DEVOIR MAISON 1 - MAP 433

Septembre 2022

Artur César Araújo Alves
João Pedro Tanaka Montalvão
Rafael Ryoma Nagai Matsutane

SOMMAIRE

1

ESTIMATION PAR MOINDRES CARRÉS DU VECTEUR β

1.1 Exercice 1

Montrons que $\arg\min_{u\in\mathbb{R}^p} J_n(u)$ est solution de $Z^TY = Z^TZu$.

Notons P la projection orthogonale de Y sur $F = Vect\left(Z^{(1)}, Z^{(2)}, \cdots, Z^{(n)}\right)$ et $\hat{u}$ le vecteur tel que $PY = Z\hat{u}$.

— Soit $u = \arg\min_{u\in\mathbb{R}^p} J_n(u)$, on peut montrer que $Zu = Z\hat{u}$:

$$J_n(u) = ||Y - Zu||^2 = ||Y - Z\hat{u} - Z(u-\hat{u})||^2$$

Comme $Y - Z\hat{u}$ est orthogonal à $Z(u-\hat{u})$:

$$||Y - Z\hat{u} - Z(u-\hat{u})||^2 = ||Y - Z\hat{u}||^2 + ||Z(u-\hat{u})||^2$$

Alors :

$$J_n(u) = J_n(\hat{u}) \Leftrightarrow Zu = Z\hat{u}$$

— Comme Z est orthogonal à $Y - Z\hat{u}$, on a que Z^T aussi est orthogonal à $Y - Z\hat{u}$ et alors :

$$Z^T(Y - Z\hat{u}) = 0 \Leftrightarrow Z^TY = Z^TZ\hat{u}$$

Donc, $Z^TY = Z^TZ\hat{u}$

Finalement, $Z\hat{u} = ZU$ pour $u = \arg\min_{u\in\mathbb{R}^p} J_n(u)$ nous donne le résultat.

1.2 Exercice 2

— $Z^{\#}Z = (Z^TZ)^{-1}Z^TZ = I_p$
— $ZZ^{\#} = H$ est le projecteur orthogonal sur F si et seulement si :

$$\forall v,\, u : \;\; \langle v - Hv,\, Zu\rangle = 0$$

Ce qui est verifié :

$$\langle v - Hv,\, Zw\rangle = u^TZ^T(v - Z(Z^TZ)^{-1}Z^Tv) = u^TZ^Tv = u^TZ^TZ(Z^TZ)^{-1}Z^Tv$$

En utilisant que $(Z^TZ)^{-1}Z^TZ = I_p$, on obtient que :

$$\langle v - Hv,\, Zw\rangle = u^TZ^Tv - u^TZ^Tv = 0$$

1.3 Exercice 3

On cherche directement des minimuns locaux de $J_n(u)$:

$$J_n(u) = ||Y^T - u^TZ^T|| = (Y^T - u^TZ^T)(Y - Zu) = ||Y||^2 - Y^TZu - u^TZ^TY + u^TZ^TZu$$

D'où $\nabla J_n(\hat{u}) = -(Y^TZ)^T - (Z^TY) + 2Z^TZu = -2Z^TY + 2Z^TZu$, et alors :

$$\nabla J_n\hat{(}u) = 0 \Leftrightarrow \hat{u} = (Z^TZ)^{-1}Z^TY = Z^{\#}Y$$

Alors, $Z^{\#}Y$ est l'unique point critique. En plus, $H_{J_n(u)} = 2Z^TZ$ qui est une matrice positive definite. Donc, $\hat{u}$ est un minimun local.

1.4 Exercice 4

Montrons que $\mathbb{E}_\theta[\hat{\beta}] = \beta$:

$$\mathbb{E}_\theta[\hat{\beta}] = \mathbb{E}_\theta[Z^{\#}Y] = Z^{\#}\mathbb{E}_\theta[Y] = Z^{\#}(Z\beta) = I_p\beta = \beta$$

1.5 Exercice 5

Montrons que $Var_\theta(\hat{\beta}) = \sigma^2(Z^TZ)^{-1}$:

$$Var_\theta(Z^{\#}Y) = Z^{\#}Var_\theta(Y)(Z^{\#})^T = \sigma^2Z^{\#}(Z^{\#})^T$$

$$Var_\theta(Z^{\#}Y) = \sigma^2(Z^TZ)^{-1}Z^T((Z^TZ)^{-1}Z^T)^T = \sigma^2(Z^TZ)^{-1}Z^TZ(Z^TZ)^{-1}$$

En utilisant que $(Z^TZ)^{-1}Z^TZ = I_p$:

$$Var_\theta(Z^{\#}Y) = \sigma^2I_p(Z^TZ)^{-1} = \sigma^2(Z^TZ)^{-1}$$

1.6 Exercice 6

$$\mathbb{E}_\theta[\tilde{\beta}] = \mathbb{E}_\theta[BY] = B\mathbb{E}_\theta[Y]BZ\beta \Leftrightarrow BZ - I_p = 0$$

La dernière equivalence peut être vue en choisissant $\beta = e_i$, car ça montre que la colonne i de $BZ - I_p$ vaut bien zéro.

1.7 Exercice 7

$$\mathbb{E}_\theta[(\tilde{\beta} - \beta)(\hat{\beta}^T - \beta^T)] = \mathbb{E}_\theta[\tilde{\beta}\hat{\beta}^T] + \beta\beta^T - \beta\mathbb{E}_\theta[\hat{\beta}^T] - \mathbb{E}_\theta[\tilde{\beta}]\beta^T$$

En utilisant que $\mathbb{E}_\theta[\hat{\beta}^T] = \beta^T$ et $\mathbb{E}_\theta[\tilde{\beta}] = \beta$, on obtient que :

$$\mathbb{E}_\theta[(\tilde{\beta} - \beta)(\hat{\beta}^T - \beta^T)] = \mathbb{E}_\theta[(BY)(Y^TZ(Z^TZ)^{-1})] - \beta\beta^T = B\mathbb{E}_\theta[YY^T]Z(Z^TZ)^{-1} - \beta\beta^T$$

Or, $\mathbb{E}_\theta[YY^T] = Var_\theta(Y) + \mathbb{E}_\theta[Y](\mathbb{E}_\theta[Y])^T = \sigma^2 I_n + Z\beta\beta^T Z^T$.

Alors :

$$\mathbb{E}_\theta[(\tilde{\beta} - \beta)(\hat{\beta}^T - \beta^T)] = B(\sigma^2 I_n + Z\beta\beta^T Z^T)Z(Z^TZ)^{-1} - \beta\beta^T$$

$$\mathbb{E}_\theta[(\tilde{\beta} - \beta)(\hat{\beta}^T - \beta^T)] = \sigma^2 BZ(Z^TZ)^{-1} + BZ\beta\beta^T Z^T Z(Z^TZ)^{-1} - \beta\beta^T = \sigma^2(Z^TZ)^{-1}$$

Où on a utilisé que $BZ = I_p$.

1.8 Exercice 8

Calculons d'abord les deux matrices de covariance :

- $Var_\theta(\tilde{\beta}) = \mathbb{E}_\theta[\tilde{\beta}\tilde{\beta}^T] - \mathbb{E}_\theta[\tilde{\beta}](\mathbb{E}_\theta[\tilde{\beta}])^T = \mathbb{E}_\theta[BYY^TB^T] - \beta\beta^T =$
  $= B(\sigma^2 I_n + Z\beta\beta^T Z^T)B^T - \beta\beta^T = \sigma^2 BB^T + BZ\beta\beta^T Z^T B^T - \beta\beta^T = \sigma^2 BB^T$
- $Var_\theta(\hat{\beta}) = \sigma^2(Z^TZ)^{-1}$

Ainsi, $Var_\theta(\tilde{\beta}) - Var_\theta(\hat{\beta}) = \sigma^2(BB^T - (Z^TZ)^{-1})$ et si on pose $N = B - Z^\#$, on obtient que $\sigma^2 NN^T = \sigma^2(BB^T - (Z^TZ)^{-1})$.

Puisque $\sigma^2 NN^T$ est une matrice positive semi-definite, on voit bien que, pour $x \in \mathbb{R}^p$

$$X^T\sigma^2NN^TX \geq 0 \Leftrightarrow X^T(Var_\theta(\tilde{\beta}) - Var_\theta(\hat{\beta})) \geq 0 \Leftrightarrow X^TVar_\theta(\tilde{\beta})X \geq X^TVar_\theta(\hat{\beta})X$$

Et, alors, on a que $Var_\theta(\tilde{\beta}) \succeq Var_\theta(\hat{\beta})$.

2
ESTIMATION DE LA VARIANCE σ^2 ET COEFFICIENT DE DÉTERMINATION

2.1 Exercice 9

Calculons l'espérance de $SSE = ||Y - HY||^2$:

$$\mathbb{E}_\theta[||Y - HY||^2] = \mathbb{E}_\theta[||(I-H)Y||] = \mathbb{E}_\theta[Tr\left((I-H)YY^T(I-H)^T\right)]$$

En utilisant que $\mathbb{E}[Tr(X)] = Tr(\mathbb{E}[X])$, on a que :

$$\mathbb{E}_\theta[Tr\left((I-H)YY^T(I-H)^T\right)] = Tr\left(\mathbb{E}_\theta[(I-H)YY^T(I-H)^T]\right) = Tr\left((I-H)\mathbb{E}_\theta[YY^T](I-H)^T\right)$$

En utilisant que $\mathbb{E}_\theta[YY^T] = \sigma^2 I + Z\hat{\beta}(\hat{\beta}Z)^T$ et le fait que Z est orthogonal à $(I-H)$, on a que :

$$Tr\left((I-H)\mathbb{E}_\theta[YY^T](I-H)^T\right) = Tr\left((I-H)(\sigma^2 I + Z\hat{\beta}(\hat{\beta}Z)^T)(I-H)^T\right) =$$

$$= Tr\left(\sigma^2(I-H)(I-H)^T\right) = \sigma^2 Tr\left((I-H)(I-H)^T\right)$$

Comme $A = I - H$ est une projection orthogonale, on a que $A = A^T$; de plus, como A est une projection, on a que $A^2 = A$ et alors :

$$\sigma^2 Tr\left((I-H)(I-H)^T\right) = \sigma^2 Tr[(I-H)^2] = \sigma^2 Tr(I-H)$$

Finalement, si A est une projection orthogonale, alors $Tr(A) = rank(A)$; comme $rank(H) = p$, on a que $rank(I_H) = n - p$ et alors :

$$\sigma^2 Tr(I - H) = \sigma^2 rank(I - H) = \sigma^2(n - p)$$

On peut conclure alors que, comme $\mathbb{E}_\theta[SSE] = \sigma^2(n - p)$, alors :

$$\mathbb{E}_\theta[\hat{\sigma}^2] = \mathbb{E}_\theta[(n - p)^{-1} SSE] = \sigma^2 \Rightarrow \mathbb{E}_\theta[\hat{\sigma}^2] = \sigma^2$$

2.2 Exercice 10

On utilise ici que $Y - HY$ est orthogonal à HY, car H est une projection orthogonale. Alors, on a que :

$$||Y||^2 = ||(Y - HY) + HY||^2 = ||Y - HY||^2 + ||HY||^2 = SSE + RSS$$

3
CAS DE LA RÉGRESSION LINÉAIRE GAUSSIENNE

3.1 Exercice 11

On calcule la vraisemblance en faisant le produit des densités de probabilité pour chaque y_i de l'observation, où l'on note u_i la moyenne correspondante :

$$L(\theta, Y) = \prod_{i=1}^{n} f(y_i) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y_i - u_i)^2}{2\sigma^2}\right)$$

D'où, la log-vraisemblance $l(\theta, Y) = n^{-1} log L(\theta, Y)$:

$$l(\theta, Y) = -n^{-1} \sum_{i=1}^{n} \left(-\frac{(y_i - u_i)^2}{2\sigma^2}\right) - log(\sigma\sqrt{2\pi}) = -\frac{||Y - Z\beta||^2}{2\sigma^2 n} - log\sigma - log\sqrt{2\pi}$$

On calcule la dérivé par rapport à β et σ :

$$\frac{\partial l(\theta, Y)}{\partial \beta} = -\frac{2Z^T(Z\beta - Y)}{2\sigma^2 n}$$

$$\frac{\partial l(\theta, Y)}{\partial \sigma} = \frac{||Y - Z\beta||^2}{n\sigma^3} - \frac{1}{\sigma}$$

Les estimateurs de EMV sont des points de maximum locaux pour la fonction de vraisemblance. Puisque la matrice hessienne est positive semi-définie il suffit de resoudre $\nabla l(\theta, Y) = \vec{0}$. On trouve alors,

$$\hat{\beta}_{emv} = (Z^T Z)^{-1} Z^T Y$$
$$\hat{\sigma}^2_{emv} = \frac{||Y - Z\hat{\beta}_{emv}||^2}{n}$$

3.2 Exercice 12

Puisque l'estimateur est une transformation affine d'une variable gaussienne $\hat{\beta} = Z^{\#}Y$, il l'est aussi.

Or, $Y \sim N(Z\beta, \sigma^2 I_n)$. Alors, $\hat{\beta} \sim N(Z^{\#}Z\beta, (Z^{\#})\sigma^2 I_n (Z^{\#})^T)$

Donc, $\hat{\beta} \sim N(\beta, \sigma^2(Z^T Z)^{-1})$.

3.3 Exercice 13

On remarque que $Y - Z\hat{\beta} \sim N(0, \sigma^2 I_n)$ est une variable gaussienne centré et en plus c'est la projection de Y sur l'espace orthogonal à $F = Vect(Z^1, ..., Z^p)$, i.e, $Y - Z\hat{\beta} = (I - P_F)Y = P_{F^\perp}Y$.

On utilise le fait que la projection orthogonal preserve la norme et puis le théorème de Cochran pour trouver la distribution de la norme carré :

$$(n-p)\hat{\sigma}^2 = \left|\left|Y - Z\hat{\beta}\right|\right|^2 = \left|\left|P_{F^\perp}(Y - Z\hat{\beta})\right|\right|^2 = ||P_{F^\perp}Y||^2 \sim \sigma^2\chi^2(dim(F^\perp))$$

Or, $dim(F^\perp) = n - p$ et donc, en divisant par σ^2,

$$(n-p)\frac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-p)$$

3.4 Exercice 14

On utilise les résultats de l'exercice 2 pour montrer que chaque estimateur est fonction unique soit de $P_F Y$ soit de $P_{F^\perp} Y$:

$\hat{\beta} = (I_p)Z^{\#}Y = (Z^{\#}Z)Z^{\#}Y = Z^{\#}(H)Y = Z^{\#}P_F Y$

$\hat{\sigma^2} = \frac{1}{n-p}||Y - Z\hat{\beta}||^2 = \frac{1}{n-p}||(I-H)Y||^2 = \frac{1}{n-p}||P_{F^\perp}Y||^2$

Par le théorème de Cochran ces projections sont indépendants et donc des transformations fixes de ces variables le seront aussi.

4 TESTS STATISTIQUES, CAS RÉGRESSION LINÉAIRE GAUSSIENNE

4.1 Exercice 15

Une loi de student à $(n-p)$ degrés de liberté est une variable $T = \frac{G}{\sqrt{K/(n-p)}}$ où G est une gaussienne centrée réduite et K suit la loi $\chi^2(n-p)$. On cherche à trouver G et K telle que T soit égale à l'expression donné.

Or,

$$\hat{\beta} \sim N(\beta, \sigma^2(Z^TZ)^{-1})$$

Alors,

$$x^T(\hat{\beta} - \beta) \sim N(0, \sigma^2 x^T(Z^TZ)^{-1}x)$$

D'où,

$$G = \frac{x^T(\hat{\beta} - \beta)}{\sqrt{\sigma^2 x^T(Z^TZ)^{-1}x}} \sim N(0,1)$$

De plus, le résultat d'une question précedante nous donne,

$$K = (n-p)\frac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-p)$$

Finalement, il est facile de montrer que,

$$\frac{x^T(\hat{\beta}-\beta)}{\hat{\sigma}\sqrt{x^T(Z^TZ)^{-1}x}} = \frac{G}{\sqrt{K/(n-p)}}$$

4.2 Exercice 16

On utilise les quantiles de la loi de student à (n-p) pour trouver un intervalle de niveau $1-\alpha$,

$$\mathbb{P}_\theta\left(\frac{x^T\hat{\beta}-x^T\beta}{\hat{\sigma}\sqrt{x^T(Z^TZ)^{-1}x}} \in [t_{\alpha/2}^{(n-p)}, t_{1-\alpha/2}^{(n-p)}]\right) = 1-\alpha$$

Alors,

$$\mathbb{P}_\theta\left((x^T\hat{\beta}-t_{1-\alpha/2}^{(n-p)})\hat{\sigma}\sqrt{x^T(Z^TZ)^{-1}x} \leq x^T\beta \leq (x^T\hat{\beta}-t_{\alpha/2}^{(n-p)})\hat{\sigma}\sqrt{x^T(Z^TZ)^{-1}x}\right) = 1-\alpha$$

Ce qui nous donne bien l'intervalle,

$$\left[\hat{\sigma}(x^T\hat{\beta}-t_{1-\alpha/2}^{(n-p)})\sqrt{x^T(Z^TZ)^{-1}x},\, \hat{\sigma}(x^T\hat{\beta}-t_{\alpha/2}^{(n-p)})\sqrt{x^T(Z^TZ)^{-1}x}\right]$$

4.3 Exercice 17

On considère le test $\phi_c : Z \to \{0,\, 1\}$ tel que :

$$\phi_c(Z) = \mathbf{1}_{\{T(Z)\geq c\}}$$

Où $T(Z) = \frac{x^T\hat{\beta}}{\hat{\sigma}\sqrt{x^T(Z^TZ)^{-1}x}}$ suit la loi de Student à $(n-p)$ degrés de liberté sous $\theta \in \Theta_0 = \{\beta^T x = 0\}$.

Comme ϕ_c est de niveau α, on a que :

$$\sup_{\theta \in \Theta_0} \mathbb{P}_\theta \left(T(Z) > c_\alpha \right) \leq \alpha \Rightarrow \mathbb{P}_{\{\beta^T x = 0\}} \left(T(Z) > c_\alpha \right) \leq \alpha$$

En considérant l'égalité :

$$\mathbb{P}_{\{\beta^T x = 0\}} \left(T(Z) > c_\alpha \right) = \alpha \Rightarrow 1 - \mathbb{P}_{\{\beta^T x = 0\}} \left(T(Z) \leq c_\alpha \right) = \alpha$$

Donc :

$$\mathbb{P}_{\{\beta^T x = 0\}} \left(T(Z) \leq c_\alpha \right) = 1 - \alpha \Rightarrow c_a = q_{1-\alpha}^{S(n-p)}$$

Où $q_{1-\alpha}^{S(n-p)}$ est la quantille d'ordre $1-\alpha$ de la loi de Student à $(n-p)$ degrés de liberté.

Alors, le test désiré est décrit par :

$$\phi(Z) = \mathbf{1}_{T(Z) \geq q_{1-\alpha}^{S(n-p)}}$$

4.4 Exercice 18

Si nous observons Z et alors la statistique $T(Z) = \frac{x^T \hat{\beta}}{\hat{\sigma}\sqrt{x^T (Z^T Z)^{-1} x}}$, le test sera rejecté si $T(Z) > q_{1-\alpha}^{S(n-p)}$, et donc la p-valeur $\alpha(Z)$ est la solution de l'équation :

$$T(Z) = q_{1-\alpha(Z)}^{S(n-p)} \Rightarrow 1 - \alpha(Z) = \mathbb{P}\left(S \leq \frac{x^T \hat{\beta}}{\hat{\sigma}\sqrt{x^T (Z^T Z)^{-1} x}} \right)$$

Alors, la p-valeur est donné par $\alpha(Z)$, où :

$$\alpha(Z) = \mathbb{P}\left(S > \frac{x^T \hat{\beta}}{\hat{\sigma}\sqrt{x^T (Z^T Z)^{-1} x}} \right)$$

Où S suit une loi de Student à $(n-p)$ degrés de liberté.

4.5 Exercice 19

Puisque $\hat{\beta} \sim N(\beta, \sigma^2 (Z^T Z)^{-1})$, $A(\hat{\beta} - \beta)$ est un vecteur gaussien centré de matrice de covariance $\sigma^2 A (Z^T Z)^{-1} A^T$ qui est de rang q car A l'est aussi par hypothèse.

Alors,

$$\frac{1}{\sigma^2}(A(\hat{\beta}-\beta))^T[A(Z^TZ)^{-1}A^T]^{-1}(A(\hat{\beta}-\beta)) \sim \chi^2(q)$$

De plus, par un résultat précedant,

$$(n-p)\frac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-p)$$

On en déduit que l'expression donnée est la raison de deux chi-deux divisés par leurs degrés de liberté q et $n-p$, ce qui conclut la démonstration.

4.6 Exercice 20

La région de confiance RC_α de niveau α est donné en prennant $A \in \mathbb{R}^{2\times p}$ tel que :

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

Donc, on a que :

$$A(\hat{\beta}-\beta) = \begin{bmatrix} \hat{\beta}_1 - \beta_1 \\ \hat{\beta}_2 - \beta_2 \end{bmatrix}$$

Alors, en prennant l'expression de l'exercice 19, la région de confiance de niveau $1-\alpha$ pour $\beta = (\beta_1,\, \beta_2)$ est donné par :

$$RC_\alpha(\beta_1,\, \beta_2) = \left\{(\beta_1,\, \beta_2) \in \mathbb{R}^2,\ \ \frac{1}{2\hat{\sigma}^2}[\hat{\beta}_1-\beta_1,\, \hat{\beta}_2-\beta_2][A(Z^TZ)^{-1}A^T]^{-1}\begin{bmatrix} \hat{\beta}_1 - \beta_1 \\ \hat{\beta}_2 - \beta_2 \end{bmatrix} \leq f_{(2,\,n-p,\,1-\alpha)}\right\}$$

Où $f_{(q,\,n-p,\,1-\alpha)}$ est le fractile de niveau $(1-\alpha)$ d'une loi de Fisher admettant $(q,\, n-p)$ degrés de liberté.

Si on note c_{ij} le terme géneral de $(Z^TZ)^{-1}$ et en développent la dernière expression, on obtient que :

$$RC_\alpha(\beta_1,\, \beta_2) = \left\{(\beta_1,\, \beta_2) \in \mathbb{R}^2,\ \ \frac{B}{2\hat{\sigma}^2(c_{11}c_{22}-c_{12}^2)} \leq f_{(2,\,n-p,\,1-\alpha)}\right\}$$

Où B est donné par :

$$B(\beta_1,\, \beta_2) = \left(c_{22}(\hat{\beta}_1-\beta_1)^2 - 2c_{12}(\hat{\beta}_1-\beta_1)(\hat{\beta}_2-\beta_2) + c_{11}(\hat{\beta}_2-\beta_2)^2\right)$$

On peut remarquer que $B(\beta_1,\, \beta_2)$ denote une ellipse, et comme la région $RC_\alpha(\beta_1,\, \beta_2)$ est donné par un produit entre $B(\beta_1,\, \beta_2)$ par $\frac{1}{2\hat{\sigma}^2(c_{11}c_{22}-c_{12}^2)}$, on a aussi que la région de confiance $RC_\alpha(\beta_1,\, \beta_2)$ de niveau $1-\alpha$ est décrit par une ellipse.