Utilització de la pseudoinversa quan $X'X$ és singular

Biplop Dey

10 de gener de 2021

Resum

En aquest treball s'ha analitzat les propietats dels estimadors de la regressió lineal quan la matriu de disseny no té rang màxima, veurem que les propietats són iguals independentment de la inversa generalitzada que es faci servir. A més és compatible per matriu de disseny amb rang màxima, l'única propietat que canvia és que β *depèn de la inversa que es fa servir i sempre és esbiaixat. Desprès compararem dues inverses: g-inv i pseudoinversa. La pseudoinversa té la propietat de que l'estimació de* $\hat{\beta} = X^+Y$ *té norma mínima i la solució general és* $\hat{\beta} = X^+Y + (I - X^+X)w, w \in \mathbb{R}^p$*, no cal fer cap selecció prèvia de columnes linealment independents com el cas de g-inv. També en g-inv, depenent de la combinació de columnes que s'agafen, donen diferents solucions. Bàsicament per aquests raons la pseudoinversa és millor. Finalment es demostra el teorema de Gauss-Markov general, i que això ens permetra trobar estimadors BLUE* $c'\beta$*, per certs* c'*. Aquests estimadors són de forma* $b'_*X\hat{\beta}$ *on* $b_* \in C(X)$ *i són BLUE de* $b'_*X\beta$ *i tenen variància* $Var(b'_*X\hat{\beta}) = b'_*Pb_*$*, on* P *es la matriu de projecció ortogonal a* $C(X)$*.*

Índex

1 Inversa generalitzada

1.1 Introducció

Considerem el sistema de regressió lineal $Y = X\beta - \epsilon$, on $\epsilon \sim N(0, \sigma^2 I)$ és el vector d'error, $Y \in \mathbb{R}^n$ és el vector resposta i $X \in \mathbb{R}^{n \times p}$ la matriu de disseny i seva columna i-esima la denotem X_i i $\beta \in \mathbb{R}^p$ és el vector de regressió de coeficients. La solució de mínim quadrat és $\hat{\beta}$ que satisfà el sistema $(X'X)\hat{\beta} = X'Y$, si $(X'X)$ és invertible aleshores $\hat{\beta} = (X'X)^{-1}X'Y$. Però en el cas de que $X'X$ no és singular fa servir servir la inversa generalitzada per resoldre l'equació normal. Veurem que hi han diverses inverses i que hi han infinites solucions. Començarem mirant propietats de la inversa generalitzada i les seves aplicacions a regressió lineal. Calcularem la g-inv i mirarem les propietats de la pseudoinversa. Desprès farem una comparació entre la g-inv i la pseudoinversa. Finalment veurem les funcions estimables per trobar BLUE de certes combinacions lineals de β.

1.2 Propietats de l'inversa generalitzada i aplicació a regressió lineal

Começem veien quan $X'X$ no és singular: Si $p \leq n$, aleshores $r(X) \leq min(n, p) = p$ i $X'X$ és invertible sí i només sí X no té columnes linealment dependents, r(X)=p. Es a dir en el cas $p \leq n$ $X'X$ és singular si només si $r(X) < p$.

Si $p > n \Rightarrow r(X) \leq min(n, p) = n \Rightarrow r(X'X) = r(X) \leq n < p$, per tant $X'X$ no és invertible. Es a dir en el cas $p > n$ $X'X$ és sempre singular.

Definició 1. *Sigui $A \in \mathbb{R}^{m \times n}$ i sigui $G \in \mathbb{R}^{n \times m}$ tal que:*

I) $AGA = A$

II) $GAG = G$

Aleshores si compleix les condició I es diu inversa generalitzada o simplement g-inv ultilitzarem la notació $G = A^-$ i si compleix I i II es diu inversa generalitzada reflexiva o simplement g-inv reflexiva, farem servir la mateixa notació que g-inv, aquestes inverses existeix però no són úniques.

Lema 1. $(A^-)' = (A')^-$ *també és cert per la pseudoinversa.*

Demostració. Veiem que primer per g-inv, propietat I: $A'(A^-)'A' = (A^-A') = A'$ i si a més és reflexiva $(A^-)'A'(A^-)' = (A^-AA^-)' = (A^-)'$. Ara si és pseudoinversa (veure definició 2), es compleix la propietat III: $((A^+)'A')' = AA^+ = (AA^+)' = (A^+)'A'$ i la propietat IV $(A'(A^+)')' = A^+A = (A^+A)' = A'(A^+)'$. □

Farem servir les propietats de la g-inv per demostrar que les propietats dels estimadors no canvien encara que $r(X) < p$. Anomenarem $C(X)$ al subespai vectorial generat per les columnes de X i $B = (X'X)$.

La solució $\hat{\beta}$ del criteri del mínim quadrat , $min_{\hat{\beta}}||Y - X\hat{\beta}||^2$, es satisfà el sistema d'equacions normals $X'Y = (X'X)\hat{\beta}$, la demostració es pot consultar a [3].

Ara si $X'X$ es no singular la solució és trivial. En el cas de ser singular la solució és $\hat{\beta} = (X'X)^{-}X$ (aquesta la solució general per qualsevol rang d'X). Mirem l'interpretació geometrica de com ha de ser la solució $\hat{\beta}$:

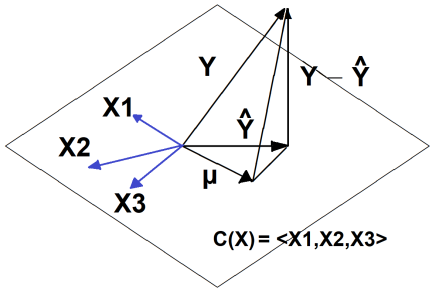

Figura 1: Interpretació geomètrica

Com estem buscant $\hat{\beta}$ tal que $||Y - X\hat{\beta}||^2$ sigui mínima és el mateix que dir $||Y - \mu||^2, \mu \in C(X)$, aleshores la distància entre Y i el espai vectorial $C(X)$ que és un pla, és mínima quan $\mu = \hat{Y}$, es a dir la projecció ortogonal de Y al pla. Això és independent de la inversa que es faci servir. I la solució $\hat{\beta}$ són coordenades de $\hat{Y}$ respecte una base formada per les columnes de $C(X)$, si $r(X) = p$ aleshores només hi hauria una base i per tant una solució, però com podem veure la figura X_1, X_2, X_3 són linealment dependent, i per tant podem agafar una base $\{X_1, X_2\}$ o $\{X_2, X_3\}$ etc i ens donarien diferents solucions.

Lema 2. *$\hat{\beta} = (X'X)^{-}X'Y$ és solució de l'equació normal $X'X\hat{\beta} = X'Y$.*

Demostració. Per simplificar posem $B = (X'X)$, $X'Y = B\beta = BB^{-}(B\beta) = BB^{-}X'Y = B(B^{-}X'Y) = B\hat{\beta} \Rightarrow X'Y = B\hat{\beta}$. □

Lema 3. *Si $PX'X = QX'X$ aleshores $PX' = QX'$.*

Demostració. Farem servir que per qualsevol matriu A, si $AA' = 0$, aleshores $A = 0$. Sigui $(PX' - QX)(PX' - QX') = (PX' - QX')X(P - Q)' = (PX'X - QX'X)(P - Q)' = 0(P - Q)' = 0 \Rightarrow PX' - QX' = 0$ □

Teorema 1. *Si G és g-inv de $X'X$, alehores:*

1. G' és també g-inv de $X'X$.
2. $G(X'X)G'$ és simètric i g-inv reflexiva de $X'X$.
3. (a) $(X'X)GX' = X'$ i (b) $XG(X'X) = X$.
4. $XGX' = XHX'$ per qualsevol g-inv H.
5. XGX' és simètric.

Demostració. Per simplificar posem $B = (X'X)$
1) Com G és g-inv $(BG'B)' = BGB = B$ fent la transposta ho tenim $BG'B = B'$

2) És clar que és simètric. Per veure que propietat I: $B(GBG')B = (BGB)G'B = BG'B = B$, l'ultma igualtat es per 1). Per veure la propietat II: $GBG'(BGB)G' = G(BG'B)G' = GBG$

3) En (a) $BGX'X = IX'X$ i per lema 3 $BGX' = IX'$, en (b) $BG'X'X = IX'X$ i ara també per el lema $BG'X' = IX'$ fem la transposta $XGB = X$.

4) Per 3.b $XGX'X = X$ i $XHX'X = X$ i pel lema 3 $XGX' = XHX'$.

5) $(XGX')' = XG'X'$, per (1) G' és g-inv i per (4) $XG'X' = XGX'$. □

El resultat (4) és important perquè definim $P = X(X'X)^{-}X'$ i més endavant veurem que es la matriu de projecció ortogonal a $C(X)$ i per (4) P és única o sigui que és independent de la g-inv que es faci servir.

1.3 Construcció de g-inv

Lema 4. *Sigui $W_{p\times p}$ amb rang $r < p$. Aleshores es pot posar com: $W = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$ on $r(A) = r$. La g-inv és $W^{-} = \begin{pmatrix} A^{-1} & 0 \\ 0 & 0 \end{pmatrix}$*

Demostració. Podem particionar la matriu, perquè $r(W) = r$ alehores exiteix una menor A amb $r(A) = r$. Veiem és una g-inv: $WW^{-}W = \begin{pmatrix} A & B \\ C & CA^{-1}B \end{pmatrix}$ Això és igual a W si $D = CA^{-1}B$, com A té rang r aleshores les columnes $\begin{pmatrix} B \\ D \end{pmatrix}$ és combinació lineal de les columnes de $\begin{pmatrix} A \\ C \end{pmatrix}$. i.e. existeix una matriu $F_{r\times(p-r)}$ tal que $\begin{pmatrix} B \\ D \end{pmatrix} = \begin{pmatrix} A \\ C \end{pmatrix} F$, aleshores $B = AF \Rightarrow F = A^{-1}B$, com $D = CF \Rightarrow D = CA^{-1}B$. Observem també que $W^{-}WW^{-} = W^{-}$, per tant és una g-inv reflexiva.

□

Càlcul de g-inv de $X'X$. Com $r(X) = r$, agafem r columnes linealment independents, una manera és esglaonant per columnes $X = CQ$, on Q és una matriu invertible i C esta esglaonat per columnes, mirem les $p-r$ últimes columnes de Q per trobar les columnes linealment independent de X. Un altre manera es resoldre un sistema. Posem $X = (X_1, X_2)$ on $r(X_1) = r$, $X'X = \begin{pmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{pmatrix}$, i com $r(X_1'X_1) = r$ i pel lema d'adalt, $(X'X)^{-} = \begin{pmatrix} (X_1'X_1)^{-1} & 0 \\ 0 & 0 \end{pmatrix}$. Una manera de calcular la projecció P és $P = X_1(X_1'X_1)^{-1}X_1'$ perquè és única, aquest resultat el farem servir en el seguent lema.

Lema 5. *Sigui $r(X) = r \leq p$ i $P = X(X'X)^{-}X'$, aleshores:*

1. *P i $I - P$ són matrius de projeccions.*

2. $r(I-P) = tr(I-P) = n - r$

3. $PX = X$

Demostració. 1) P és simètric per Teorema 1 part 5. P és idempotent: Sigui $G = (X'X)^-$ i $P = XGX' \Rightarrow P^2 = (XGX')(XGX') = (XG)(X'XGX') = XGX'$ per Teorema 1 part 3a. P és simètric i idempotent. És fàcil veure que $I - P$ també. 2) Per (1) $I - P$ és un projecció per tant $r(I-P) = tr(I-P)$, posem $P = X_1(X_1'X_1)^{-1}X_1'$ alehores $r(I-P) = tr(I-P) = tr(I) - tr(P) = n - r$, ja que $tr(P) = tr(X_1(X_1'X_1)^{-1}X_1') = tr(X_1'X_1(X_1'X_1)^{-1}) = tr(I_r) = r$.
3) $PX = XGX'X = X$ per teorema 1 part 3b.

□

Observem que $X^- = (X'X)^-X'$, ja que $XX^-X = X(X'X)^-X'X = PX = X$ i també es compleix la propietat reflexiva: $X^-XX^- = (X'X)^-X'X(X'X)X' = (X'X)^-X' = X^-$.

Lema 6. *$P = XX^-$ és la matriu de projecció ortogonal a $C(X)$.*

Demostració. Sabem que P és idempotent i simètric, ara veiem que és projeció a $C(X)$, per qualsevol vector x, es descomposa com $x = Px + (I-P)x$, i si $\forall x, y$ tenim $Px = x, (I-P)y = y$, tenim $x'y = (Px)'(I-P)y = x'P'(I-P) = x'P(I-P)y = 0$. I si $y \in C(A) \Rightarrow \exists v$ tal que $y = Xv$ i $Py = PXv = Xv = y$, reciprocament si $Py = y \Rightarrow y = XX^-y \Rightarrow y \in C(X)$. □

1.4 Propietats dels estimadors

En general $\hat{\beta}$ no és únic però les propietats de $\hat{Y} = PY = X\hat{\beta}$ són uniques, veiem exemples:

Lema 7. *$\hat{Y}$ és un estimador no esbiaixat de $\mu = X\beta$*

Demostració. $E(\hat{Y}) = E[PY] = PX\beta = X\beta = \mu$ □

Teorema 2. *Per qualsevol combinació lineal $c'\mu$, $c'\hat{Y}$ és l'únic estimador BLUE de $c'\mu$.*

Demostració. Per lema 7 $E(c'\hat{Y}) = c'\mu$ no és esbiaixat, i sigui $d'Y(f'\mu = (f'P)Y = d'Y)$ un altre estimador lineal esbiaixat de $c'\mu$, esbiaixat implica que $E(d'Y) = c'\mu$, i també tenim $E(d'Y) = d'\mu$. Aleshores $d'\mu = c'\mu$, com μ és un vector de $C(X)$, no sabem que és, però sense saber el valor sempre es compleix $d'\mu = c'\mu$, per tant $d'\mu = c'\mu$ per tot $\mu \in C(X)$. S'obté desprès que $(c-d)'\mu = 0, \forall \mu \in C(X)$, per tant $c - d$ és ortogonal a $C(X)$. Aleshores, $P(c-d) = 0$ i $Pc = Pd$. El $Var(c'\hat{Y})$ és: $Var(c'\hat{Y}) = Var(c'PY) = Var((Pc)'Y) = Var((Pd)'Y) = \sigma^2(Pd)'Pd = \sigma^2 d'Pd$ i $Var(d'Y) = \sigma^2 d'd$, i tenim $Var(d'Y) - Var(c'\hat{Y}) = \sigma^2 d'd - \sigma^2 d'Pd = \sigma^2 d'(I-P)d = \sigma^2 d'(I-P)^2 d = \sigma^2[(I-P)d]'(I-P)d \geq 0$, ens dona el que voliem. Observem que $Var(d'Y) - Var(c'\hat{Y}) = 0 \Leftrightarrow (I-P)d = 0$ i.e., $d = Pd = Pc$, i.e $d'Y = (Pc)'Y = c'PY = c'\hat{Y}$, establint unicitat. □

Aquest teorema ens diu que $\hat{Y}$ és BLUE de Y ja que podem agafar $c_i = (0, \cdots, 1, \cdots, 0)$ on 1 esta a posició i-esima, i $c_i'\hat{Y} = \hat{Y}_i$ és BLUE de Y_i per tot i.

Corol·lari 1. *Si $r(X) = p$, alehores $a'\hat{\beta}$ és BLUE de $a'\beta$, per qualsevol a.*

Demostració. Com $r(X) = p \Rightarrow X'X$ invertible i tenim : $a'\beta = a'I\beta = a'B^(- 1)B\beta = a'B^{-1}X'\mu = c'\mu$, on $c' = a'B^{-1}X'$. i també $a'\hat{\beta} = a'B^{-1}B\hat{\beta} = aB^{-1}X'\hat{Y} = c'\mu$, i per teorema d'abans $c'\hat{Y}$ és BLUE de $c'\mu$, ja el tenim. □

El teorema 6 de Gauss-Markov generaltizarà aquest resultat pel cas en que X no tingui rang màxim. Observem que quan X no té rang màxim el $\hat{\beta}$ no és únic, i no serveix parlar de que $\hat{\beta}$ és òptim en cap sentit.

Proposició 1 (Estimació de σ^2)**.** *Sigui $r(X) = r$, S^2 és una estimació no esbiaixada de σ^2, on*

$$S^2 = \frac{1}{n-r}(Y - X\hat{\beta})'(Y - X\hat{\beta}) = \frac{SSE}{n-r}$$

Demostració. $(n-r)S^2 = SSE = (Y - PY)'(Y - PY) = Y'(I - P)Y$, fem servir els resultats de teoria: $\Sigma_Y = \sigma^2 I$ i $E[X'AX] = tr(A\Sigma_X) + \mu_X' A\mu_X$.
$E[Y'(I-P)Y] = tr(\sigma^2 I(I-P)) + \mu'(I-P)\mu = \sigma^2 tr(I-P) + \mu'(I-P)\mu = \sigma^2(n-r) + \beta'X'(I-P)X\beta = \sigma^2(n-r)$ perquè $(I-P)X = 0$. □

Teorema 3. *Sigui $Y \sim N(X\beta, \sigma^2 I)$, aleshores:*

1. $E(\hat{Y}) = X\beta$ i $Var(\hat{Y}) = \sigma^2 P$
2. $E(e) = 0$ i $Var(e) = \sigma^2(I - P)$
3. $E(\hat{\beta}) = (X'X)^-(X'X)\beta$ i $Var(\hat{\beta}) = \sigma^2(X'X)^-$
4. $\hat{\beta}$ i e són independents
5. $(n-r)S^2/\sigma^2 = \epsilon'(I-P)\epsilon/\sigma^2 \sim \chi^2_{n-r}$

Demostració. 1) $E(\hat{Y}) = E(PY) = PX\beta = X\beta$ i $Var(\hat{Y}) = \Sigma_{PY} = P\Sigma_Y P' = \sigma^2 PIP = \sigma^2 P$.

2) $E(Y - \hat{Y}) = E(Y) - E(\hat{Y}) = X\beta - X\beta = 0$ i $Var(e) = Var((I-P)Y) = \Sigma_{(I-P)Y} = (I-P)\Sigma_Y(I-P)' = \sigma^2(I-P)$.

3)$E(B^-X'Y) = B^-X'X\beta = B^-B\beta$ i $Var(\hat{\beta}) = \Sigma_{B^-X'Y} = B^-X'\Sigma_Y(B^-X')' = \sigma^2 B^-X'XB^- = \sigma^2 B^-$, hem fet servir que B^- és g-inv reflexiva.

4) Com $\hat{\beta} = AY$, on $A = B^-X'Y$ i $e = (I-P)Y$, per [3] apendix B4, hem de veure que A i $(I-P)$ són ortogonals, $A(I-P)' = A - AP = A - B^-X'P = A - B^-(PX)' = A - B^-X' = 0$.

5) Per [3] apèndix B5, $\frac{\epsilon'}{\sigma}(I-P)\frac{\epsilon'}{\sigma} \sim \chi^2_d$, on $d = r(I-P) = n - r$. □

Notem per 1), 2) i 3) que $I-P, P, (X'X)^{-}$ són simetriques (per $I-P, P$ ho sabiem) i semidefinides no-negatives, ja que, excepte la constant σ^2, són matrius de covariància de vectors aleatoris.
En aquest apartat fent servir les propietats de la g-inv tenim els mateixos resultats que en cas de $X'X$ no singular. Remarquem també que els altres estimadors no depenen de la g-inv, només $\hat{\beta}$ depèn de la inversa que es faci servir.
En el cas de $X'X$ singular, les distribucions de $\hat{Y}=PY, e=(I-P)Y, \hat{\beta}=X^{-}Y$, no seran normals ja que $P, I-P, X^{-}$ no tenen rang màxim i per tant no són invertibles.

2 Pseudoinversa

2.1 Introducció

Definició 2. *Sigui $A \in \mathbb{R}^{m\times n}$ i sigui $A^{+} \in \mathbb{R}^{n\times m}$, A^{+} és pseudoinversa de A sí i només sí:*

I) $AA^{+}A = A$

II) $A^{+}AA^{+} = A^{+}$

III) $(AA^{+})' = AA^{+}$

IV) $(A^{+}A)' = A^{+}A$

També es diu inversa de Moore-Penrose , veurem que aquesta existeix i és única.

Teorema 4. *Donada una matriu $A_{m\times n}$ existeix una única pseudoinversa $A^{+}_{n\times m}$*

Demostració. Demostrarem primer l'unicitat, sigui A_1^{+}, A_2^{+} pseudoinverses d'A. Observem que, $AA_1^{+} = (AA_2^{+}A)A_1^{+} = (AA_2^{+})(AA_1^{+}) = (AA_2^{+})'(AA_1^{+})' = (A_2^{+})A'(AA_1^{+})' = (A_2^{+})'(AA_1^{+}A)' = (A_2^{+})'A' = AA_2^{+}$, hem fet servir la propietat III. Analogament per fent servir la propietat IV: $A_1^{+}A = AA_2^{+} \Rightarrow A_1^{+}A = A_1^{+}(AA_2^{+}A) = (A_1^{+}A)(A_2^{+}A) = (A_1^{+}A)'(A_2^{+}A)' = (A_1^{+}A)'A'(A_2^{+})' = (AA_1^{+}A)'(A_2^{+})' = A'(A_2^{+})' = (A_2^{+}A)' = A_2^{+}A$. Ara fent servir aquest dos resultats i la propietat II , tenim $A_1^{+} = A_1^{+}(AA_1^{+}) = A_1^{+}AA_2^{+} = A_2^{+}AA_2^{+} = A_2^{+}$, es demostra l'unicitat.

En existència, farem servir descomposició en valors singulars DVS (SVD en angles) per decomposar A com $A = U\Sigma V'$, on $U_{m\times m}, V_{n\times n}$ són matrius unitaris i $\Sigma_{m\times n}$ és una matriu amb diagonal no negativa i la resta és 0. Aleshores definim $A^{+} = V\Sigma^{+}U'$, on $\Sigma^{+}_{n\times m}$ és Σ' però substituim cada element de la diagonal per la seva inversa si és diferent de 0, és fàcil veure que Σ^{+} és la pseudoinversa (veure [6]).Veiem que A^{+} és la psuedoinversa d'A:

I. $AA^{+}A = U\Sigma V'V\Sigma^{+}U'U\Sigma V' = U\Sigma\Sigma^{+}\Sigma V' = U\Sigma V' = A$

II. $A^{+}AA^{+} = V\Sigma^{+}U'U\Sigma V'V\Sigma^{+}U' = V\Sigma^{+}\Sigma\Sigma^{+}U' = V\Sigma^{+}U' = A^{+}$

III. $(AA^{+})' = (U\Sigma V'V\Sigma^{+}U')' = (U\Sigma\Sigma^{+}U')' = U(\Sigma\Sigma^{+})'U' = U(\Sigma\Sigma^{+})U' = U\Sigma V'V\Sigma^{+}U' = AA^{+}$

IV. $(A^+A)' = (V\Sigma^+U'U\Sigma V')' = (V\Sigma^+\Sigma V')' = V(\Sigma^+\Sigma)'V' = V(\Sigma^+\Sigma)V' = V\Sigma^+U'U\Sigma V' = A^+A$

$\square$

Observem que podem posar $A^+ = (A'A)^+A'$, ja que com la pseudoinversa és una g-inv reflexiva i hem vist que compleix la propietat I i II. La propietat III: $(A((A'A)^+A')' = P' = P = A(A'A)^+A'$ i propietat IV: $(((A'A)^+A')A)' = ((A'A)^+(A'A))' = (A'A)^+(A'A)$ i per unicitat $A^+ = (A'A)^+A'$.

2.2 Aplicacions a regressió lineal

Lema 8. *$Q = A^+A$ és una matriu de projecció ortogonal a $C(A')$.*

Demostració. Sigui $Q = A^+A$, és idempotent $Q^2 = (A^+AA^+)A = A^+A$ i és simètrica $Q' = (A^+A)' = A^+A = Q$, per tant és una matriu de projecció ortogonal, veiem que és projecció a $C(A')$: Farem servir $AQ = A \Rightarrow (AQ)' = QA' = A'$ i si $y \in C(A') \Rightarrow \exists x$ tal que $A'x = y \Rightarrow Qy = QA'x = A'x = y$ i si $Qy = y \Rightarrow y = (A^+A)'y = A'(A^+)'y \Rightarrow y \in C(A')$. Per tant Q és una projecció ortogonal a $C(A')$ $\square$

Observem que pel lema anterior, $I - Q$ és projecció ortogonal a $C(A')^\perp = Ker(A)$.

Teorema 5. *Sigui $A \in \mathbb{R}^{m\times n}$, i fixem $b \in \mathbb{R}^m$ aleshores $\forall x \in \mathbb{R}^n$, tenim $||Ax-b||^2 \geq ||Az-b||^2$ on $z = A^+b$, i l'igualtat es dona sí i només sí $x = A^+b + (I - A^+A)w$, on $w \in \mathbb{R}^m$ o $x = A^+b + v$ on $v \in Ker(A)$.*

Demostració. Farem servir les propietats de la projecció P, observem que $A'(Az - b) = A'(AA'b - b) = A'(Pb - b) = A'P'b - A'b = (PA)'b - A'b = 0$. Ara $||Ax - b - (Az-b) + Az - b||^2 = ||A(x-z)||^2 + 2(A(x-z))'(Az-b) + ||Az-b||^2 = ||A(x-z)||^2 + ||Az-b||^2 \geq ||Az-b||^2$, ja que $(A(x-z))'(Az-b) = (x-z)'A'(Az-b) = 0$. L'igualtat és dona si $||Ax-b||^2 = ||Az-b||^2 \Leftrightarrow ||A(x-z)||^2 = 0 \Leftrightarrow A(x-z) = 0$, si A és injectiva i.e. $Ker(A) = 0$, A té columnes linealment independent, aleshores $x = z$, en general $x = z + v$ on $v \in Ker(A)$, o simplement $x = z + (I - A^+A)w, w \in \mathbb{R}^m$ fent servir la projecció ortogonal de $Ker(A)$ pel lema d'abans. $\square$

Notem que la primera part de la demostració no hem fet servir propietat III i IV, aleshores compleix per g-inv també.

Corol·lari 2. *El sistema $Ax = b$ té solució per algun x. Aleshores $z = A^+b$ també és solució del sistema i $||x||^2 \geq ||z||^2$ per qualsevol solució x, i l'igualtat es dona sí i només sí $x = z$.*

Demostració. Farem servir les propietats de $Q = A^+A$: $Qz = A^+AA^+b = A^+b = z$ i $Q' = Q$. Suposem que $Ax = b$ per algun x. Pel teorema d'abans $z = A^+b$ també és solucio, ja que com $Ax - b = 0$ tenim $||Az-b||^2 \leq ||Ax-b||^2 = 0 \Rightarrow Az = b$. Alehores, $z'(x-z) = (Qz)'(x-z) = z'Q(x-z) = z'(A^+Ax - z) = z'(A^+b - z) = 0$. Ara $||x - z + z||^2 = ||z||^2 + 2z'(x-z) + ||x-z||^2 = ||z||^2 + ||x-z||^2 \geq ||z||^2$. i l'igualtat és dona sí i només sí $||x-z||^2 = 0 \Leftrightarrow x = z$. $\square$

La solució $\hat{\beta} = (X'X)^{-}X'Y$ que resol l'equació normal també és un mínim quadrat, i hi ha moltes maneres de calcular $\hat{\beta}$ perquè depenen de les inverses que es facin servir. En el següent corol·lari ens serveix per classificar els $\hat{\beta}$.

Corol·lari 3. *Donat el sistema de regressió $X\beta = Y + \epsilon$, $\hat{\beta} = X^{+}Y + (I - X^{+}X)w, w \in \mathbb{R}^p$ són totes les solucions que satisfà el criteri del mínim quadrat i $\hat{\beta} = X^{+}Y$ és la solució amb mínima norma.*

Demostració. Pel teorema 5, $||X\beta - Y|| \geq ||X\hat{\beta} - Y||$, on $\hat{\beta} = X^{+}Y$ i la solució d'equació normal és $XB^{-}X'Y \Rightarrow ||X\beta - Y|| \geq ||XB^{-}X'Y - Y||$ també es un minim, aleshores tenim igualtat $||X\hat{\beta} - Y|| = ||XB^{-}X'Y - Y||$ i un altre cop pel teorema 5 tenim que $\hat{\beta} = X^{+}Y + (I - X^{+}X)w = B^{-}X'Y$, per algun $w\mathbb{R}^p$, per tant la solució general és $\hat{\beta} = X^{+}Y + (I - X^{+}X)w$. Ara per veure que $\hat{\beta} = X^{+}Y$ té mínima norma, sabem que $v = B^{-}X'Y$ satisfà l'equació normal: $(X'X)v = X'Y$, apliquem el corol·lari d'adalt a aquest sistema $A = (X'X), x = v, b = X'Y$, i tenim que $\hat{\beta} = (X'X)^{+}X'Y = X^{+}Y$ és solució amb norma mínima. □

Per tant la propietat de la pseudoinversa és que $\hat{\beta} = X^{+}Y$ és la solució amb mínima norma. I que per qualsevol mínim quadrat és de la forma $\hat{\beta} = X^{+}Y + (I - X^{+}X)w, w \in \mathbb{R}^p$, i per qualsevol solució $(X'X)^{-}X'Y$, independent de la g-inv que es faci servir, existeix w tal que $X^{+}Y + (I - X^{+}X)w = (X'X)^{-}X'Y$.

3 Propietats de les diferents estimacions de β

Sabem pel teorema que les propietats dels estimadors són iguals per qualsevol g-inv, i només canvien en $\hat{\beta}$ on suposant $(X'X)^{-}$ és g-inv reflexiva tenim $Var(\hat{\beta}) = \sigma^2(X'X)^{-}$ i $E[\hat{\beta}] = (X'X)^{-}(X'X)\beta$, veiem que en el cas de $X'X$ singular $(X'X)^{-}(X'X) \neq I$ i per tant sempre serà esbiaixat.

Primer mirem com és la esperança i la variància de $\hat{\beta}$ fent servir la g-inv reflexiva: Tenim $(X'X)^{-}(X'X) = \begin{pmatrix} I_r & (X_1'X_1)^{-1}X_1'X_2 \\ 0 & 0 \end{pmatrix}$, posem com $\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$ on β_1 són les primeres r elements de β i β_2 les $p-r$ últimes, mateix amb $E(\hat{\beta}) = \begin{pmatrix} E(\hat{\beta}_1) \\ E(\hat{\beta}_2) \end{pmatrix}$, aleho-res $E(\hat{\beta}) = (X'X)^{-}(X'X)\beta \Rightarrow E(\hat{\beta}_1) = \beta_1 + (X_1'X_1)^{-1}X_1'X_2\beta_2$ i $E(\hat{\beta}_2) = 0$, com que les columnes de X_2 són linealment dependents de X_1 existeix una matriu $F_{r\times(p-r)}$ tal que $X_2 = X_1F$, ara $(X_1'X_1)^{-1}X_1'X_2 = (X_1'X_1)^{-1}X_1'X_1F = F \Rightarrow E(\hat{\beta}_1) = \beta_1 + F\beta_2$, on $F\beta_2$ és el biaix de l'estimador $\hat{\beta}_1$ i com $E(\hat{\beta}_2) = 0$ el biaix de $\hat{\beta}_2$ és β_2, per tant $\hat{\beta}$ és un estimador esbiaixat, com ja sabiem. En $Var(\hat{\beta}_i) = 0, i = r+1, \cdots, p \Rightarrow \hat{\beta}_i = 0$ són constants.
En cas de pseudoinversa, no podem desenvolupar com vam fer abans: $\hat{\beta} = X^{+}Y \Rightarrow E(\hat{\beta}) = X^{+}X\beta$, i $Var(\hat{\beta}) = \sigma^2(X'X)^{+}$.

3.1 Quina inversa utilitzem quan $X'X$ és singular?

Doncs desprès de veure tots els apartats anteriors el tipus d'inversa només afecta a l'estimació de β. En el cas de la g-inv hem de seleccionar quines columnes linealment independent per calcular la g-inv, això és un inconvenient ja que primer hem de trobar les columnes linealment independents i a més d'aquestes combinacions n'hi han moltes i cada una ens donarien diferents solucions, les columnes que no es fan servir per calcular g-inv tenen les betes igual a zero. Això dificulta l'explicació del model. Mentre que en la pseudoinversa hem de calcular directament X^+ que es fa amb SVD i és computacionalment eficient. A més podem dir que $\hat{\beta} = X^+Y$ és "òptima"en el sentit de que té la norma mínima i a més ens assegurem que, a diferència d'altres inverses, els elements de $\hat{\beta}$ no es fan arbitràriament grans. Com a conclusió és millor ultilitzar la pseudoinversa.

3.2 Funcions estimables i teorema de Gauss-Markov

El problema que hem vist a l'apartat anterior és que els estimadors de β són esbiaixats quan $X'X$ és singular. En aquesta part veurem que, lo millor que podem fer és estimar $c'\beta$ per certs c' i a més veurem aquests estimadors són únics BLUE.

Definició 3. *Una combinació lineal $a'\beta$ és estimable si té un estimador lineal no esbiaxat, i.e $E[b'Y] = a'\beta$ per algun b, per tot β.*

Lema 9.
1. *$a'\beta$ és estimable sí i només sí $a \in C(X')$*
2. *Si $a'\beta$ és estimable, existeix un únic $b_* \in C(X)$ tal que $a = X'b_*$*

Demostració. 1)$E(b'Y) = b'X\beta = a'\beta \Leftrightarrow a = X'b \Leftrightarrow a \in C(X')$
2) $a'\beta$ és estimable, per (1) $\Rightarrow a = X'b, b \in \mathbb{R}^n$ i $b = b_* + \tilde{b}$ és una descomposició única on $b_* \in C(X)$ i $\tilde{b} \in C(X)^\perp \Rightarrow a = X'b = X'b_* + X'\tilde{b} = X'b_*$ □

Teorema 6 (Gauss-Markov)**.** *Si $a'\beta$ és estimable, aleshores:*
1. *$a'\hat{\beta}$ és únic*
2. *$a'\hat{\beta}$ és BLUE de $a'\beta$*

Notem que el resultat del teorema és independent de la inversa que es faci servir per calcular $\hat{\beta}$.

Demostració. 1) Per lema 9, $a = X'b_*$ per un únic $b_* \in C(X)$. Per tant $a'\hat{\beta} = b_*'X\hat{\beta} = b_*'\hat{Y}$, és únic perquè $\hat{Y}$ és únic.
2) Per teorema 2 $b_*'\hat{Y}$ és BLUE de $b_*'\mu$ per $a'\hat{\beta} = b_*'\hat{Y}$ per (i) i $a'\beta = b_*'X\beta = b_*'\mu$. □

Lema 10. *si $a'\beta$ és estimable aleshores $a'(X'X)^-X'X = a'$ per qualsevol g-inv $(X'X)^-$.*

Demostració. si $a'\beta$ és estimable , aleshores $a = X'b_*, b_* \in C(X)$. Aleshores $a'B^-B = b_*'XB^-X'X = b_*'PX = b_*'X = a'$, independentment de la inversa.
□

Teorema 7. *Si $a'\beta$ és estimable, aleshores* $Var(a'\hat{\beta}) = \sigma^2 a'(X'X)^- a = \sigma^2 b_*' P b_*$

Demostració. $Var(a'\hat{\beta}) = Var(a'B^- X'Y) = a'B^- X'(\sigma^2 I)XB^- a = \sigma^2 a'B^- BB^- a = \sigma^2 a'B^- a$, l'última igualtat per anterior lema. Notem que $a'B^- a = b_*' X(X'X)^- X'b_*' = b_*' P b_*$. $\square$

Observacio: $a = X'b_*, b_* \in C(X) \Leftrightarrow b_* = Xd, d \in \mathbb{R}^p$, podem posar $a = X'Xd, d \in \mathbb{R}^p$ qualsevol i $a'\beta = d'(X'X)\beta$ i el seu blue és $d'(X'X)\hat{\beta}$, per tant només podem estimar combinacions que donen $d'(X'X)$ i $Var(d'(X'X)\hat{\beta}) = d'Pd$.

Referències

[1] I. Ruczinski. *Advanced Methods in Biostatistics 2 (140.752) - Notes* `http://www.biostat.jhsph.edu/~iruczins/teaching/140.751/notes/notes.html`, 2021.

[2] K. Kerr. *Theory of Linear Models.* `http://courses.washington.edu/b533/lecture_notes.html`, 2021

[3] M. Farré. *Regressió múltiple.* Curs d'Estadística Matemàtica (UAB), 2020-2021.

[4] *Generalized inverse.* `https://en.wikipedia.org/wiki/Generalized_inverse`, 2021

[5] *Moore-Penrose inverse.* `https://en.wikipedia.org/wiki/Moore%E2%80%93Penrose_inverse`, 2021

[6] *Proofs involving Moore-Penrose inverse.* `https://en.wikipedia.org/wiki/Proofs_involving_the_Moore%E2%80%93Penrose_inverse`, 2021

[7] George A.F. Seber, Alan J. Lee. *Linear Regression Analysis Second Edition.* WILEY-INTERSCIENCE, Auckland, 2003.