ME 233 Advance Control II

Lecture 7
Least Squares Estimation

(ME233 Class Notes pp. LS1-LS5)

Notation

Let X and Y be continuous random vectors with joint PDF $p_{XY}(x,y)$

Let x and y be respectively outcomes of X and Y and

$$x \in R_x \subset R^{n_x} \qquad y \in R_y \subset R^{n_y}$$

$$p_{XY} : R_x \times R_y \to R_+$$

Marginal Expectation

Let X and Y be continuous random vectors with joint PDF $p_{XY}(x,y)$

Marginal Expectation (mean) of X

$$m_X = E\{X\}$$

$$= \int_{R_x} \underbrace{\int_{R_y} x\, p_{XY}(x,y)\,dy}_{x p_X(x)}\,dx$$

Marginal Expectation

Let X and Y be continuous random variables with joint PDF $p_{XY}(x,y)$

Marginal Expectation (mean) of X

$$m_X = E\{X\} = \int_{R_x} x\, p_X(x)\,dx$$

$$= \hat{x}$$

new notation
(following the ME233 class notes)

Marginal Expectation $\widehat{x}$

$\widehat{x}$ is the minimum least squares **marginal estimator** of X

- For any deterministic vector $\boldsymbol{z}$

$$E\{\|X-\widehat{x}\|^2\} \leq E\{\|X-z\|^2\}$$

$$\|z\|^2 = z^T z$$

$$E\{\|X-\widehat{x}\|^2\} \leq E\{\|X-z\|^2\}$$

- Proof:

$$E\{\|X-z\|^2\} = \int_{R_x} \|x-z\|^2 p_X(x)dx$$

$$= \int_{R_x} \|x-\widehat{x}+\widehat{x}-z\|^2 p_X(x)dx$$

$$= \int_{R_x} \|x-\widehat{x}\|^2 p_X(x)dx + \int_{R_x} \|\widehat{x}-z\|^2 p_X(x)dx$$

$$+\ 2\int_{R_x} (x-\widehat{x})^T(\widehat{x}-z)p_X(x)dx \quad \text{“cross term”}$$

$$E\{\|X-\widehat{x}\|^2\} \leq E\{\|X-z\|^2\}$$

- Consider the “cross term”

$$\int_{R_x} (x-\widehat{x})^T \underbrace{(\widehat{x}-z)}_{\text{deterministic}} p_X(x)dx =$$

$$= (\widehat{x}-z)^T \int_{R_x} (x-\underbrace{\widehat{x}}_{\text{deterministic}})p_X(x)dx$$

$$= (\widehat{x}-z)^T \left[\underbrace{\int_{R_x} x p_X(x)dx}_{\widehat{x}} - \widehat{x}\right]$$

$$= 0$$

$$E\{\|X-\widehat{x}\|^2\} \leq E\{\|X-z\|^2\}$$

- Proof:

$$E\{\|X-z\|^2\} = \int_{R_x} \|x-\widehat{x}\|^2 p_X(x)dx + \int_{R_x} \|\widehat{x}-z\|^2 p_X(x)dx$$

$$= E\{\|X-\widehat{x}\|^2\} + \underbrace{E\{\|\widehat{x}-z\|^2\}}_{\geq 0}$$

$$\geq E\{\|X-\widehat{x}\|^2\}$$

QED

Conditional Expectation

Let X and Y be continuous random vectors with joint PDF $p_{XY}(x,y)$

Conditional Expectation (conditional mean) of X given and outcome $Y=y$

$$m_{X|Y}(y) = E\{X|Y=y\}$$

$$= \int_{R_x} x\, p_{X|Y}(x|y)dx$$

Conditional Expectation

Conditional Expectation (conditional mean) of X given and outcome $Y=y$

$$m_{X|Y}(y) = \int_{R_x} x\, p_{X|Y}(x|y)dx$$

$$= \int_{R_x} x \left(\frac{p_{XY}(x,y)}{p_Y(y)}\right) dx$$

$$= \hat{x}|_y$$

new notation (following the ME233 class notes)

Conditional Expectation $\hat{x}|_y$

Notice that the conditional expectation $\hat{x}|_y$

$$\hat{x}|_y = \int_{R_x} x \frac{p_{XY}(x,y)}{p_Y(y)}dx$$

can be interpreted as a function of <u>the random variable Y</u>.

$$\hat{X}|_Y = \int_{R_x} x \frac{p_{XY}(x,Y)}{p_Y(Y)}dx$$

Conditional Expectation $\hat{X}|_Y$

Lemma:

For any function $f(\cdot)$ of the random vector Y, with the appropriate dimensions

$$E\{X\, f(Y)\} = E\{\hat{X}|_Y\, f(Y)\}$$

we can replace $\boldsymbol{X}$ *by its conditional expectation* $\hat{X}|_Y$

$$E\{X\, f(Y)\} = E\{\widehat{X}|_Y\, f(Y)\}$$

Proof

$$E\{X\, f(Y)\} =$$

$$= \int_{R_y}\int_{R_x} x\, f(y)\, \underbrace{p_{XY}(x,y)}_{p_{X|Y}(x|y)p_Y(y)} dxdy$$

(Baye's rule)

$$= \int_{R_y}\int_{R_x} x\, p_{X|Y}(x|y)\, f(y)\, p_Y(y) dxdy$$

$$E\{X\, f(Y)\} = E\{\widehat{X}|_Y\, f(Y)\}$$

Proof

$$E\{X\, f(Y)\} =$$

$$= \int_{R_y}\int_{R_x} x\, p_{X|Y}(x|y)\, f(y)\, p_Y(y) dxdy$$

$$= \int_{R_y} \underbrace{\left(\int_{R_x} x\, p_{X|Y}(x|y)\, dx\right)}_{\widehat{x}|_y} f(y)\, p_Y(y) dy$$

$$E\{X\, f(Y)\} = E\{\widehat{X}|_Y\, f(Y)\}$$

Proof

$$E\{X\, f(Y)\} = \int_{R_y} \widehat{x}|_y\, f(y)\, p_Y(y) dy$$

$$= E\{\widehat{X}|_Y\, f(Y)\}$$

QED

Conditional Expectation $\widehat{X}|_Y$

Theorem:

$\widehat{X}|_Y$ is the least squares minimum estimator of ***X*** given ***Y***, *i.e.*

$$E\{\|X - \widehat{X}|_Y\|^2\} \leq E\{\|X - f(Y)\|^2\}$$

for all functions $f(\cdot)$ of Y of appropriate dimensions

$$\|X\|^2 = X^T X$$

$$E\{\|X-\hat{X}|_Y\|^2\} \leq E\{\|X-f(Y)\|^2\}$$

Proof:

$$E\{\|X-f(Y)\|^2\}$$

$$= \int_{R_y}\int_{R_x} \|x-f(y)\|^2 \, p_{XY}(x,y)dxdy$$

$$= \int_{R_y}\int_{R_x} \|x-\hat{x}|_y + \hat{x}|_y - f(y)\|^2 \, p_{XY}(x,y)dxdy$$

$$= E\{\|X-\hat{X}|_Y + \hat{X}|_Y - f(Y)\|^2\}$$

$$E\{\|X-\hat{X}|_Y\|^2\} \leq E\{\|X-f(Y)\|^2\}$$

Proof:

$$E\{\|X-f(Y)\|^2\} = E\{\|X-\hat{X}|_Y + \hat{X}|_Y - f(Y)\|^2\}$$

$$= E\{\|X-\hat{X}|_Y\|^2\} + E\{\|\hat{X}|_Y - f(Y)\|^2\}$$

$$+ \quad 2E\{(X-\hat{X}|_Y)^T(\hat{X}|_Y - f(Y))\}$$

$$E\{\|X-\hat{X}|_Y\|^2\} \leq E\{\|X-f(Y)\|^2\}$$

Proof:

$$E\{\|X-f(Y)\|^2\} = E\{\|X-\hat{X}|_Y + \hat{X}|_Y - f(Y)\|^2\}$$

$$= E\{\|X-\hat{X}|_Y\|^2\} + E\{\|\hat{X}|_Y - f(Y)\|^2\}$$

$$\boxed{+ \quad \underbrace{2E\{(X-\hat{X}|_Y)^T\underbrace{(\hat{X}|_Y - f(Y))}_{h(Y)}\}}_{2(E\{(X^T h(Y)\} - E\{\hat{X}|_Y^T h(Y)\})}}$$

0

$$E\{\|X-\hat{X}|_Y\|^2\} \leq E\{\|X-f(Y)\|^2\}$$

Proof:

$$E\{\|X-f(Y)\|^2\} = E\{\|X-\hat{X}|_Y + \hat{X}|_Y - f(Y)\|^2\}$$

$$= E\{\|X-\hat{X}|_Y\|^2\} + \underbrace{E\{\|\hat{X}|_Y - f(Y)\|^2\}}_{\geq 0}$$

$$\geq E\{\|X-\hat{X}|_Y\|^2\}$$

QED

Marginal Variance of X

$$\sigma_X^2 = \text{trace}\,(\Lambda_{XX})$$

$$\Lambda_{XX} = E\{(X-\hat{x})(X-\hat{x})^T\}$$

$$\hat{x} = \int_{R_x} x\, p_X(x) dx$$

$$\Lambda_{XX} = \int_{R_x} (x-\hat{x})(x-\hat{x})^T p_X(x) dx$$

Conditional Variance of X given Y

$$\sigma_{X|Y}^2(Y) = \text{trace}\left(\Lambda_{X|Y}(Y)\right)$$

$$\Lambda_{X|Y}(Y) = E\{(X-\hat{X}|_Y)(X-\hat{X}|_Y)^T|Y\}$$

$$\hat{X}|_Y = \int_{R_x} x\, p_{X|Y}(x,Y) dx$$

$$\Lambda_{X|Y}(Y) = \int_{R_x} (x-\hat{X}|_Y)(x-\hat{X}|_Y)^T p_{X|Y}(x,Y) dx$$

Expectation of Conditional Variance of X

$$E\{\sigma_{X|Y}^2(Y)\} = \text{trace}\left(E\{\Lambda_{X|Y}(Y)\}\right)$$

$$E\{\sigma_{X|Y}^2(Y)\} = E\{\|X-\hat{X}|_Y\|^2\}$$

$$\hat{X}|_Y = \int_{R_x} x\, p_{X|Y}(x,Y) dx$$

$$\Lambda_{X|Y}(Y) = \int_{R_x} (x-\hat{X}|_Y)(x-\hat{X}|_Y)^T p_{X|Y}(x,Y) dx$$

Expectation of Conditional mean of X

$$E\{\hat{X}|_Y\} = \hat{x}$$

$$E\{\hat{X}|_Y\} = \int_{R_y} \hat{x}|_y\, p_Y(y) dy$$

$$= \int_{R_y} \left(\int_{R_x} x\, p_{X|Y}(x,y) dx\right) p_Y(y) dy$$

(by Baye's rule)

$$= \int_{R_x} x \left(\int_{R_y} p_{XY}(x,y) dy\right) dx$$

$$= \int_{R_x} x\, p_X(x) dx \quad = \hat{x}$$

Variance of the Conditional mean of X given Y

$$\sigma^2_{\widehat{X}|_Y} = \mathrm{trace}\left(\Lambda_{\widehat{X}|_Y \widehat{X}|_Y}\right)$$

$$\Lambda_{\widehat{X}|_Y \widehat{X}|_Y} = E\{(\widehat{X}|_Y - \widehat{x})(\widehat{X}|_Y - \widehat{x})^T\}$$

$$\widehat{X}|_Y = \int_{R_x} x\, p_{X|Y}(x,Y)dx$$

$$\Lambda_{\widehat{X}|_Y \widehat{X}|_Y} = \int_{R_y} (\widehat{x}|_y - \widehat{x})(\widehat{x}|_y - \widehat{x})^T p_Y(y)dy$$

Law of variances

marginal variance = expected value of conditional variance

+ variance of conditional mean

$$\sigma^2_X = E\{\sigma^2_{X|Y}(Y)\} + \sigma^2_{\widehat{X}|_Y}$$

$$\Lambda_{XX} = E\{\Lambda_{X|Y}(Y)\} + \Lambda_{\widehat{X}|_Y \widehat{X}|_Y}$$

$$\sigma^2_X = E\{\sigma^2_{X|Y}(Y)\} + \sigma^2_{\widehat{X}|_Y}$$

Proof:

$$E\{\|X - \widehat{x}\|^2\} = E\{\|X - \widehat{X}|_Y + \widehat{X}|_Y - \widehat{x}\|^2\}$$

$$= E\{\|X - \widehat{X}|_Y\|^2\} + E\{\|\widehat{X}|_Y - \widehat{x}\|^2\}$$

$$\boxed{\underbrace{+\ 2E\{(X - \widehat{X}|_Y)^T \underbrace{(\widehat{X}|_Y - \widehat{x})}_{h(Y)}\}}_{2(E\{(X^T h(Y)\} - E\{\widehat{X}|_Y^T h(Y)\})} \quad \mathbf{0}}$$

$$\sigma^2_X = E\{\sigma^2_{X|Y}(Y)\} + \sigma^2_{\widehat{X}|_Y}$$

Proof:

$$\underbrace{E\{\|X - \widehat{x}\|^2\}}_{\sigma^2_X} = \underbrace{E\{\|X - \widehat{X}|_Y\|^2\}}_{E\{\sigma^2_{X|Y}(Y)\}} + \underbrace{E\{\|\widehat{X}|_Y - \widehat{x}\|^2\}}_{\sigma^2_{\widehat{X}|_Y}}$$

QED

Conditional Expectation for Gaussians

When $\boldsymbol{X}$ and $\boldsymbol{Y}$ are jointly Gaussians

The conditional probability $p_{X|Y}(x|y)$

and

conditional expectations $\widehat{x}|_y$
(for any outcome $\boldsymbol{y}$)

can be calculated very easily!

Conditional Expectation for Gaussians

When $\boldsymbol{X}$ and $\boldsymbol{Y}$ are jointly Gaussians

The conditional covariance $\Lambda_{X|Y}$

is not a function of $\boldsymbol{Y}$ *!*

$$\Lambda_{XX} = \Lambda_{X|Y} + \Lambda_{\widehat{X}|_Y \widehat{X}|_Y}$$

and

$$\Lambda_{\widehat{X}|_Y \widehat{X}|_Y} = \Lambda_{XY} \Lambda_{YY}^{-1} \Lambda_{YX}$$

Conditional expectation for Gaussians

- The conditional expectation of X given $Y = y$

$$\widehat{x}|_y = E\{X|y\}$$

$$\widehat{x}|_y = \widehat{x} + \Lambda_{XY}\,\Lambda_{YY}^{-1}\,(y - \widehat{y})$$

affine function of the outcome y !!

$\widehat{x} = E\{X\}$

$\widehat{y} = E\{Y\}$

Conditional PDF for Gaussians

- The conditional PDF of X given $Y = y$

$$p_{X|Y}(x|y) \;=\; \frac{1}{(2\pi)^{n/2}\sqrt{|(\Lambda_{X|Y}|}} e^{-\frac{1}{2}(x-\widehat{x}|_y)^T \Lambda_{X|Y}^{-1}(x-\widehat{x}|_y)}$$

$$\Lambda_{X|Y} = \Lambda_{XX} - \Lambda_{XY}\,\Lambda_{YY}^{-1}\Lambda_{YX}$$

Independent of y !!

Normal distribution

Conditional Mean

Let X be a random n vector and Y be a random m vector

- The conditional mean of X given outcome Y=y depends on the outcome

$$\widehat{x}_{|y} = m_{X|y} = E\{X|Y = y\}$$

- The conditional estimation error of X given outcome Y=y is:

$$\tilde{X}_{|y} = X - \widehat{x}_{|y}$$

Conditional Mean for Gaussians

Let $X \sim N(\widehat{x}, \Lambda_{XX})$ $\quad$ $Y \sim N(\widehat{y}, \Lambda_{YY})$

- The conditional mean of X given outcome Y=y is linear

$$\widehat{x}_{|y} = \widehat{x} + \Lambda_{XY}\,\Lambda_{YY}^{-1}\,(y - \widehat{y})$$

- The conditional estimation error of X given outcome Y=y is:

$$\tilde{X}_{|y} = \tilde{X} - \Lambda_{XY}\,\Lambda_{YY}^{-1}\,\tilde{y}$$

$$\tilde{X} = X - \widehat{x} \qquad \tilde{y} = y - \widehat{y}$$

$$\widehat{X}_{|Y} = \widehat{x} + \Lambda_{XY}\,\Lambda_{YY}^{-1}\,(Y - \widehat{y})$$

- The expected value of the conditional mean $\widehat{X}_{|Y}$

$$E\{\widehat{X}_{|Y}\} = \widehat{x} + \Lambda_{XY}\,\Lambda_{YY}^{-1}\,\underbrace{(\underbrace{E\{Y\}}_{\widehat{y}} - \widehat{y})}_{E\{\tilde{Y}\} = 0}$$

$$= \widehat{x}$$

$$\widehat{X}_{|Y} = \widehat{x} + \Lambda_{XY}\,\Lambda_{YY}^{-1}\,(Y - \widehat{y})$$

- The covariance of the conditional mean $\widehat{X}_{|Y}$

$$\Lambda_{\widehat{X}_{|Y}\widehat{X}_{|Y}} = E\{(\widehat{X}_{|Y} - \widehat{x})(\widehat{X}_{|Y} - \widehat{x})^T\}$$

$$= E\{\Lambda_{XY}\,\Lambda_{YY}^{-1}\tilde{Y})(\Lambda_{XY}\,\Lambda_{YY}^{-1}\tilde{Y})^T\}$$

$$= \Lambda_{XY}\,\Lambda_{YY}^{-1}E\{\tilde{Y}\tilde{Y}^T\}\,\Lambda_{YY}^{-1}\Lambda_{YX}$$

$$= \Lambda_{XY}\,\Lambda_{YY}^{-1}\Lambda_{YX}$$

Least Squares Estimation: Property 1

- The conditional estimation error $\tilde{X}_{|Y}$ and Y are ***uncorrelated***

$$E\{\tilde{X}_{|Y}\tilde{Y}^T\} = 0$$

- $\tilde{X}_{|Y}$ and $\widehat{X}_{|Y}$ are ***orthogonal***

$$E\{\tilde{X}_{|Y}\widehat{X}^T_{|Y}\} = 0 \quad \text{and} \quad E\{\tilde{X}^T_{|Y}\widehat{X}_{|Y}\} = 0$$

$$E\{\tilde{X}_{|Y}\tilde{Y}^T\} = 0$$

Proof

$$\tilde{X}_{|Y} = \tilde{X} - \Lambda_{XY}\Lambda_{YY}^{-1}\tilde{Y}$$

$$\begin{aligned} E\{\tilde{X}_{|Y}\tilde{Y}^T\} &= E\left\{\left[\tilde{X} - \Lambda_{XY}\Lambda_{YY}^{-1}\tilde{Y}\right]\tilde{Y}^T\right\} \\ &= E\{\tilde{X}\tilde{Y}^T\} - \Lambda_{XY}\Lambda_{YY}^{-1}E\{\tilde{Y}\tilde{Y}^T\} \\ &= \Lambda_{XY} - \Lambda_{XY}\underbrace{\Lambda_{YY}^{-1}\Lambda_{YY}}_{I} \\ &= 0 \end{aligned}$$

$$E\{\tilde{X}_{|Y}\widehat{X}^T_{|Y}\} = 0$$

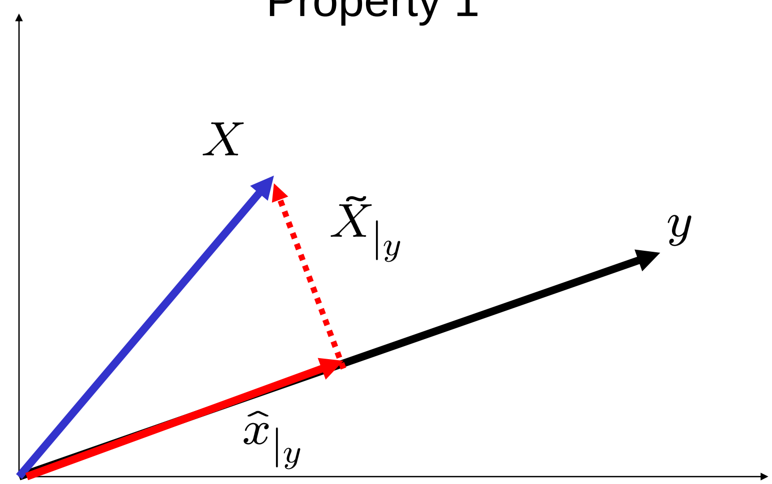
Proof

$$\widehat{X}_{|Y} = \widehat{x} + \Lambda_{XY}\Lambda_{YY}^{-1}\tilde{Y}$$

$$\begin{aligned} E\{\widehat{X}_{|Y}\tilde{X}^T_{|Y}\} &= E\left\{\left[\widehat{x} + \Lambda_{XY}\Lambda_{YY}^{-1}\tilde{Y}\right]\tilde{X}^T_{|Y}\right\} \\ &= \widehat{x}\underbrace{E\{\tilde{X}^T_{|Y}\}}_{=0} + \Lambda_{XY}\Lambda_{YY}^{-1}\underbrace{E\{\tilde{Y}\tilde{X}^T_{|Y}\}}_{=0} \end{aligned}$$

QED

Deterministic interpretation of Property 1



Recursive LS Estimation

Let X , Y and Z be a random Gaussian vectors

$X \sim N(\hat{x}, \Lambda_{XX})$ $\quad X \in \mathcal{R}^n \quad \} \; n$

$Y \sim N(\hat{y}, \Lambda_{YY})$ $\quad Y \in \mathcal{R}^M \quad \} \; M >> n, \, p$

$Z \sim N(\hat{z}, \Lambda_{ZZ})$ $\quad Z \in \mathcal{R}^p \quad \} \; p$

Recursive LS Estimation

1. Assume that we already know of outcome $\boldsymbol{Y = y}$

and we have obtained

$$\hat{x}_{|y} = E\{X|Y = y\}$$

$$\hat{x}_{|y} = \hat{x} + \Lambda_{XY} \underbrace{\Lambda_{YY}^{-1}}_{\text{inversion of an } M \times M \text{ matrix}} (y - \hat{y})$$

n $\qquad$ M

Recursive LS Estimation

1. Assume that we already know of outcome $\boldsymbol{Y = y}$

and we have obtained

$$\hat{x}_{|y} = \hat{x} + \Lambda_{XY} \Lambda_{YY}^{-1} (y - \hat{y})$$

2. Now we also know the outcome $\boldsymbol{Z = z}$

How do we obtain efficiently?

$$\hat{x}_{|yz} = E\{X|Y = y, \, Z = z\} \;\; ?$$

None-Recursive LS Estimation

1) Define the vector $\quad W = \begin{bmatrix} Z \\ Y \end{bmatrix} \qquad \hat{w} = \begin{bmatrix} \hat{z} \\ \hat{y} \end{bmatrix}$

2) Compute $\quad \hat{x}_{|yz} = E\{X|Y = y, \, Z = z\}$

$$\hat{x}_{|yz} = \hat{x} + \Lambda_{XW} \underbrace{\Lambda_{WW}^{-1}}_{\text{inversion of an } (p+M) \times (p+M) \text{ matrix}} (w - \hat{w})$$

n $\qquad$ $p + M$

Least Squares Estimation: Property 2

Assume that Z and Y are uncorrelated, i.e.

$$\Lambda_{ZY} = E\{\tilde{Z}\tilde{Y}^T\} = 0$$

Then,

$$\hat{x}_{|yz} = \hat{x}_{|}y + E\{\tilde{X}_{|Y}|z\}$$

Least Squares Estimation: Property 2

If Z and Y are uncorrelated,

$$\hat{x}_{|yz} = \hat{x}_{|}y + E\{\tilde{X}_{|Y}|z\}$$

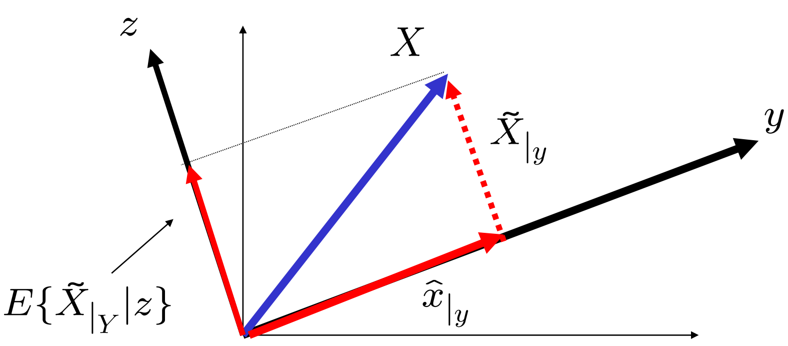
where

$$\hat{x}_{|y} = \hat{x} + \Lambda_{XY}\Lambda_{YY}^{-1}(y-\hat{y})$$

$$E\{\tilde{X}_{|Y}|z\} = \Lambda_{XZ}\Lambda_{ZZ}^{-1}(z-\hat{z}) \qquad (E\{\tilde{X}_{|Y}\}=0)$$

$$(E\{\tilde{Y}\tilde{Z}^T\}=0)$$

Deterministic interpretation of Property 2



$$\hat{x}_{|yz} = \hat{x}_{|y} + E\{\tilde{X}_{|Y}|z\}$$

$$E\{\tilde{Y}\tilde{Z}^T\} = 0 \Rightarrow E\{\tilde{X}_{|Y}|z\} = E\{\tilde{X}|z\}$$

- According to least squares estimation:

$$E\{\tilde{X}_{|Y}|z\} = \underbrace{E\{\tilde{X}_{|Y}\}}_{=0} + \Lambda_{\tilde{X}_{|Y}Z}\Lambda_{ZZ}^{-1}(z-\hat{z})$$

$$\Lambda_{\tilde{X}_{|Y}Z} = E\{\tilde{X}_{|Y}\tilde{Z}^T\} = E\left\{\left[\tilde{X} - \Lambda_{XY}\Lambda_{YY}^{-1}\tilde{Y}\right]\tilde{Z}^T\right\}$$

$$= \underbrace{E\{\tilde{X}\tilde{Z}^T\}}_{\Lambda_{XZ}} - \Lambda_{XY}\Lambda_{YY}^{-1}\underbrace{E\{\tilde{Y}\tilde{Z}^T\}}_{=0}$$

because Z and Y are uncorrelated

$$E\{\tilde{Y}\tilde{Z}^T\} = 0 \Rightarrow E\{\tilde{X}_{|Y}|z\} = E\{\tilde{X}|z\}$$

- Therefore,

$$E\{\tilde{X}_{|Y}|z\} = \Lambda_{XZ}\Lambda_{ZZ}^{-1}(z - \hat{z}) \ = E\{\tilde{X}|z\}$$

QED

Proof of LS property 2

Since $\boldsymbol{Z}$ and $\boldsymbol{Y}$ are uncorrelated,

$$\hat{x}_{|yz} = \hat{x} + \underbrace{\Lambda_{XW}}_{\begin{bmatrix} \Lambda_{XZ} & \Lambda_{XY} \end{bmatrix}} \underbrace{\Lambda_{WW}^{-1}}_{\begin{bmatrix} \Lambda_{ZZ}^{-1} & 0 \\ 0 & \Lambda_{YY}^{-1} \end{bmatrix}} \underbrace{(w - \hat{w})}_{\begin{bmatrix} \tilde{z} \\ \tilde{y} \end{bmatrix}}$$

$$\hat{x}_{|yz} = \hat{x} + \Lambda_{XY}\Lambda_{YY}^{-1}\tilde{y} \ + \Lambda_{XZ}\Lambda_{ZZ}^{-1}\tilde{z}$$

Proof of LS property 2

Since $\boldsymbol{Z}$ and $\boldsymbol{Y}$ are uncorrelated,

$$\hat{x}_{|yz} = \underbrace{\hat{x} + \Lambda_{XY}\Lambda_{YY}^{-1}\tilde{y}}_{\hat{x}_{|y}} \ + \underbrace{\Lambda_{XZ}\Lambda_{ZZ}^{-1}\tilde{z}}_{E\{\tilde{X}|z\}}$$

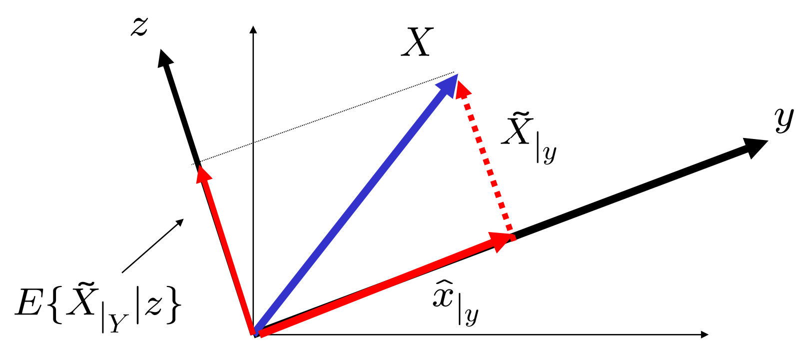
However, since

$$E\{\tilde{Y}\tilde{Z}^T\} = 0 \Rightarrow E\{\tilde{X}|z\} = E\{\tilde{X}_{|Y}|z\}$$

$$\hat{x}_{|yz} = \hat{x}_{|}y + E\{\tilde{X}_{|Y}|z\}$$

QED

Deterministic interpretation of Property 2



$$\hat{x}_{|yz} = \ \hat{x}_{|y} \ + \ E\{\tilde{X}_{|Y}|z\}$$

Least Squares Estimation : Property 3

What happens when Z and Y are **correlated**?

$$\Lambda_{ZY} = E\{\tilde{Z}\tilde{Y}^T\} \neq 0$$

Then,

$$\widehat{x}_{|yz} = \widehat{x}_{|y} + \underbrace{E\{\tilde{X}_{|Y} | \tilde{z}_{|y}\}}$$

We need to explain what this means

Recursive LS Estimation

Assume that we already know of outcome $Y = y$
Then we compute:

The conditional mean of X

$$\widehat{x}_{|y} = \widehat{x} + \Lambda_{XY}\,\Lambda_{YY}^{-1}\,\tilde{y}$$

The conditional mean of Z

$$\widehat{z}_{|y} = \widehat{z} + \Lambda_{ZY}\,\Lambda_{YY}^{-1}\,\tilde{y}$$

The corresponding conditional estimation errors are:

$$\tilde{X}_{|Y} = X - \widehat{X}_{|Y} \qquad \tilde{Z}_{|Y} = Z - \widehat{Z}_{|Y}$$

Both are random signals that are **uncorrelated** with Y

Recursive LS Estimation

We now get a new outcome $Z = z$ in addition to $Y = y$

We still have:

The conditional mean of X

$$\widehat{x}_{|y} = \widehat{x} + \Lambda_{XY}\,\Lambda_{YY}^{-1}\,\tilde{y}$$

The conditional mean of Z

$$\widehat{z}_{|y} = \widehat{z} + \Lambda_{ZY}\,\Lambda_{YY}^{-1}\,\tilde{y}$$

But now:

$$\tilde{X}_{|y} = X - \widehat{x}_{|y}$$

This is still random

$$\boxed{\tilde{z}_{|y} = z - \widehat{z}_{|y}}$$

This is now an outcome

Recursive LS Estimation

We get a new outcome $Z = z$ in addition to $Y = y$

$$\tilde{X}_{|Y} = X - \widehat{X}_{|Y}$$

This is still random

$$\tilde{z}_{|y} = z - \widehat{z}_{|y}$$

This is now an outcome

Compute:

The expected value of $\tilde{X}_{|Y}$ given the outcome $\tilde{z}_{|y}$

$$E\{\tilde{X}_{|Y} | \tilde{z}_{|y}\}$$

Recursive Least Squares

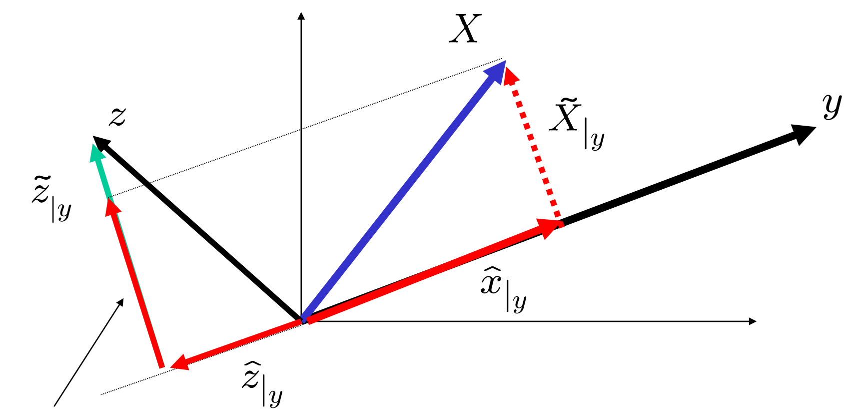
The expected value of $\boldsymbol{X}$ given outcomes $\boldsymbol{y}$ and $\boldsymbol{z}$

$$\hat{x}_{|yz} = \underbrace{\hat{x}_{|y}}_{\text{The expected value of } X \text{ given outcome } y} + \underbrace{E\{\tilde{X}_{|Y} | \tilde{z}_{|y}\}}_{\text{The expected value of } \tilde{X}_{|Y} \text{ given the outcome } \tilde{z}_{|y}}$$

The expected value of $\boldsymbol{X}$ given outcome $\boldsymbol{y}$

The expected value of $\tilde{X}_{|Y}$ given the outcome $\tilde{z}_{|y}$

Deterministic interpretation of Property 3



$$\hat{x}_{|yz} = \hat{x}_{|y} + E\{\tilde{X}_{|Y} | \tilde{z}_{|y}\}$$

Computation of $E\{\tilde{X}_{|Y} | \tilde{z}_{|y}\}$

$$\boxed{E\{\tilde{X}_{|Y} | \tilde{z}_{|y}\} = (\Lambda_{XZ} - \Lambda_{XY}\Lambda_{YY}^{-1}\Lambda_{YZ})\, \Lambda_{Z|Y}^{-1}\, \tilde{z}_{|y}}$$

where $$\Lambda_{Z|Y} = \Lambda_{ZZ} - \Lambda_{ZY}\Lambda_{YY}^{-1}\Lambda_{YZ}$$

The covariance of the conditional PDF $p_{Z|Y}(z, y)$

Derivation of Recursive LS Estimation

1) Define the vector $W = \begin{bmatrix} Z \\ Y \end{bmatrix} \quad \hat{w} = \begin{bmatrix} \hat{z} \\ \hat{y} \end{bmatrix}$

2) Compute $\hat{x}_{|yz} = E\{X | Y = y,\, Z = z\}$

$$\hat{x}_{|yz} = \hat{x} + \Lambda_{XW}\, \underbrace{\Lambda_{WW}^{-1}}_{\text{inversion of an } (p+M) \times (p+M) \text{ matrix}} (w - \hat{w})$$

inversion of an (p+M) × (p+M) matrix

Solution: use Schur complement

- Given

$$\Lambda_{WW} = \begin{bmatrix} \Lambda_{ZZ} & \Lambda_{ZY} \\ \Lambda_{YZ} & \Lambda_{YY} \end{bmatrix} \qquad \text{and} \qquad \Lambda_{YY}^{-1}$$

- Compute the Schur complement of Λ_{YY}

$$\Delta = \Lambda_{ZZ} - \Lambda_{ZY}\Lambda_{YY}^{-1}\Lambda_{YZ}$$

$$= \Lambda_{Z|Y}$$

which is the conditional covariance

Solution: use Schur complement of Λ_{YY}

- Given

$$\Lambda_{WW} = \begin{bmatrix} \Lambda_{ZZ} & \Lambda_{ZY} \\ \Lambda_{YZ} & \Lambda_{YY} \end{bmatrix} \qquad \Lambda_{Z|Y} = \Lambda_{ZZ} - \Lambda_{ZY}\Lambda_{YY}^{-1}\Lambda_{YZ}$$

- Then

$$\Lambda_{WW}^{-1} = \begin{bmatrix} \Lambda_{Z|Y}^{-1} & -\Lambda_{Z|Y}^{-1}F \\ -F^T\Lambda_{Z|Y}^{-1} & \Lambda_{YY}^{-1} + F^T\Lambda_{Z|Y}^{-1}F \end{bmatrix}$$

$$F = \Lambda_{ZY}\Lambda_{YY}^{-1}$$

None-Recursive LS Estimation

$$\hat{x}_{|yz} = \hat{x} + \underbrace{\Lambda_{XW}}_{\begin{bmatrix} \Lambda_{XZ} & \Lambda_{XY} \end{bmatrix}} \underbrace{\Lambda_{WW}^{-1}}_{\begin{bmatrix} \Lambda_{ZZ} & \Lambda_{ZY} \\ \Lambda_{YZ} & \Lambda_{YY} \end{bmatrix}^{-1}} \underbrace{(w - \hat{w})}_{\tilde{w} = \begin{bmatrix} \tilde{z} \\ \tilde{y} \end{bmatrix}}$$

$$W = \begin{bmatrix} Z \\ Y \end{bmatrix}$$

Use Schur complement

$$\hat{x}_{|yz} = \hat{x} + \begin{bmatrix} \Lambda_{XZ} & \Lambda_{XY} \end{bmatrix} \underbrace{\begin{bmatrix} \Lambda_{ZZ} & \Lambda_{ZY} \\ \Lambda_{YZ} & \Lambda_{YY} \end{bmatrix}^{-1}}_{\begin{bmatrix} \Lambda_{Z|Y}^{-1} & -\Lambda_{Z|Y}^{-1}F \\ -F^T\Lambda_{Z|Y}^{-1} & \Lambda_{YY}^{-1} + F^T\Lambda_{Z|Y}^{-1}F \end{bmatrix}} \begin{bmatrix} \tilde{z} \\ \tilde{y} \end{bmatrix}$$

$$\Lambda_{Z|Y} = \Lambda_{ZZ} - \Lambda_{ZY}\Lambda_{YY}^{-1}\Lambda_{YZ} \qquad F = \Lambda_{ZY}\Lambda_{YY}^{-1}$$

Use Schur complement

$$\widehat{x}_{|yz} = \widehat{x} + \Lambda_{XY}\,\Lambda_{YY}^{-1}\widetilde{y}$$

$$+ \; (\Lambda_{XZ} - \Lambda_{XY}\Lambda_{YY}^{-1}\Lambda_{YZ})\,\Lambda_{Z|Y}^{-1}\,(\widetilde{z} - \Lambda_{ZY}\Lambda_{YY}^{-1}\widetilde{y})$$

$$\Lambda_{Z|Y} = \Lambda_{ZZ} - \Lambda_{ZY}\Lambda_{YY}^{-1}\Lambda_{YZ}$$

Use Schur complement

$$\widehat{x}_{|yz} = \underbrace{\widehat{x} + \Lambda_{XY}\,\Lambda_{YY}^{-1}\widetilde{y}}_{\widehat{x}_{|y} \;\longleftarrow\; \textit{expected value of } \boldsymbol{X} \textit{ given outcome } \boldsymbol{y}}$$

$$+ \; (\Lambda_{XZ} - \Lambda_{XY}\Lambda_{YY}^{-1}\Lambda_{YZ})\,\Lambda_{Z|Y}^{-1}\,(\widetilde{z} - \Lambda_{ZY}\Lambda_{YY}^{-1}\widetilde{y})$$

Use Schur complement

We will now show that

$$\widehat{x}_{|yz} = \widehat{x}_{|y}$$

$$+ \; \underbrace{(\Lambda_{XZ} - \Lambda_{XY}\Lambda_{YY}^{-1}\Lambda_{YZ})\,\Lambda_{Z|Y}^{-1}\,(\widetilde{z} - \Lambda_{ZY}\Lambda_{YY}^{-1}\widetilde{y})}_{E\{\widetilde{X}_{|Y} | \widetilde{z}_{|y}\}}$$

The expected value of $\widetilde{X}_{|y}$ given the outcome $\widetilde{z}_{|y}$

Computation of $\widetilde{z}_{|y}$

The conditional mean of $\boldsymbol{Z}$ given $\boldsymbol{Y = y}$:

$$\widehat{z}_{|y} = \widehat{z} + \Lambda_{ZY}\,\Lambda_{YY}^{-1}\,\widetilde{y}$$

$$\widetilde{z}_{|y} = z - \widehat{z}_{|y}$$

$$\widetilde{z}_{|y} = \underbrace{z - \widehat{z}}_{\widetilde{z}} + \Lambda_{ZY}\,\Lambda_{YY}^{-1}\,\widetilde{y}$$

Computation of $E\{\tilde{X}_{|Y}|\tilde{z}_{|y}\}$

Therefore, $$\tilde{z}_{|y} = \tilde{z} + \Lambda_{ZY}\,\Lambda_{YY}^{-1}\,\tilde{y}$$

We will now compute $E\{\tilde{X}_{|Y}|\tilde{z}_{|y}\}$ using the LS result:

$$E\{\tilde{X}_{|Y}|\tilde{z}_{|y}\} = E\{\tilde{X}_{|Y}\} + E\{\tilde{X}_{|Y}\tilde{Z}_{|Y}^T\}E\{\tilde{Z}_{|Y}\tilde{Z}_{|Y}^T\}^{-1}\,\tilde{z}_{|y}$$

to verify that

$$E\{\tilde{X}_{|Y}|\tilde{z}_{|y}\} = (\Lambda_{XZ} - \Lambda_{XY}\Lambda_{YY}^{-1}\Lambda_{YZ})\,\Lambda_{Z|Y}^{-1}\,\underbrace{(\tilde{z} - \Lambda_{ZY}\Lambda_{YY}^{-1}\tilde{y})}_{\tilde{z}_{|y}}$$

Computation of $E\{\tilde{X}_{|Y}|\tilde{z}_{|y}\}$

Using Gaussian least squares results:

$$E\{\tilde{X}_{|Y}|\tilde{z}_{|y}\} = \cancel{E\{\tilde{X}_{|Y}\}}^{\,\boldsymbol{0}}$$

$$+\ E\{\tilde{X}_{|Y}\tilde{Z}_{|Y}^T\}\,E\{\tilde{Z}_{|Y}\tilde{Z}_{|Y}^T\}^{-1}\,\tilde{z}_{|y}$$

Estimation errors always have zero means

Computation of $E\{\tilde{X}_{|Y}|\tilde{z}_{|y}\}$

Using Gaussian least squares results:

$$E\{\tilde{X}_{|Y}|\tilde{z}_{|y}\} = E\{\tilde{X}_{|Y}\tilde{Z}_{|Y}^T\}\,\underbrace{E\{\tilde{Z}_{|Y}\tilde{Z}_{|Y}^T\}}^{-1}\,\tilde{z}_{|y}$$

$$E\{\tilde{Z}_{|Y}\tilde{Z}_{|Y}^T\} = \Lambda_{\tilde{Z}_{|Y}\tilde{Z}_{|Y}} = \Lambda_{Z|Y}$$

$$= \Lambda_{ZZ} - \Lambda_{ZY}\Lambda_{YY}^{-1}\Lambda_{YZ}$$

the conditional covariance

Computation of $E\{\tilde{X}_{|Y}|\tilde{z}_{|y}\}$

Using Gaussian least squares results:

$$E\{\tilde{X}_{|Y}|\tilde{z}_{|y}\} = E\{\tilde{X}_{|Y}\tilde{Z}_{|Y}^T\}\ \Lambda_{Z|Y}^{-1}\,\tilde{z}_{|y}$$

Notice that, from the Schur decomposition result,

$$E\{\tilde{X}_{|Y}|\tilde{z}_{|y}\} = (\Lambda_{XZ} - \Lambda_{XY}\Lambda_{YY}^{-1}\Lambda_{YZ})\,\Lambda_{Z|Y}^{-1}\,\tilde{z}_{|y}$$

Computation of $E\{\tilde{X}_{|Y}|\tilde{z}_{|y}\}$

Using Gaussian least squares results:

$$E\{\tilde{X}_{|Y}|\tilde{z}_{|y}\} = \underbrace{E\{\tilde{X}_{|Y}\tilde{Z}^T_{|Y}\}}\,\Lambda^{-1}_{Z|Y}\,\tilde{z}_{|y}$$

$$E\{(\tilde{X} - \Lambda_{XY}\Lambda^{-1}_{YY}\tilde{Y})\tilde{Z}^T_{|Y}\}$$

$$E\{\tilde{X}\tilde{Z}^T_{|Y}\} + \Lambda_{XY}\Lambda^{-1}_{YY}\cancelto{0}{E\{\tilde{Y}\tilde{Z}^T_{|Y}\}}$$

Computation of $E\{\tilde{X}_{|Y}|\tilde{z}_{|y}\}$

Using Gaussian least squares results:

$$E\{\tilde{X}_{|Y}|\tilde{z}_{|y}\} = \underbrace{E\{\tilde{X}\tilde{Z}^T_{|Y}\}}\,\Lambda^{-1}_{Z|Y}\,\tilde{z}_{|y}$$

$$E\{\tilde{X}\tilde{Z}^T_{|Y}\} = E\{\tilde{X}(\tilde{Z} - \Lambda_{ZY}\Lambda^{-1}_{YY}\tilde{Y})^T\}$$

$$= E\{\tilde{X}\tilde{Z}^T\} - E\{\tilde{X}\tilde{Y}^T\}\Lambda^{-1}_{YY}\Lambda_{YZ}$$

$$= \Lambda_{XZ} - \Lambda_{XY}\Lambda^{-1}_{YY}\Lambda_{YZ}$$

Computation of $E\{\tilde{X}_{|Y}|\tilde{z}_{|y}\}$

Therefore,

$$E\{\tilde{X}_{|Y}\tilde{Z}^T_{|Y}\} = \Lambda_{XZ} - \Lambda_{XY}\Lambda^{-1}_{YY}\Lambda_{YZ}$$

and

$$E\{\tilde{X}_{|Y}|\tilde{z}_{|y}\} = (\Lambda_{XZ} - \Lambda_{XY}\Lambda^{-1}_{YY}\Lambda_{YZ})\,\Lambda^{-1}_{Z|Y}\,\tilde{z}_{|y}$$

QED

Recursive Least Squares

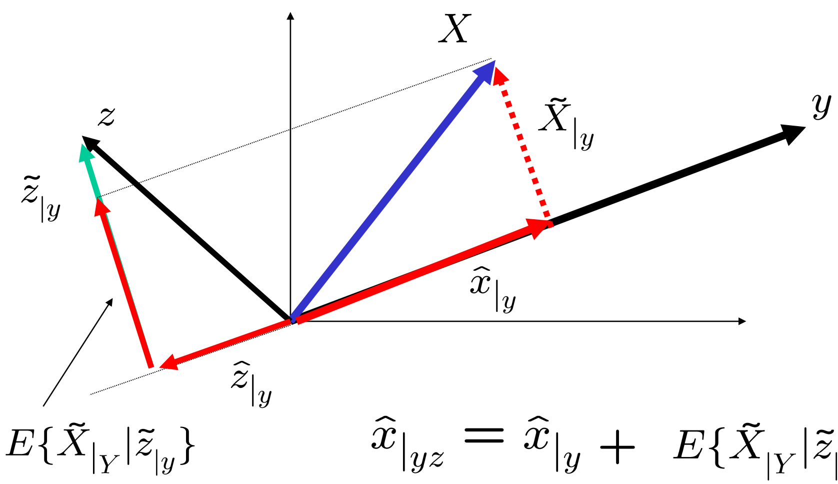
The expected value of $\boldsymbol{X}$ given outcomes $\boldsymbol{y}$ and $\boldsymbol{z}$

$$\widehat{x}_{|yz} = \widehat{x}_{|y} + E\{\tilde{X}_{|Y}|\tilde{z}_{|y}\}$$

The expected value of $\boldsymbol{X}$ given outcome $\boldsymbol{y}$

The expected value of $\tilde{X}_{|y}$ given the outcome $\tilde{z}_{|y}$

Deterministic interpretation of Property 3



$$\hat{x}_{|yz} = \hat{x}_{|y} + E\{\tilde{X}_{|Y}|\tilde{z}_{|y}\}$$

Summary

- The conditional mean is the least squares estimator:

$$E\{\|X - \hat{X}|_Y\|^2\} \leq E\{\|X - f(Y)\|^2\}$$

- For Gaussians, the conditional mean is a linear function

$$\hat{x}|_y = \hat{x} + \Lambda_{XY}\Lambda_{YY}^{-1}(y - \hat{y})$$

Summary

The conditional mean can be computed recursively:

1. If we first know of outcome $\boldsymbol{Y} = \boldsymbol{y}$

$$\hat{x}_{|y} = \hat{x} + \Lambda_{XY}\Lambda_{YY}^{-1}\tilde{y}$$

Summary

The conditional mean can be computed recursively:

2 If we afterwards know of outcome $\boldsymbol{Z} = \boldsymbol{z}$

$$\hat{z}_{|y} = \hat{z} + \Lambda_{ZY}\Lambda_{YY}^{-1}\tilde{y}$$

$$\tilde{z}_{|y} = z - \hat{z}_{|y}$$

and

$$\boxed{\hat{x}_{|yz} = \hat{x}_{|y} + E\{\tilde{X}_{|Y}|\tilde{z}_{|y}\}}$$