MMSE} = E[X|Y]$ minimizes the conditional MSE $E[(X - \hat{X})^2|Y]$.

Proof.

$$E[(X - \hat{X})^2\big|Y] = \int_{-\infty}^{+\infty} (x - \hat{X})^2 f_{X|Y}(x|y)dx.$$

To minimize this conditional MSE, we determine $\hat{X}$ such that the derivative of the MSE is zero.

$$\frac{\partial}{\partial \hat{X}} E[(X - \hat{X})^2\big|Y] = -2\int_{-\infty}^{+\infty} (x - \hat{X}) f_{X|Y}(x|y)dx.$$

$$\frac{\partial}{\partial \hat{X}} E[(X - \hat{X})^2\big|Y] = 0 \Rightarrow -2\int_{-\infty}^{+\infty} (x - \hat{X}) f_{X|Y}(x|y)dx = 0.$$

Therefore,

$$\hat{X}\int_{-\infty}^{+\infty} f_{X|Y}(x|y)dx = \int_{-\infty}^{+\infty} x f_{X|Y}(x|y)dx \quad \Rightarrow \quad \hat{X}_{MMSE} = E[X|Y].$$

□

Corollary 1. *The conditional MSE corresponding to $\hat{X}_{MMSE}$ is $E[(X - \hat{X}_{MMSE})^2|Y] = \sigma^2_{X|Y}$.*

Proof. The proof directly follows from the definition of the variance.

$$E[(X - \hat{X}_{MMSE})^2|Y] = E[(X - E[X|Y])^2|Y] = Var[X|Y] = \sigma^2_{X|Y}.$$

□

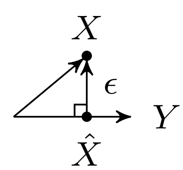
Sometimes $E[X|Y]$ is difficult to find and a linear MMSE (LMMSE) is used instead, i.e. $\hat{X}_{LMMSE} = \alpha Y + \beta$.

Theorem 2. *The LMMSE (linear minimum mean squared estimate) of X given Y that minimize the conditional MSE is given by*

$$\hat{X}_{LMMSE} = \frac{cov(X,Y)}{\sigma_Y^2}(Y - \mu_Y) + \mu_X.$$

Remark 1. *(Orthogonality Principle) The $\hat{X}$ that minimizes the MSE is given by $\hat{X} \perp \epsilon$, i.e., $\hat{X} \perp (X - \hat{X})$.*

Proof. (sketch) Assume WLOG that random variables X and Y are zero mean.



Denote by $\hat{X}$ the estimate of X given Y. In order to minimize $E[\epsilon^2] = ||\epsilon||^2 = E[||X - \hat{X}||^2]$, the error ϵ should be orthogonal to the observation Y as shown in the figure above. $\epsilon \perp Y$, therefore,

$$E[(X - \hat{X})Y] = 0.$$
$$E[(X - \alpha Y)Y] = 0,$$
$$E[XY] - \alpha E[Y^2] = 0.$$

Hence,

$$\alpha = \frac{E[XY]}{E[Y^2]} = \frac{cov(X,Y)}{\sigma_Y^2}.$$

Therefore,

$$\hat{X}_{LMMSE} = \frac{cov(X,Y)}{\sigma_Y^2}Y.$$

□

The result above is for any two zero mean random variables. The general result, i.e. when $\mu_X, \mu_Y \neq 0$, can be obtained by the same reasoning and is given by,

$$\hat{X}_{LMMSE} = \frac{cov(X,Y)}{\sigma_Y^2}(Y - \mu_Y) + \mu_X.$$

Example 1. *Suppose that in a room the temperature is given by a RV $X \sim N(\mu_X, \sigma_X^2)$. A sensor in this room observes $Y = X + W$, where W is the additive noise given by $N(0, \sigma_W^2)$. Assume X and W are independent.*

1. *Find the MMSE of X given Y.*

$$\hat{X}_{MMSE} = E[X|Y].$$

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{f_X(x) f_{Y|X}(y|x)}{f_Y(y)}.$$

Since Y is the sum of two independent gaussian RVs X and W, we know from homework 3 that,

$$Y \sim N(\mu_X, \sigma_X^2 + \sigma_W^2).$$

Furthermore,

$$\begin{aligned}
E[Y|X] &= E[X+W|X] = X + E[W] = X. \\
Var[Y|X] &= E[Y^2|X] - E[Y|X]^2 \\
&= E[X^2 + 2XW + W^2|X] - X^2 \\
&= E[X^2|X] + 2E[XW|X] + E[W^2|X] - X^2 \\
&= X^2 + 0 + \sigma_W^2 - X^2 \\
&= \sigma_W^2.
\end{aligned}$$

Therefore,

$$f_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi\sigma_W^2}} \exp\left[-\frac{(y-x)^2}{2\sigma_W^2}\right].$$

Therefore,

$$\begin{aligned}
f_{X|Y}(x|y) &= \frac{1}{\sqrt{2\pi \frac{\sigma_X^2 \sigma_W^2}{\sigma_X^2 + \sigma_W^2}}} \exp\left[-\frac{(x-\mu_X)^2}{2\sigma_X^2} - \frac{(y-x)^2}{2\sigma_W^2} + \frac{(y-\mu_X)^2}{2(\sigma_X^2 + \sigma_W^2)}\right] \\
&= \frac{1}{\sqrt{2\pi\sigma'^2}} \exp\left[-\frac{(x-\mu')^2}{2\sigma'^2}\right].
\end{aligned}$$

Where,

$$\sigma'^2 = \frac{\sigma_X^2 \sigma_W^2}{\sigma_X^2 + \sigma_W^2}.$$

We are interested in,

$$\hat{X}_{MMSE} = E[X|Y] = \mu'.$$

To determine μ' take $x = 0$:

$$\begin{aligned}
\frac{-\mu'^2}{2\frac{\sigma_X^2 \sigma_W^2}{\sigma_X^2 + \sigma_W^2}} &= \frac{-\mu_X^2}{2\sigma_X^2} - \frac{y^2}{2\sigma_W^2} + \frac{(y-\mu_X)^2}{2(\sigma_X^2 + \sigma_W^2)} \\
\mu'^2 &= \frac{\sigma_W^2 \mu_X^2}{\sigma_X^2 + \sigma_W^2} + \frac{\sigma_X^2 y^2}{\sigma_X^2 + \sigma_W^2} - \frac{\sigma_X^2 \sigma_W^2 (y-\mu_X)^2}{(\sigma_X^2 + \sigma_W^2)^2} \\
&= \frac{\sigma_X^4 y^2 + 2\mu_X \sigma_X^2 \sigma_W^2 y + \sigma_W^4 \mu_X^2}{(\sigma_X^2 + \sigma_W^2)^2} \\
&= \left(\frac{\sigma_X^2 y + \sigma_W^2 \mu_X}{\sigma_X^2 + \sigma_W^2}\right)^2.
\end{aligned}$$

Therefore,

$$\hat{X}_{MMSE} = E[X|Y] = \mu' = \frac{\sigma_X^2}{\sigma_X^2 + \sigma_W^2} Y + \frac{\sigma_W^2 \mu_X}{\sigma_X^2 + \sigma_W^2}.$$

2. *Find the linear MMSE of X given Y.*

$$\begin{aligned} cov(X,Y) &= E[XY] - E[X]E[Y] \\ &= E[X(X+W)] - E[X]E[X+W] \\ &= E[X^2] + E[XW] - \mu_X^2 \\ &= \sigma_X^2 + \mu_X^2 + 0 - \mu_X^2 \\ &= \sigma_X^2. \end{aligned}$$

Applying the general formula of LMMSE,

$$\begin{aligned} \hat{X}_{LMMSE} &= \frac{cov(X,Y)}{\sigma_Y^2}(Y - \mu_Y) + \mu_X \\ &= \frac{\sigma_X^2}{\sigma_X^2 + \sigma_W^2}(Y - \mu_X) + \mu_X \\ &= \frac{\sigma_X^2}{\sigma_X^2 + \sigma_W^2} Y + \frac{\sigma_W^2 \mu_X}{\sigma_X^2 + \sigma_W^2}. \end{aligned}$$

Remark 2. *Notice that $\hat{X}_{LMMSE} = \hat{X}_{MMSE}$, in fact this is always the case if the random variable to estimate X, and the observation Y, are jointly gaussian.*

3. *Find the MSE.*

Method 1: (Orthogonality principle)

$$\begin{aligned} E[\epsilon^2] &= E[(\hat{X} - X)^2] \\ &= |E[\epsilon(\hat{X} - X)]| \\ &= |\cancelto{0}{E[\epsilon\hat{X}]} - E[\epsilon X]| \\ &= |E[\epsilon X]| \\ &= \left| E[\hat{X}X] - E[X^2] \right| \\ &= \left| E\left[\frac{\sigma_X^2}{\sigma_X^2 + \sigma_W^2} XY + \frac{\sigma_W^2 \mu_X}{\sigma_X^2 + \sigma_W^2} X \right] - \sigma_X^2 - \mu_X^2 \right| \\ &= \left| \frac{\sigma_X^2(\sigma_X^2 + \mu_X^2)}{\sigma_X^2 + \sigma_W^2} + \frac{\sigma_W^2 \mu_X^2}{\sigma_X^2 + \sigma_W^2} - \sigma_X^2 - \mu_X^2 \right| \\ &= \frac{\sigma_X^2 \sigma_W^2}{\sigma_X^2 + \sigma_W^2}. \end{aligned}$$

Method 2: (Towering property + Corollary 1)

$$
\begin{aligned}
E[\epsilon^2] &= E[E[\epsilon^2|Y]] \\
&= E[E[(X - \hat{X}_{MMSE})^2|Y]] \\
&= E[\sigma^2_{X|Y}] \quad \textit{(corollary 1)} \\
&= E[\sigma'^2] \quad \textit{(part 2)} \\
&= E\left[\frac{\sigma_X^2 \sigma_W^2}{\sigma_X^2 + \sigma_W^2}\right] \\
&= \frac{\sigma_X^2 \sigma_W^2}{\sigma_X^2 + \sigma_W^2}.
\end{aligned}
$$

2 MMSE Based on Vector Observation

Theorem 3. *The Linear Minimum Mean-Square Estimate LMMSE $\hat{X}_{LMMSE}$ of X given an observed random vector $\underline{Y} = (Y_1, \ldots, Y_n)^T$ is given by*

$$\hat{X}_{LMMSE} = K_{XY}^T K_{YY}^{-1}(\underline{Y} - \underline{\mu}_Y) + \mu_X,$$

where,

$$
\begin{aligned}
\mu_X &= E[X], \\
\underline{\mu}_Y &= (E[Y_1], E[Y_2], \ldots, E[Y_n]), \\
K_{YY} &= E[\underline{Y}\,\underline{Y}^T] - \mu_Y \mu_Y^T, \\
\textit{and } K_{XY} &= (Cov[XY_1], Cov[XY_2], \ldots, Cov[XY_n])^T,
\end{aligned}
$$

where K_{YY} is the covariance matrix of Y.

Proof. First, let us assume that $\mu_X = 0$ and $\underline{\mu}_Y = \underline{0}$. Then, we can write

$$
\begin{aligned}
\hat{X}_{LMMSE} &= a_1 Y_1 + a_2 Y_2 + \cdots + a_n Y_n \\
&= \underline{a}^t \underline{Y}.
\end{aligned}
$$

By the orthogonality principle: $(X - \hat{X}_{LMMSE}) \perp Y_i \quad i = 1, 2, \ldots, n$,

$$E[\underline{a}^t \underline{Y} \cdot Y_i] = E[XY_i] \quad i = 1, 2, \ldots, n,$$

$$E[(a_1 Y_1 + a_2 Y_2 + \cdots + a_n Y_n) Y_i] = E[XY_i] \quad i = 1, 2, \ldots, n.$$

So, we get the following $n \times n$ linear system with n unknowns, $a_1, \ldots, a_n$:

$$
\begin{aligned}
a_1 E[Y_1^2] + a_2 E[Y_1 Y_2] + \cdots + a_n E[Y_1 Y_n] &= E[XY_1], \\
a_1 E[Y_2 Y_1] + a_2 E[Y_2^2] + \cdots + a_n E[Y_2 Y_n] &= E[XY_2], \\
&\vdots \\
a_1 E[Y_n Y_1] + a_2 E[Y_n Y_2] + \cdots + a_n E[Y_n^2] &= E[XY_n].
\end{aligned}
$$

In matrix form, this can be written as

$$\underline{a}^t K_{YY} = K_{XY}^t,$$

$$\underline{a}^t = K_{XY}^t K_{YY}^{-1}.$$

Where,

$$K_{YY} = \begin{bmatrix} E[Y_1^2] & E[Y_1Y_2] & \dots & E[Y_1Y_n] \\ E[Y_2Y_1] & E[Y_2^2] & \dots & E[Y_2Y_n] \\ \vdots & \vdots & & \vdots \\ E[Y_nY_1] & E[Y_nY_2] & \dots & E[Y_n^2] \end{bmatrix},$$

and,

$$K_{XY} \stackrel{\text{def}}{=} \begin{bmatrix} Cov[XY_1] \\ Cov[XY_2] \\ \vdots \\ Cov[XY_n] \end{bmatrix} = \begin{bmatrix} E[XY_1] \\ E[XY_2] \\ \vdots \\ E[XY_n] \end{bmatrix}.$$

So,

$$\hat{X}_{LMMSE} = K_{XY}^T K_{YY}^{-1} \underline{Y}.$$

In general, if $\mu_X \neq 0$ and $\underline{\mu}_Y \neq \underline{0}$,
Apply the same method above to $X' = X - \mu_X$ and $\underline{Y}' = \underline{Y} - \underline{\mu}_Y$, then we get

$$\hat{X}_{LMMSE} = K_{XY}^T K_{YY}^{-1} (\underline{Y} - \underline{\mu}_Y) + \mu_X.$$

□

Example 2. *Multiple Antenna Receiver*

Assume 2 antennas receive signals independently. $Y_1 = X + N_1, \quad Y_2 = X + N_2,$
$X \sim N(0,2), \qquad N_1, N_2 \sim N(0,1)$. *Assume they are all independent.*

1. *Find the LMMSE of X given Y_1.*

$$\hat{X}_{LMMSE} = \frac{Cov(XY_1)}{V(Y_1)} Y_1.$$

$$\begin{aligned} Cov(XY_1) &= E[XY_1] - E[X]E[Y_1] \qquad \textit{Note that } E[X]E[Y_1] = 0 \\ &= E[X(X+N_2)] \\ &= E[X^2] + E[XN_1] = 2 + 0 = 2. \end{aligned}$$

$$V(Y_1) = V(X) + V(N_1) = 2 + 1 = 3.$$

Therefore, $\hat{X}_{LMMSE} = \frac{2}{3} Y_1$

2. *Find the LMMSE of X given Y_1 and Y_2.*

Usually, we want to find that $\hat{X} = a_1 Y_1 + a_2 Y_2 + C$.
In this case, $C = 0$.
While $X - \hat{X} \perp Y_1$, *and* $X - \hat{X} \perp Y_2$,
we can obtain,

$$\begin{aligned} E[(X - aY_1 - a_2 Y_2)Y_1] &= 0. \\ E[(X - aY_1 - a_2 Y_2)Y_2] &= 0. \\ a_1 E[Y_1^2] + a_2 E[Y_1 Y_2] &= E[XY_1]. \\ a_1 E[Y_1 Y_2] + a_2 E[Y_2^2] &= E[XY_2]. \end{aligned}$$

$$K_{Y_1 Y_2} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = K_{XY}.$$

Therefore,

$$\begin{aligned} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = K_{Y_1 Y_2}^{-1} K_{XY} &= \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}^{-1} \begin{bmatrix} 2 \\ 2 \end{bmatrix} \\ &= \frac{1}{5} \begin{bmatrix} 3 & -2 \\ -2 & 3 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \end{bmatrix} \end{aligned}$$

$\Longrightarrow \quad \hat{X}_{LMMSE} = \frac{2}{5}(Y_1 + Y_2).$

3. *Compare the MSE of part 1 and part 2*
From part 1 we have $\hat{X}_1 = \frac{2}{3} Y_1$, *then*

$$\begin{aligned} MSE_1 &= E[(X - \hat{X})^2] \\ &= E[X^2] - 2E[X\hat{X}] + E[\hat{X}^2] \\ &= 2 - (2)\frac{2}{3} E[XY_1] + \frac{4}{9} E[Y_1^2] \\ &= 2 - \frac{4}{3}(2) + \frac{4}{9}(3) \\ &= \frac{2}{3} \\ &= 0.66. \end{aligned}$$

From part 2 we have $\hat{X}_2 = \frac{2}{5}(Y_1 + Y_2)$, *then*

$$\begin{aligned} MSE_2 &= E[X^2] - 2E[X\hat{X}] + E[\hat{X}^2] \\ &= 2 - (2)\frac{2}{5}(E[XY_1] + E[XY_2]) + \frac{4}{25}\left(E[Y_1^2] + 2E[Y_1 Y_2] + E[Y_2^2]\right) \\ &= 2 - \frac{4}{5}(2 + 2) + \frac{4}{25}(3 + 2(2) + 3) \\ &= \frac{2}{5} = 0.4 \end{aligned}$$

Therefore, $MSE_2 < MSE_1$, which is intuitive since in part 2 we can benefit from the additional observation Y_2 to improve our estimation.

3 Finding The MMSE Using The Orthogonality Principle

Theorem 4 (The Orthogonality Principle)**.** *The MMSE of $\hat{X}$ of X given Y, where $\hat{X} = g(Y)$, where $g(*) \in \Gamma$ and (Γ^* is all functions, linear functions, constants), is found when $\hat{X} = \min E[(X - g(Y))^2]$ where the minimization is over $g(*) \in \Gamma$. The MMSE $= E[X^2] - E[\hat{X}^2]$. In this case, $\hat{X}$ is unique and the error is orthogonal to the observation (($X - \hat{X}) \perp Y$). The * indicates there are some technical conditions on gamma but they are not discussed here.*

Proof. Proof is omitted. □

Example 3. *$X = (X_1, X_2, X_3)$ are jointly Gaussian and, $\underline{\mu}_X = (0, 0, 0)$,*

$$K_{XX} = R_{XX} = \begin{bmatrix} 1 & 0.2 & 0.1 \\ 0.2 & 2 & 0.3 \\ 0.1 & 0.3 & 4 \end{bmatrix}.$$

1. *Find the LMMSE of X_3 Given X_1 and X_2.*

   Usually, we write $\hat{X} = a_1Y_1 + a_2Y_2 + c$. When $\mu_x = 0$, we have $c = 0$. Therefore, we write $\hat{X}_3 = a_1X_1 + a_2X_2$. By the Orthogonality Principle, the error $\perp$ observation space, that is

$$\begin{aligned}
(X_3 - \hat{X}_3) \perp X_1 &\implies E[(X_3 - \hat{X}_3)X_1] = 0 \\
&\implies a_1E[X_1^2] + a_2E[X_1X_2] = E[X_1X_3] \\
(X_3 - \hat{X}_3) \perp X_2 &\implies E[(X_3 - \hat{X}_3)X_2] = 0 \\
&\implies a_1E[X_1X_2] + a_2E[X_2^2] = E[X_2X_3]
\end{aligned}$$

   Denote $Y = (X_1, X_2)^T$. In matrix form,

$$\begin{bmatrix} E[X_1^2] & E[X_1X_2] \\ E[X_1X_2] & E[X_2^2] \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} E[X_1X_3] \\ E[X_2X_3] \end{bmatrix} \implies \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = K_{YY}^{-1}K_{X_3Y}$$

$$K_{YY} = \begin{bmatrix} 1 & 0.2 \\ 0.2 & 2 \end{bmatrix} \Rightarrow K_{YY}^{-1} = \begin{bmatrix} 1.0204 & -0.102 \\ -0.102 & 0.5102 \end{bmatrix}.$$

$$K_{X_3Y}^T = [Cov(X_3X_1) \quad Cov(X_3X_2)] = [0.1 \quad 0.3].$$

   Therefore,

$$\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 0.0714 \\ 0.1429 \end{bmatrix}$$

Hence,

$$\hat{X}_3 = 0.0714X_1 + 0.1429X_2.$$

2. *Find the MSE corresponding to* $\hat{X}_3$.

$$\begin{aligned} MSE &= E[(X_3 - \hat{X}_3)^2] = E[X_3^2] - E[\hat{X}_3^2] \\ &= 4 - E[(a_1X_1 + a_2X_2)^2] \\ &= 4 - a_1^2E[X_1^2] - a_2^2E[X_2^2] - 2a_1a_2E[X_1X_2] \\ &= 3.95. \end{aligned}$$