LINEAR MMSE ESTIMATION

TERM PAPER FOR EE 602 – STATISTICAL SIGNAL PROCESSING

By,

DHEERAJ KUMAR

VARUN KHAITAN

Introduction

Linear MMSE estimators are chosen in practice because they are simpler than the optimal Bayesian estimators and retain the MMSE criterion. Here the constraint on the estimator is assumed to be linear. We will also show how this estimator depends just on the correlation between the variable to be estimated and the observations. This is a very important advantage as it does not require us to know the joint probability function of the observation and the estimate.

Linear MMSE Estimation

Let us assume that we have to estimate a scalar parameter θ based on the data set { x[0],x[l], . . . , x[N – 1] } or in vector form X = [x[0] x[l] . . . x[N - l]] ^{T}. We do not assume any specific knowledge of the joint PDF p(x,θ), because as we shall see, a knowledge of the first two moments suffices. We show that the correlation between θ and X is sufficient to estimate θ .
We now consider the class of all linear estimators of the form

$$\hat{\theta} = \sum_{n=0}^{N-1} a_n x[n] + a_N$$

The task is to choose the weighting coefficients a_N to minimize the Bayesian MSE (Mean Square Error) ;

$$Bmse(\hat{\theta}) = E\left[\left(\theta - \hat{\theta}\right)^2\right]$$

Here the expectation is with respect to the PDF p(x, θ). The resultant estimator is termed the linear minimum mean square error (LMMSE) estimator. Here the term a_N has been incorporated to allow for a mismatch between the means of θ and x. If the means are equal then this coefficient is zero and may be omitted.
A LMMSE will be suboptimal unless the MMSE estimator happens to be linear. Otherwise, better estimators will exist which will of course be nonlinear. It is important to note that since the LMMSE estimator relies on the correlation between random variables, a parameter uncorrelated with the data cannot be linearly estimated.

We now try to find the optimal values of the coefficients by minimizing the BMSE error. We differentiate the error with respect to a_N and set it to zero.

$$\frac{d}{d_{a_N}} E\left[\left(\theta - \sum_{n=0}^{N-1} a_n x[n] - a_N\right)^2\right] = -2E\left[\theta - \sum_{n=0}^{N-1} a_n x[n] - a_N\right] = 0$$

$$a_N = E(\theta) - \sum_{n=0}^{N-1} a_n x[n]$$

Substituting this into the BMSE expression we now need to minimize

$$Bmse(\hat{\theta}) = E\{[\sum_{n=0}^{N-1} a_n\left(x[n] - E(x[n])\right) - (\theta - E(\theta))]^2\}$$

over the other N-1 coefficients. Letting **a** = $[\boldsymbol{a_0}\ \boldsymbol{a_1} \dots .\ \boldsymbol{a_{N-1}}]^T$**,** we have

$$\begin{aligned}
\mathrm{Bmse}(\hat{\theta}) &= E\left\{\left[\mathbf{a}^T(\mathbf{x} - E(\mathbf{x})) - (\theta - E(\theta))\right]^2\right\} \\
&= E\left[\mathbf{a}^T(\mathbf{x} - E(\mathbf{x}))(\mathbf{x} - E(\mathbf{x}))^T\mathbf{a}\right] - E\left[\mathbf{a}^T(\mathbf{x} - E(\mathbf{x}))(\theta - E(\theta))\right] \\
&\quad - E\left[(\theta - E(\theta))(\mathbf{x} - E(\mathbf{x}))^T\mathbf{a}\right] + E\left[(\theta - E(\theta))^2\right] \\
&= \mathbf{a}^T\mathbf{C}_{xx}\mathbf{a} - \mathbf{a}^T\mathbf{C}_{x\theta} - \mathbf{C}_{\theta x}\mathbf{a} + C_{\theta\theta}
\end{aligned}$$

Where C_{xx} is the NxN covariance matrix of X and $C_{x\theta}$ is the Nx1 cross covariance vector. Taking the derivative and equating to zero we get

$$\frac{\partial \mathrm{Bmse}(\hat{\theta})}{\partial \mathbf{a}} = 2\mathbf{C}_{xx}\mathbf{a} - 2\mathbf{C}_{x\theta}$$

$$\mathbf{a} = \mathbf{C}_{xx}^{-1}\mathbf{C}_{x\theta}.$$

$$\begin{aligned}
\hat{\theta} &= \mathbf{a}^T\mathbf{x} + a_N \\
&= \mathbf{C}_{x\theta}^T\mathbf{C}_{xx}^{-1}\mathbf{x} + E(\theta) - \mathbf{C}_{x\theta}^T\mathbf{C}_{xx}^{-1}E(\mathbf{x})
\end{aligned}$$

Finally the value of θ is $\hat{\theta} = \mathbf{C}_{\theta x}\mathbf{C}_{xx}^{-1}\mathbf{x}.$ and for this the value of the BMSE error is

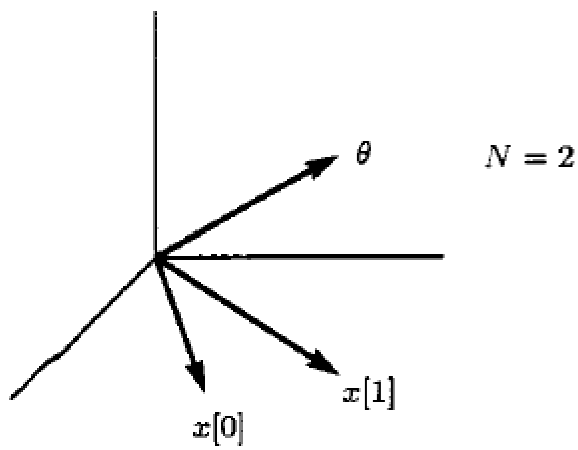
$$\mathrm{Bmse}(\hat{\theta}) = C_{\theta\theta} - \mathbf{C}_{\theta x}\mathbf{C}_{xx}^{-1}\mathbf{C}_{x\theta}.$$

GEOMETRICAL INTERPRETATION

For this interpretation we first assume that θ and x are zero mean. If that is not the case we subtract their means from the variables and thus make them zero mean random variables.

Let $\hat{\theta} = \sum_{n=0}^{N-1} a_n x[n]$ and we want to minimize the error $Bmse(\hat{\theta}) = E\left[(\theta - \hat{\theta})^2\right]$

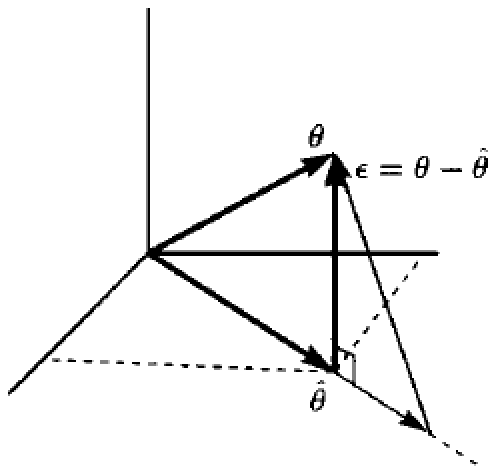
Let us now think of the random variables θ, x[0], x[1], . . . , x[N – 1] as elements in a vector space as shown symbolically.



If θ can be expressed as a linear combination of the observations then it will lie in the plane spanned by them. If it cannot, as is usually the case, θ we picture θ as only partially lying in the subspace spanned by the x[n]'s. To complete the description of our vector space we require the notion of an inner product between two vectors. We define the inner product as <x,y> = E[xy].

$$\begin{aligned} E[(\theta - \hat{\theta})^2] &= E\left[\left(\theta - \sum_{n=0}^{N-1} a_n x[n]\right)^2\right] \\ &= \left\| \theta - \sum_{n=0}^{N-1} a_n x[n] \right\|^2 . \end{aligned}$$

This means that minimization of the MSE is equivalent to a minimization of the squared length of the error vector E = $(\theta - \hat{\theta})$. As can be seen in the figure below the length of the is minimized when E is orthogonal to the subspace spanned by the $x_n s$.



If it had a component in the plane then that could always be taken care of by changing the values of the coefficients. Hence, we require

$$E\left[(\theta - \hat{\theta})x[n]\right] = 0 \qquad n = 0, 1, \ldots, N-1$$

This is the important orthogonality principle or projection theorem. It says that in estimating the realization of a random variable by a linear combination of data samples, the optimal estimator is obtained when the error is orthogonal to each data sample.
Using the orthogonality principle, the weighting coefficients are easily found as

$$E\left[\left(\theta - \sum_{n=0}^{N-1} a_m x[m]\right) x[n]\right] = 0 \qquad n = 0, 1, \ldots, N-1$$

$$\sum_{n=0}^{N-1} a_m E(x[m]x[n]) = E(\theta x[n]) \qquad n = 0, 1, \ldots, N-1$$

In the matrix form these N equations can be written as

$$\begin{bmatrix} E(x^2[0]) & E(x[0]x[1]) & \ldots & E(x[0]x[N-1]) \\ E(x[1]x[0]) & E(x^2[1]) & \ldots & E(x[1]x[N-1]) \\ \vdots & \vdots & \ddots & \vdots \\ E(x[N-1]x[0]) & E(x[N-1]x[1]) & \ldots & E(x^2[N-1]) \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_{N-1} \end{bmatrix} = \begin{bmatrix} E(\theta x[0]) \\ E(\theta x[1]) \\ \vdots \\ E(\theta x[N-1]) \end{bmatrix}$$

Recognizing these matrices as familiar forms we see that the matrix equation can be written as $\mathbf{C}_{xx}\mathbf{a} = \mathbf{C}_{x\theta}$ or $\mathbf{a} = \mathbf{C}_{xx}^{-1}\mathbf{C}_{x\theta}$.

Thus the LMMSE is given by $\hat{\theta} = \mathbf{a}^T\mathbf{x} = \mathbf{C}_{x\theta}^T\mathbf{C}_{xx}^{-1}\mathbf{x}$. Finally $\hat{\theta} = \mathbf{C}_{\theta x}\mathbf{C}_{xx}^{-1}\mathbf{x}$

The minimum Bayesian MSE is the squared length of the error vector or

$$\begin{aligned}
\mathrm{Bmse}(\hat{\theta}) &= ||\epsilon||^2 \\
&= \left\|\theta - \sum_{n=0}^{N-1} a_n x[n]\right\|^2 \\
&= E\left[\left(\theta - \sum_{n=0}^{N-1} a_n x[n]\right)^2\right]
\end{aligned}$$

After a series of manipulations like

$$\begin{aligned}
\mathrm{Bmse}(\hat{\theta}) &= E\left[\left(\theta - \sum_{n=0}^{N-1} a_n x[n]\right)\left(\theta - \sum_{m=0}^{N-1} a_m x[m]\right)\right] \\
&= E\left[\left(\theta - \sum_{n=0}^{N-1} a_n x[n]\right)\theta\right] - E\left[\left(\theta - \sum_{n=0}^{N-1} a_n x[n]\right)\sum_{m=0}^{N-1} a_m x[m]\right] \\
&= E(\theta^2) - \sum_{n=0}^{N-1} a_n E(x[n]\theta) - \sum_{m=0}^{N-1} a_m E\left[\left(\theta - \sum_{n=0}^{N-1} a_n x[n]\right) x[m]\right].
\end{aligned}$$

We see that the last term is zero because of the orthogonality condition. Thus the estimate simplifies to

$$\begin{aligned}
\mathrm{Bmse}(\hat{\theta}) &= C_{\theta\theta} - \mathbf{a}^T \mathbf{C}_{x\theta} \\
&= C_{\theta\theta} - \mathbf{C}_{x\theta}^T \mathbf{C}_{xx}^{-1} \mathbf{C}_{x\theta} \\
&= C_{\theta\theta} - \mathbf{C}_{\theta x} \mathbf{C}_{xx}^{-1} \mathbf{C}_{x\theta}
\end{aligned}$$

This is in agreement with the analytical expression we obtained earlier and thus we have shown that the process of obtaining the LMMSE has a very simple geometrical interpretation.