$$\Pi_X^2 = \Pi_X$$

$$\mathrm{rank}\ \Pi_X = \mathrm{tr}\ \Pi_X = \mathrm{tr}\ \{X(X'X)^{-1}X'\} = p$$

さて、

$$\min\ Q = Q(\widehat{\boldsymbol{\beta}}) = ||\boldsymbol{e}||^2 = \boldsymbol{y}'(I - \Pi_X)\boldsymbol{y}$$ 証明終

[注意]　定数項がある場合、Prop.1.2.4 は、Prop.1.2.1 と同じ。
2 つの問題の解と最小値は、一見違っているが、実は同じであることが確かめられる。

Def.1.2.2.

$\widehat{\boldsymbol{y}} := X\widehat{\boldsymbol{\beta}} = \Pi_X \boldsymbol{y}$：回帰ベクトル

$\boldsymbol{e} := \boldsymbol{y} - \widehat{\boldsymbol{y}} = (I - \Pi_X)\boldsymbol{y}$：残差ベクトル

$\widehat{\boldsymbol{\beta}} := (X'X)^{-1}X'\boldsymbol{y}$：最小二乗解

幾何学的意味：データ $\boldsymbol{y}$ を、回帰空間（$X\boldsymbol{\beta}$ の作る空間）$M(X)$ に射影する。Π_X は、この回帰空間への射影であり、$I - \Pi_X$ は $M(X)^{\perp}$（誤差空間）への射影である。

Prop.1.2.5.　$||\widehat{\boldsymbol{y}}||^2 = \boldsymbol{y}'\Pi_X\boldsymbol{y}$

$$||\boldsymbol{e}||^2 = \boldsymbol{y}'(I - \Pi_X)\boldsymbol{y}$$

$$\widehat{\boldsymbol{y}}'\boldsymbol{e} = 0$$

[注意]：　$\boldsymbol{y} = \widehat{\boldsymbol{y}} + \boldsymbol{e}$

長さ　$||\boldsymbol{y}||^2 = \boldsymbol{y}'\Pi_X\boldsymbol{y} + \boldsymbol{y}'{\Pi_X}^{\perp}\boldsymbol{y}$

ただし、${\Pi_X}^{\perp} := I - \Pi_X$

1.3. 母数モデルの回帰分析

確率を導入する。

$$y_i = \boldsymbol{\beta}'\boldsymbol{x_i} + \varepsilon_i, \quad (i = 1, \cdots, n)$$

ここで、$\varepsilon_1, \cdots, \varepsilon_n$ は r.v.（確率変数）で **誤差** と呼ばれる。
$E(\varepsilon_i) = 0, \quad E(\varepsilon_i \varepsilon_j) = \sigma^2 \delta_{ij}, \quad (i, j = 1, \cdots, n)$
$\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n$: given , $\quad \sigma > 0$

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$E(\boldsymbol{\varepsilon}) = \boldsymbol{0}, \quad \mathrm{Var}(\boldsymbol{\varepsilon}) = E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \sigma^2 I_n$$

X : known $\qquad$ (rank $X = p$)

$\boldsymbol{\beta}, \ \sigma^2$: unknown

問題 $\quad \boldsymbol{\beta}, \ \sigma^2$ に関する推測

Def.1.3.1. $\quad \widehat{\boldsymbol{\beta}}, \ \widehat{\boldsymbol{y}}, \ \boldsymbol{e}$: 1.2 節の Def.1.2.2. の通り

$\widehat{\boldsymbol{y}} := X\widehat{\boldsymbol{\beta}} = \Pi_X \boldsymbol{y}$: 回帰ベクトル

$\boldsymbol{e} := \boldsymbol{y} - \widehat{\boldsymbol{y}} = (I - \Pi_X)\boldsymbol{y}$: 残差ベクトル

$\widehat{\boldsymbol{\beta}} := (X'X)^{-1}X'\boldsymbol{y}$: 最小二乗解

$\widehat{\sigma}^2 := Q(\widehat{\boldsymbol{\beta}})/(n-p)$

Prop.1.3.1.

(i) $\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (X'X)^{-1}X'\boldsymbol{\varepsilon}$

$$E(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta}, \quad \mathrm{Var}(\widehat{\boldsymbol{\beta}}) = \sigma^2 (X'X)^{-1}$$

(ii) $\boldsymbol{e} = (I - \Pi_X)\boldsymbol{\varepsilon}$

$$E(\boldsymbol{e}) = \boldsymbol{0}, \quad \mathrm{Var}(\boldsymbol{e}) = \sigma^2 (I - \Pi_X)$$

(iii) $\mathrm{Cov}(\widehat{\boldsymbol{\beta}}, \boldsymbol{e}) = O$

(iv) $(n-p)\widehat{\sigma}^2 = ||e||^2 = \boldsymbol{\varepsilon}'(I - \Pi_X)\boldsymbol{\varepsilon}$

$$E(\widehat{\sigma}^2) = \sigma^2$$

補足 : (i) と (iv) より、$\widehat{\boldsymbol{\beta}}, \ \widehat{\sigma}^2$ は $\boldsymbol{\beta}, \ \sigma^2$ の不偏推定量であることがわかる。

証明 : (i) $\widehat{\boldsymbol{\beta}} = (X'X)^{-1}X'\boldsymbol{y} = (X'X)^{-1}X'(X\boldsymbol{\beta} + \boldsymbol{\varepsilon})$

$$= \boldsymbol{\beta} + (X'X)^{-1}X'\boldsymbol{\varepsilon}$$

ゆえに、$E(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$

$$\mathrm{Var}(\widehat{\boldsymbol{\beta}}) = (X'X)^{-1}X' \ \mathrm{Var}(\boldsymbol{\varepsilon}) \ X(X'X)^{-1} = \sigma^2 (X'X)^{-1}$$

(ii) Prop.1.2.4 の証明において、$H := I - \Pi_X$ とすれば、

$$\boldsymbol{e} = H\boldsymbol{y} = H(X\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = H\boldsymbol{\varepsilon}$$

ゆえに、$E(\boldsymbol{e}) = \boldsymbol{0}$

$$\mathrm{Var}(\boldsymbol{e}) = H\ \mathrm{Var}(\boldsymbol{\varepsilon})\ H = \sigma^2 H^2 = \sigma^2 H$$

(iii) $\mathrm{Cov}(\widehat{\boldsymbol{\beta}}, \boldsymbol{e}) = E\left\{(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\boldsymbol{e}'\right\} = (X'X)^{-1}X'\ E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')\ H = \sigma^2(X'X)^{-1}X'H = O$

(iv) $(n-p)\widehat{\sigma}^2 = Q(\widehat{\boldsymbol{\beta}}) = ||e||^2 = \boldsymbol{\varepsilon}'H\boldsymbol{\varepsilon}$

$$E(\boldsymbol{\varepsilon}'H\boldsymbol{\varepsilon}) = \mathrm{tr}\ E(H\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}) = \sigma^2 \mathrm{tr}\ H = (n-p)\sigma^2$$

ゆえに、$E(\widehat{\sigma}^2) = \sigma^2$　　証明終

Def.1.3.2.　$\boldsymbol{c} : p \times 1$

(i) $L = \boldsymbol{a}'\boldsymbol{y},\ \ \boldsymbol{a} : n \times 1$ が、

$$E(L|\boldsymbol{\beta}) = \boldsymbol{c}'\boldsymbol{\beta},\ \ \forall\boldsymbol{\beta}$$

を満たすとき、$\boldsymbol{c}'\boldsymbol{\beta}$ の 線形不偏推定量 という。

(ii) L_0 が、$\boldsymbol{c}'\boldsymbol{\beta}$ の **最良線形不偏推定量 (BLUE)** とは、
線形不偏推定量の中で、分散最小　であること。

Prop.1.3.2.(Gauss-Markov)　$\forall\boldsymbol{c} : n \times 1$　$\boldsymbol{c}'\widehat{\boldsymbol{\beta}}$ は $\boldsymbol{c}'\boldsymbol{\beta}$ の BLUE である。

証明：(i) **linear**　$L_0 := \boldsymbol{c}'\widehat{\boldsymbol{\beta}} = \boldsymbol{c}'(X'X)^{-1}X'\boldsymbol{y}$

(ii) 不偏　$E(L_0|\boldsymbol{\beta}) = E(\boldsymbol{c}'\widehat{\boldsymbol{\beta}}|\boldsymbol{\beta}) = \boldsymbol{c}'\boldsymbol{\beta},\ \ \forall\boldsymbol{\beta}$

(iii) 最小分散　問題 $\begin{cases} \min\limits_{\boldsymbol{a}}\ \mathrm{Var}\ L \\ 条件 \begin{cases} L = \boldsymbol{a}'\boldsymbol{y},\ \ \boldsymbol{a} : n \times 1 \\ E(L|\boldsymbol{\beta}) = \boldsymbol{c}'\boldsymbol{\beta},\ \ \forall\boldsymbol{\beta} \end{cases} \end{cases}$

を解いた解が、$\boldsymbol{c}'\widehat{\boldsymbol{\beta}}$ であることを言えばよい。

$$E(L|\boldsymbol{\beta}) = \boldsymbol{a}'E(\boldsymbol{y}|\boldsymbol{\beta}) = \boldsymbol{a}'X\boldsymbol{\beta}$$

$$L : 不偏 \quad \Longleftrightarrow \quad \boldsymbol{a}'X = \boldsymbol{c}' \quad \Longleftrightarrow \quad X'\boldsymbol{a} = \boldsymbol{c}$$

$$\mathrm{Var}(L) = \boldsymbol{a}'\ \mathrm{Var}(\boldsymbol{y})\ \boldsymbol{a} = \sigma^2\boldsymbol{a}'\boldsymbol{a}$$

問題 $\begin{cases} \min\limits_{\boldsymbol{a}}\ ||\boldsymbol{a}|| \\ 条件：X'\boldsymbol{a} = \boldsymbol{c} \end{cases}$

Prop.B1 より、

解：$\boldsymbol{a} = X(X'X)^{-1}\boldsymbol{c}$

ゆえに、$L = \boldsymbol{a}'\boldsymbol{y} = \boldsymbol{c}'(X'X)^{-1}X'\boldsymbol{y} = \boldsymbol{c}'\widehat{\boldsymbol{\beta}}$　　証明終

以後、この節では、

$$\varepsilon_i \quad i.i.d \quad N(0, \sigma^2) \quad \text{とする。}$$

Prop.1.3.3. $\boldsymbol{\beta},\ \sigma^2$ の MLE(最尤推定量) は、それぞれ、$\widehat{\boldsymbol{\beta}},\ \dfrac{1}{n}||e||^2 =: \widehat{\sigma}^2_M$ である。

証明：$$L(\boldsymbol{\beta},\ \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\ \sigma} \exp\left\{ -\frac{1}{2\sigma^2}(y_i - \boldsymbol{\beta}'\boldsymbol{x_i})^2 \right\}$$

$$= K\sigma^{-n} \exp\left\{ -\frac{1}{2\sigma^2} Q(\boldsymbol{\beta}) \right\} \qquad K := (2\pi)^{-\frac{n}{2}}$$

これを、$\boldsymbol{\beta}$ について最大にするには、$Q(\boldsymbol{\beta})$ を $\boldsymbol{\beta}$ について最小にすることであり、Prop.1.2.4 より、$\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}$

このとき、$L(\widehat{\boldsymbol{\beta}},\ \sigma^2) = K\sigma^{-n} \exp\left\{ -\dfrac{1}{2\sigma^2}||e||^2 \right\}$

これを、σ^2 について最大にするには、$t = \sigma^2$ と置いて（対数を取り）微分すれば分かる。

$$\sigma^2 = \frac{1}{n}||e||^2$$

証明終

[注意]：尤度関数の最大値 $= (2\pi)^{-\frac{n}{2}} \widehat{\sigma}_M^{-n} \exp(-\frac{n}{2})$

補足：最尤推定量, MLE(Maximum Likelihood Estimator) について

密度関数 $f(x;\theta)$ は、通常、母数 θ が定数で与えられており、x の起こりやすさを表していると解釈される。この発想を逆転して、この関数を、x が得られたとき、母数 θ の関数とみて、$L(\theta) = L(\theta; x) = f(x;\theta)$ とおくとき、この関数 $L(\theta)$ を **尤度関数** (Likelihood Function) という。また、この関数が最大になる θ の値を、尤度を最大にする推定量（**最尤推定量**）という。このような方法を、**最尤法**という。

（後で行う、**尤度比検定** にも繋がる考え方である。）

Prop.1.3.4. (i) $\widehat{\boldsymbol{\beta}} \ \sim\ N_p(\boldsymbol{\beta}, \sigma^2(X'X)^{-1})$

(ii) $\boldsymbol{e} \ \sim\ N_n(\boldsymbol{0}, \sigma^2(I - \Pi_X))$

(iii) $\widehat{\boldsymbol{\beta}} \perp\!\!\!\perp \boldsymbol{e}$

(iv) $(n-p)\widehat{\sigma}^2 \ \sim\ \sigma^2 \chi^2_{n-p}$

(v) $\widehat{\boldsymbol{\beta}} \perp\!\!\!\perp \widehat{\sigma}^2$

証明：Prop.1.3.1 より、

$$\begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \boldsymbol{e} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{0} \end{bmatrix} + \begin{bmatrix} (X'X)^{-1}X' \\ I - \Pi_X \end{bmatrix} \boldsymbol{\varepsilon}$$

Prop.C14 より、

$$\begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \boldsymbol{e} \end{bmatrix} \quad \sim \quad N_{p+n}\left(\begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{0} \end{bmatrix}, \Phi \right)$$

ここで、$\Phi = \begin{bmatrix} (X'X)^{-1}X' \\ I - \Pi_X \end{bmatrix} \mathrm{Var}(\boldsymbol{\varepsilon}) \begin{bmatrix} (X'X)^{-1}X' \\ I - \Pi_X \end{bmatrix}' = \sigma^2 \begin{bmatrix} (X'X)^{-1} & O \\ O & I - \Pi_X \end{bmatrix}$