ECE531 Screencast 6.1: Introduction to Linear MMSE Bayesian Estimation

D. Richard Brown III

Worcester Polytechnic Institute

Introduction

Our focus here is on Bayesian MMSE estimators. We know that such estimators are given as the conditional mean of the unknown parameter(s):

$$\hat{\theta}_{\mathsf{MMSE}}(y) = \mathrm{E}[\Theta \,|\, Y = y]$$

The conditional mean can often be difficult to compute. Two approaches to getting useful results:

- **Restrict the model**, e.g. linear and Gaussian

  $$Y = H\Theta + W$$

  with H known and Θ and W both Gaussian.
- **Restrict the class of estimators**, e.g. linear (actually affine) estimators

  $$\hat{\theta}(y) = c + A^\top y$$

Linear estimators may not be as good as general Bayesian estimators, but they are interesting since they can be computationally convenient.

Review: MMSE Estimation in the Linear Gaussian Model

In the linear Gaussian model, we have

$$\begin{aligned}\hat{\theta}_{\text{MMSE}}(y) &= \mathrm{E}[\Theta \,|\, Y = y] \\ &= \mathrm{E}[\Theta] + \Sigma_\Theta H^\top \left(H \Sigma_\Theta H^\top + \Sigma_W\right)^{-1} (y - H\mathrm{E}[\Theta]) \\ &= c + A^\top y\end{aligned}$$

In this case, the conditional mean is linear in the observations with

$$\begin{aligned}A^\top &= \Sigma_\Theta H^\top \left(H \Sigma_\Theta H^\top + \Sigma_W\right)^{-1} \\ c &= \mathrm{E}[\Theta] - \Sigma_\Theta H^\top \left(H \Sigma_\Theta H^\top + \Sigma_W\right)^{-1} H\mathrm{E}[\Theta]\end{aligned}$$

Hence, there is no loss of optimality in restricting ourselves to a **linear MMSE estimator** in the context of linear Gaussian models.

Scalar Linear MMSE Estimation (1/3)

Assume $\Theta \in \mathbb{R}$. Our LMMSE estimator must be of the form

$$\hat{\theta}(y) = c + A^\top y$$

with $A = [a_0, \dots, a_{N-1}]^\top \in \mathbb{R}^{N \times 1}$ and $c \in \mathbb{R}$. The problem here is to find the coefficients $a_0, \dots, a_{N-1}$ and c to minimize the MSE.

The MSE can be written as a function of A and c as follows:

$$\begin{aligned}
J(A,c) = \mathrm{E}\left[\left(\hat{\theta}(Y) - \Theta\right)^2\right] &= \mathrm{E}\left[\left(c + A^\top Y - \Theta\right)^2\right] \\
&= \mathrm{E}\left[\left((c - \mathrm{E}[\Theta]) + A^\top Y - (\Theta - \mathrm{E}[\Theta])\right)^2\right] \\
&= (c - \mathrm{E}[\Theta])^2 + A^\top \mathrm{E}\left[YY^\top\right] A + \mathrm{var}\left[\Theta^2\right] \\
&\quad + 2(c - \mathrm{E}[\Theta]) A^\top \mathrm{E}\left[Y\right] - 2(c - \mathrm{E}[\Theta])\mathrm{E}\left[\Theta - \mathrm{E}[\Theta]\right] \\
&\quad - 2A^\top \mathrm{E}\left[Y(\Theta - \mathrm{E}[\Theta])\right]
\end{aligned}$$

Scalar Linear MMSE Estimation (2/3)

We have an expression for the MSE

$$J(A,c) = (c - \mathrm{E}[\Theta])^2 + A^\top \mathrm{E}\left[YY^\top\right] A + \mathrm{var}\left[\Theta^2\right] + 2(c - \mathrm{E}[\Theta])A^\top \mathrm{E}\left[Y\right] - 2A^\top \mathrm{E}\left[Y(\Theta - \mathrm{E}[\Theta])\right]$$

We want to find A and c to minimize J. We can take the gradient of $J(A,c)$ with respect to $[c, a_0, \dots, a_N - 1]^\top$ and set it equal to zero...

$$\begin{bmatrix} 2(c - \mathrm{E}[\Theta]) + 2A^\top \mathrm{E}[Y] \\ 2\mathrm{E}\left[YY^\top\right] A + 2(c - \mathrm{E}[\Theta])\mathrm{E}\left[Y\right] - 2\mathrm{E}\left[Y(\Theta - \mathrm{E}[\Theta])\right] \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

How to solve:

- First equation implies $c = \mathrm{E}[\Theta] - A^\top \mathrm{E}[Y] = \mathrm{E}[\Theta] - \mathrm{E}[Y^\top]A$.
- Plug this result into the second equation and solve for A.

Scalar Linear MMSE Estimation (3/3)

The second equation now becomes

$$\begin{aligned} \mathrm{E}\left[YY^\top\right]A + \left(\mathrm{E}[\Theta] - \mathrm{E}[Y^\top]A - \mathrm{E}[\Theta]\right)\mathrm{E}\left[Y\right] - \mathrm{E}\left[Y(\Theta - \mathrm{E}[\Theta])\right] &= 0 \\ \mathrm{E}\left[YY^\top\right]A - \mathrm{E}\left[Y\right]\mathrm{E}[Y^\top]A - \mathrm{E}\left[Y\Theta\right] + \mathrm{E}\left[Y\right]\mathrm{E}[\Theta] &= 0 \\ \left(\mathrm{E}\left[YY^\top\right] - \mathrm{E}\left[Y\right]\mathrm{E}[Y^\top]\right)A &= \mathrm{E}\left[Y\Theta\right] - \mathrm{E}\left[Y\right]\mathrm{E}[\Theta] \end{aligned}$$

Recalling that

$$\begin{aligned} \mathrm{cov}(Y,Y) &= \mathrm{E}\{YY^\top\} - \mathrm{E}\{Y\}\mathrm{E}\{Y^\top\} \\ \mathrm{cov}(Y,X) &= \mathrm{E}\{YX\} - \mathrm{E}\{Y\}\mathrm{E}\{X\} \text{ (when } X \text{ is a scalar)} \end{aligned}$$

we can write

$$\boxed{A_{\mathsf{LMMSE}} = \left[\mathrm{cov}(Y,Y)\right]^{-1}\mathrm{cov}(Y,\Theta)}$$

and c_{LMMSE} follows from $c = \mathrm{E}[\Theta] - A_{\mathsf{LMMSE}}^\top \mathrm{E}[Y]$.

Summary and Remarks

Putting it all together, we have

$$\begin{aligned}
\hat{\theta}_{\mathsf{LMMSE}}(y) &= c + A_{\mathsf{LMMSE}}^{\top} Y \\
&= \mathrm{E}[\Theta] - A_{\mathsf{LMMSE}}^{\top} \mathrm{E}[Y] + A_{\mathsf{LMMSE}}^{\top} Y \\
&= \mathrm{E}[\Theta] + A_{\mathsf{LMMSE}}^{\top} (Y - \mathrm{E}[Y]) \\
&= \mathrm{E}[\Theta] + \mathrm{cov}(\Theta, Y) \left[\mathrm{cov}(Y, Y)\right]^{-1} (Y - \mathrm{E}[Y])
\end{aligned}$$

Remarks:

- This is the same form as we saw with the linear Gaussian model. Our derivation did not assume a linear Gaussian model, however. We only assumed a linear estimator.
- Computation of $\hat{\theta}_{\mathsf{LMMSE}}(y)$ only requires knowledge of means and covariances. We do not need full knowledge of the joint distributions.
- This result easily extends to $p \times 1$ vector parameters (the MSE for each parameter can be minimized separately). A becomes $N \times p$ and c becomes $p \times 1$. $\mathrm{E}[\Theta]$ also becomes $p \times 1$ and $\mathrm{cov}(\Theta, Y)$ becomes $p \times N$.