# ECE531 Screencast 5.5: Bayesian Estimation for the Linear Gaussian Model

D. Richard Brown III

Worcester Polytechnic Institute

# Bayesian Estimation for the Linear Gaussian Model

Recall the linear Gaussian model

$$Y = H\Theta + W$$

where the observation $Y \in \mathbb{R}^n$, the "mixing matrix" $H \in \mathbb{R}^{n \times m}$ is known, the unknown parameter vector $\Theta \in \mathbb{R}^m$ is distributed as $\mathcal{N}(\mu_\Theta, \Sigma_\Theta)$, and the unknown noise vector $W \in \mathbb{R}^n$ is distributed as $\mathcal{N}(0, \Sigma_W)$.

Unless otherwise specified, we always assume the noise and the unknown parameters are independent of each other.

What are the Bayesian MMSE/MMAE/MAP estimators in this case?

# Linear Gaussian Model: Posterior Distribution Analysis

To develop an expression for the posterior distribution $\pi_y(\theta)$, we first note that $\pi_y(\theta) = \frac{p_{Y,\Theta}(y,\theta)}{p_Y(y)}$. To find the joint distribution $p_{Y,\Theta}(y,\theta)$ let

$$Z = \begin{bmatrix} Y \\ \Theta \end{bmatrix} = \begin{bmatrix} H & I \\ I & 0 \end{bmatrix} \begin{bmatrix} \Theta \\ W \end{bmatrix}$$

Since $\Theta$ and $W$ are independent of each other and each is Gaussian, they are jointly Gaussian. Furthermore, since $Z$ is a linear transformation of a jointly Gaussian random vector, it too is jointly Gaussian.

To fully characterize $Z \in \mathcal{N}(\mu_Z, \Sigma_Z)$, we just need its mean and covariance:

$$\begin{aligned} \mu_Z &:= \mathrm{E}[Z] = \begin{bmatrix} H\mu_\Theta \\ \mu_\Theta \end{bmatrix} \\ \Sigma_Z &:= \mathrm{cov}[Z] = \begin{bmatrix} H\Sigma_\Theta H^\top + \Sigma_W & H\Sigma_\Theta \\ \Sigma_\Theta H^\top & \Sigma_\Theta \end{bmatrix} \end{aligned}$$

# Linear Gaussian Model: Posterior Distribution Analysis

To compute the posterior, we can write

$$\pi_y(\theta) = \frac{p_Z(z)}{p_Y(y)} = \frac{\frac{1}{(2\pi)^{(m+n)/2}|\Sigma_Z|^{1/2}} \exp\left\{\frac{-(z-\mu_Z)^\top \Sigma_Z^{-1}(z-\mu_Z)}{2}\right\}}{\frac{1}{(2\pi)^{n/2}|\Sigma_Y|^{1/2}} \exp\left\{\frac{-(y-\mu_Y)^\top \Sigma_Y^{-1}(y-\mu_Y)}{2}\right\}}$$

To simplify the terms **outside** of the exponentials, note that

$$\Sigma_Z \quad := \quad \mathrm{cov}[Z] = \begin{bmatrix} H\Sigma_\Theta H^\top + \Sigma_W & H\Sigma_\Theta \\ \Sigma_\Theta H^\top & \Sigma_\Theta \end{bmatrix} = \begin{bmatrix} \Sigma_Y & \Sigma_{Y,\Theta} \\ \Sigma_{\Theta,Y} & \Sigma_\Theta \end{bmatrix}$$

The determinant of a partitioned matrix $P = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$ can be written as $|P| = |A| \cdot |D - CA^{-1}B|$ if $A$ is invertible. Covariance matrices are invertible, hence the terms outside the exponentials can be simplified to

$$\frac{\frac{1}{(2\pi)^{(m+n)/2}|\Sigma_Z|^{1/2}}}{\frac{1}{(2\pi)^{n/2}|\Sigma_Y|^{1/2}}} \quad = \quad \frac{1}{(2\pi)^{m/2}|\Sigma_\Theta - \Sigma_{\Theta,Y}\Sigma_Y^{-1}\Sigma_{Y,\Theta}|^{1/2}}$$

# Linear Gaussian Model: Posterior Distribution Analysis

To simplify the terms **inside** the exponentials, we can use a matrix inversion formula for partitioned matrices ($A$ must be invertible)

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} (A - BD^{-1}C)^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{bmatrix}$$

and the matrix inversion lemma

$$(A - BD^{-1}C)^{-1} = A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1}.$$

Skipping all the algebraic details, we can write

$$\frac{\exp\left\{\frac{-(z-\mu_Z)^\top \Sigma_Z^{-1}(z-\mu_Z)}{2}\right\}}{\exp\left\{\frac{-(y-\mu_Y)^\top \Sigma_Y^{-1}(y-\mu_Y)}{2}\right\}} = \exp\left\{\frac{-(\theta - \alpha(y))^\top \Sigma^{-1}(\theta - \alpha(y))}{2}\right\}$$

where $\alpha(y) = \mu_\Theta + \Sigma_{\Theta,Y}\Sigma_Y^{-1}(y - \mu_Y)$ and $\Sigma = \Sigma_\Theta - \Sigma_{\Theta,Y}\Sigma_Y^{-1}\Sigma_{Y,\Theta}$.

# Linear Gaussian Model: Posterior Distribution Analysis

Putting it all together, we have the posterior distribution

$$\pi_y(\theta) = \frac{1}{(2\pi)^{m/2}|\Sigma|^{1/2}} \exp\left\{ \frac{-(\theta - \alpha(y))^\top \Sigma^{-1}(\theta - \alpha(y))}{2} \right\}$$

where $\alpha(y) = \mu_\Theta + \Sigma_{\Theta,Y}\Sigma_Y^{-1}(y - \mu_Y)$ and $\Sigma = \Sigma_\Theta - \Sigma_{\Theta,Y}\Sigma_Y^{-1}\Sigma_{Y,\Theta}$ with

$$\begin{aligned}
\Sigma_{\Theta,Y} &= \mathrm{cov}(\Theta, Y) = \mathrm{E}\left[(\Theta - \mu_\Theta)(H\Theta + W - H\mu_\Theta)^\top\right] = \Sigma_\Theta H^\top \\
\Sigma_{Y,\Theta} &= \Sigma_{\Theta,Y}^\top = H\Sigma_\Theta \\
\Sigma_Y &= \mathrm{cov}(Y, Y) = H\Sigma_\Theta H^\top + \Sigma_W \\
\mu_Y &= \mathrm{E}[H\Theta + W] = H\mu_\Theta
\end{aligned}$$

What can we say about the posterior distribution of the random parameter $\Theta$ conditioned on the observation $Y = y$?

# Linear Gaussian Model: Bayesian Estimators

## Lemma

*In the linear Gaussian model, the parameter vector* $\Theta$ *conditioned on the observation* $Y = y$ *is jointly Gaussian distributed with*

$$\begin{aligned} \mathrm{E}[\Theta \,|\, Y = y] &= \mu_\Theta + \Sigma_\Theta H^\top \left( H \Sigma_\Theta H^\top + \Sigma_W \right)^{-1} (y - H\mu_\Theta) \\ \mathrm{cov}[\Theta \,|\, Y = y] &= \Sigma_\Theta - \Sigma_\Theta H^\top \left( H \Sigma_\Theta H^\top + \Sigma_W \right)^{-1} H \Sigma_\Theta \end{aligned}$$

## Corollary

*In the linear Gaussian model*

$$\hat{\theta}_{mmse}(y) = \hat{\theta}_{mmae}(y) = \hat{\theta}_{map}(y)$$

# Linear Gaussian Model: Bayesian Estimator Remarks

- All of the estimators are linear (actually affine) in the observation $y$.
- Recall that the performance of the Bayesian MMSE estimator is

$$\begin{aligned}\text{MMSE} &= \mathrm{E}\left[\|\Theta - \hat{\theta}_{\text{mmse}}(Y)\|_2^2\right] \\ &= \int \text{trace}\left\{\text{cov}(\Theta \,|\, Y = y)\right\} p(y)\, dy.\end{aligned}$$

In the linear Gaussian model, we see that $\text{cov}[\Theta \,|\, Y = y]$ does not depend on $y$. Hence, we can move the trace outside of the integral and write the MMSE as

$$\begin{aligned}\text{MMSE} &= \text{trace}\left\{\text{cov}[\Theta \,|\, Y = y]\right\} \int p(y)\, dy \\ &= \text{trace}\left\{\Sigma_\Theta\right\} - \text{trace}\left\{\Sigma_\Theta H^\top \left(H \Sigma_\Theta H^\top + \Sigma_W\right)^{-1} H \Sigma_\Theta\right\}.\end{aligned}$$