1 Review

\[ h(y) \triangleq \hat{X}_{\text{mmse}} = E[X|Y] \]

Estimating a scalar from a vector; given the joint distribution function \( p(X,Y) \)

\[ X \in \mathbb{R}(\text{scalar}) \]
\[ Y \in \mathbb{R}^n(\text{vector}) \]

\[ h : \mathbb{R}^n \mapsto \mathbb{R} \]

\[ e = x - h(y) \]

\( h(y) \) minimizes \( E[e^2] \). This is considered as the optimal solution. We notice that by mapping the \( y \) into \( x \) we get,

\[ h(y) = E[X|Y] = \sum_i x_i P(x_i|y) \]

What if \( X \in \mathbb{R}^m \). How would you estimate \( X \) given \( Y \), given the joint distribution \( P(X,Y) \)

\[ h(y) : \mathbb{R}^n \mapsto \mathbb{R}^m \]
\[ h(Y) = \begin{bmatrix} E[X_1|Y] \\ E[X_2|Y] \\ \vdots \\ E[X_m|Y] \end{bmatrix} \triangleq \begin{bmatrix} h_1(y) \\ h_2(y) \\ \vdots \\ h_m(y) \end{bmatrix} \]

We get a different \( h \) every time. The process is simple: First, Assume a scalar. Then derive the equation. Next show the optimal estimator. Finally there will be a different \( h \) for every different \( Y \).

**Example 1**

\[
f_{XY}(x, y) = \begin{cases} kxy & 0 < x < y < 1 \\ 0 & \text{Otherwise} \end{cases}
\]

The Fig. 1 shows \( f_{XY}(x, y) \). To find the best estimator. Find \( \hat{X}_{\text{mmse}} \triangleq h_{\text{mmse}}(y) = ? \)

\[
E[X|Y] = \hat{X}_{\text{mmse}}
\]

![Figure 1: Plot of \( f_{XY}(x, y) \)](image)

\[
f_{X|Y}(x|y) = \frac{f_{XY}(x, y)}{f_Y(y)}
\]

\[
f_Y(y) = \int_0^y kxy \, dx
\]

\[
= \frac{kx^2y}{2} \bigg|_0^y = \frac{ky^3}{2}
\]

\[
f_{X|Y}(X|Y) = \frac{kxy}{k\frac{y^3}{2}}
\]
\[
\begin{aligned}
&= \begin{cases} 
\frac{2x}{y^2} & 0 < x < y < 1 \\
x & \text{Otherwise}
\end{cases} \\
E[X|Y = y] = \int_0^y \frac{2x^2}{y^2} dx = \frac{2x^3}{3y^2} = \frac{2y}{3}
\end{aligned}
\]

if \(y_1 = \frac{1}{2}\) then \(x_1 = \frac{1}{3}\) and \(y_2 = \frac{1}{3}\) then \(x_2 = \frac{2}{9}\)

The \(E[e^2]\) will be minimized

Example 2

\(Y = X^3\) Estimate \(Y\) given \(X\) a random variable

\[\hat{Y} = E[Y|X] = E[X^3|X] = X^3\]

Example 3

\(X \sim N(0, K_{xx})\)

\[
X = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}, \quad K_{xx} = \begin{bmatrix} 3 & -1 & 1 \\ -1 & 3 & 1 \\ 1 & 1 & 3 \end{bmatrix}, \quad \mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}
\]

Determine the MMSE estimator of \(X_3\) given \(X_1\) and \(X_2\)

\[
Y = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}
\]

\[
\hat{X}_3 = h_{\text{mmse}}(Y) = E[X_3|Y]
\]

Recall that for a jointly Gaussian Random Vector we have

\[f_{X_3|Y} \sim N(\tilde{\mu}, \tilde{K})\]

\[\tilde{\mu} = \mu_2 + K_{21}K_{11}^{-1}(x_1) - \mu_1\]

\[K_{xx} = \begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix} = \begin{bmatrix} 3 & -1 & 1 \\ -1 & 3 & 1 \\ 1 & 1 & 3 \end{bmatrix}\]

\[X = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}, \quad X_1 = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad X_2 = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad K_{11} = \begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix}, \quad K_{21} = \begin{bmatrix} 1 & 1 \end{bmatrix}\]
Theorem 1.1 Orthogonality Principle
The orthogonality principle says that the error vector of the optimal estimator (mean square error) is orthogonal to any possible estimator which can be seen in Fig. 2 for one-dimensional case.

\[
\begin{align*}
X_1 &= Y, & \mu_1 &= \begin{bmatrix} 0 \\ 0 \end{bmatrix}, & \mu_2 &= \begin{bmatrix} 0 \end{bmatrix}, \\
X_3 &= \begin{bmatrix} 1 & 1 \\ -1 & 3 \end{bmatrix}^{-1} Y \\
    &= \begin{bmatrix} 1 & 1 \\ -1 & 3 \end{bmatrix}^{-1} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}
\end{align*}
\]

\[e \triangleq x - h_{\text{mmse}}(Y)\]
\[E[e h(Y)] = 0\]

\[x = \hat{X} - X\]

\[W\]

Figure 2: Orthogonality Principle

\[
\begin{align*}
e &\triangleq x - h_{\text{mmse}}(Y) \\
E[e h(Y)] &\triangleq 0
\end{align*}
\]

\textbf{Proof} : The orthogonality principle can be verified, as follows

\[
E[(X - E[X|Y])h(Y)]
\]
\[
= E[(Xh(Y)) - E[X|Y]h(Y)]
\]
\[
= E[Xh(Y)] - E[Xh(Y)]
\]
\[
= 0
\]
Note:

\[ E[E[X|Y]] = E[X] \]

because,

\[
E[E[X|Y]] = \sum_{j=-\infty}^{\infty} \sum_{i=-\infty}^{\infty} x_i p(x_i|y_j) p(y_j) \\
= \sum_{j=-\infty}^{\infty} \sum_{i=-\infty}^{\infty} x_i p(x_i, y_j) = E[X]
\]

Often times to compute \( E[X|Y] \) is hard and time consuming, because if \( Y \) is a multi-dimensional vector then we have to compute many integrals. So, instead we can try our best to compute a linear estimator LMMSE. Specifically,

\[
\hat{Y} = h(X) = WX + b
\]

The goal is to find \( W \) and \( b \) such that the error is minimized

\[
E[||Y - \hat{Y}||^2] = E[||Y - (WX + b)||^2]
\]

To find \( W \) and \( b \), we will take derivative \( E[e^T e] \) and set to zero.

\[
E[((Y - (WX) + b))^T (Y - (WX) + b)] \\
= E[Y^T Y - Y^T WX - Y^T b - X^T W^T Y + X^T W^T WX + X^T W^T b - b^T Y + b^T W X + b^T b] = A
\]

\[
\frac{dA}{db} = E[-Y + WX - Y + WX + 2b] = 0
\]

\[
b^* = E[Y] - WE[X]
\]

Note: Derivative of a scalar w.r.t vector

\[
\frac{df(x)}{dx} = \begin{bmatrix}
\frac{df(x)}{dx_1} \\
\frac{df(x)}{dx_2} \\
\vdots \\
\frac{df(x)}{dx_n}
\end{bmatrix}
\]

\[
\frac{dA}{dW} = E[-YX^T - YX^T + 2WX^T X + 2bX^T X] = 0
\]

\[
\frac{dA}{dW} = -RYX + WRRX + bE[X^T X] = 0
\]
\[
\frac{dA}{dW} = -R_{XX} + WR_{XX} + (E[Y] - WE[X])E[X^T] = 0
\]

Note: Derivative of a scalar w.r.t matrix

\[
\frac{d\mathbf{f}(W)}{d\mathbf{W}} = \begin{bmatrix}
\frac{d\mathbf{f}(W)}{dW_1} & \cdots & \frac{d\mathbf{f}(W)}{dW_n}
\frac{d\mathbf{f}(W)}{dW_2} & \cdots & \frac{d\mathbf{f}(W)}{dW_n}
\vdots & \ddots & \vdots
\frac{d\mathbf{f}(W)}{dW_n} & \cdots & \frac{d\mathbf{f}(W)}{dW_{nn}}
\end{bmatrix}
\]

Let

\[
\mu_X = E[X]
\]
\[
\mu_Y = E[Y]
\]
\[
W(R_{XX} - \mu_X\mu_Y^T) = R_{YY} - \mu_Y^T\mu_X
\]
\[
W^* = K_{YY}(K_{XX})^{-1}
\]
\[
h_{\text{lmmse}}(X) = W\mathbf{x} + b = k_{YY}k_{XX}^{-1}X + \mu_Y - k_{YY}k_{XX}^{-1}\mu_X
\]

**Example 4**

\(X, Y\) are jointly Gaussian, \(X \in \mathbb{R}, Y \in \mathbb{R}\), with:

\[
K = \begin{bmatrix}
1 & -1 \\
-1 & 2
\end{bmatrix}, \quad \text{mean} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.
\]

Find \(\hat{x}_2 = h_{\text{lmmse}}(x_1)\)

From the conditions, we get \(K_{XY} = -1, K_{XX} = 1, \mu_X = 0, \mu_Y = 0\).

Therefore, \(h_{\text{lmmse}}(x_1) = (-1)(1)x_1 + (-1)(1)0 = -x_1 = y\)

**Theorem 1.2** If \(X\) and \(Y\) are jointly Gaussian, then the LMMSE estimator \(\hat{X} = h_{\text{lmmse}}(Y)\) of is also the MMSE estimator.

**Proof** (Sketch:) From the previous lecture that for a jointly Gaussian vectors \(X\) and \(Y\), the conditional distribution of \(X\) given \(Y\) is also a Gaussian. Furthermore, \(E[X|Y]\) which is the MMSE estimator of \(X\), was shown to be a linear function of \(Y\). Therefore, the MMSE estimator is also the LMMSE estimator.
2 Random Sequence

Definition 2.1 Given a sample space $\Omega$ with event and probability $p$ defined for each event, a random sequence $X[n, \omega]$ is a function that is assigned to each outcome of the sample space. Let us now observe how to map a Random Variable Event Fig. 3

In Fig. 4 we see the Functions of a Random Sequence

Example 1
$X(n, \omega) \triangleq A(\omega) f[n]$ where $A(\omega)$ is a random variable $A(\omega) \sim Bern(p)$ and $f[n] = \frac{1}{n}$.

$A(\omega) = \begin{cases} 1 \\ 0 \end{cases}$

$X(n, 0) = 0 \frac{1}{n} = 0$

$X(n, 1) = 1 \frac{1}{n} = \frac{1}{n}$

Considering the Fig. 5 below
Example 2
\(X(n, \omega) = A(\omega)\sin(\frac{n\pi}{10} + \theta(\omega))\) where \(A(\omega)\) and \(\theta(\omega)\) are r.v.s.

The Fig. 6 below illustrates \(X(n, \omega)\)

2.1 Mean Function

2.1.1 Continuous Random Variable (C.R.V)

\[
\mu_X[n] \triangleq E[X[n]] = \int_{-\infty}^{+\infty} x f_X(x; n) \, dx_n = \int_{-\infty}^{+\infty} x f_{X^n}(x_n) \, dx_n
\]
We can look at the mean function in the Fig. 7 below

![Mean Function](image)

**Figure 7: Mean Function**

2.1.2 Discrete Random Variable (D.R.V)

\[
\sum_{i=-\infty}^{\infty} x_i P(X[n] = x_i) = \sum_{i=-\infty}^{\infty} x_i P_n(x_n)
\]