DEVOIR MAISON #1 - MAP 433

16 septembre 2021

BARRÉ Théo
CHEN Longteng
C. A. F. Bruno

ÉCOLE POLYTECHNIQUE
IP PARIS

1
THÉORIE

1.1 Estimation par moindres carrés du vecteur $\boldsymbol{\beta}$

1. Montrer que toute solution $\hat{\mathbf{u}} \in \arg\min_{\mathbf{u}\in\mathbb{R}^p} J_n(\mathbf{u})$ est solution des *équations d'estimation* en $\mathbf{u} \in \mathbb{R}^p$:
$$\mathbf{Z}^T\mathbf{Y} = \mathbf{Z}^T\mathbf{Z}\mathbf{u}$$
Comme l'espace d'image de $\mathbf{Z}$, Im$\mathbf{Z}$ est un espace fermé, on peux faire la décomposition orthogonale. On suppose que $\mathbf{Y} = \mathbf{Y_1} + \mathbf{Y_2}$, où $\mathbf{Y_1} \in \mathrm{Im}\mathbf{Z}^\perp, \mathbf{Y_2} \in \mathrm{Im}\mathbf{Z}$. Et $\mathbf{Y_2}$ est la projection de $\mathbf{Y}$ sur Im$\mathbf{Z}$, alors on a $\mathbf{Z}^T(\mathbf{Y} - \mathbf{Y_2}) = 0, \mathbf{Y_2} = \mathbf{Z}\hat{\mathbf{u}}$.
Pour tout $\mathbf{u}$, $\|\mathbf{Y} - \mathbf{Z}\mathbf{u}\|^2 = \langle \mathbf{Y} - \mathbf{Z}\mathbf{u}, \mathbf{Y} - \mathbf{Z}\mathbf{u}\rangle = \|\mathbf{Y_1}\|^2 + \|\mathbf{Z}(\hat{\mathbf{u}} - \mathbf{u})\|^2$, donc si $\mathbf{u}$ minimise $J_n(\mathbf{u})$, alors $\mathbf{u} = \hat{\mathbf{u}}$ et il satisfait :
$$\mathbf{Z}^T\mathbf{Y} = \mathbf{Z}^T\mathbf{Z}\mathbf{u}.$$
2. Montrer que $\mathbf{Z}^\#\mathbf{Z} = \mathbf{I}_p$ et $\mathbf{Z}\mathbf{Z}^\# = H$ où H est le projecteur orthogonal sur l'espace vectoriel engendré par les colonnes de la matrice $\mathbf{Z}$.
$\mathbf{Z}^\#\mathbf{Z} = (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{Z} = \mathbf{I}_p$, et $\mathbf{Z}\mathbf{Z}^\# = \mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T$, on se souvient que lorsque $\mathbf{Z}$ est de rang p, le projecteur sur Im$\mathbf{Z}$ est $\mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T$.
3. Montrer que l'estimateur des moindres carrés est unique et a pour expression :
$$\hat{\boldsymbol{\beta}} := (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{Y} = \mathbf{Z}^\#\mathbf{Y}$$
Si on considère l'estimateur des moindres carrés $\hat{\boldsymbol{\beta}}$, il doit minimiser $J_n(\mathbf{u})$, donc il satisfait $\mathbf{Z}^T\mathbf{Y} = \mathbf{Z}^T\mathbf{Z}\hat{\boldsymbol{\beta}}$, car $\mathbf{Z}^T\mathbf{Z}$ est inversible, $\hat{\boldsymbol{\beta}} = \mathbf{Z}^\#\mathbf{Y}$
4. Montrer que l'estimateur des moindres carrés est un estimateur sans biais de $\boldsymbol{\beta}$.
$$\mathbb{E}_\theta\left[\hat{\boldsymbol{\beta}}\right] = \mathbf{Z}^\#\mathbb{E}_\theta\left[\mathbf{Y}\right] = \mathbf{Z}^\#\mathbf{Z}\boldsymbol{\beta} = \boldsymbol{\beta}.$$
5. Montrer que pour tout $\theta \in \Theta$, la matrice de covariance de cet estimateur est donnée par :
$$\mathrm{Var}_\theta(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{Z}^T\mathbf{Z})^{-1}$$
$$\mathrm{Var}_\theta(\hat{\boldsymbol{\beta}}) = \mathbf{Z}^\#\,\mathrm{Var}_\theta(\mathbf{Z}^\#)^T = \sigma^2(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^{-1} = (\mathbf{Z}^T\mathbf{Z})^{-1}.$$
6. Montrer que l'estimateur $\tilde{\boldsymbol{\beta}}$ est sans biais si et seulement si $\mathbf{B}\mathbf{Z} = \mathbf{I}_p$.
Si on a $\boldsymbol{\beta} = \mathbb{E}_\theta\left[\tilde{\boldsymbol{\beta}}\right] = \mathbf{B}\mathbb{E}_\theta\left[\mathbf{Y}\right] = \mathbf{B}\mathbf{Z}\boldsymbol{\beta}$, ce résultat doit être vrai pour tout $\boldsymbol{\beta}$, donc $\mathbf{B}\mathbf{Z} = \mathbf{I}_p$, et si $\mathbf{B}\mathbf{Z} = \mathbf{I}_p$, on a naturellement que $\tilde{\boldsymbol{\beta}}$ est sans biais.

7. Montrer que pour tout $\theta \in \Theta$,
$$\mathbb{E}_\theta\left[(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta})(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})^T\right] = \sigma^2\mathbf{B}(\mathbf{Z}^{\#})^T = \sigma^2(\mathbf{Z}^T\mathbf{Z})^{-1}$$
Noter que cette quantité est la matrice de covariance $\mathrm{Cov}_\theta(\tilde{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}})$

$$\mathbb{E}_\theta\left[(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta})(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})^T\right] = \mathrm{Cov}_\theta(\tilde{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}) = \mathbf{B}\,\mathrm{Var}_\theta(\mathbf{Y})(\mathbf{Z}^{\#})^T = \sigma^2\mathbf{B}\mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^{-1} = \sigma^2(\mathbf{Z}^T\mathbf{Z})^{-1}$$

8. Montrer que pour tout $\theta \in \Theta$, $\mathrm{Var}_\theta(\tilde{\boldsymbol{\beta}}) \succeq \mathrm{Var}_\theta(\hat{\boldsymbol{\beta}})$.
Pour tout $x \in \mathbb{R}^p$, soit $h_1 = x^T(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}), h_2 = x^T(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})$, on va montrer que $\mathrm{Var}_\theta(h_1) \geq \mathrm{Var}_\theta(h_2)$, on considère
$$r = \frac{\mathrm{Cov}_\theta(h_1, h_2)}{\sqrt{\mathrm{Var}_\theta(h_1)\,\mathrm{Var}_\theta(h_2)}}$$
On sait que $|r| \leq 1$ et $r = \dfrac{x^T\,\mathrm{Cov}_\theta(\tilde{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}})x}{\sqrt{\mathrm{Var}_\theta(h_1)\,\mathrm{Var}_\theta(h_2)}} = \dfrac{x^T\,\mathrm{Var}_\theta(\hat{\boldsymbol{\beta}})x}{\sqrt{x^T\,\mathrm{Var}_\theta(\hat{\boldsymbol{\beta}})xx^T\,\mathrm{Var}_\theta(\tilde{\boldsymbol{\beta}})x}}$, on a $x^T\,\mathrm{Var}_\theta(\hat{\boldsymbol{\beta}})x \leq x^T\,\mathrm{Var}_\theta(\tilde{\boldsymbol{\beta}})x$. Alors, $\mathrm{Var}_\theta(\tilde{\boldsymbol{\beta}}) \succeq \mathrm{Var}_\theta(\hat{\boldsymbol{\beta}})$.

1.2 Estimation de la variance σ^2 et Coefficient de détermination

9. En observant que pour tout vecteur $w \in \mathbb{R}^l$, on a $Tr(ww^T) = ||w||^2$, montrer que
$$\hat{\sigma}^2 = (n-p)^{-1}SSE = \frac{1}{n-p}||Y - \mathbf{Z}\hat{\beta}||^2$$
est un estimateur sans biais de la variance σ^2.
Donc, on va montrer que $\hat{\sigma}^2$ est un estimateur ainsi que $E_\theta[\hat{\sigma}^2] = \sigma^2$.

On a :
$$E_\theta[\hat{\sigma}^2] = \frac{1}{n-p}Tr(E_\theta(Y^T(I_n - H)^T(I_n - H)Y)) = \frac{1}{n-p}E_\theta(Y^T(I_n - H)Y)$$
Et on a $Y = Y_1 + Y_2$ avec où $Y_1 \in \mathrm{Im}\mathbf{Z}^\perp, Y_2 \in \mathrm{Im}\mathbf{Z}, Y_2 = HY$, donc $E_\theta[\hat{\sigma}^2] = \frac{1}{n-p}E_\theta(||Y_1||^2) =$, or $Y_1 = Y - HY = (I_n - H)(\mathbf{Z}\beta + \beta\epsilon(\theta)) = \sigma(I_n - H)\epsilon(\theta)$. Alors
$$E_\theta(||Y_1||^2) = \sigma^2 Tr(I_n - H)$$
Car H est un projecteur de rang p, donc $Tr(H) = p$. On a alors $E_\theta[\hat{\sigma}^2] = \sigma^2$, $\hat{\sigma}^2$ est non-biaisé.

10. Montrer que
$$||Y||^2 = RSS + SSE$$

$$RSS + SSE = ||Y_1||^2 + ||Y_2||^2 = ||Y||^2$$

1.3 Cas de la régression linéaire gaussienne

11. Déterminer l'estimateur du maximum de vraisemblance du paramètre θ

    Alors on va trouver le vecteur θ tel que la fonction
    $$n \log \sigma + \sigma^{-2}\frac{1}{2}(Y - \mathbf{Z}\beta)^T(Y - \mathbf{Z}\beta)$$
    atteigne son minimum. On a les équations suivantes :
    $$\sigma^2 = \frac{1}{n}(Y - \mathbf{Z}\beta)^T(Y - \mathbf{Z}\beta)$$
    $$(Y - \mathbf{Z}\beta)^T\mathbf{Z} = 0$$
    Donc $\hat{\beta} = (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T Y$,$\hat{\sigma}^2 = \frac{1}{n}||Y - \mathbf{Z}\hat{\beta}||^2$
12. Pour tout $\theta \in \Theta$, déterminer la distribution de l'estimateur des moindres carrés $\hat{\beta}$ sous $\mathbb{P}_\theta$

    $\hat{\beta} = (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T Y$, donc $\hat{\beta}$ est un vecteur gaussien avec l'espérance $(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\mathbb{E}_\theta(Y) = \beta$, la matrice de covariance $(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T Cov_\theta(Y)((\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T)^T = \sigma^2(\mathbf{Z}^T\mathbf{Z})^{-1}$

    Donc, $\hat{\beta} \sim N(\beta, \sigma^2(\mathbf{Z}^T\mathbf{Z})^{-1})$
13. Pour tout $\theta \in \Theta$, déterminer la distribution de $\hat{\sigma^2}$ sous $\mathbb{P}_\theta$

    $\hat{\sigma}^2 = \frac{1}{n}||Y - \mathbf{Z}\hat{\beta}||^2 = \frac{1}{n}||(I_n - H)(Y - \mathbf{Z}\beta)||^2$, ici $H = \mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T$, $Y - \mathbf{Z}\beta \sim N(0, \sigma^2 I_n)$

    De plus $H^2 = H$ est un projecteur de rang p, d'après le théorème de Cochran on a alors que $\frac{n}{(n-p)\sigma^2}\hat{\sigma^2} \sim \chi^2(n-p)$
14. Pour tout $\theta \in \Theta$, montrer que $\hat{\beta}$ et $\hat{\sigma}^2$ sont indépendants sous $\mathbb{P}_\theta$

    D'après le théorème de Cochran, $Y - HY$ et HY sont indépendants, donc $\hat{\sigma}^2 = g(Y - HY)$ et $\hat{\beta} = (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T HY = f(HY)$ sont indépendants.

1.4 Tests statistiques, cas régression linéaire gaussienne

15. Soit $x \in \mathbb{R}^p$. Montrer que sous $\mathbb{P}_\theta$,

$$\eta = \frac{x^T\hat{\beta} - x^T\beta}{\hat{\sigma}\sqrt{x^T(Z^TZ)^{-1}x}}$$

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

$X = \frac{x^T\hat{\beta}-x^T\beta}{\sigma\sqrt{x^T(Z^TZ)^{-1}x}}$ suit la loi de $N(0,1)$, $S = \frac{\hat{\sigma^2}}{\sigma^2}$ suit la loi de $\chi^2(n-p)$, de plus d'après la conclusion précédente X, S sont indépendants,donc $\eta = \frac{X}{\sqrt{S}}$ suit la loi de Student à degrés de liberté.

16. Soit $\alpha \in (0,1)$. Déterminer un intervalle de confiance bilatéral de niveau de couverture $1-\alpha$ pour $\beta^T x$.

$\beta^T x = x^T\beta$, et η suit une loi de Student à degrés de liberté.

$$\mathbb{P}_\theta[|\eta| \leq t_{n-p}(1-\alpha/2)] = 1-\alpha$$

Donc on a un intervalle de confiance bilatéral I de niveau de couverture $1-\alpha$ pour $\beta^T x$, $I = [x^T\hat{\beta} - t_{n-p}(1-\alpha/2)\hat{\sigma}\sqrt{x^T(Z^TZ)^{-1}x}, x^T\hat{\beta} + t_{n-p}(1-\alpha/2)\hat{\sigma}\sqrt{x^T(Z^TZ)^{-1}x}]$.

$$\mathbb{P}_\theta[\beta^T x \in I] = \mathbb{P}_\theta[|\eta| \leq t_{n-p}(1-\alpha/2)] = 1-\alpha$$

17. Soit $\alpha \in (0,1)$. Construire un test de l'hypothèse

$$H_0 : \beta^T x = 0, \text{ contre } H_1 : \beta^T x \neq 0$$

de niveau α.

On rejette H_0 si et seulement si $|T| > c$, ici $T = \frac{x^T\hat{\beta}}{\hat{\sigma}\sqrt{x^T(Z^TZ)^{-1}x}}$ donc on doit avoir

$$\mathbb{P}_{\beta^T x=0}(|T| > c) \leq \alpha$$

Quand $\beta^T x = 0$, T suit une loi de Student à $(n-p)$ degrés de liberté. On peut prendre $c = t_{n-p}(1-\alpha/2)$. i.e. Quand $|T| > t_{n-p}(1-\alpha/2)$, on rejette H_0.

18. Déterminer la p-valeur de ce test.
$\hat{\alpha}(Z) = \inf\{\alpha \in (0,1) : Z \in \mathcal{R}_\alpha\}$, ici $Z \in \mathcal{R}_\alpha$ si et seulement si $|T(Z)| > t_{n-p}(1-\alpha/2)$ C'est-à-dire que $\alpha > 2 - 2\phi(|T(Z)|)$, ϕ est la loi de distribution de Student à (n-p) degrés de liberté. Alors la p-valeur est $2 - \phi(|\frac{x^T\hat{\beta}}{\hat{\sigma}\sqrt{x^T(Z^TZ)^{-1}x}}|)$

19. Soit A une matrice de taille $q \times p$ de rang $q \leq p$. Montrer que sous $\mathbb{P}_\theta$

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

suit une loi de Fisher à $(q, n-p)$ degrés de liberté.
Soit $v = A(\hat{\beta}-\beta)$, v suit la loi de $N(0, \sigma^2 A(Z^TZ)^{-1}A^T)$, et

$$B = A(Z^TZ)^{-1}A^T = (A(Z^TZ)^{-1}Z^T)(A(Z^TZ)^{-1}Z^T)^T = P^TP$$

B est inversible. De plus $B^T = B$, B est définie positive. Donc on peut trouver une matrice inversible Q de taille $q \times q$ tel que $QQ^T = B$. Soit $w = Q^{-1}v \sim N(0, \sigma^2 Q^{-1}B(Q^T)^{-1}) = N(0, \sigma^2 I_q)$. Donc $\frac{w^Tw}{\sigma^2}$ suit la loi de $\chi^2(q)$, de plus $(n-p)\frac{\hat{\sigma}^2}{\sigma^2}$ suit la loi de $\chi^2(n-p)$, et w est une fonction de $\hat{\beta}$, donc w est indépendant de $\hat{\sigma}$. Donc $\frac{w^Tw}{q\hat{\sigma}^2}$ suit une loi de Fisher à $(q, n-p)$ degrés de liberté. $w^Tw = v^T(QQ^T)^{-1}v = vB^{-1}v$, donc $\frac{1}{q\hat{\sigma}^2}\{A(\hat{\beta}-\beta)\}^T[A(Z^TZ)^{-1}A^T]\{A(\hat{\beta}-\beta)\}$ suit une loi de Fisher à $(q, n-p)$ degrés de liberté.

20. Déterminer une région de confiance pour le vecteur (β_1, β_2)(on a posé $\beta = [\beta_1, ..., \beta_p]^T$).

Avec la conclusion précédente, on peut poser que $A = (a_{ij})_{2\times p} = [I_2, 0]$, avec cela on a $\{A(\beta))\}^T = (\beta_1, \beta_2)$

Soit $v = A(\hat{\beta}-\beta))$, $T = \frac{v^TB^{-1}v}{2\hat{\sigma}^2}$ suit une loi de Fisher à $(2, n-p)$ degrés de liberté. Car

$$\mathbb{P}_\theta(T \leq F_{2,n-p}(1-\alpha)) = 1-\alpha$$

Donc la région de confiance pour le vecteur (β_1, β_2) est la zone

$$\Omega = \{(\beta_1, \beta_2) | a(\hat{\beta}_1-\beta_1)^2 + b(\hat{\beta}_2-\beta_2)^2 + 2c(\hat{\beta}_1-\beta_1)(\hat{\beta}_2-\beta_2) \leq 2\hat{\sigma}^2 F_{2,n-p}(1-\alpha)\}$$

Ici $a = z_{11}, b = z_{22}, c = z_{12}$, si on suppose que $(Z^TZ)^{-1} = (z_{ij})_{p\times p}$, comme elle est inversible et définie positive, on a $ab > c^2$, donc Ω est une ellipse.

2 PRATIQUE

cf jupyter notebook