Estimation de la valeur génétique des candidats reproducteurs

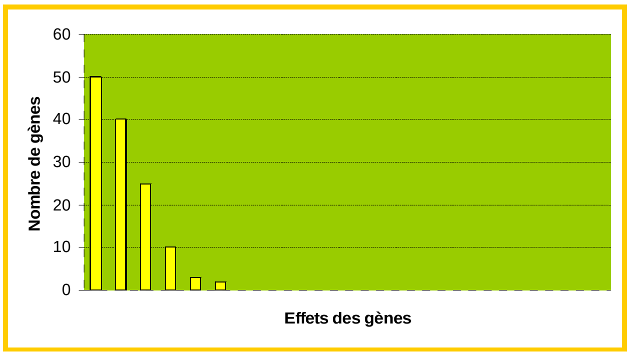
BLUP et Equations du Modèle Mixte d'Henderson

ALIMENTATION
AGRICULTURE
ENVIRONNEMENT

C. Robert-Granié, Novembre 2011

Modèle Polygénique infinitésimal

Le caractère est gouverné par une infinité (un très grand nombre) de gènes, chaque gène ayant un effet infinitésimal



ALIMENTATION
AGRICULTURE
ENVIRONNEMENT

C. Robert-Granié, Novembre 2011

Modèle Polygénique infinitésimal

La valeur phénotypique (P) s'exprime comme : $P = G + E$

G : valeur génétique (effet moyen du génotype sur un caractère donné)
E : effet de l'environnement

En l'absence d'interaction génotype x environnement, on a :

$$\mathrm{Var}(P) = \mathrm{Var}(G) + \mathrm{Var}(E) \qquad \mathrm{Cov}(G,E) = 0$$

$$G = A + D + I$$

A : valeur génétique additive
D : valeur de dominance
I : valeur de l'épistasie

Seule A se transmet de parent aux descendants

Modèle Polygénique infinitésimal

Valeur génétique additive d'un descendant connaissant celles de ses parents

Aléa de méiose : d'un descendant à l'autre, ce ne sont pas exactement les mêmes gènes qui sont transmis par les parents E(w) = 0

$$A_{desc} = \tfrac{1}{2} A_{père} + \tfrac{1}{2} A_{mère} + w$$

Le père (la mère) transmet la moitié de ses gènes, la moitié de sa valeur génétique additive

Le sélectionneur s'intéresse à A

Un caractère quantitatif est gouverné par un grand nb de loci à effets individuels faibles et indépendants.
Par application du théorème limite centrale, la somme des effet moyens de ces gènes suit une distribution normale : $A \sim N(0, \sigma_a^2)$

Objectif de l'amélioration génétique des animaux

Sélectionner au sein d'une population, c'est mettre à la reproduction les « meilleures » femelles avec les « meilleurs » mâles.

Les « meilleur(e)s » sont ceux (celles) qui portent les meilleurs gènes, c'est-à-dire ceux (celles) qui ont les plus fortes valeurs génétiques additives (*A*) pour les caractères que l'on souhaite améliorer.

Les « meilleur(e)s » sont ceux (celles) qui transmettront la plus forte supériorité.

Il faut **estimer / prédire** les valeurs génétiques additives (*A*) .

Indexation : technique de calcul permettant d'obtenir l'estimation de la valeur génétique additive des reproducteurs

(ou index ou EBV=Estimated Breeding value)

Informations utilisées pour l'évaluation Génétique

- Les données (y) : performances ou phénotypes
- Pedigree et généalogie des animaux (individu, père, mère)
- Les effets de milieu identifiés (ex: troupeau, année, saison, …)

- Des effets de milieu non identifiés → résidus d'un modèle
- La valeur génétique additive → inconnue

Performances = valeur génétique additive + milieux identifiés + milieux non identifiés

$$\mathbf{y} \qquad \mathbf{a} \sim N(0, \mathbf{A}\sigma_a^2) \qquad \boldsymbol{\beta} \qquad \mathbf{e} \sim N(0, \mathbf{I}\sigma_e^2)$$

Perf. du candidat

Perf. des ascendants, descendants, collatéraux

Evaluation Génétique

Le BLUP modèle animal est actuellement la méthode d'évaluation génétique de référence. Il présente deux propriétés très intéressantes :

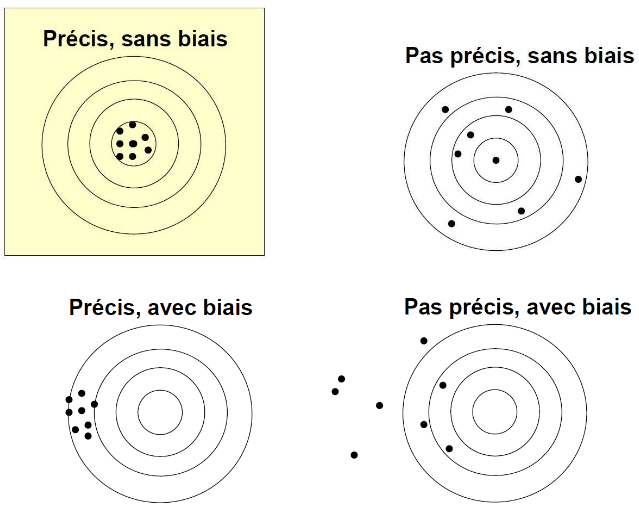
☞ Estimation simultanée des valeurs génétiques et des effets du milieu

PAS DE BIAIS

☞ Utilisation de l 'ensemble des informations disponibles

PRECISION MAXIMALE

Qu'est-ce qu'un estimateur précis et sans biais ?



C. Robert-Granié, Novembre 2011
ALIMENTATION AGRICULTURE ENVIRONNEMENT

Afin d'avoir la meilleure prédiction possible,

Il est nécessaire de collecter l'ensemble des informations le plus précisément possible et de manière exhaustive.

Ignorer certains effets environnementaux qui ont une influence sur la performance conduit à des estimations biaisées des autres effets (effets fixes et valeur génétique) et à des surestimations de la précision de l'évaluation → erreur sur le classement final des animaux.

Si on introduit des effets inutiles dans le modèles, il n'y aura pas d'impact sur le biais mais des impacts sur la précision des valeurs génétiques.

Difficulté essentielle de l'évaluation = bien séparer les effets génétiques des effets de l'environnement.

Rappel sur la notion de meilleur prédicteur

Soit 2 variables aléatoires de distributions quelconques, Z1 et Z2 .

Z_1 : variable observée et Z_2 variable à prédire.

La relation entre le prédicteur $(\hat{Z}_2)$ que l'on cherche et la vraie valeur est donnée par:

$$Z_2 = \hat{Z}_2 + e \text{ avec } e : \text{résidus (ou erreurs)}$$

Le meilleur prédicteur est celui qui minimise l'espérance du carré des erreurs.

Le meilleur prédicteur est l'espérance de Z_2 sachant les données Z_1.

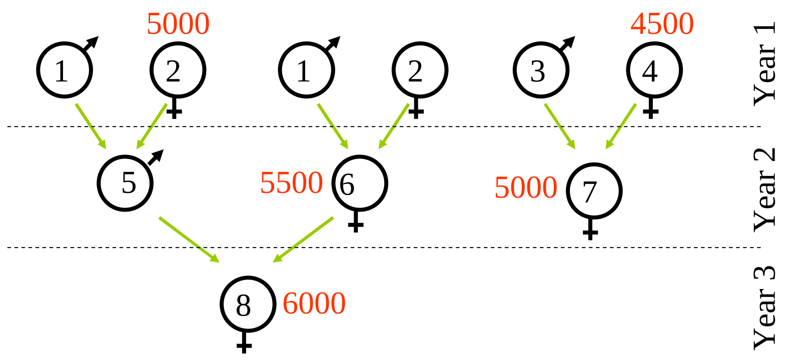
$$\hat{Z}_2 : \text{ min de } \mathrm{E}(\mathrm{e}^2) = E\left[(\hat{Z}_2 - Z_2)^2\right]$$

$$\hat{Z}_2 = E(Z_2 | Z_1)$$

Par construction, E(e)=0 : estimateur sans biais ; Var(e) minimale ; e et $\hat{Z}_2$ indépendants

Exemple

Caractère : Production laitière mesurée sur une lactation de 5 femelles, réalisées au cours de 3 années différentes



On souhaite prédire la valeur génétique additive de l'ensemble de ces animaux

Modèle

$$y_i = \beta_j + a_i + e_i \text{ ou matriciellement } \mathbf{y}=\mathbf{X\beta}+\mathbf{Za}+\mathbf{e}$$

$$y_2(=5000) = \beta_1 + a_2 + e_2 \qquad y_6(=5500) = \beta_2 + a_6 + e_6$$
$$y_4(=4500) = \beta_1 + a_4 + e_4 \qquad y_7(=5000) = \beta_2 + a_7 + e_7$$
$$y_8(=6000) = \beta_3 + a_8 + e_8$$

$$\begin{bmatrix} y_2 \\ y_4 \\ y_6 \\ y_7 \\ y_8 \end{bmatrix} = \begin{bmatrix} 5000 \\ 4500 \\ 5500 \\ 5000 \\ 6000 \end{bmatrix} = \begin{bmatrix} 100 \\ 100 \\ 010 \\ 010 \\ 001 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{array}{c} \text{Ind. } 1\;2\;3\;4\;5\;6\;7\;8 \\ \begin{bmatrix} 01000000 \\ 00010000 \\ 00000100 \\ 00000010 \\ 00000001 \end{bmatrix} \end{array} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \\ a_7 \\ a_8 \end{bmatrix} + \begin{bmatrix} e_2 \\ e_4 \\ e_6 \\ e_7 \\ e_8 \end{bmatrix}$$

Meilleur prédicteur

Nous voulons estimer/prédire la valeur génétique additive (a_i) des 8 animaux, à partir des observations (performances) (**y**) des 5 femelles.

On considère la distribution jointe des données (**y**)
et des effets génétiques additifs (**a**)

- Cas 1: la distribution jointe est entièrement connue, (distribution normale)

➔ $\hat{\mathbf{a}} = E[\mathbf{a}|\mathbf{y}]$ = Meilleur Prédicteur de **a** (Cochran, 1951)

Cas 1: la distribution jointe de y et a est entièrement connue, distribution multinormale. On connaît la loi, les espérances et les variances.

Dans ce cadre, on a :

Cov(a_i,y*)

$$\begin{bmatrix} a_i \\ \mathbf{y}^* \end{bmatrix} = \begin{bmatrix} a_i \\ \mathbf{y} - \mathbf{X}\beta \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_a^2 & c_i \\ c_i' & V \end{bmatrix} \right)$$

Var(y*)

D'après ces connaissances, on a la propriété suivante :

$$(a_i|\mathbf{y}^*) \sim N(\mu_a + c_i' V^{-1}(\mathbf{y} - X\beta), V - c_i' \sigma_a^{-2} c_i)$$

$$\hat{a}_i = c_i' V^{-1}(\mathbf{y} - X\beta)$$

Dans ce cas, on suppose que l'on connaît avec exactitude les effets fixes et les distributions des observations et valeurs génétiques → ce qui est rarement le cas

Cas 1: la distribution jointe de y et a est entièrement connue, distribution multinormale. On connaît la loi, les espérances et les variances.

Exemple :

Dans le cas de la prédiction de la valeur génétique d'un individu à partir de sa performance propre, on a :

$$y_i = y, X\beta = \mu, V = \sigma_y^2, c_i = \mathrm{cov}(y, a_i) = \mathrm{cov}(a_i + e_i, a_i) = \sigma_a^2$$

$$\hat{a}_i = \frac{\sigma_a^2}{\sigma_y^2}(\mathbf{y} - \mu) = h^2(\mathbf{y} - \mu)$$

Meilleur prédicteur linéaire (BLP)

Cas 2: la distribution jointe est inconnue mais seuls les 2 premiers moments de la loi jointe sont connus

$$E\begin{bmatrix} a_i \\ \mathbf{y}^* \end{bmatrix} = E\begin{bmatrix} a_i \\ \mathbf{y} - \mathbf{X}\beta \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{et Var}\begin{bmatrix} a_i \\ \mathbf{y} \end{bmatrix} = \begin{bmatrix} \sigma_a^2 & \mathbf{c}_i \\ \mathbf{c}_i' & \mathbf{V} \end{bmatrix}$$

avec σ_a^2, $\mathbf{c}_i$ et $\mathbf{V}$ sont supposés connus

Objectif : recherche d'un prédicteur d'une variable aléatoire non observable à partir d'observations **y**.

➔ $\hat{\mathbf{a}}$ = Best Linear Prediction de **a** (BLP)

Selection index theory (Smith, 1936; Hazel, 1943)

Best Linear Prediction
(= Selection Index Theory)

$$\mathbf{y}^* = \mathbf{Z}\ \mathbf{a} + \mathbf{e} \quad (\text{or } \mathbf{y}^* = \mathbf{y} - \mathbf{X}\beta = \mathbf{Z}\ \mathbf{a} + \mathbf{e})$$

Se restreindre à une classe de prédicteur : Prédicteurs Linéaires

➔ Le meilleur $\hat{a}_i = b_1 y_1 + b_2 y_2 + \ldots + b_n y_n = \mathbf{b}'\mathbf{y}^*$

➔ minimise $E\left[(\hat{a}_i - a_i)^2\right]$

La valeur minimale de $E\left[(\hat{a}_i - a_i)^2\right]$ est obtenue lorsque ses dérivées partielles par rapport aux inconnues (**b**) sont nulles.

Calcul du Meilleur Prédicteur Linéaire

Nous voulons $\dfrac{\partial E\left[(\hat{a}_i - a_i)^2\right]}{\partial \mathbf{b}} = 0$

$$\begin{aligned}
E\left[(\hat{a}_i - a_i)^2\right] &= E\left[(\mathbf{b}'\mathbf{y} - a_i)^2\right] \\
&= E\left[(\mathbf{b}'\mathbf{y} - a_i)(\mathbf{b}'\mathbf{y} - a_i)'\right] \\
&= E\left[\mathbf{b}'\mathbf{y}\mathbf{y}'\mathbf{b} - a_i\mathbf{y}'\mathbf{b} - \mathbf{b}'\mathbf{y}a_i + a_i^2\right] \\
&= \mathbf{b}'E\left[\mathbf{y}\mathbf{y}'\right]\mathbf{b} - 2\mathbf{b}'E[\mathbf{y}a_i] + E[a_i^2] \\
&= \mathbf{b}'\mathbf{V}\mathbf{b} - 2\mathbf{b}'\mathbf{c}_i + \sigma_a^2
\end{aligned}$$

car :

$\mathbf{V} = E[\mathbf{y}\mathbf{y}'] - \cancel{E[\mathbf{y}]}\cancel{E[\mathbf{y}']} = E[\mathbf{y}\mathbf{y}']$

$\mathbf{c}_i = E[\mathbf{y}a_i] - \cancel{E[\mathbf{y}]}\cancel{E[a_i]} = E[\mathbf{y}a_i]$

$\sigma_a^2 = E[a_i^2] - \cancel{E[a_i]^2} = E[a_i^2]$

$$\frac{\partial E\left[(\hat{a}_i - a_i)^2\right]}{\partial \mathbf{b}} = 0 \ \text{➔}\ 2\mathbf{V}\mathbf{b} - 2\mathbf{c}_i = 0 \Rightarrow \mathbf{b} = \mathbf{V}^{-1}\mathbf{c}_i$$

➡ $\boxed{\hat{a}_i = \mathbf{b}'\mathbf{y}^* = \mathbf{c}_i'\mathbf{V}^{-1}\mathbf{y}^* = \mathbf{c}_i'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\beta)}$

Application du BLP (1/3)

- Exemple 1:

  y = propre perf. (sélection massale)

$$\mathbf{V} = \mathrm{Var}(\mathbf{y}) = \sigma_y^2 \;, \qquad c_i^{'} = \mathrm{cov}(a_i, y_i) = \mathrm{cov}(a_i, a_i + e_i) = \mathrm{var}(a_i) = \sigma_a^2$$

$$\hat{a}_i = \mathbf{c}_i^{'}\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\beta)$$

$$\hat{a}_i = \frac{\sigma_a^2}{\sigma_y^2}(y_i - \mathbf{X}\beta) = h^2(y_i - \mathbf{X}\beta)$$

Application du BLP (2/3)

- Example 2: évaluation père = observations de n filles du père i

  $\mathbf{y} = (y_1, y_2, \ldots y_n)'$ avec $y_j = \frac{1}{2}a_i + e_j^*$

Nous voulons : $\hat{a}_i = \mathbf{b}'\mathbf{y} = b_1 y_1 + b_2 y_2 + \ldots + b_n y_n$

$= b_0\, y_1 + b_0 y_2 + \ldots + b_0 y_n$ (toutes les obs jouent le même rôle)

$= b\left(\frac{y_1 + y_2 + \ldots + y_n}{n}\right) = b\bar{y}$ (avec $b = nb_0$)

($\bar{y}$ résume toute l'information)

Quel est ce b ?

Application du BLP (3/3)

$$\text{car } h^2 = \frac{\sigma_a^2}{\sigma_y^2}$$

$$c_i = \mathrm{cov}(a_i, \bar{y}) = \mathrm{cov}\left(a_i, \frac{y_1 + y_2 + \ldots + y_n}{n}\right)$$

$$= \frac{1}{n}\left(\sum_{j=1}^{n} \mathrm{cov}(a_i, y_j)\right) = \mathrm{cov}(a_i, y_j) = \mathrm{cov}(a_i, \frac{1}{2}a_i + e_j^*) = \frac{1}{2}\sigma_a^2 = \frac{1}{2}h^2\sigma_y^2$$

(car j est une fille de i)

$$\mathbf{V} = \mathrm{Var}(\bar{y}) = \mathrm{Var}\left(\frac{y_1 + y_2 + \ldots + y_n}{n}\right)$$

$$= \frac{1}{n^2}\left(\sum_{j=1}^{n} \mathrm{var}(y_j) + \sum_{j=1}^{n} \sum_{k=1, j \neq k}^{n} \mathrm{cov}(y_j, y_k)\right) = \frac{1}{n^2}\left(n\,\mathrm{var}(y_j) + n(n-1)\mathrm{cov}(y_j, y_k)\right)$$

$$= \frac{1}{n}\left(\sigma_y^2 + (n\text{-}1)\frac{1}{4}\sigma_a^2\right) = \frac{1 + (n-1)\frac{h^2}{4}}{n}\sigma_y^2$$

(car j et k sont filles de i)

$$\boxed{\mathrm{cov}(y_j, y_k) = \mathrm{cov}(\frac{1}{2}a_i, \frac{1}{2}a_i) = \frac{1}{4}\sigma_a^2}$$

$$\Rightarrow \quad \hat{a}_i = \mathbf{c}_i' \mathbf{V}^{-1} \bar{y} = \frac{0.5\ nh^2}{1 + \ 0.25\,(n-1)h^2}\bar{y}$$

Propriétés du BLP

$$\hat{a}_i = \mathbf{b}'\mathbf{y}^* = \mathbf{c}_i'\mathbf{V}^{-1}\mathbf{y}^* = \mathbf{c}_i'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\beta)$$

C'est un prédicteur non biaisé : $E[\hat{a}_i] = E[a_i]$

Si on calculait un grand nombre de fois $\hat{a}_i$ à partir d'observations indépendantes, en moyenne cette valeur génétique estimée serait égale à la vraie valeur

L'estimée $\hat{a}_i$ maximise la corrélation entre valeur prédite et valeur vraie

Prédiction qui maximise la probabilité d'un classement correct des valeurs génétiques des animaux pris 2 à 2.

Inconvénients du BLP

Les hypothèses peuvent être mise à défaut

- connaissance a priori de l'espérance des observations est nulle → cela suppose une correction parfaite des données pour les effets fixes.

- différences de niveau génétique moyen des vaches accouplées aux divers taureaux

Meilleur prédicteur Linéaire Non Biaisé (BLUP)

Cas 3: la distribution jointe et les espérances sont inconnues ; seules les variances et covariances sont connues

→ paramètres génétiques (h^2…)

➔ $\hat{\mathbf{a}}$ = Best Linear Unbiased Prediction de **a** (BLUP)

Différence avec BLP : abandon de l'hypothèse de connaissance de l'espérance des observations $E[y^*]=0$

BLUP : hypothèses

- Modèle mixte: $\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\ \mathbf{a}\ +\ \mathbf{e}$
- Var(**y**) = **V** , Var(**a**) = **G** et Cov(**y**, **a'**) = **C** sont connues
- Objectif : on veut estimer des combinaisons linéaires $\omega_i = k_i^{'}\boldsymbol{\beta} + m_i^{'}\mathbf{a}$

$$\begin{bmatrix}\omega_1\\ \omega_2\\ \vdots\\ \omega_n\end{bmatrix} = \begin{bmatrix}k_{11} & k_{12} & \cdots & k_{1p}\\ k_{21} & k_{22} & \cdots & k_{2p}\\ \vdots & \vdots & \ddots & \vdots\\ k_{n1} & k_{n2} & \cdots & k_{np}\end{bmatrix}\begin{bmatrix}\beta_1\\ \beta_2\\ \vdots\\ \beta_n\end{bmatrix} + \begin{bmatrix}m_{11} & m_{12} & \cdots & m_{1p'}\\ m_{21} & m_{22} & \cdots & m_{2p'}\\ \vdots & \vdots & \ddots & \vdots\\ m_{n1} & m_{n2} & \cdots & m_{np'}\end{bmatrix}\begin{bmatrix}a_1\\ a_2\\ \vdots\\ a_{p'}\end{bmatrix}$$

- Exemples:

$$\omega_i = [0 \ \cdots \ 0 \ \cdots \ 0]\boldsymbol{\beta} + [0 \ \cdots \ 1 \ \cdots \ 0]\mathbf{a} = a_i$$

$$\omega_i = [0 \ \cdots \ 0]\boldsymbol{\beta} + \begin{bmatrix}0 & 0 & \cdots & \frac{1}{N} & \frac{1}{N} & \frac{1}{N} & \cdots & 0 & 0\end{bmatrix}\mathbf{a} = \bar{a}_{\text{year } q}$$

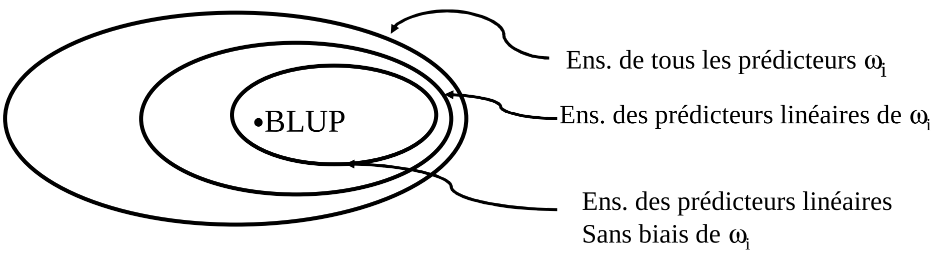
$$\omega_i = [0 \ \cdots \ 1 \ \cdots \ 0]\boldsymbol{\beta} + [0 \ \cdots \ 1 \ \cdots \ 0]\mathbf{a} = \beta_z + a_i$$

BLUP : objectif

On cherche à prédire : $\omega_i = k_i^{'}\boldsymbol{\beta} + m_i^{'}\mathbf{a}$ par $\hat{\omega}_i$

Le prédicteur $\hat{\omega}_i$ de ω_i est recherché :

- parmi les prédicteurs linéaires ➔ $\hat{\omega}_i = \mathbf{b}'\mathbf{y}$
- d'erreur quadratique moyenne minimum ➔ minimize $E\left[(\hat{\omega}_i - \omega_i)^2\right]$
- sans biais ➔ $E[\hat{\omega}_i] = E[\omega_i]$



BLUP : quelques remarques

- Aucune hypothèse n'est faite sur la distribution des obs. y
- Estimateur de $\boldsymbol{\beta}$ = Best Linear Unbiased Estimator (BLUE)
- $\mathrm{Var}(\hat{\omega}_i - \omega_i) = \mathrm{E}\left[(\hat{\omega}_i - \omega_i)^2\right] - \mathrm{E}\left[\hat{\omega}_i - \omega_i\right]^2$

  mais $\mathrm{E}[\hat{\omega}_i] = \mathrm{E}[\omega_i] \quad \Rightarrow \mathrm{E}[\hat{\omega}_i - \omega_i] = 0$

  $\Rightarrow$ Minimiser $\mathrm{E}\left[(\hat{\omega}_i - \omega_i)^2\right] \Leftrightarrow$ Minimiser $\mathrm{Var}(\hat{\omega}_i - \omega_i)$

La fonction à minimiser $\mathrm{E}\left[(\hat{\omega}_i - \omega_i)^2\right]$ sous la contrainte $\mathrm{E}[\hat{\omega}_i] = \mathrm{E}[\omega_i]$

Rappel : $\mathrm{Var}(z) = \mathrm{E}(z^2) - \mathrm{E}(z)^2$

Calcul du BLUP (1/3)

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\ \mathbf{a}\ +\ \mathbf{e}$$

- On sait que: $\mathrm{E}[\mathbf{y}] = \mathbf{X}\boldsymbol{\beta}\ ,\ \mathrm{E}[\mathbf{a}] = \mathbf{0}$

  donc

$$\mathbf{V} = \mathrm{Var}(\mathbf{y}) = \mathrm{E}[\mathbf{yy}'] - \mathrm{E}[\mathbf{y}]\mathrm{E}[\mathbf{y}]' = \mathrm{E}[\mathbf{yy}'] - \mathbf{X}\boldsymbol{\beta\beta}'\mathbf{X}'$$

$$\mathbf{G} = \mathrm{Var}(\mathbf{a}) = \mathrm{E}[\mathbf{aa}'] - \mathrm{E}[\mathbf{a}]\mathrm{E}[\mathbf{a}]' = \mathrm{E}[\mathbf{aa}']$$

$$\mathbf{C} = \mathrm{Cov}(\mathbf{y},\mathbf{a}') = \mathrm{E}[\mathbf{ya}'] - \mathrm{E}[\mathbf{y}]\mathrm{E}[\mathbf{a}]' = \mathrm{E}[\mathbf{ya}']$$

- Contrainte de non biais :

$$\mathrm{E}[\hat{\omega}_i] = \mathrm{E}[\omega_i] \quad \Rightarrow \mathrm{E}\left[\mathbf{b}_i^{'}\mathbf{y}\right] = \mathrm{E}\left[\mathbf{k}_i^{'}\boldsymbol{\beta} + \mathbf{m}_i^{'}\mathbf{a}\right]$$

$$\Rightarrow \mathbf{b}_i^{'}\mathrm{E}[\mathbf{y}] = \mathbf{k}_i^{'}\boldsymbol{\beta} + \mathbf{m}_i^{'}\mathrm{E}[\mathbf{a}] = \mathbf{k}_i^{'}\boldsymbol{\beta}$$

$$\Rightarrow \mathbf{b}_i^{'}\mathbf{X}\boldsymbol{\beta} = \mathbf{k}_i^{'}\boldsymbol{\beta} \text{ pour tout } \boldsymbol{\beta}$$

$$\Rightarrow \mathbf{b}_i^{'}\mathbf{X} = \mathbf{k}_i^{'}$$

Calcul du BLUP (2/3)

- Erreur quadratique moyenne :

$$E\left[(\hat{\omega}_i-\omega_i)^2\right]=E\left[(\mathbf{b}_i'\mathbf{y}-(\mathbf{k}_i'\boldsymbol{\beta}+\mathbf{m}_i'\mathbf{a}))^2\right]$$
$$=\mathbf{b}_i'E[\mathbf{yy}']\mathbf{b}_i\ \ -2\mathbf{b}_i'E[\mathbf{y}]\boldsymbol{\beta}'\mathbf{k}_i\ \ -2\mathbf{b}_i'E[\mathbf{ya}']\mathbf{m}_i$$
$$+\mathbf{k}_i'\boldsymbol{\beta\beta}'\mathbf{k}_i\ +\ 2\mathbf{k}_i'\boldsymbol{\beta}E[\mathbf{a}']\mathbf{m}_i+\mathbf{m}_i'E[\mathbf{aa}']\mathbf{m}_i$$

$$E\left[(\hat{\omega}_i-\omega_i)^2\right]=\ \mathbf{b}_i'(\mathbf{V}+\mathbf{X}\boldsymbol{\beta\beta}'\mathbf{X}')\mathbf{b}_i\ \ -2\,\mathbf{b}_i'\mathbf{X}\boldsymbol{\beta\beta}'\mathbf{X}'\mathbf{b}_i\ \ -2\mathbf{b}_i'\mathbf{C}\mathbf{m}_i$$
$$+\mathbf{b}_i'\mathbf{X}\boldsymbol{\beta\beta}'\mathbf{X}'\mathbf{b}_i\ +\ \mathbf{0}\ +\mathbf{m}_i'\mathbf{G}\mathbf{m}_i$$

$$E\left[(\hat{\omega}_i-\omega_i)^2\right]=\ \mathbf{b}_i'\mathbf{V}\mathbf{b}_i\ -2\mathbf{b}_i'\mathbf{C}\mathbf{m}_i+\mathbf{m}_i'\mathbf{G}\mathbf{m}_i$$

$\uparrow$

$\boxed{\text{La fonction à minimiser}}$

Calcul du BLUP (3/3)

- Rappel : Contrainte de minimisation:
- Trouver le minimum d'une fonction h(**x**) sous la contrainte g(**x**)=0:

Soit **Q(x, λ)** = h(**x**) + **λ** g(**x**) (où λ = Lagrange multiplier)

$$\frac{\partial\mathbf{Q}(\mathbf{x},\lambda)}{\partial\mathbf{x}}=0\quad\text{et}\quad\frac{\partial\mathbf{Q}(\mathbf{x},\lambda)}{\partial\lambda}=0$$

- Ici, la contrainte est : $E[\hat{\omega}_i]=E[\omega_i]\ \ \Rightarrow\mathbf{b}_i'\mathbf{X}=\mathbf{k}_i'$ et la fonction à minimiser est : $E\left[(\hat{\omega}_i-\omega_i)^2\right]=\ \mathbf{b}_i'\mathbf{V}\mathbf{b}_i\ -2\mathbf{b}_i'\mathbf{C}\mathbf{m}_i+\mathbf{m}_i'\mathbf{G}\mathbf{m}_i$

$$\mathbf{Q}(\mathbf{b}_i,\lambda)=\ \mathbf{b}_i'\mathbf{V}\mathbf{b}_i\ -2\mathbf{b}_i'\mathbf{C}\mathbf{m}_i+\mathbf{m}_i'\mathbf{G}\mathbf{m}_i+2\left(\mathbf{b}_i'\mathbf{X}-\mathbf{k}_i'\right)\lambda$$

$$\begin{cases}\dfrac{\partial\mathbf{Q}}{\partial\lambda}=\mathbf{b}_i'\mathbf{X}-\mathbf{k}_i'=0\\[2ex]\dfrac{\partial\mathbf{Q}}{\partial\mathbf{b}_i}=\mathbf{V}\mathbf{b}_i-\mathbf{C}\mathbf{m}_i+\mathbf{X}\lambda=0\end{cases}$$

Solution du BLUP

- Dans l'équation 2, exprimer $\mathbf{b}_i$ en fonction de λ
- Remplacer $\mathbf{b}_i$ dans l'équation 1
- Résoudre l'équation 1, pour obtenir λ
- Remplacer λ dans la seconde équation par sa valeur ➔ obtenir $\mathbf{b}_i$

➢ Solution:

$$\hat{\omega} = \mathbf{b}_i^{'}\mathbf{y} = \mathbf{k}_i^{'}\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} + \mathbf{m}_i^{'}\mathbf{C}'\mathbf{V}^{-1}\left[\mathbf{y}-\mathbf{X}\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}\right]$$

Relation entre BLUP et BLP

➢ Soit le modèle à effet fixe $\mathbf{y}=\mathbf{X}\boldsymbol{\beta} + \mathbf{e}$ avec $\mathrm{Var}(\mathbf{e})=\mathbf{V}$

- La solution des moindres carrés généralisés pour β est:

$$\boldsymbol{\beta}^{\circ} = \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$$

Rappel

➢ De plus :

$$\hat{\omega} = \mathbf{k}_i^{'}\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} + \mathbf{m}_i^{'}\mathbf{C}'\mathbf{V}^{-1}\left[\mathbf{y}-\mathbf{X}\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}\right]$$

$$\boxed{\begin{array}{l}\hat{\omega} = \mathbf{k}_i^{'}\boldsymbol{\beta}^{\circ} + \mathbf{m}_i^{'}\hat{\mathbf{a}} \\ \text{avec } \boldsymbol{\beta}^{\circ} = \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \text{ et } \hat{\mathbf{a}} = \mathbf{C}'\mathbf{V}^{-1}\left[\mathbf{y}-\mathbf{X}\boldsymbol{\beta}^{\circ}\right]\end{array}}$$

$\hat{\omega}_i = \mathrm{BLUP}(\omega_i)$ = 2 étape: 1) obtenir $\boldsymbol{\beta}^{\circ}$ par moindres carrés généralisés

2) obtenir $\hat{\mathbf{a}}$ = BLP sur données corrigées $\mathbf{y}-\mathbf{X}\boldsymbol{\beta}^{\circ}$

Equations du modèle mixte

- Henderson a montré que $\boldsymbol{\beta}^{\circ}$ et $\hat{\mathbf{a}}$ dans le BLUP sont solutions des équations du modèle mixte (MME):

$$\boxed{\begin{bmatrix} \mathbf{X'R^{-1}X} & \mathbf{X'R^{-1}Z} \\ \mathbf{Z'R^{-1}X} & \mathbf{Z'R^{-1}Z+G^{-1}} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} \mathbf{X'R^{-1}y} \\ \mathbf{Z'R^{-1}y} \end{bmatrix}}$$

- Avantages comparés aux BLP:
  - les effets fixes et aléatoires sont estimés/prédits en même temps
  - $\mathbf{R^{-1}}$ et $\mathbf{G^{-1}}$ ont en général une structure simple
  - Taille du système = taille du vecteur $\boldsymbol{\beta}$ + taille de **a** (et non taille de **y**)

C. Robert-Granié, Novembre 2011

ALIMENTATION AGRICULTURE ENVIRONNEMENT

Equations du modèle mixte : cas simple

- Dans le cas d'une analyse unicaractère, modèle avec un effet génétique additive : $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\ \mathbf{a}\ +\ \mathbf{e}$

  on suppose les résidus iid (indépendent et identiquement distribués)

  $\Rightarrow \mathbf{R} = \mathbf{I}\sigma_e^2$

  Var(**a**=effet génétique additive) = $\mathbf{G} = \mathbf{A}\sigma_a^2$ (**A** = matrice de parenté)

- En multipliant des 2 côtés par σ_e^2 :

$$\boxed{\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z+\alpha A^{-1}} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \end{bmatrix}}$$

où $\alpha = \frac{\sigma_e^2}{\sigma_a^2} = \frac{1-h^2}{h^2}$

C. Robert-Granié, Novembre 2011

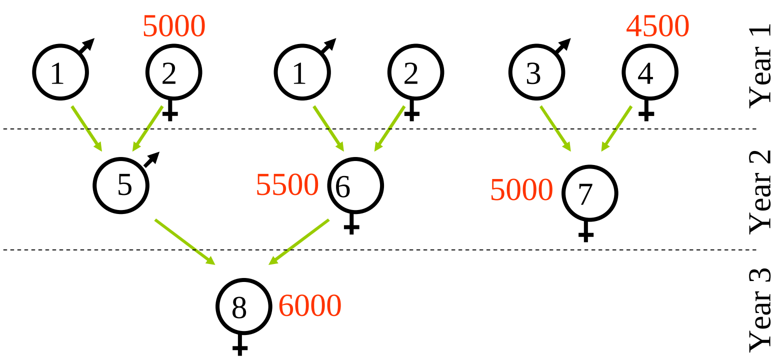
ALIMENTATION AGRICULTURE ENVIRONNEMENT

Meilleur prédicteur Linéaire Non Biaisé

<u>Cas 4</u>: La distribution jointe, ainsi que les espérances et variances covariances sont inconnues

Nous devons alors estimer les paramètres génétiques à partir des données; puis on considère qu'ils sont connus (on revient au cas 3)

Exemple



$$y_i = \beta_j + a_i + e_i$$

Modèle

$y_i = \beta_j + a_i + e_i$ ou matriciellement $\mathbf{y}=\mathbf{X\beta}+\mathbf{Za}+\mathbf{e}$

$$\begin{bmatrix} y_2 \\ y_4 \\ y_6 \\ y_7 \\ y_8 \end{bmatrix} = \begin{bmatrix} 5000 \\ 4500 \\ 5500 \\ 5000 \\ 6000 \end{bmatrix} = \begin{bmatrix} 1&0&0 \\ 1&0&0 \\ 0&1&0 \\ 0&1&0 \\ 0&0&1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} 0&1&0&0&0&0&0&0 \\ 0&0&0&1&0&0&0&0 \\ 0&0&0&0&0&1&0&0 \\ 0&0&0&0&0&0&1&0 \\ 0&0&0&0&0&0&0&1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \\ a_7 \\ a_8 \end{bmatrix} + \begin{bmatrix} e_2 \\ e_4 \\ e_6 \\ e_7 \\ e_8 \end{bmatrix}$$

$\mathbf{y}$ $\quad$ $\mathbf{X}$ $\quad$ $\mathbf{Z}$

$$y_2 (= 5000) = \beta_1 + a_2 + e_2$$
$$y_7 (= 5000) = \beta_2 + a_7 + e_7$$

Les équations du modèle mixte

$y_i = \beta_j + a_i + e_i$ ou matriciellement $\mathbf{y}=\mathbf{X\beta}+\mathbf{Za}+\mathbf{e}$

On suppose connues :

$$\mathrm{E}[\mathbf{e}] = \mathbf{0} \ , \ \mathrm{E}[\mathbf{a}] = \mathbf{0} \qquad \mathrm{V}[\mathbf{e}] = \mathbf{R} = \mathbf{I}\sigma_e^2 \ , \ \mathrm{V}[\mathbf{a}] = \mathbf{G} = \mathbf{A}\sigma_a^2$$

donc $\mathrm{E}[\mathbf{y}] = \mathbf{X\beta}$ et $\mathrm{V}[\mathbf{y}] = \mathbf{V} = \mathbf{ZGZ'} + \mathbf{R} = \mathbf{ZAZ'}\sigma_a^2 + \mathbf{I}\sigma_e^2$

$$\boxed{\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z}+\alpha\mathbf{A}^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \end{bmatrix}} \qquad \text{où } \alpha = \frac{\sigma_e^2}{\sigma_a^2} = \frac{1-h^2}{h^2}$$

Construction des équations du modèle mixte

$$\mathbf{X'X}=\begin{bmatrix}2&0&0\\0&2&0\\0&0&1\end{bmatrix}\quad \mathbf{X'Z}=\begin{bmatrix}0&1&0&1&0&0&0&0\\0&0&0&0&0&1&1&0\\0&0&0&0&0&0&0&1\end{bmatrix}\quad \mathbf{X'y}=\begin{bmatrix}5000+4500\\5500+5000\\6000\end{bmatrix}$$

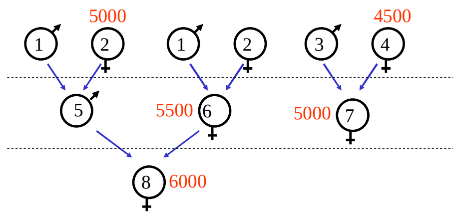
$$\mathbf{Z'X}=\begin{bmatrix}0&0&0\\1&0&0\\0&0&0\\1&0&0\\0&0&0\\0&1&0\\0&1&0\\0&0&1\end{bmatrix}\quad \mathbf{Z'Z}=\begin{bmatrix}0&0&0&0&0&0&0&0\\0&1&0&0&0&0&0&0\\0&0&0&0&0&0&0&0\\0&0&0&1&0&0&0&0\\0&0&0&0&0&0&0&0\\0&0&0&0&0&1&0&0\\0&0&0&0&0&0&1&0\\0&0&0&0&0&0&0&1\end{bmatrix}\quad \mathbf{Z'y}=\begin{bmatrix}0\\5000\\0\\4500\\0\\5500\\5000\\6000\end{bmatrix}$$

Construction des équations du modèle mixte

- Lecture d'un fichier de données contenant:

  n° animal, année, observation

  2 1 5000
  4 1 4500
  6 2 5500
  7 2 5000
  8 3 6000



- Lecture d'un fichier de généalogie pour construire la matrice de parenté **A** : animal, père, mère

Construire la matrice de parenté

$$\begin{bmatrix}\text{génotype}\\ \text{de i}\end{bmatrix} = \begin{bmatrix}\text{gamète du}\\ \text{père de i}\end{bmatrix} + \begin{bmatrix}\text{gamète de}\\ \text{la mère de i}\end{bmatrix}$$

$$= \begin{bmatrix}\text{effect moyen}\\ \text{des gamètes}\\ \text{de s}\end{bmatrix} + \begin{bmatrix}\text{écart à la moy. des effets}\\ \text{des gamètes de s}\end{bmatrix} + \begin{bmatrix}\text{effet moyen}\\ \text{des gamètes}\\ \text{de d}\end{bmatrix} + \begin{bmatrix}\text{écart à la moy. des effets}\\ \text{des gamètes de d}\end{bmatrix}$$

- Si deux parents sont connus

$$a_i = \frac{1}{2}a_s + \frac{1}{2}a_d + \varphi_i \quad \text{avec} \quad E(\varphi_i) = 0 \quad \text{et} \quad var(\varphi_i) = \left(\frac{1}{2} - \frac{F_s + F_d}{4}\right)\sigma_a^2$$

où F_i : coefficient de consanguinité de l'individu i

Construire la matrice de parenté

- si un seul parent connu, (ex le père)

Soit $\gamma_i = \frac{1}{2}a_d + \varphi_i \Rightarrow a_i = \frac{1}{2}a_s + \gamma_i$

Si $E(a_d) = 0$, $E(\gamma_i) = 0$

$var(\gamma_i) = \left(\frac{3}{4} - \frac{1}{4}F_s\right)\sigma_a^2$

Si mères non apparentées et non sélectionnées

- si les 2 parents sont inconnus,

Soit $\Psi_i = a_i$

$E(\Psi_i) = 0$

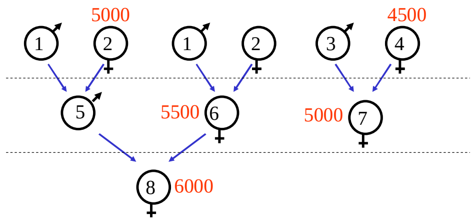
$var(\Psi_i) = \sigma_a^2$

Parents non apparentés et non sélectionnés

Exemple et matrice de parenté

- $a_1 = \Psi_1$; $a_2 = \Psi_2$; $a_3 = \Psi_3$; $a_4 = \Psi_4$

$$a_5 = \frac{1}{2}a_1 + \frac{1}{2}a_2 + \varphi_5 = \frac{1}{2}\Psi_1 + \frac{1}{2}\Psi_2 + \varphi_5$$
$$a_6 = \frac{1}{2}a_1 + \frac{1}{2}a_2 + \varphi_6 = \frac{1}{2}\Psi_1 + \frac{1}{2}\Psi_2 + \varphi_6$$
$$a_7 = \frac{1}{2}a_3 + \frac{1}{2}a_4 + \varphi_7 = \frac{1}{2}\Psi_3 + \frac{1}{2}\Psi_4 + \varphi_7$$
$$a_8 = \frac{1}{2}a_5 + \frac{1}{2}a_6 + \varphi_8 = \frac{1}{2}\Psi_1 + \frac{1}{2}\Psi_2 + \frac{1}{2}\varphi_5 + \frac{1}{2}\varphi_6 + \varphi_8$$



- $E(\Psi_i) = E(\varphi_j) = 0$, $i = 1..4$; $j = 5..8$

2 parents inconnus

2 parents connus

$$\mathrm{var}(\Psi_i) = \sigma_a^2 \quad ; \quad \mathrm{var}(\varphi_j) = \frac{1}{2}\sigma_a^2$$

➔ $\boxed{\mathbf{a} = \mathbf{T}\Psi}$

On peut exprimer tous les a_i en fonction de ψ

Exemple et matrice de parenté

- $$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \\ a_7 \\ a_8 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 1 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 1 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 1 \end{bmatrix} \begin{bmatrix} \Psi_1 \\ \Psi_2 \\ \Psi_3 \\ \Psi_4 \\ \varphi_5 \\ \varphi_6 \\ \varphi_7 \\ \varphi_8 \end{bmatrix}$$

T est simple à construire en partant des + vieux animaux

$$\left\{ \begin{array}{l} \mathbf{T} = \{t_{ij}\} \text{ matrice triangulaire inf.} \\ t_{ii} = 1 \\ t_{ij} = 0.5\, t_{sj} + 0.5\, t_{dj} \text{ si s et d parents de i} \\ t_{ij} = 0 \text{ si i n'a pas de parents connus} \end{array} \right.$$

➔ $\mathbf{a} = \mathbf{T}\Psi$

Quelques résultats sur a = Tψ qui vont nous permettre de construire A

- $\mathbf{a} = \mathbf{T\Psi} \Rightarrow \boxed{\begin{array}{l} E[\mathbf{a}] = \mathbf{T}E[\mathbf{\Psi}] = 0 \\ Var[\mathbf{a}] = Var[\mathbf{T\Psi}] = \mathbf{T}Var[\mathbf{\Psi}]\mathbf{T}' \\ \qquad = \mathbf{TDT}'\,\sigma_a^2 = \mathbf{A}\sigma_a^2 \end{array}} \Rightarrow \boxed{\mathbf{A} = \mathbf{TDT}'}$

- $\mathbf{A} = \mathbf{TDT}' = \left(\mathbf{TD}^{1/2}\right)\left(\mathbf{D}^{1/2}\mathbf{T}'\right) = \mathbf{LL}'$

  $\mathbf{D} = Var(\mathbf{\Psi})$

- $\mathbf{A}^{-1} = (\mathbf{TDT}')^{-1} = \left(\mathbf{T}^{-1}\right)'\mathbf{D}^{-1}\mathbf{T}^{-1}$

D = Var(Ψ)

- **D** = matrice diagonale de terme diagonal d_i

$$\mathbf{D}\sigma_a^2 = \begin{bmatrix} var(\Psi_1) & 0 & \cdots & 0 \\ 0 & var(\Psi_2) & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & var(\varphi_8) \end{bmatrix} = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & d_8 \end{bmatrix}\sigma_a^2$$

$$d_i = \underbrace{\begin{cases} 1 & \text{si 0 parent connu} \\ \frac{3}{4} & \text{si 1 parent connu} \\ \frac{1}{2} & \text{si 2 parents connus} \end{cases}}_{\text{Pas de consanguinité}} \qquad \underbrace{\begin{cases} 1 \\ \frac{3}{4} - \frac{F_s}{4} \\ \frac{1}{2} - \frac{1}{4}(F_s + F_d) \end{cases}}_{\text{si consanguinité}}$$

Pour la construction des équations du modèle mixte, A^{-1} est nécessaire :

$$\mathbf{A}^{-1} = \left(\mathbf{T}^{-1}\right)'\mathbf{D}^{-1}\mathbf{T}^{-1}$$

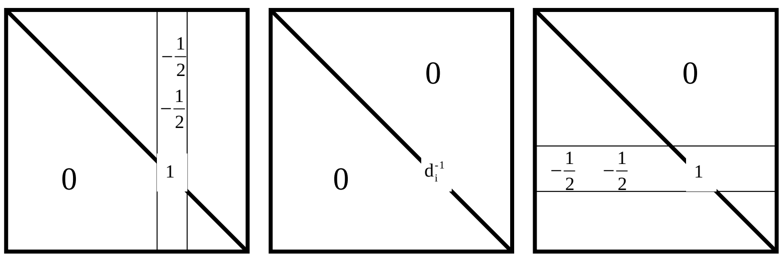
- $\mathbf{D}^{-1}$=matrice diagonale de terme diagonal d_i^{-1}
  = 1, 4/3 ou 2 si pas de consanguinité
- $\mathbf{T}^{-1}$ = matrice triangulaire inférieure, d'éléments diagonal =1
  sur chaque ligne, au plus 2 éléments non nuls
  = -1/2 dans la colonne des parents

1 2

- $\mathbf{D}^{-1} = \begin{bmatrix} 1&0&0&0&0&0&0&0\\ 0&1&0&0&0&0&0&0\\ 0&0&1&0&0&0&0&0\\ 0&0&0&1&0&0&0&0\\ 0&0&0&0&2&0&0&0\\ 0&0&0&0&0&2&0&0\\ 0&0&0&0&0&0&2&0\\ 0&0&0&0&0&0&0&2 \end{bmatrix}$ $\mathbf{T}^{-1} = \begin{bmatrix} 1&0&0&0&0&0&0&0\\ 0&1&0&0&0&0&0&0\\ 0&0&1&0&0&0&0&0\\ 0&0&0&1&0&0&0&0\\ \boxed{-0.5 \quad -0.5}&&0&0&\boxed{1}&0&0&0\\ -0.5&-0.5&0&0&0&1&0&0\\ 0&0&-0.5&-0.5&0&0&1&0\\ 0&0&0&0&-0.5&-0.5&0&1 \end{bmatrix}$

$$\mathbf{A}^{-1} = \left(\mathbf{T}^{-1}\right)'\mathbf{D}^{-1}\mathbf{T}^{-1}$$

- $\mathbf{A}^{-1}$=



Règles d'Henderson pour construire A^{-1}

(1976)

- lire le fichier pedigree : i , s , d   (animal, sire, dam)
- si 2 parents connus
  - add d_i^{-1} to A(i,i)
  - add -0.5 d_i^{-1} to A(i,s), A(s,i), A(i,d), A(d,i)
  - add 0.25 d_i^{-1} to A(s,s), A(s,d), A(d,s), A(d,d)
- si un seul parent connu (par ex, le père s)
  - add d_i^{-1} to A(i,i)
  - add -0.5 d_i^{-1} to A(i,s), A(s,i)
  - add 0.25 d_i^{-1} to A(s,s)
- si 2 parents inconnus
  - add d_i^{-1} to A(i,i)

Exemple et calcul de A-1

- Fichier pedigree: animal n°, sire n°, dam n°

1 0 0
2 0 0
3 0 0
4 0 0
5 1 2
6 1 2
7 3 4
8 5 6

$$\Rightarrow \mathbf{A}^{-1} = \begin{bmatrix} 1+\frac{1}{2}+\frac{1}{2} & \frac{1}{2}+\frac{1}{2} & 0 & 0 & -1 & -1 & 0 & 0 \\ \frac{1}{2}+\frac{1}{2} & 1+\frac{1}{2}+\frac{1}{2} & 0 & 0 & -1 & -1 & 0 & 0 \\ 0 & 0 & 1+\frac{1}{2} & \frac{1}{2} & 0 & 0 & -1 & 0 \\ 0 & 0 & \frac{1}{2} & 1+\frac{1}{2} & 0 & 0 & -1 & 0 \\ -1 & -1 & 0 & 0 & 2+\frac{1}{2} & \frac{1}{2} & 0 & -1 \\ -1 & -1 & 0 & 0 & \frac{1}{2} & 2+\frac{1}{2} & 0 & -1 \\ 0 & 0 & -1 & -1 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & -1 & -1 & 0 & 2 \end{bmatrix}$$

Construction des équations du modèle mixte

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z+\alpha A^{-1} \end{bmatrix} \begin{bmatrix} \beta \\ a \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

$$\left(\begin{array}{ccc:cccccccc}
2 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
\hdashline
0 & 0 & 0 & 0+2\alpha & 0+\alpha & 0 & 0 & 0-\alpha & 0-\alpha & 0 & 0 \\
1 & 0 & 0 & 0+\alpha & 1+2\alpha & 0 & 0 & 0-\alpha & 0-\alpha & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0+1.5\alpha & 0+0.5\alpha & 0 & 0 & 0-\alpha & 0 \\
1 & 0 & 0 & 0 & 0 & 0+0.5\alpha & 1+1.5\alpha & 0 & 0 & 0-\alpha & 0 \\
0 & 0 & 0 & 0-\alpha & 0-\alpha & 0 & 0 & 0+2.5\alpha & 0+0.5\alpha & 0 & 0-\alpha \\
0 & 1 & 0 & 0-\alpha & 0-\alpha & 0 & 0 & 0+0.5\alpha & 1+2.5\alpha & 0 & 0-\alpha \\
0 & 1 & 0 & 0 & 0 & 0-\alpha & 0-\alpha & 0 & 0 & 1+2\alpha & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0-\alpha & 0-\alpha & 0 & 1+2\alpha
\end{array}\right)
\begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \hdashline a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \\ a_7 \\ a_8 \end{bmatrix}
=
\left(\begin{array}{c} 5000+4500 \\ 5500+5000 \\ 6000 \\ \hdashline 0 \\ 5000 \\ 0 \\ 4500 \\ 0 \\ 5500 \\ 5000 \\ 6000 \end{array}\right)$$

Solution des équations du modèle mixte

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z+\alpha A^{-1} \end{bmatrix} \begin{bmatrix} \beta \\ a \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

$$\begin{bmatrix} \hat\beta_1 \\ \hat\beta_2 \\ \hat\beta_3 \\ \hat a_1 \\ \hat a_2 \\ \hat a_3 \\ \hat a_4 \\ \hat a_5 \\ \hat a_6 \\ \hat a_7 \\ \hat a_8 \end{bmatrix} = \begin{bmatrix} 4750 \\ 5250 \\ 5930 \\ 28 \\ 83 \\ -28 \\ -83 \\ 56 \\ 83 \\ -83 \\ 70 \end{bmatrix}$$

$\hat a_1$ à $\hat a_4$: Somme des « pop de base » =0

$\hat a_5$ à $\hat a_7$: Moyenne = 18.6

$\hat a_8$: Moyenne = 70

$\underbrace{\text{Moyenne} = 70}_{\text{Progrès génétique / génération}}$

$\hat\beta_1$: représente la moyenne des lactations des vaches 2 et 4 traites pendant l'année 1

$\hat\beta_1 = \frac{1}{2}[(y_2 - \hat a_2)(y_4 - \hat a_4)]$

$y_8 = 6000$; $\hat y_8 = \hat\beta_3 + \hat a_8 = 5930 + 70 = 6000$

On peut prédire des valeurs génétiques et des phénotypes pour des animaux sans performance

$\hat y_1 = \hat\beta_1 + \hat a_1 = 4750 + 28 = 4778$

BLUP - modèle animal : principes

1) modéliser les performances (identifier les principaux facteurs de variation)

2) constituer deux fichiers

→ un fichier «performances»
→ un fichier «généalogies»

règles simples

logiciel

3) construire un système d'équations

algorithmes

4) résoudre le système d'équations

5) valider les résultats

Un index est toujours accompagné d'un indicateur de sa PRECISION

La précision de l'évaluation génétique (de l'indexation)

☞ s'apprécie par le CD (coef. de détermination)

☞ dépend essentiellement du volume d'informations utilisées et de l'héritabilité

Propriétés du BLUP

Les effets de milieux et les effets génétiques sont estimés en même temps → pas de biais

Avec un modèle animal : toutes les informations (perf+pedigree) sont utilisées en même temps

L'évolution de la variance génétique au cours des générations, suite à la sélection et à la consanguinité est directement prise en compte à travers l'inverse de la matrice de variance-covariance génétique G

Mais il faut :

- bien décrire les effets de milieux et les effets génétiques
- connaître la variance additive dans la population de base supposée non consanguine et non sélectionnée
- bien identifier les relations de parenté
- toutes les données ayant servi aux opérations de sélection doivent être incluses

Merci à Vincent Ducrocq (INRA-GABI, Jouy-en-Josas).

Quelques références:

Ducrocq V. (1990) Les techniques d'évaluation génétique des bovins laitiers. INRA Productions Animales, 3 (1), 3-16

Henderson CR. (1984) Applications of Linear Models in Animal Breeding. University of Guelph, Guelph, Third Edition edited by L.R. Schaeffer
http://cgil.uoguelph.ca/pub/Henderson.html

Minvielle F. (1990) Principes d'amélioration génétique des animaux domestiques. Mieux comprendre,Institut de la recherche Agronomique. Éditeur Presses Université Laval.

Compléments

Propriétés du BLUP (1/3)

$$\boldsymbol{\beta}^\circ = \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \qquad \hat{\mathbf{a}} = \mathbf{C}'\mathbf{V}^{-1}\left[\mathbf{y}-\mathbf{X}\boldsymbol{\beta}^\circ\right]$$

- $\mathrm{Var}\left(\mathbf{k}_i'\boldsymbol{\beta}^\circ\right) = \mathbf{k}_i'\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-}\mathbf{k}_i$
- $\mathrm{Var}(\mathbf{m}_i'\hat{\mathbf{a}}) = \mathbf{m}_i'\mathrm{Var}\left(\mathbf{C}'\mathbf{V}^{-1}\left[\mathbf{y}-\mathbf{X}\boldsymbol{\beta}^\circ\right]\right)\mathbf{m}_i$

$$= \mathbf{m}_i'\mathbf{C}'\ \mathrm{Var}\left(\left[\mathbf{V}^{-1}\text{-}\mathbf{V}^{-1}\mathbf{X}\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-}\mathbf{X}'\mathbf{V}^{-1}\right]\mathbf{y}\right)\mathbf{C}\mathbf{m}_i$$

$$= \mathbf{m}_i'\mathbf{C}'\mathrm{Var}(\mathbf{P}\mathbf{y})\mathbf{C}\mathbf{m}_i \ \text{ with } \mathbf{P} = \mathbf{V}^{-1}\text{-}\mathbf{V}^{-1}\mathbf{X}\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-}\mathbf{X}'\mathbf{V}^{-1}$$

(on peut montrer que P est idempotent, **PX=0, X'P=0, PVP=P)**

$$\mathrm{Var}(\mathbf{m}_i'\hat{\mathbf{a}}) = \mathbf{m}_i'\mathbf{C}'\mathbf{P}\mathbf{C}\mathbf{m}_i$$

$$= \mathbf{m}_i'\mathbf{C}'\mathbf{V}^{-1}\mathbf{C}\mathbf{m}_i - \mathbf{m}_i'\mathbf{C}'\mathbf{V}^{-1}\mathbf{X}\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{C}\mathbf{m}_i$$

$$\le \mathbf{m}_i'\mathbf{C}'\mathbf{V}^{-1}\mathbf{C}\mathbf{m}_i$$

(les estimés BLUP sont moins variables que les estimés BLP)

Propriétés du BLUP (2/3)

- $\mathrm{Cov}\left(\mathbf{k}_i'\beta^\circ, \hat{\mathbf{a}}'\mathbf{m}_i\right) = \mathbf{k}_i'\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-}\mathbf{X}'\mathbf{V}^{-1}\mathrm{Cov}(\mathbf{y},\mathbf{y}')\mathbf{PCm}_i$

$$= \mathbf{k}_i'\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-}\mathbf{X}'\mathbf{PCm}_i = \mathbf{0} \quad \text{car } \mathbf{X}'\mathbf{P} = \mathbf{0}$$

- $\mathrm{Cov}(\mathbf{m}_i'\hat{\mathbf{a}},\ \mathbf{a}'\mathbf{m}_i) = \mathbf{m}_i'\mathrm{Cov}(\mathbf{C}'\mathbf{Py},\mathbf{a})\mathbf{m}_i = \mathbf{m}_i'\mathbf{C}'\mathbf{PCm}_i = \mathrm{Var}(\mathbf{m}_i'\hat{\mathbf{a}})$
- $\mathrm{Var}\left[\mathbf{m}_i'(\hat{\mathbf{a}}-\mathbf{a})\right] = \mathbf{m}_i'\mathrm{Var}(\hat{\mathbf{a}}-\mathbf{a})\mathbf{m}_i$

$$= \mathbf{m}_i'\left[\mathrm{Var}(\hat{\mathbf{a}}) - 2\mathrm{Cov}(\hat{\mathbf{a}},\mathbf{a}) + \mathrm{Var}(\mathbf{a})\right]\mathbf{m}_i$$

$$= \mathbf{m}_i'\left[\mathrm{Var}(\mathbf{a}) - \mathrm{Var}(\hat{\mathbf{a}})\right]\mathbf{m}_i$$

$$= \underbrace{\mathbf{m}_i'\mathbf{Gm}_i - \mathbf{m}_i'\mathbf{C}'\mathbf{V}^{-1}\mathbf{Cm}_i}_{\text{Variance d'erreur de Prédiction du BLP}} + \underbrace{\mathbf{m}_i'\mathbf{C}'\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Cm}_i}_{\text{« Coût supp. » lié à l'estimation de } \beta}$$

Propriétés du BLUP (3/3)

$$\mathrm{Cov}\left(\mathbf{k}_i'\left(\beta^\circ - \beta\right), (\hat{\mathbf{a}}-\mathbf{a})'\mathbf{m}_i\right) = -\mathrm{Cov}(\mathbf{k}_i'\beta^0, \mathbf{a}'\mathbf{m}_i)$$

$$= -\mathbf{k}_i'\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Cm}_i \leq \mathbf{0}$$

Le BLUP maximise la probabilité d'un classement correct des valeurs génétiques des animaux pris 2 à 2.