Prédiction des effets aléatoires

- le cas le plus simple. Dans un modèle Gaussien

$Y = (Y_1 - Y_n)$ $\qquad Y_i = \mu + E_i$

$E_i \sim N(0, \sigma^2)$.

E_i : terme d'erreur aléatoire : ce qui reste $\Rightarrow$ résidus après avoir enlevé une partie déterministe

↳ Quel est le meilleur "prédicteur" pour les E_i ?.

$\boxed{\hat{E}_i = Y_i - \overline{Y}}$ = prédicteur linéaire, ss biais

→ Situation triviale, les ennuis commencent lorsqu'il y a plusieurs sources de variabilité.

→ Question intéressante : Quels sont les résidus dans un modèle mixte ? $\left\{ \begin{array}{l} Z\hat{U} + \hat{E} \\ \hat{U} \ ? \\ \hat{\mathbb{E}} \ ? \end{array} \right.$

- Dans le cadre plus général que le modèle mixte : On considère un couple (Y, U) tel que Y est observé, mais pas U.

↳ Comment obtenir de l'info sur U à partir des obs. de Y ?.

→ Dans le cadre des modèles mixtes : On veut estimer G mais aussi l'écart $\hat{U}_i$ au profil moyen $\Rightarrow$ prédict°.

→ On peut substituer une variable aléatoire inobservable par une fonction d'une variable aléatoire observable.

↳ $\hat{U} = f(Y)$, t.q "l'écart" (à définir) entre $\hat{u}$ et u soit faible

→ Dépend des connaissances que l'on a sur la loi jointe de (Y,U), et/ou sur les moments 1er ordre ? 2e ordre ? . .

• Si on connait la loi jointe on peut calculer
$$\mathbb{E}(\hat{U}-U)^2 = \iint (\hat{u}-u) f(u,y)\, dy\, du.$$
(u scalaire).

• mais on peut être ds des situat° où seuls les moments sont connus.

→ 3 méthodes de Prediction:

• Best Prediction : quand on connait la loi jointe de (Y,U)

Pas d'hyp. sur la loi (pas d'hyp de N)

• Best Linear Prediction : qd on connait les 1e & 2e moments

• Best Linear Unbiased Pred : quand on connait le 2e moment (mais pas le 1e).

1)- Meilleur Prédicteur

→ On se donne le critère de l'erreur quadratique.

$$MSE = \mathbb{E}\left(\left[\hat{U} - U\right]^2\right)$$
$$= \mathbb{V}\left(\hat{U} - U\right) + \left[\mathbb{E}(U) - \mathbb{E}(\hat{U})\right]^2$$

↳ On fait la distinction par rapport @ l'estimateur d'un paramètre fixe β.

- pour $\hat{\beta}$ estimateur de β, on souhaite minimiser un critère de variance (max précision)
  ↳ Concerne la variabilité autour d'une valeur fixe β.

- pour $\hat{u}$ prédicteur de u : u est aléatoire !
  ↳ Il faut donc que l'erreur de prédiction prenne en compte la variabilité de la VAR @ prédire

→ $\mathbb{V}(\hat{U} - U)$ prend en compte la variabilité de u.

- On montre que le meilleur prédicteur de U est la $\mathbb{E}(U|Y)$.

→ Revient @ estimer l'espérance conditionnelle.

Preuve : $\mathbb{E}(\hat{U}-U)^2 = \underbrace{V(\hat{U}-U)} + \left[\mathbb{E}(\hat{U}) - \mathbb{E}(U)\right]^2$

On utilise : $V(X) = \mathbb{E}_Y\left(V(X|Y)\right) + V_Y\left(\mathbb{E}(X|Y)\right)$

avec $X = \hat{U} - U$

Or $V_Y\left(\mathbb{E}(U|Y)\right) = 0$ car $\mathbb{E}(U|Y)$ est une fct de Y fixe.

$$\mathbb{E}(\hat{U}-U)^2 = \underbrace{\mathbb{E}_Y\left(V(U|Y)\right)}_{\text{ne dépend pas de } \hat{u}} + \underbrace{V_Y\left(\hat{U} - \mathbb{E}(U|Y)\right)}_{\boxed{\text{nul ssi } \hat{U} = \mathbb{E}(U|Y)}} + \left[\mathbb{E}(\hat{U}) - \mathbb{E}(U)\right]$$

→ $\mathbb{E}(U|Y)$ est un prédicteur sans biais de U

$$\mathbb{E}_Y\left(\mathbb{E}(U|Y)\right) = \mathbb{E}(U)$$

→ $\boxed{MSE\left(\mathbb{E}(U|Y)\right) = \mathbb{E}_Y\left(V(U|Y)\right) = W(\hat{U}-U)}$.

↳ l'erreur de prédict prenant en compte la variabilité de U est égale à la valeur moyenne des variances de U ayant obs. Y.

Autres Propriétés de l'Esp. Conditionnelle.

$$\begin{cases} \mathrm{cov}\left(\mathbb{E}(U|Y),\, u'\right) = \mathbb{V}\left(\mathbb{E}(U|Y)\right) \\ \mathrm{cov}\left(\hat{u}, u\right) = \mathbb{V}(\hat{U}) \\ \mathrm{cov}\left(\mathbb{E}(U|Y),\, Y'\right) = \mathrm{cov}(U, Y') \end{cases}$$

(Searle P263).

$$\boxed{\mathbb{V}(\hat{U}-U) = V(U) - V(\hat{U})}$$

$V(\hat{U}-U) = \mathbb{E}_Y(V(U|Y))$ or $V(U) = \mathbb{E}(V(U|Y)) + V(\mathbb{E}(U|Y))$

2. Meilleurs Prédicteurs linéaires

↳ Dans le BP on ne spécifie pas la forme du prédicteur.

↳ On peut se restreindre aux prédicteurs linéaires

→ On s'intéresse au BLP dans le sens où $\hat{U}$ est une fonction linéaire de la VAR observée.

$$\boxed{\hat{U} = \underset{N\times 1}{a} + b' \left(\underbrace{Y - \mathbb{E}(Y)}_{N\times 1}\right).}$$

⚠ On suppose que $\mathbb{E}(U)$, $\mathbb{E}(Y)$, $\mathbb{V}(U)$, $V(Y)$ $\mathrm{cov}(U,Y)$ sont connus!.

- $\mathbb{E}(\hat{U}-U) = a - \mathbb{E}(U)$
- $\mathbb{V}(\hat{U}-U) = b' V_Y b - 2b' \operatorname{cov}(Y,U) + V_U$.

En général on note $\begin{cases} V(Y) = V \\ V(U) = G \end{cases}$.

- $V(\hat{U}-U) = b'Vb - 2b \operatorname{cov}(Y,U) + G$.

↳ à minimiser par rapport à (a,b).

$$\begin{cases} \dfrac{\partial\, \mathbb{E}(\hat{U}-U)^2}{\partial a} = 2(a - \mathbb{E}(U)) = 0. \\[2ex] \dfrac{\partial\, \mathbb{E}(\hat{U}-U)^2}{\partial b} = 2Vb - 2\operatorname{cov}(Y,U) = 0. \end{cases}$$

$$\begin{cases} a = \mathbb{E}(U) \quad (\text{non biais}). \\ b = V^{-1} \operatorname{cov}(Y,U) \end{cases}$$

$$\boxed{\hat{U} = \mathbb{E}(U) + \operatorname{cov}(U,Y)\, V^{-1} (Y - E(Y))}$$

↳ Valable sans hypothèse de normalité mais avec les moments connus.

Propriétés du BLP :

$$\mathbb{V}(\hat{U}-U) = b'Vb - 2b'\operatorname{cov}(Y,U) + G.$$

avec $b = V^{-1}\operatorname{cov}(Y,U)$.

$$\mathbb{V}(\hat{U}-U) = G + \operatorname{cov}(U,Y)V^{-1}\operatorname{cov}(Y,U)$$

or On sait aussi par la propriété du BP

que $\mathbb{V}(\hat{U}-U) = \mathbb{V}(U) - \mathbb{V}(\hat{U})$

$$\underbrace{\mathbb{V}(\hat{U}-U)}_{G+\operatorname{cov}(U,Y)V^{-1}\operatorname{cov}(Y,U)} = \underbrace{\mathbb{V}(U)}_{G} - \mathbb{V}(\hat{U})$$

$$\rightarrow \boxed{\mathbb{V}(\hat{U}) = \operatorname{cov}(U,Y)V^{-1}\operatorname{cov}(Y,U)}$$

<u>Application du cas Gaussien</u> :

$$\begin{pmatrix} U \\ Y \end{pmatrix} \sim N\left(\begin{matrix}\mu_u \\ \mu_y\end{matrix}, \begin{bmatrix} G & \operatorname{cov}(U,Y) \\ \operatorname{cov}(Y,U) & V \end{bmatrix}\right)$$

$$\mathbb{E}(U|Y) = \mu_u + \operatorname{Cov}(Y,U)V^{-1}(Y-\mu_y)$$

$$\mathbb{V}(\hat{U}-U) = G - \operatorname{Cov}(Y,U)V^{-1}\operatorname{Cov}(U,Y)$$

$$\mathbb{V}(U|Y) = G - \operatorname{cov}(Y,U)V^{-1}\operatorname{cov}(U,Y).$$

↳ découle de : $\mathbb{V}(\hat{U}-U) = E_Y(\mathbb{V}(U|Y))$

la propriété de BP.

3. BLUP

↳ le BLUP est proposé pour relâcher l'hypothèse des 1ers moments connus.

BLP : $\hat{U} = \mathbb{E}(U) + \mathrm{cov}(U,Y)\, V^{-1} \left(Y - \mathbb{E}(Y)\right)$

BLUP : $\mathbb{E}(U) \to k'\beta$, $\mathbb{E}(Y) \to X\beta$.

On suppose que les 1ers moments sont des fct linéaires d'un paramètre β (Inconnu!)

On considère $W = k\beta' + m'U$

↳ W est une combinaison d'effets fixes et d'effets aléatoires.

→ Dans un premier tps on n'a pas besoin des hypothèses de normalité.

→ On souhaite déterminer (k, m) t.q $\hat{W}$ est un BLUP.

. Predicteur linéaire : $\hat{W} = b'Y$.

. " sans biais $\mathbb{E}(\hat{W}) = b'\mathbb{E}(Y)$

$= b'X\beta$

mais $W = k'\beta^* + m'u$

$\mathbb{E}(W) = k'\beta$ $\Rightarrow$ $b'X\beta = k'\beta$

$\Rightarrow$ $\boxed{X'b = k}$ $\Rightarrow$ Contrainte.

. De variance minimale :

$$V(\hat{W} - W) = V(b'Y - m'u)$$

$$= bVb' + mGm' - 2b\,\mathrm{cov}(Y, U)m'.$$

↳ On veut minimiser $V(\hat{W} - W)$ sous la contrainte $X'b = k$ → Multiplicateurs de Lagrange.

. $Q = V(\hat{W} - W) - 2\theta'(X'b - k)$

↑ multiplicateur

$$\begin{cases} \dfrac{\partial Q}{\partial b} = 2bV - 2\operatorname{cov}(Y,U)m - 2\theta' X' = 0. & (1) \\[2ex] \dfrac{\partial Q}{\partial m} = 2mG - 2b\operatorname{cov}(Y,U) = 0. & (2) \\[2ex] \dfrac{\partial Q}{\partial \theta} = -2(X'b - k) = 0. & (3) \end{cases}$$

- <u>On cherche θ</u> : de (1) : $b = V^{-1}\left(\operatorname{cov}(YU)m - X\theta\right)$.

↳ dans (3) : $X'\left(V^{-1}\left[\underbrace{\operatorname{cov}(Y,U)}_{C}m - X\theta\right]\right) = k$.

↳ $X'V^{-1}X\theta = X'V^{-1}Cm - k$.

$$\boxed{\theta = (X'V^{-1}X)^{-}\left[X'V^{-1}Cm - k\right]}$$

On cherche b : t.q.

$b'Y = \left[V^{-1}\left(Cm - X\theta\right)\right]'Y$ (à partir de (1)).

$= m'C'V^{-1}Y - \underbrace{\theta'X'V^{-1}Y}$.

on remplace la valeur de θ.

$$\underbrace{[X'V^{-1}Cm - k]'(X'V^{-1}X)^{-}}_{\theta'} \times X'V^{-1}Y.$$

⚠ $\hat{\beta} = (X'V^{-1}X)^{-}X'V^{-1}Y$

$= $ OLS estimateur de β. (Inconnu).

$\rightarrow \quad \theta' X' V^{-1} Y = -k'\hat{\beta} + m'C'V^{-1}(X\hat{\beta})$

c'est un estimateur de $\mathbb{E}(Y)$!.

↳ Retour sur $\hat{W} = b'Y = (Cm - X\theta)' V^{-1} Y$

$= m'C'V^{-1}Y + k\hat{\beta} - m'C'V^{-1}X\hat{\beta}$.

$$\boxed{\hat{W} = \underbrace{k\hat{\beta}}_{\text{BLUE}(\beta)} + m' \underbrace{C'V^{-1}(Y - X\hat{\beta})}_{\hat{u} = \text{BLUP}(U)}}$$

Mais ⚠ . On n'a supposé que les 2^e moments connus uniquement.

. les 1^{er} moments sont estimés $X\hat{\beta}$

. V, C sont connus.

. V^{-1} à calculer !!!

4- Méthode d'Henderson et Équations du modèle Mixte

- On se place dans le modèle Gaussien.

$$Y = X\beta + ZU + E \qquad U \perp\!\!\!\perp E.$$

$U \to N(0,G)$, $\quad E \to N(0,R)$

$$V = V(Y) = Z'GZ + R.$$

- l'Objectif est de developper une méthode pour estimer BLUE $(X\beta)$ et BLUP (U)

- Idée : Maximiser $\log f(Y,U) = \log f(Y|U) + \log f(U)$

$$\begin{cases} Y|U \sim N(X\beta + ZU, R) \\ U \sim N(0,G) \end{cases}$$

- $-2 \log f(Y/U) = N \log 2\pi + \log|R| + \|Y - X\beta - ZU\|^2_{R^{-1}}$

$-2 \log f(U) = q \log 2\pi + \log|G| + u'G^{-1}u.$

$$\log f(Y,U) \propto \log|R| + \log|G| + \|Y - X\beta - ZU\|^2_{R^{-1}} + u'G^{-1}u.$$

Idée : ne pas considérer U comme une Var mais comme un paramètre !

$$\frac{\partial -2\log f(Y|U)}{\partial \beta} = -2X'R^{-1}(Y - X\beta - ZU) = 0$$

$$\frac{\partial -2\log f(U)}{\partial u} = -2Z'R^{-1}(Y - X\beta - ZU) + 2G^{-1}U = 0.$$

$\hookrightarrow$
$$\begin{bmatrix} X'R^{-1}X & XR^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'R^{-1}Y \\ Z'R^{-1}Y \end{bmatrix}.$$

$\rightarrow$ Système d'Henderson ou équations du modèle mixte ;

Intérêt calculatoire : pas d'inverse de V !

Solutions du système : $\begin{cases} GLS : \hat{\beta} \\ BLUP : \hat{U} \end{cases}$!

Rq : si G^{-1} n'était pas là, on aurait les équat° du ML pour $Y = X\beta + ZU + E$ avec U fixé.

$\hookrightarrow$ Ce système est "régularisé".

. Pour resoudre le système on doit inverser une matrice structurée par bloc.

↳ On utilise des resultats d'Algèbre sur les Compléments de Schur (Searle 453 Harville 101).

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}^{-1} = \begin{bmatrix} A^{11} & A^{12} \\ A^{21} & A^{22} \end{bmatrix}$$

A_{11}, A_{22} : carrées non singulières.

1). $A^{11} = \left(A_{11} - A_{12} A_{22}^{-1} A_{21} \right)^{-1}$

$= A_{11}^{-1} + A_{11}^{-1} A_{12} A^{22} A_{21} A_{11}^{-1}$.

2) $A^{12} = - A_{11}^{-1} A_{12} A^{22} = - A^{11} A_{12} A_{22}^{-1}$

Système d'Henderson

$$\begin{bmatrix} R^{-1} & R^{-1} Z \\ Z' R^{-1} & Z' R^{-1} Z + G^{-1} \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}.$$

- terme $A^{22} = \left[A_{22} - A_{21} A_{11}^{-1} A_{12}\right]^{-1}$

$$= \left[Z'R^{-1}Z + G - Z'R^{-1} R R^{-1} Z\right]^{-1}$$

$$= G. \rightarrow \text{donne le terme } \begin{bmatrix} A^{11} & A^{12} \\ A^{21} & G \end{bmatrix}$$

- terme $A^{11} = \underbrace{\left[R^{-1} - R^{-1} Z \left(Z'R^{-1}Z + G^{-1}\right)^{-1} Z'R^{-1}\right]^{-1}}_{\left(A_{11} - A_{12} A_{22}^{-1} A_{21}\right)^{-1}.}$

$$A^{11} = \underbrace{R + RR^{-1} Z G Z' R^{-1} R}_{A_{11}^{-1} + A_{11}^{-1} A_{12} A^{22} A_{21} A_{11}^{-1}.} = R + ZGZ' = V.$$

$$\hookrightarrow \boxed{V = \left[R^{-1} - R^{-1}Z\left[Z'R^{-1}Z + G^{-1}\right]^{-1} Z'R^{-1}\right]^{-1}.}$$

$$= W^{-1}.$$

- $A^{12} = -A_{11}^{-1} A_{12} A^{22} = -RR^{-1} Z G = -ZG$

$$= -A^{11} A_{12} A_{22}^{-1} = -VR^{-1} Z \left(Z' R^{-1} Z + G^{-1}\right)^{-1}$$

$$\boxed{G Z' V^{-1} = \left(Z' R^{-1} Z + G^{-1}\right)^{-1} Z' R^{-1}}$$

Retour sur les Equations du modèle mixte.

$$\begin{cases} X'R^{-1}X\hat{\beta} = X'R^{-1}(Y - Z\hat{U}) & \text{(1) : "résidus" de } \hat{U} \\ (Z'R^{-1}Z + G^{-1})\hat{u} = Z'R^{-1}(Y - X\hat{\beta}) & \text{"résidus" de } \hat{\beta} \end{cases}$$

(1). On veut se débarasser de $Z\hat{U}$.

↳ on prend la 2^e equation

↳ Donne $X'WX\hat{\beta} = X'WY$.

avec $W = R^{-1} - R^{-1}Z(Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1}$.

Or ! $W = V^{-1}$. $\Rightarrow X'V^{-1}X\hat{\beta} = X'V^{-1}Y$.

$$\boxed{\hat{\beta} = (X'V^{-1}X)^{-}X'V^{-1}Y} \rightarrow \text{gls for } \beta.$$

↳ BLUE(β)

(2) Pour trouver $\hat{u}$:

$$\hat{u} = \underbrace{(Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1}}(Y - X\hat{\beta}).$$

$$= GZ'V^{-1}(Y - X\hat{\beta}) \quad \text{(d'après le comp. de Schur)}.$$

↳ c'est bien le BLUP !

Interprétation du prédicteur

$$\hat{Y} = X\hat{\beta} + Z\hat{U}$$
$$= X\hat{\beta} + Z\left((Z'R^{-1}Z + G^{-1})^{-1} Z'R^{-1} (Y - X'\hat{\beta}) \right)$$

or $V^{-1} = R^{-1} - R^{-1}Z\,(Z'R^{-1}Z + G^{-1})^{-1} Z'R^{-1}$

$$RV^{-1} = I - Z(\ )^{-1} Z'R^{-1}.$$

$\hookrightarrow$ $\hat{Y} = \underbrace{RV^{-1}}\, X\hat{\beta} + (I - RV^{-1})Y.$

c'est un poids

$$= \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + \sigma_u^2} = \frac{\text{V. de l'erreur}}{\text{V. totale}}.$$

↗ peu de variabilité interindividuelle : $\sigma_u^2 \ll \sigma_\varepsilon^2$

→ On prend l'effet fixe $X\hat{\beta}$

↘ bcp de dispersion

→ On garde la donnée Y.