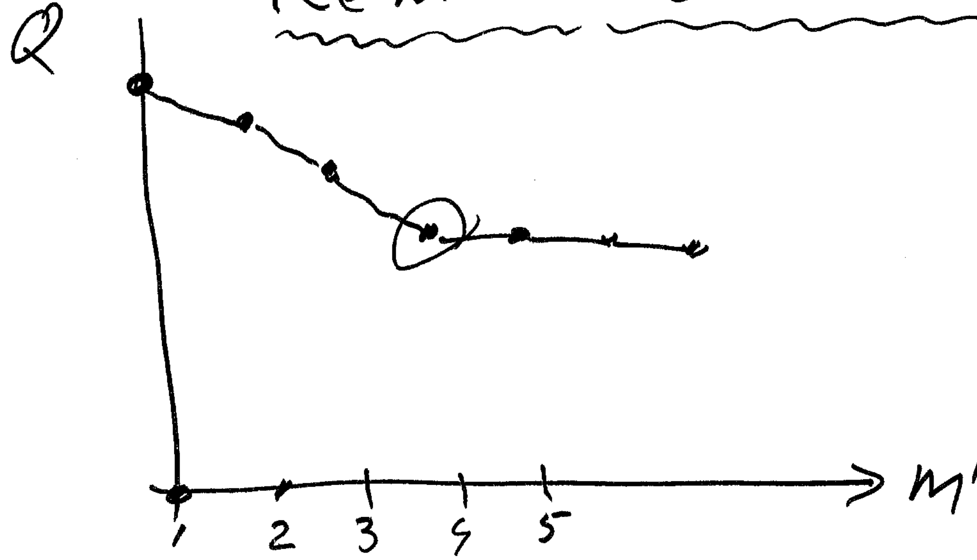
Remark on Linear Regression



$\hat{a}_1 \sim a_1$
$\hat{a}_2 \sim a_2$
$\hat{a}_3 \sim a_3$
$\hat{a}_4 \sim a_4$
$\hat{a}_4 \sim 0$

$f(x) = a + a x + \cdots + a_m x^{m-1}$

m - not known

assume: m'

true $a_1 + a_2 x + a_3 x^2 + a_4 x^3$

assume: $a_1 + a_2 x + a_3 x^2 + a_4 x^3 + a_5 x^4$

Suppose that S - not invertible

i.e., $\exists x \neq 0 : Sx = 0$

$\Rightarrow \langle \underbrace{Sx}_{=0}, x \rangle = 0$ - cont with $S > 0$.

Show that $(S^{-1})^*$ is inverse of S^*:

$$(S^{-1})^* S^* = (\underbrace{S \cdot S^{-1}}_{=I})^* = I^* = I$$

Inner product in ONB: $e_1 \dots e_n$

$x, y \in \mathcal{R}$ $\quad \langle x, y \rangle = \sum_{i=1}^{n} x_i y_i$

$$\langle x, y \rangle = \langle \sum_i x_i e_i , \sum_j y_j e_j \rangle$$

$$= \sum_{ij} x_i y_j \underbrace{\langle e_i , e_j \rangle}_{= \delta_{ij} = \begin{cases} 1 & i=j \\ 0 & i \neq j \end{cases}} = \sum_i x_i y_i$$

---

show $E \langle \nu, x \rangle = \langle E\nu , x \rangle$

$$E \langle \nu, x \rangle = E \sum_i \nu_i x_i = \sum_i (E\nu_i) \cdot x_i$$

$$= \sum_i (E\nu)_i \, x_i = \langle E\nu , x \rangle .$$

ν, μ - indep:

$E \langle \nu, \mu \rangle =$

$$= E \sum_i \nu_i \mu_i = \sum_i E(\nu_i \mu_i)$$

$$= \sum_i E\nu_i \cdot E\mu_i = \langle E\nu, E\mu \rangle$$

$$\nu' = \nu - E\nu \qquad E\nu' = 0$$

$$Sx = E\langle \nu', x\rangle \nu'$$

$$s_{ij} = \langle e_i, Se_j\rangle = \langle e_i, E\langle \nu', e_j\rangle \nu'\rangle$$

$$= E\langle e_i, \langle \nu', e_j\rangle \nu'\rangle = E\langle e_i, \nu'\rangle \cdot \langle e_j, \nu'\rangle$$

$$= E\, \nu_i' \nu_j' = \sum (\nu_i - E\nu_i)(\nu_j - E\nu_j)$$

$$= \mathrm{Cov}(\nu_i, \nu_j)$$

Show $S \geq 0$

$$\forall x \quad \langle Sx, x\rangle = \langle E\langle \nu', x\rangle \nu', x\rangle$$

$$= E\langle \nu', x\rangle \cdot \langle \nu', x\rangle = E\langle \nu', x\rangle^2 \geq 0.$$

$T = \mathrm{Var}(B\nu) \qquad EB\nu = BE\nu \qquad B\nu - EB\nu = B(\nu - E\nu) = B\nu'$

$$\forall x \quad Tx = E\langle B\nu', x\rangle B\nu'$$

$$= EB\langle \nu', B^*x\rangle \nu$$

$$= B\, \underbrace{E\langle \nu', B^*x\rangle \nu'}_{= S(B^*x)} = \underbrace{BSB^*}x$$

$$\Longrightarrow T = BSB^*.$$

$$E\|\nu\|^2 = E\sum_i \nu_i^2 = \sum_i \underbrace{E\nu_i^2}_{= s_{ii}} = \sum_i s_{ii} = \mathrm{tr}\, S.$$

Linear Experiment.

$x \in \mathcal{D}$ – unknown vector

$$y = Ax + \nu$$

$A : \mathcal{D} \to \mathcal{R}$ – linear mapping

$\nu \in \mathcal{R}$ – random vector

$E\nu = 0$ $\quad$ $\mathrm{var}(\nu) = S$

S – variance operator. $S : \mathcal{R} \to \mathcal{R}$

$y \in \mathcal{R}$ – random vector

Experiment: (A, S)

Raw data: (y, A, S)

Examples

(a) $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^2 = \mathcal{D}$

$y_1 = x_1 + v_1$
$y_2 = x_2 + v_2$
$\quad y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \quad v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I$$

$E v_1 = E v_2 = 0 \qquad E v_i^2 = \sigma^2$

v_1 and v_2 – independent

$$S = \mathrm{Var}(v) = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix} = \sigma^2 I$$

Experiment: $(A, S) = (I, \sigma^2 I)$

(b) $x \in \mathbb{R}^2$

$$y_1 = x_1 + v_1$$
$$y_2 = x_2 + v_2$$
$$y_3 = x_1 + x_2 + v_3$$

$y \in \mathbb{R}^3$

$v \in \mathbb{R}^3$ rand ver.

$$y = Ax + v$$

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}$$

v_i – i.i.d.

$S = \sigma^2 I_3$

(c)

$$y_1 = x_1 + x_2 + v_1$$
$$y_2 = x_1 - x_2 + v_2$$

$$A = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \qquad Ax = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 + x_2 \\ x_1 - x_2 \end{bmatrix}$$

$$S = \sigma^2 I$$

$$v_1 = \varepsilon_1 + \varepsilon_0$$
$$v_2 = \underbrace{\varepsilon_2}_{\text{fast}} + \underbrace{\varepsilon_0}_{\text{slow}}$$

$\varepsilon_0, \varepsilon_1, \varepsilon_2$ indep

$\varepsilon_1, \varepsilon_2$ – id distr

$E\varepsilon_i = 0$

$$Var(\varepsilon_1) = Var(\varepsilon_2) = \sigma_1^2$$
$$Var(\varepsilon_0) = \sigma_0^2 \implies Var(v_1) = Var(\varepsilon_1) + Var(\varepsilon_0) = \sigma_1^2 + \sigma_0^2$$

$$S = Var(\nu) = E\begin{bmatrix} \nu_1^2 & \nu_1\nu_2 \\ \nu_2\nu_1 & \nu_2^2 \end{bmatrix}$$

$$= E\begin{bmatrix} (\varepsilon_1+\varepsilon_0)^2 & (\varepsilon_1+\varepsilon_0)(\varepsilon_2+\varepsilon_0) \\ (\varepsilon_2+\varepsilon_0)(\varepsilon_1+\varepsilon_0) & (\varepsilon_2+\varepsilon_0)^2 \end{bmatrix}$$

$$E(\varepsilon_1+\varepsilon_0)^2 = E(\varepsilon_1^2 + 2\varepsilon_1\varepsilon_0 + \varepsilon_0^2)$$

$$= \sigma_1^2 + 2\underbrace{E\varepsilon_1}_{=0}\cdot\underbrace{E\varepsilon_0}_{=0} + \sigma_0^2 = \sigma_1^2 + \sigma_0^2$$

$$E(\varepsilon_1+\varepsilon_0)(\varepsilon_2+\varepsilon_0) = E(\varepsilon_1\varepsilon_2 + \varepsilon_1\varepsilon_0 + \varepsilon_0\varepsilon_2 + \varepsilon_0\varepsilon_0)$$

$$= E\varepsilon_0^2 = \sigma_0^2$$

$$S = \begin{bmatrix} \sigma_1^2+\sigma_0^2 & \sigma_0^2 \\ \sigma_0^2 & \sigma_1^2+\sigma_0^2 \end{bmatrix}$$

correlation between ν_1 and ν_2

$$r = \frac{cov(\nu_1, \nu_2)}{\sqrt{Var(\nu_1)\cdot Var(\nu_2)}} = \frac{\sigma_0^2}{\sigma_1^2+\sigma_0^2}$$

$0 \le r \le 1 \qquad \sigma^2 = \sigma_1^2 + \sigma_2^2$

$$S = \sigma^2\begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix}$$

(d) A: same as in (a) but with correlated noise

$$A = I_2 \qquad S = \sigma^2 \begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix}$$

(e) A: same as in (c)

$$A = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \qquad S = \sigma^2 \begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix}$$

Optimal Estimation

$$y = Ax + \nu$$

$x \in \mathcal{D}$ - unknown

$y \in \mathcal{R}$ - observation.

$\hat{x} = ?$ estimate of x

Find linear transf. $R : \mathcal{R} \to \mathcal{D}$

$$\hat{x} = Ry$$

$$\hat{x} - x = Ry - x = R(Ax + \nu) - x$$

$$= \underbrace{(RA - I)x}_{\substack{\text{Systematic} \\ \text{Error.} \\ \Rightarrow \text{Bias.}}} + \underbrace{R\nu}_{\substack{\text{Random} \\ ER\nu = 0}}$$

$$E\|Ry - x\|^2 = E\langle (RA - I)x + R\nu, (RA - I)x + R\nu \rangle$$

$$= E\left[\|(RA - I)x\|^2 + 2\langle (RA - I)x, R\nu \rangle + \|R\nu\|^2\right]$$

$$= \|(RA - I)x\|^2 + \operatorname{tr}\left(\underbrace{\operatorname{Var}(R\nu)}_{= RSR^*}\right)$$

$$= \|(RA - I)x\|^2 + \operatorname{tr}(RSR^*)$$

$$H(R) = \sup_{x \in \mathcal{D}} E\|Ry - x\|^2$$

$$= \sup_x \left[\|(RA - I)x\|^2 + \mathrm{tr}\, RSR^*\right]$$

$$= \begin{cases} +\infty & \text{if } RA \neq I \\ \mathrm{tr}(RSR^*) & \text{if } RA = I \end{cases}$$

Constraint $RA = I$

$\Rightarrow \hat{x} = Ry$ – unbiased

$$Ry - x = (RA - I)x + Rv$$

$$\mathbb{E}(Ry - x) = (RA - I)x = 0$$

iff $\underline{RA - I = 0}$

<u>Optimization problem</u>!

$$\mathrm{tr}(RSR^*) \sim \min_R \quad : RA = I$$

minimize RSR^* ?

$$\mathrm{Var}(Rv) = RSR^* \sim \min_R \; : \; RA = I$$

Show that

$$R = (A^* S^{-1} A)^{-1} A^* S^{-1}$$

Assumptions:

* S - inventible.

$$v = \begin{bmatrix} v_1 \\ 0 \end{bmatrix} \quad E v_1 = 0 \quad \text{when } E v_1^2 = \sigma^2 \quad \mathrm{Var}\, v = \begin{bmatrix} \sigma^2 & 0 \\ 0 & 0 \end{bmatrix}$$

* $A^* S^{-1} A$ - inventible.

$$A = {}_{n}\Big[\;\Big]_{m} \qquad n \geq m$$

columns of A are independent.

$$RA = (A^* S^{-1} A)^{-1} A^* S^{-1} A = I$$

constraint satisfied.

$\bar{R} = R + D$ (disturbed).:

$$\bar{R}A = I \Rightarrow (R+D)A = I$$

$$RA + DA = I \Rightarrow \underline{DA = 0}$$

$$\mathrm{Var}(\bar{R}v) = \bar{R} S \bar{R}^*$$

$$= \left[(A^* S^{-1} A)^{-1} A^* S^{-1} + D \right] S \cdot$$

$$\left[S^{-1} A \underbrace{(A^* S^{-1} A)}_{}^{-1} + D^* \right]$$

$$= (\,)^{-1} A^* S^{-1} \underbrace{S S^{-1}}_{=I} A (\,)^{-1} +$$

$$+ (\,)^{-1} A^* \underbrace{S^{-1} S}_{=I} D^* + D S S^{-1} A (\,)^{-1} + D S D^*$$

$$= (A^* S^{-1} A)^{-1} + (A^* S^{-1} A)^{-1} \underbrace{A^* D^*}_{=0} + \underbrace{DA}_{=0} (\,)^{-1} + DSD^*$$

$$= (A^* S^{-1} A)^{-1} + \underbrace{D S D^*}_{\geq 0} \sim \min_{D}$$

if $D = 0$

$$\Rightarrow R = (A^* S^{-1} A)^{-1} A^* S^{-1}$$

provides min to

$$H(R) = \sup_{x} E \| R y - x \|^2$$

$\hat{x} = R y$

Best (minimizes var of error)
Linear
Unbiased
Estimate

BLUE

$$R = (A^* S^{-1} A)^{-1} A^* S^{-1}$$

$$\mathrm{Var}(\hat{x}) = \mathrm{Var}\, Ry = (A^* S^{-1} A)^{-1}$$

$$E\|Ry - x\|^2 = \mathrm{tr}(A^* S^{-1} A)^{-1}$$

Examples

(a) $A = I_2 \qquad S = \sigma^2 I$

$$R = \left(I \cdot (\sigma^2 \cdot I)^{-1} \cdot I\right)^{-1} I \cdot (\sigma^2 I)^{-1}$$

$$= (\sigma^{-2} I)^{-1} \cdot \sigma^{-2} = \sigma^2 \cdot \sigma^{-2} \cdot I = I$$

$y_1 = x_1 + v_1$

$y_2 = x_2 + v_2$

$\mathrm{Var}\, \hat{x} = \sigma^2 I \implies \mathrm{Var}\, \hat{x}_1 = \mathrm{Var}\, \hat{x}_2 = \sigma^2$

(b) $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} \qquad S = \sigma^2 I_3$

$$R = \Big(\underbrace{\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}}_{A^*} \underbrace{(\sigma^2 I)^{-1}}_{S^{-1}} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} \Big)^{-1} A^* (\sigma^2 I)^{-1}$$

$$= \sigma^2 \sigma^{-2} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

$$= \frac{1}{3} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 2 & -1 & 1 \\ -1 & 2 & 1 \end{bmatrix}$$

$$\hat{x} = Ry = \frac{1}{3} \begin{bmatrix} 2 & -1 & 1 \\ -1 & 2 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

$$= \frac{1}{3} \begin{bmatrix} 2y_1 - y_2 + y_3 \\ -y_1 + 2y_2 + y_3 \end{bmatrix}$$

$$\mathrm{Var}(\hat{x}) = \frac{\sigma^2}{3} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} = \begin{bmatrix} \frac{2}{3}\sigma^2 & -\frac{\sigma^2}{3} \\ -\frac{\sigma^2}{3} & \frac{2}{3}\sigma^2 \end{bmatrix}$$

$$\mathrm{Var}(\hat{x}_1) = \mathrm{Var}(\hat{x}_2) = \frac{2}{3}\sigma^2 < \sigma^2$$

(c) $A = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ $\quad S = \sigma^2 I$

$$\mathrm{Var}(\hat{x}) = (A^* S^{-1} A)^{-1}$$

$$= \sigma^2 \left(\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \right)^{-1}$$

$$= \sigma^2 \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}^{-1} = \sigma^2 \begin{bmatrix} 1/2 & 0 \\ 0 & 1/2 \end{bmatrix} = \begin{bmatrix} \frac{\sigma^2}{2} & 0 \\ 0 & \frac{\sigma^2}{2} \end{bmatrix}$$

$$\mathrm{Var}(\hat{x}_1) = \mathrm{Var}(\hat{x}_2) = \underline{\underline{\frac{\sigma^2}{2} < \sigma^2}}$$

$$R = \frac{\sigma^2}{2} I \cdot A^* (\sigma^2 I)^{-1} = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

$$\hat{x} = R \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} y_1 + y_2 \\ y_1 - y_2 \end{bmatrix}.$$