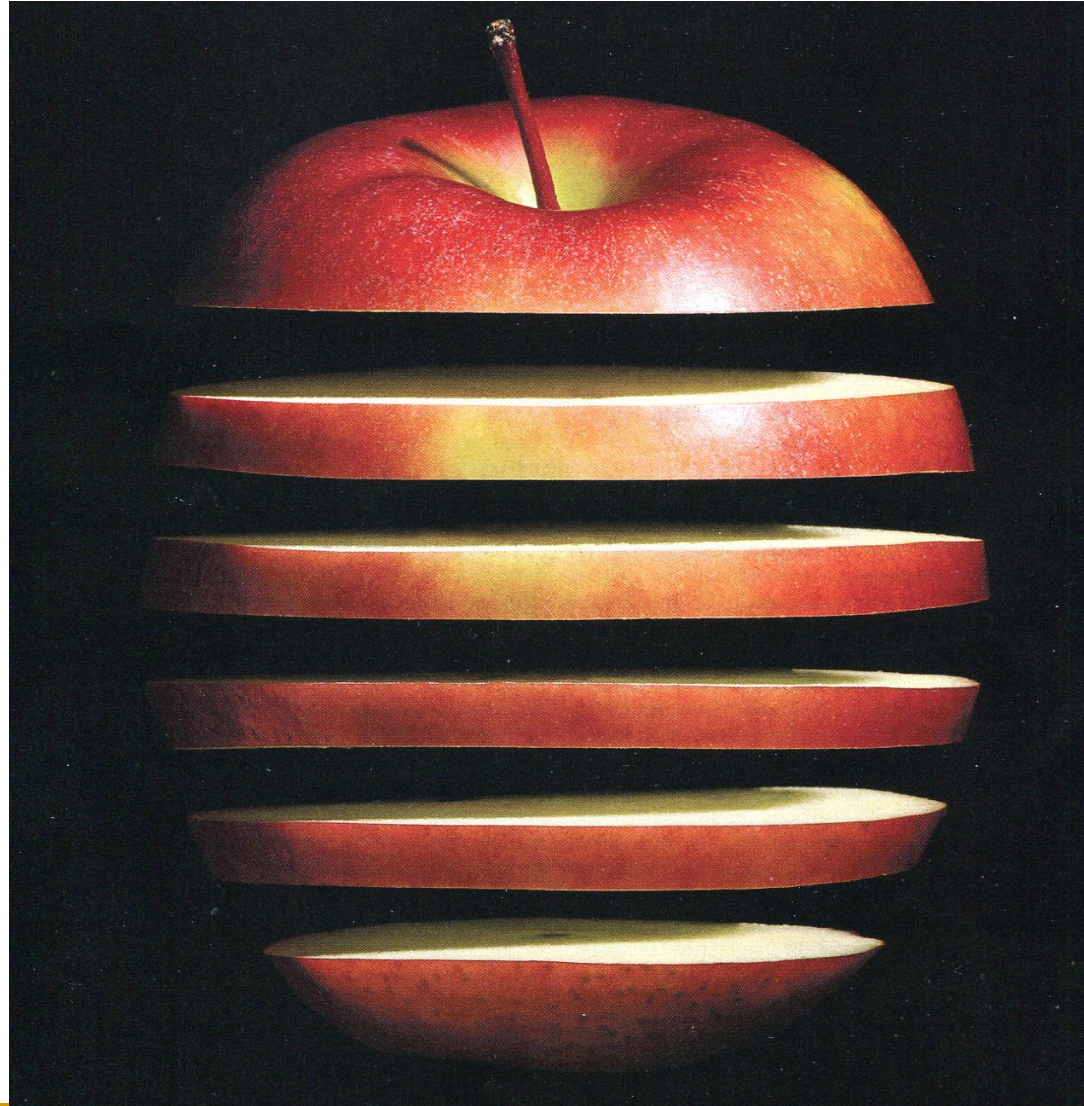


## Carnet de route & vagabondage au du modèle mixte



## Sommaire

- Introduction
- Les fondamentaux
  - Henderson:1959, 1973
  - Dempster, Laird & Rubin, 1977
- EM, Scaled EM et PX-EM
- Hétéroscédasticité
- Modélisation des variances
- Modèles non linéaires et représentation graphique
- Données discrètes ordinales
- Sélection de variables: détection de gènes sélectionnés
- A priori de variances
- Déviance marginale et comparaison de modèles

## Les gammes

- Vogt & Brocca: prépa Carnot Dijon
- Guy Lefort: cours INA + DEA Génétique quantitative 1967-1970
- Albert Jacquard: cours de probabilité au DEA de Génétique quantitative et appliquée 1969-1970
- Charles Henderson (1973) Sire evaluation & genetic trends
- Larry Schaeffer: cours Guelph 1979
- Charles Henderson: cours Landquart 1981
- Daniel Gianola: Jouy 1982; cours St Ferreol, 1984; Urbana: 1985-1986 ; Madison 1995

## Modèle linéaire mixte

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

$$\mathbf{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$$

$$\mathbf{u} \sim \mathcal{N}_q(\mathbf{0}, \mathbf{G}), \mathbf{e} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{R}), \text{Cov}(\mathbf{u}, \mathbf{e}') = \mathbf{0}$$

$$\text{Var}(\mathbf{y}) = \mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R} = \sum_{m=1}^M \gamma_m \mathbf{V}_m$$

## Modèle linéaire mixte

### THE ESTIMATION OF ENVIRONMENTAL AND GENETIC TRENDS FROM RECORDS SUBJECT TO CULLING\*

C. R. HENDERSON

*Animal Husbandry Department, Cornell University, Ithaca, N. Y., U.S.A.*

OSCAR KEMPTHORNE

*Statistical Laboratory, Iowa State College, Ames, Iowa, U.S.A.*

S. R. SEARLE

*Animal Husbandry Department, Cornell University, Ithaca, N. Y., U.S.A.*

AND

C. M. VON KROSIGK

*Animal Husbandry Department, Iowa State College, Ames, Iowa, U.S.A.*

*Biometrics*, Vol. 15, No. 2 (Jun., 1959), pp. 192-218

## Modèle linéaire mixte/Equations d'Henderson

Prédire:  $W = \mathbf{k}'\boldsymbol{\beta} + \mathbf{m}'\mathbf{u}$  à partir de  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}$

parmi  $\hat{W} = \mathbf{a}'\mathbf{y}$  tq  $E(\hat{W}) = E(W)$  et  $E(\hat{W} - W)^2$  min

$BLUP(\hat{W}) = \mathbf{k}'\hat{\boldsymbol{\beta}} + \mathbf{m}'\hat{\mathbf{u}}$  où  $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{GLS}$   $\hat{\mathbf{u}} = \mathbf{G}\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})$

$$\begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{pmatrix}$$

$\begin{pmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{12}' & \mathbf{C}_{22} \end{pmatrix}$  = Inverse de la matrice des coefts

$\text{Var}(\hat{\boldsymbol{\beta}}) = \mathbf{C}_{11}$ ;  $\text{Var}(\hat{\mathbf{u}} - \mathbf{u}) = \mathbf{C}_{22}$ ;  $\text{cov}[\hat{\boldsymbol{\beta}}, (\hat{\mathbf{u}} - \mathbf{u})] = \mathbf{C}_{12}$

## EMM/justification de CRH

$$f(\mathbf{y}, \mathbf{u}) = g(\mathbf{y} | \mathbf{u})h(\mathbf{u})$$

$$\mathbf{y} | \mathbf{u} \sim \mathcal{N}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}, \mathbf{R}); \quad \mathbf{u} \sim \mathcal{N}(\mathbf{0}, \mathbf{G})$$

$$g(\mathbf{y} | \mathbf{u}) = (2\pi)^{-N/2} |\mathbf{R}|^{-1/2} \exp[-1/2(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u})' \mathbf{R}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u})]$$

$$h(\mathbf{u}) = (2\pi)^{-q/2} |\mathbf{G}|^{-1/2} \exp[-1/2\mathbf{u}' \mathbf{G}^{-1} \mathbf{u}]$$

$$\ell(\boldsymbol{\beta}, \mathbf{u}; \mathbf{y}) = -2 \ln f(\mathbf{y}, \mathbf{u})$$

$$\begin{aligned} \ell(\boldsymbol{\beta}, \mathbf{u}; \mathbf{y}) &= (N + q) \ln(2\pi) + \ln |\mathbf{R}| + \ln |\mathbf{G}| \\ &\quad + (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u})' \mathbf{R}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u}) + \mathbf{u}' \mathbf{G}^{-1} \mathbf{u} \end{aligned}$$

## EMM/justification de CRH(suite)

$$\ell(\boldsymbol{\beta}, \mathbf{u}; \mathbf{y}) = Cste + \ln|\mathbf{R}| + \ln|\mathbf{G}| \\ + (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u})' \mathbf{R}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u}) + \mathbf{u}' \mathbf{G}^{-1} \mathbf{u}$$

$$(I) \quad \frac{\partial(-2\ell)}{\partial\boldsymbol{\beta}} = -2\mathbf{X}'\mathbf{R}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u})$$

$$(II) \quad \frac{\partial(-2\ell)}{\partial\mathbf{u}} = -2\mathbf{Z}'\mathbf{R}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u}) + 2\mathbf{G}^{-1}\mathbf{u}$$

(I)=0 et (II)=0 donnent :

$$\begin{cases} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\tilde{\boldsymbol{\beta}} + \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}\tilde{\mathbf{u}} = \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X}\tilde{\boldsymbol{\beta}} + (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})\tilde{\mathbf{u}} = \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{cases}$$

\* $\ell$  n'est pas une logvraisemblance (du moins au sens classique)

\*  $\mathbf{u}$  n'est pas un paramètre



## EMM/justification de CRH(suite)

Soit  $\boldsymbol{\theta} = (\boldsymbol{\beta}', \mathbf{u}')'$

$$f(\mathbf{y} | \boldsymbol{\theta}) = g(\mathbf{y} | \mathbf{u})$$

$$\pi(\boldsymbol{\theta}) = \underbrace{\pi(\boldsymbol{\beta})}_{=Cste} h(\mathbf{u}) \propto h(\mathbf{u})$$

$$\pi(\boldsymbol{\theta} | \mathbf{y}) \propto f(\mathbf{y} | \mathbf{u}) h(\mathbf{u}) = f(\mathbf{y}, \mathbf{u})$$

Solutions MME=MAP( $\boldsymbol{\theta}$ ) =  $E(\boldsymbol{\theta} | \mathbf{y})$

sous l'hypothèse de normalité et d'a priori plat sur les effets fixes

---

# Contribution de CR Henderson

## **SIRE EVALUATION AND GENETIC TRENDS** C. R. Henderson

*J ANIM SCI* 1973, 1973:10-41.

## ML des variances/ Algorithme de CRH (1973) type EM

$$\begin{bmatrix} Z_1'Z_1 + 1 \frac{\sigma_e^2}{\sigma_1^2} & \dots & Z_1'Z_s \\ \vdots & & \vdots \\ Z_s'Z_1 & \dots & Z_s'Z_s + 1 \frac{\sigma_e^2}{\sigma_s^2} \end{bmatrix}^{-1} = \begin{bmatrix} T_{11} & T_{12} & \dots & T_{1s} \\ T_{12}' & T_{22} & \dots & T_{2s} \\ \vdots & \vdots & \dots & \vdots \\ T_{1s}' & T_{2s}' & \dots & T_{ss} \end{bmatrix} \quad (10b)$$

Then the ML estimates are as follows

$\widehat{K}'\beta = K'\hat{\beta}$ , where  $\hat{\beta}$  is any solution to (10a) and  $K'\beta$  is estimable.

$$\hat{\sigma}_e^2 = (y'y - \hat{\beta}'X'y - \hat{u}'Z'y)/n.$$

$$\hat{\sigma}_i^2 = (\hat{u}_i'\hat{u}_i + \hat{\sigma}_e^2 \text{tr } T_{ii})/q_i.$$

These solutions require, of course, an iterative solution. The starting point is to substitute first estimates of  $\sigma_e^2/\sigma_i^2$  in (10b) and solve for  $\hat{\beta}$ ,  $\hat{u}$ . Then these results are used to find second estimates of  $\sigma_e^2$ ,  $\sigma_i^2$ , which in turn are substituted in (10b) to find second estimates of  $\hat{\beta}$ ,  $\hat{u}$ .

# Algorithm EM

## Maximum Likelihood from Incomplete Data via the *EM* Algorithm

By A. P. DEMPSTER, N. M. LAIRD and D. B. RUBIN

*Harvard University and Educational Testing Service*

[Read before the ROYAL STATISTICAL SOCIETY at a meeting organized by the RESEARCH SECTION on Wednesday, December 8th, 1976, Professor S. D. SILVEY in the Chair]

### SUMMARY

A broadly applicable algorithm for computing maximum likelihood estimates from incomplete data is presented at various levels of generality. Theory showing the monotone behaviour of the likelihood and convergence of the algorithm is derived. Many examples are sketched, including missing value situations, applications to grouped, censored or truncated data, finite mixture models, variance component estimation, hyperparameter estimation, iteratively reweighted least squares and factor analysis.

*Keywords:* MAXIMUM LIKELIHOOD; INCOMPLETE DATA; EM ALGORITHM; POSTERIOR MODE

*Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 39, No. 1. (1977), pp. 1-38.

## Algorithme EM

- EM s'appuie sur le **concept de données manquantes**
- **Cadre d'applications très vaste**
  - données incomplètes
  - censure, modèles mixtes, données discrètes,
  - Mélanges (ex QTL), Chaînes de Markov
- **Nombreuses variantes**
  - ECM; ECME, Stochastic EM, SAEM, Monte Carlo EM, Nested EM, Slice EM,
  - GEM, Gradient EM, EI (iterated)M, Metropolis EM,
  - Structural EM,
  - Supplemented EM,
  - Scaled EM, PX-EM,.....
  - Variational EM, etc...

## EM/formulation générale

-Considérer à l'instar de 1) la vraisemblance mais ici de données augmentées ou complètes (observées  $\mathbf{y}$ +manquantes  $\mathbf{z}$ )  $L(\phi; \mathbf{y}, \mathbf{z}) = \ln p(\mathbf{y}, \mathbf{z}; \phi)$

Comme  $L(\phi; \mathbf{y}, \mathbf{z})$  dépend de données non observées  $\mathbf{z}$

**E**-Remplacer  $L(\phi; \mathbf{y}, \mathbf{z})$  par sa meilleure prédiction  $E[L(\cdot) | \mathbf{y}, \phi]$

On le fait par itération:  $E[L(\cdot) | \mathbf{y}, \phi = \phi^{[n]}]$  appelée  $Q(\phi; \phi^{[n]})$

**M**-Maximiser  $Q(\phi; \phi^{[n]})$  la fonction obtenue par rapport à  $\phi$

## EM/justification

$\mathbf{y}_{N \times 1}$  = va observables ,  $\mathbf{z}_{q \times 1}$  = va manquantes (cachées, latentes)  
dont les densités marginales  $g(\mathbf{y} | \phi)$ , conjointe  $f(\mathbf{y}, \mathbf{z} | \phi)$  et  
conditionnelle  $h(\mathbf{z} | \mathbf{y}, \phi)$  dépendent d'un vect. de paramètres  $\phi$

$$\frac{\partial \ln g(\mathbf{y} | \phi)}{\partial \phi} = E_c \left[ \frac{\partial \ln f(\mathbf{y}, \mathbf{z} | \phi)}{\partial \phi} \right]$$

l'espérance notée  $E_c(\cdot)$  étant prise par rapport à  $h(\mathbf{z} | \mathbf{y}, \phi)$

Efron (1977) attribue cette identité à Fisher (1925) p717

## EM/justification

Comme  $h(\mathbf{z} | \mathbf{y}, \phi = \phi^{[t]})$  ne dépend pas de  $\phi$ ,  
les opérateurs sont interchangeables d'où

$$\mathbf{E}_c^{[t]} \left[ \frac{\partial \ln f(\mathbf{y}, \mathbf{z} | \phi)}{\partial \phi} \right] = \frac{\partial}{\partial \phi} \left\{ \mathbf{E}_c^{[t]} \left[ \ln f(\mathbf{y}, \mathbf{z} | \phi) \right] \right\}$$

Application à l'obtention des équations du REML

avec  $\mathbf{z} = (\boldsymbol{\beta}, \mathbf{u})$  qui permet d'éliminer automatiquement

les effets fixes par intégration (Foulley, 1993, J Dairy Sci).

Technique efficace pour obtenir REML en GLMM et NLMM



## EM standard

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

$$\mathbf{u}_{qx1} = \{u_j\} \sim \mathcal{N}(\mathbf{0}, \sigma_1^2 \mathbf{I}_q); \quad \mathbf{e}_{Nx1} = \{e_i\} \sim \mathcal{N}(\mathbf{0}, \sigma_0^2 \mathbf{I}_q)$$

Si  $\mathbf{u}$  était observable

$$\tilde{\sigma}_1^2 = \left( \sum_{j=1}^q u_j^2 \right) / q = \mathbf{u}'\mathbf{u} / q; \text{ idem pour } \tilde{\sigma}_0^2 = \left( \sum_{i=1}^N e_i^2 \right) / N$$

$$\tilde{\sigma}_1^2 = \left( \sum_{j=1}^q u_j^2 \right) / q = \mathbf{u}' \mathbf{u} / q$$

-En fait  $\mathbf{u}$  n'est pas observable

-On procède par itération

-Phase E:  $E(\mathbf{u}' \mathbf{u} \mid \mathbf{y}, \phi = \phi^{[t]})$

-Phase M:  $\sigma_1^{2[n+1]} = E_c^{[n]}(\mathbf{u}' \mathbf{u}) / q$

$$E_c(\mathbf{u}' \mathbf{u}) = \hat{\mathbf{u}}' \hat{\mathbf{u}} + \text{tr} \left[ \text{var}(\hat{\mathbf{u}} - \mathbf{u}) \right]$$

## Scaled EM

Foulley & Quaas, GSE 27 (1995) 211-228

Meng & van Dyk, JRSS 60 (1998) 559-578

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \sigma_1 \mathbf{Z}\mathbf{u}^* + \mathbf{e};$$

$$\mathbf{u}^* \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_q); \mathbf{e} \sim \mathcal{N}(\mathbf{0}, \sigma_0^2 \mathbf{I}_N); \mathbf{u}^* \perp \mathbf{e}$$

Si  $\mathbf{u}^*$  était observé, modèle linéaire en  $\boldsymbol{\beta}, \sigma_1$

$$\mathbf{y} = \underbrace{\mathbf{X}}_{Cov} \underbrace{\boldsymbol{\beta}}_{Par} + \underbrace{\mathbf{Z}\mathbf{u}^*}_{Cov} \underbrace{\sigma_1}_{Par} + \mathbf{e}; \mathbf{y} = \underbrace{\left[ \begin{array}{c} \mathbf{X}, \mathbf{Z}\mathbf{u}^* \\ \mathbf{w} \end{array} \right]}_{\mathbf{w}} \underbrace{\left[ \begin{array}{c} \boldsymbol{\beta} \\ \sigma_1 \\ \phi \end{array} \right]}_{\boldsymbol{\phi}} + \mathbf{e}$$

## Scaled EM

$$\mathbf{y} = \underbrace{\begin{bmatrix} \mathbf{X} & \mathbf{Z}\mathbf{u}^* \\ \mathbf{W} \end{bmatrix}}_{\mathbf{W}} \begin{bmatrix} \boldsymbol{\beta} \\ \underbrace{\sigma_1}_{\phi} \end{bmatrix} + \mathbf{e} \Rightarrow \mathbf{W}'\mathbf{W}\tilde{\boldsymbol{\phi}} = \mathbf{W}'\mathbf{y}$$

$$\boxed{\left[ \mathbf{E}_c^{[n]}(\mathbf{W}'\mathbf{W}) \right] \boldsymbol{\phi}^{[n+1]} = \mathbf{E}_c^{[n]}(\mathbf{W}'\mathbf{y})}$$

$$\mathbf{E}_c^{[n]}(\mathbf{v}) = \text{espérance par rapport à } \mathbf{v} \mid \mathbf{y}, \phi = \boldsymbol{\phi}^{[n+1]}$$

## Scaled EM par étapes

$$\mathbf{y} - \mathbf{X}\boldsymbol{\beta} = \sigma_1 \mathbf{Z}\mathbf{u}^* + \mathbf{e}$$

$$1) E_c^{[n]} \left( \mathbf{u}^{*'} \mathbf{Z}' \mathbf{Z} \mathbf{u}^* \right) \sigma_1^{[n+1]} = E_c^{[n]} \left[ \mathbf{u}^{*'} \mathbf{Z}' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^{[n]}) \right]$$

$$\sigma_1^{[n+1]} \geq 0 \text{ si } \sigma_1^{[n]} \geq 0 \text{ et } \sigma_0^{[n]} > 0 \text{ (FQ)}$$

2) Résolution en  $\boldsymbol{\beta}^{[n+1]}$  par MME d'Henderson

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \lambda^{[n]}\mathbf{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}^{[n+1]} \\ \mathbf{u}^{[n+1]} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix}$$

$$\lambda^{[n]} = \sigma_0^{2[n+1]} / \sigma_1^{2[n+1]}; \mathbf{u}^{[n+1]} = \sigma_1^{[n+1]} \mathbf{u}^{*[n+1]}$$

## Scaled vs EM Standard

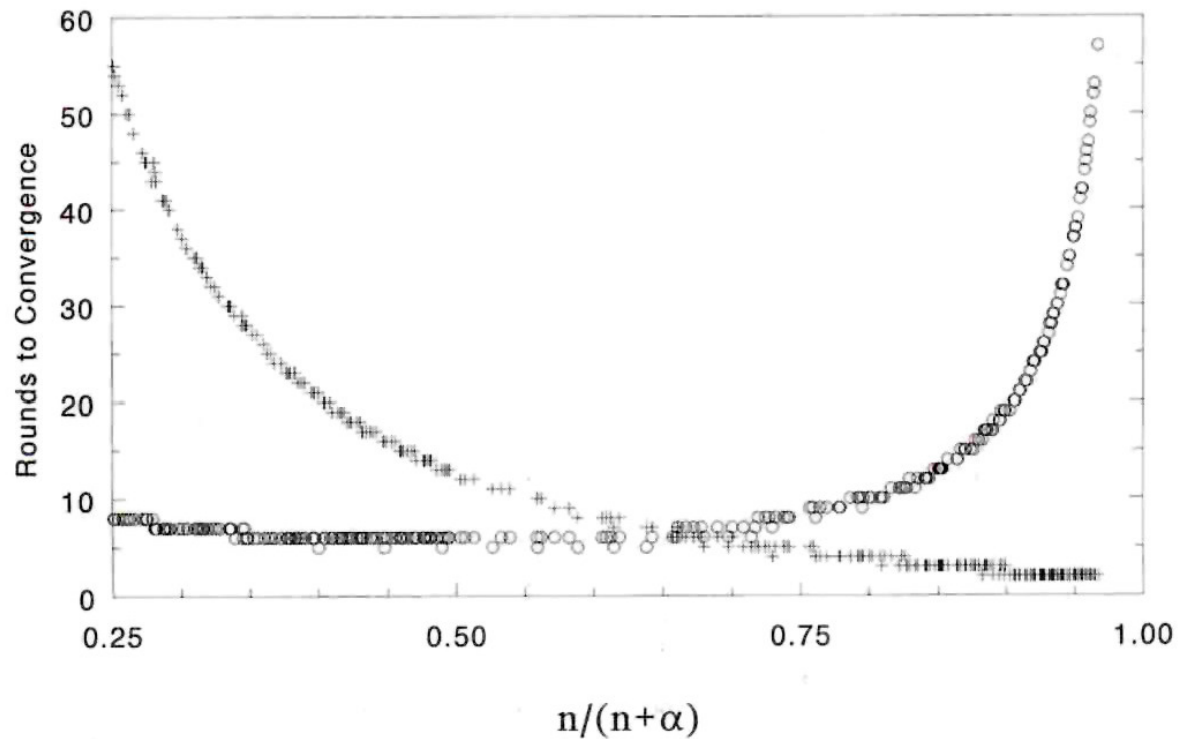


Fig1. Vitesse de convergence (nombre d'itérations) pour les algorithmes EM classique (croix) et « Scaled » (ronds) dans un dispositif de 100 familles de demi-frères de même taille ( $n$ ) en fonction du rapport  $R^2 = n / (n + \alpha)$  où  $\alpha = \sigma_0^2 / \sigma_1^2$  est le ratio de la variance résiduelle à la variance entre familles.

## Scaled EM: effets aléatoires corrélés

Pour l'unité expérimentale  $i$  (ex individu)

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \sum_{k=1}^K \mathbf{Z}_{ik} u_{ik} + \mathbf{e}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{u}_i + \mathbf{e}_i$$

$$\mathbf{u}_{i(K \times 1)} = (u_{ik})_{k=1:K} \quad \mathbf{Z}_i = (\mathbf{Z}_{i1}, \dots, \mathbf{Z}_{ik}, \dots, \mathbf{Z}_{iK})$$

$K$  effets aléatoires  $u_{ik}$  corrélés (ex intercept, pente)

$$\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{ij}, \dots, y_{in_i})' \quad \text{var}(\mathbf{u}_i) = \mathbf{G}_0 = \{g_{kl}\}$$

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \boxed{\mathbf{Z}_i \mathbf{T} \mathbf{u}_i^*} + \mathbf{e}_i \quad \text{var}(\mathbf{u}_i^*) = \mathbf{I}_K \quad \text{var}(\mathbf{e}_i) = \sigma_e^2 \mathbf{I}_{n_i}$$

$\mathbf{T}_{K \times K}$  transformation de Cholesky tq  $\mathbf{G}_0 = \mathbf{T} \mathbf{T}'$

## Scaled EM: effets aléatoires corrélés

$$\text{Ex: } K = 2 \quad \mathbf{u}_i = \mathbf{T}\mathbf{u}_i^* \quad \begin{pmatrix} u_{i,1} \\ u_{i,2} \end{pmatrix} = \begin{pmatrix} t_{11} & 0 \\ t_{21} & t_{22} \end{pmatrix} \begin{pmatrix} u_{i,1}^* \\ u_{i,2}^* \end{pmatrix}$$

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_{i1}u_{i1} + \mathbf{Z}_{i2}u_{i2} + \mathbf{e}_i$$

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_{i1}t_{11}u_{i1}^* + \mathbf{Z}_{i2}(t_{21}u_{i1}^* + t_{22}u_{i2}^*) + \mathbf{e}_i$$

Si  $\mathbf{u}_i^*$  observé, on a le modèle linéaire  $\mathbf{y}_i = \mathbf{W}_i\boldsymbol{\phi} + \mathbf{e}_i$

$$\mathbf{W}_i = (\mathbf{X}_i, \mathbf{Z}_{i1}u_{i1}^*, \mathbf{Z}_{i2}u_{i1}^*, \mathbf{Z}_{i2}u_{i2}^*) \quad \boldsymbol{\phi} = (\boldsymbol{\beta}', t_{11}, t_{21}, t_{22})'$$

$$\mathbf{W}'\mathbf{W} = \sum_i \mathbf{W}_i'\mathbf{W}_i \quad \mathbf{W}'\mathbf{y} = \sum_i \mathbf{W}_i'\mathbf{y}_i$$

$$\boxed{\left[ \mathbf{E}_c^{[t]} (\mathbf{W}'\mathbf{W}) \right] \boldsymbol{\phi}^{[t+1]} = \mathbf{E}_c^{[t]} (\mathbf{W}'\mathbf{y})}$$



## Scaled vs EM standard

*Scaled – EM*

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \boxed{\mathbf{Z}_i \mathbf{T} \mathbf{u}_i^*} + \mathbf{e}_i \quad \text{var}(\mathbf{u}_i^*) = \mathbf{I}_K \quad \text{var}(\mathbf{e}_i) = \sigma_e^2 \mathbf{I}_{n_i}$$

$\mathbf{T}_{K \times K}$  transformation de Cholesky tq  $\mathbf{G}_0 = \mathbf{T} \mathbf{T}'$

---

*PX – EM* (Liu, Rubin & Wu, 1998)

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \boxed{\mathbf{Z}_i \tilde{\mathbf{T}} \tilde{\mathbf{u}}_i} + \mathbf{e}_i \quad \text{var}(\tilde{\mathbf{u}}_i) = \tilde{\mathbf{G}}_0 \quad \text{var}(\mathbf{e}_i) = \sigma_e^2 \mathbf{I}_{n_i}$$

$\tilde{\mathbf{T}} = (t_{kl})_{k,l=1,\dots,K}$  tq  $\mathbf{G}_0 = \tilde{\mathbf{T}} \tilde{\mathbf{G}}_0 \tilde{\mathbf{T}}'$

## PX-EM

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + t_{11}\mathbf{Z}_1\tilde{\mathbf{u}}_1 + t_{12}\mathbf{Z}_1\tilde{\mathbf{u}}_2 + t_{21}\mathbf{Z}_2\tilde{\mathbf{u}}_1 + t_{22}\mathbf{Z}_2\tilde{\mathbf{u}}_2 + \mathbf{e}$$

$\mathbf{u}_i^*$  n'est plus standardisé mais  $\text{var}(\mathbf{u}_i^*) = \tilde{\mathbf{G}}_0$

$\mathbf{T} = \{t_{kl}\}$  : paramètres auxiliaires ou de travail

Idée:

-1) calculer  $\tilde{\mathbf{G}}_0$  par un EM standard

-2) calculer  $t_{11}, t_{12}, t_{21}, t_{22}$  comme avec scaled EM

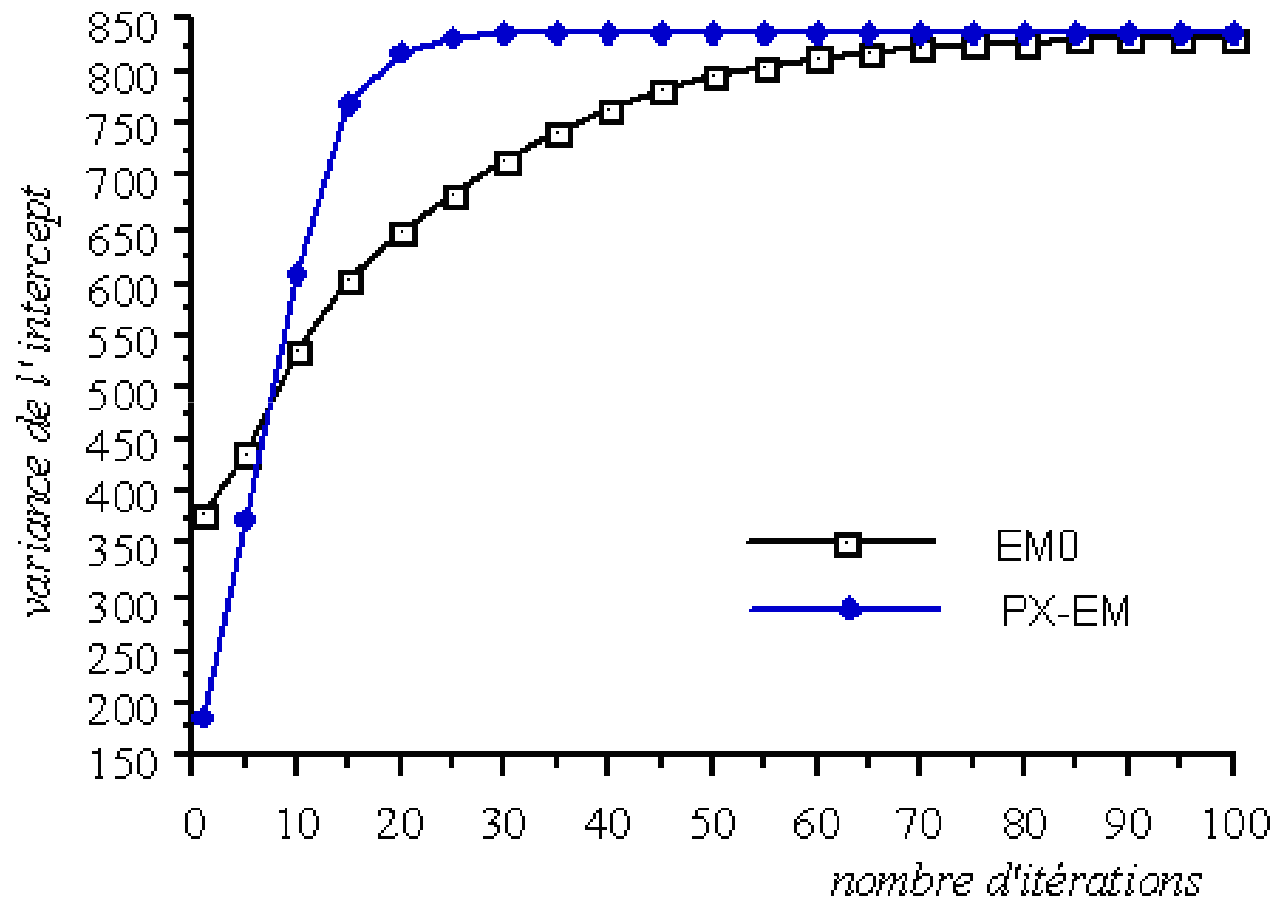
-3) revenir à  $\mathbf{G}_0$  par  $\mathbf{G}_0^{[t+1]} = \tilde{\mathbf{T}}^{[t+1]}\tilde{\mathbf{G}}_0^{[t+1]}\tilde{\mathbf{T}}^{[t+1]}$

En fait, à chaque itération EM, on va calculer les espérances condit.

et donc  $\tilde{\mathbf{G}}_0$ , en considérant la loi de  $\tilde{\mathbf{u}}$  conditionnellement à  $\mathbf{T} = \mathbf{I}_2$

ie  $\tilde{\mathbf{u}}$  m distribution que  $\mathbf{u}$

## Exemple PX vs EM0



## PX-EM

- Algorithme PX-EM REML à partir des équations du modèle mixte d'Henderson (Foulley & van Dyk, 2000)
- One step late PX-EM (van Dyk, Tang, 2003)
- A priori de Gelmann (2006) sur les composantes de variance
- PX-EM SAEM pour les modèles non linéaires mixtes (Lavielle, Meza, 2007) et modèles probit-normal (Meza, Jaffrezic & Foulley, 2009)
- Estimation REML de matrice de variance covariance après réduction de dimension (Meyer, 2008)
- Approche bayésienne des modèles à équations structurales (Demeyer, 2011)
- Algorithme PX-EM REML pour les modèles linéaires mixtes d'analyse factorielle avec application aux études GxE (Diffey, 2012; Diffey et al, 2014)

## Hétérogénéité des composantes u: GxE

Variabilité génétique des caractères dans différents milieux

Foulley, Hébert, Quaas, 1994; Robert-Granié, Foulley & Ducrocq, 1995

$$y_{ijk} = \mu + h_i + \sigma_{s,i} s_j^* + \sigma_{hs,i} hs_{ij}^* + \sigma_{e,i} e_{ijk}^*$$

$h_i$  est l'effet "fixe" du milieu  $i = 1, 2, \dots, I$

$\sigma_{s,i} s_j^*$  est la contribution aléatoire de la famille  $j$  élevée dans le milieu  $i$

$\sigma_{hs,i} hs_{ij}^*$  est un terme homologue au précédent relatif à l'interaction milieu  $i$  x famille  $j$

$e_{ijk} = \sigma_{e,i} e_{ijk}^*$  est la résiduelle

tq  $s_j^* \sim_{iid} \mathcal{N}(0,1)$ ,  $hs_{hs,j}^* \sim_{iid} \mathcal{N}(0,1)$ ,  $\text{Cov}(s_j^*, hs_{hs,j}^*) = 0$ ,  $e_{ijk}^* \sim_{iid} \mathcal{N}(0,1)$

Même milieu  $\text{Cov}(y_{ijk}, y_{ijk'}) = \sigma_{s,i}^2 + \sigma_{hs,i}^2$  Milieux difts  $\text{Cov}(y_{ijk}, y_{i'jk'}) = \sigma_{s,i} \sigma_{s,i'}$

Plus parcimonieux que le modèle multidimensionnel  $J > 3$

Liens étroits avec les modèles d'analyse factorielle  $g_{ij} = \lambda_t f_j + \delta_{ij}$  (ex Meyer, 2009)

## Modèles structuraux et relations fonctionnelles

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \sigma_{u_i} \mathbf{Z}_i \mathbf{u}^* + \sigma_{e_i} \mathbf{e}_i^* ; \ln \sigma_{e_i}^2 = \mathbf{p}_i' \boldsymbol{\delta}; \tau_i = \sigma_{u_i} / \sigma_{e_i}$$

1) modèle structural:  $\ln \tau_i = \mathbf{h}_i' \boldsymbol{\lambda}$

2) modèle fonctionnel:  $\boxed{\sigma_{u_i} / \sigma_{e_i}^b = cst} \Leftrightarrow \ln \sigma_{u_i} = a + b \ln \sigma_{e_i}$

(Foulley & Quaas, 1997)  $b = 0 \Leftrightarrow \sigma_{u_i}^2 = cst;$

$$b = 1 \Leftrightarrow \tau_i = cst \text{ ou } \rho_i (\text{intraclass}) = \frac{\sigma_{u_i}^2}{\sigma_{u_i}^2 + \sigma_{e_i}^2} = cst$$

Application à l'expérience de toxicologie de Dempster et al 1984  
sur le poids de souriceaux de 3 groupes (H,L,C )

Foulley (1999) Interstat

Extension du modèle structural à des effets "aléatoires"

$$\log \sigma_i^2 = \mathbf{p}'_i \boldsymbol{\delta} + \mathbf{q}'_i \mathbf{v}$$

$$\mathbf{v} = (\mathbf{v}'_1, \dots, \mathbf{v}'_k, \dots, \mathbf{v}'_K)' \quad \mathbf{v}_k \sim_{id} \mathcal{N}(0, \xi_k) \text{ ou } \mathbf{v} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Xi})$$

Foulley, San Cristobal, Gianola & Im 1992, CSDA

Foulley & Quaas, 1995, GSE

## Modèle mixte structural/autres références

Même idée dans

-X. Lin, 1997, Biometrics

-Goldstein's multi-level analysis: Browne et al, 2002, CSDA

-Double Hierarchical GLM: Lee & Nelder, 2006, JRSS

-Generalized Additive Model for Location, Scale and Shape:  
Rigby & Stasinopoulos, 2005, Appl. Stat



## Application à l'analyse de gènes différentiellement exprimés

Model for expression level of gene  $i = 1, \dots, I$ , in replication  $j = 1, \dots, n_{ik}$  and condition  $k$  (typically 1,2)

$$y_{ijk} = \mu_{ik} + e_{ijk} \text{ with } e_{ijk} \sim \mathcal{N}(0, \sigma_{ik}^2)$$

Hypothesis testing  $H_0$  " $\mu_{i1} = \mu_{i2}$ " vs  $H_1$  " $\mu_{i1} \neq \mu_{i2}$ "

Moment estimation by gene  $\hat{\sigma}_{ik}^2 = s_{ik}^2 = \sum_{j=1}^J (y_{ijk} - y_{i.k})^2 / (n_{ik} - 1)$

too imprecise & low power of detection of differently expressed genes

## Application à l'analyse de gènes différentiellement exprimés

$$\ln \sigma_{ik}^2 = \eta_k + v_{ik} \quad \text{with} \quad v_{ik} \sim \mathcal{N}(0, \tau_k^2)$$

$$\ln \hat{\sigma}_{ik}^2 = \eta_k + \lambda_{ik} (\ln s_{ik}^2 - \eta_k) \quad \text{with} \quad \lambda_{ik} = \tau_k^2 / (\tau_k^2 + \omega_{ik}^2); \quad \omega_{ik}^2 \propto 2 / (n_{ik} - 1)$$

$$\text{Welsch's stat } T_i = \frac{\hat{\mu}_{i1} - \hat{\mu}_{i2}}{\sqrt{n_{i1}^{-1} \hat{\sigma}_{i1}^2 + n_{i2}^{-1} \hat{\sigma}_{i2}^2}} \approx T_{v_i} \quad \text{Jaffrezic, Marot \& Foulley (2007)}$$

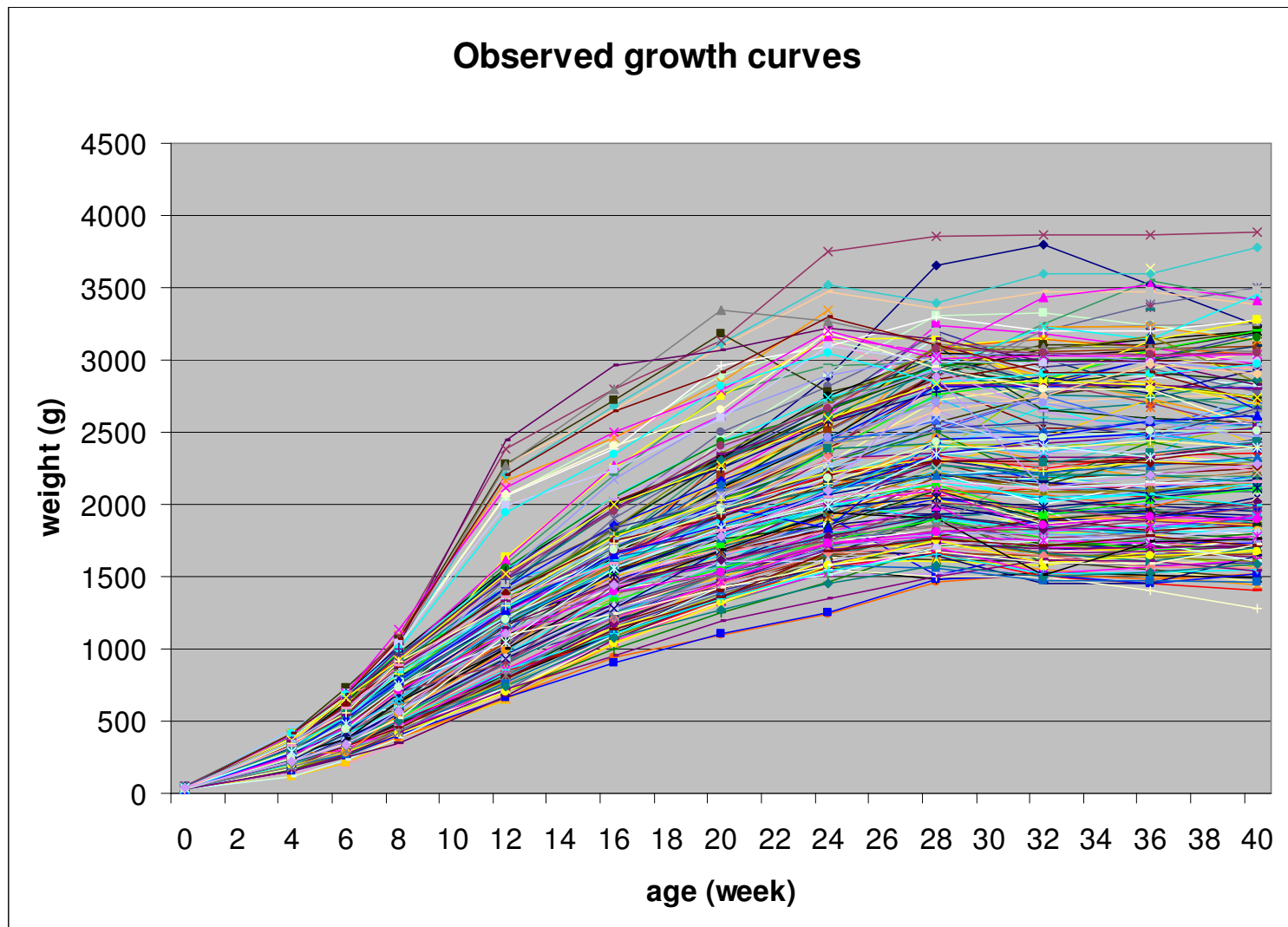
**SMVar**: R package <http://cran.r-project.org/> compared with

SAM (Tusher et al, 2001), VarMixt (Delmar et al, 2005), Limma (Smyth, 2004)

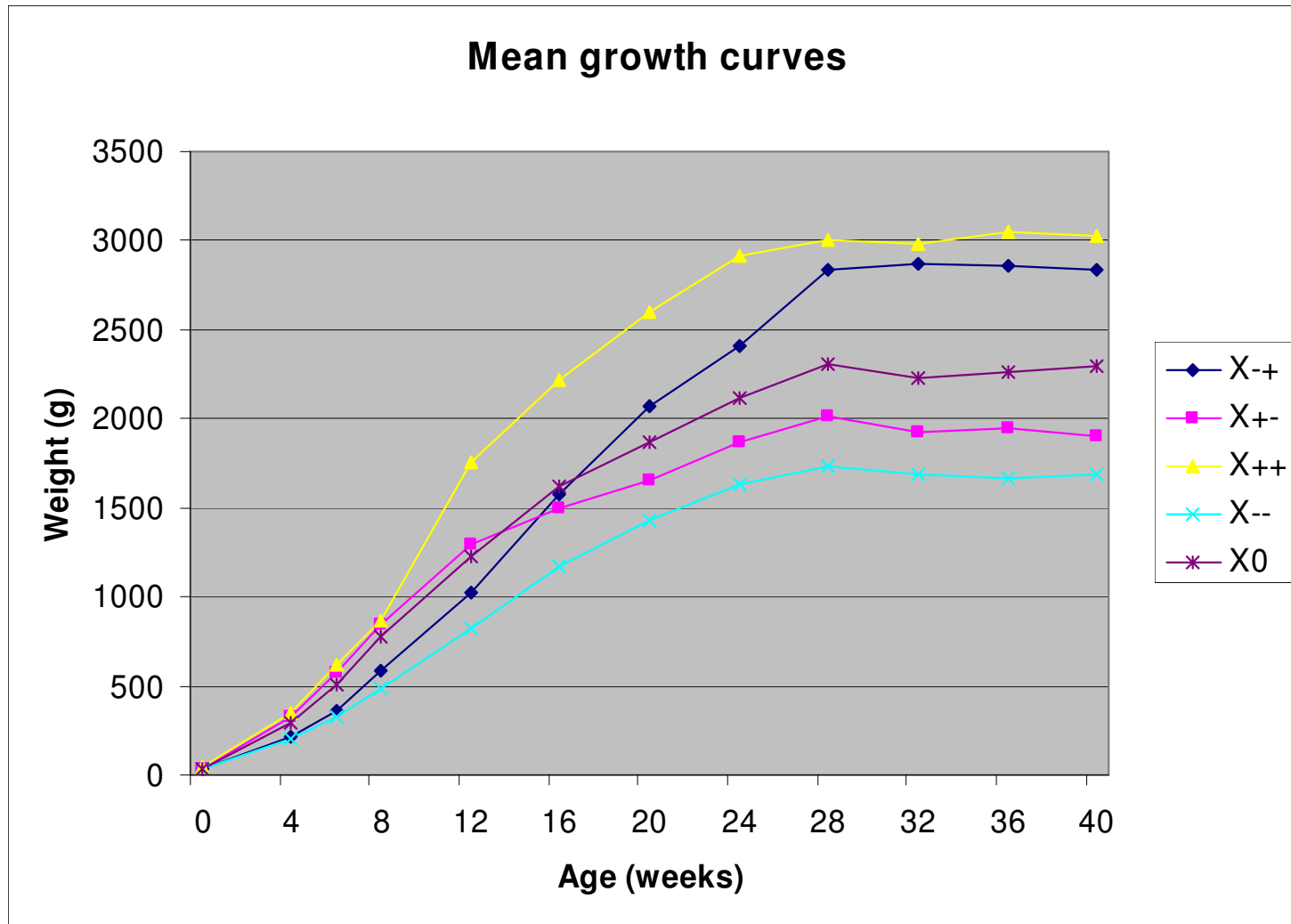
## Modèles non linéaires mixtes (NLMM): Ex données de croissance

- Expérience de Ricard à l'INRA de sélection divergente sur le poids à 2 âges: 8 et 36 semaines
- 5 lignées:
  - +-(HL); -+ (LH)
  - ++(HH); --(LL)
  - Témoin (C)
- Données=3058 contrôles effectués sur 265 femelles à la dernière génération

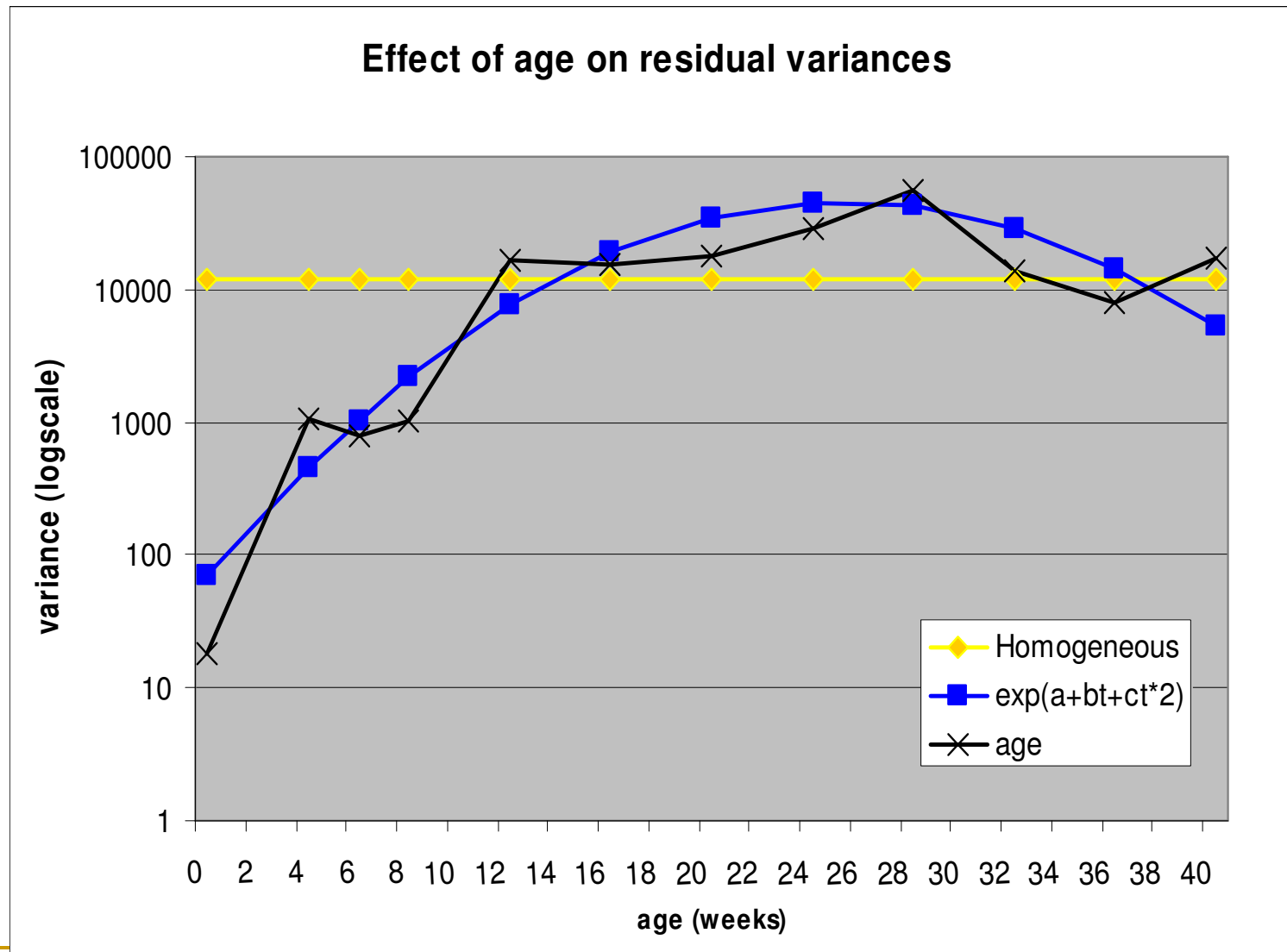
## NLMM: Ex données de croissance/profils d'ensemble



## NLMM: Ex données de croissance/variabilité entre lignées



## NLMM: Ex données de croissance/variance résiduelle



## NLMM: Ex données de croissance/modèle

### Modèles

$$y_{hij} = A_{hi} \exp\left[-B_{hi} \exp(-C_{hi} T_j)\right] + e_{hij}$$

avec  $T_j = t_j / 100$  (multiplication du coef  $C$  par 100)

$$A_{hi} = \underbrace{\alpha_h}_{\substack{\text{strain} \\ \text{(fixed)}}} + \underbrace{a_{hi}}_{\substack{\text{animal within strain} \\ \text{(aléatoire)}}} \quad a_{hi} \sim_{iid} \mathcal{N}(0, \sigma_a^2) \Leftrightarrow A_{hi} \sim_{iid} \mathcal{N}(\alpha_h, \sigma_a^2)$$

$$B_{hi} = \beta_h + b_{hi} \quad b_{hi} \sim_{iid} \mathcal{N}(0, \sigma_b^2) \Leftrightarrow B_{hi} \sim_{iid} \mathcal{N}(\beta_h, \sigma_b^2)$$

$$C_{hi} = \gamma_h + c_{hi} \quad c_{hi} \sim_{iid} \mathcal{N}(0, \sigma_c^2) \Leftrightarrow C_{hi} \sim_{iid} \mathcal{N}(\gamma_h, \sigma_c^2)$$

$$\sigma_{ab} = \text{Cov}(a_{hi}, b_{hi}) \quad \sigma_{ac} = \text{Cov}(a_{hi}, c_{hi}) \quad \sigma_{bc} = \text{Cov}(b_{hi}, c_{hi}) ;$$

$$e_{hij} \sim_{iid} \mathcal{N}(0, \sigma_{e_{t(j)}}^2)$$

$h = 1(MP), 2(PM), 3(PP), 4(MM), 5(Témoin C)$

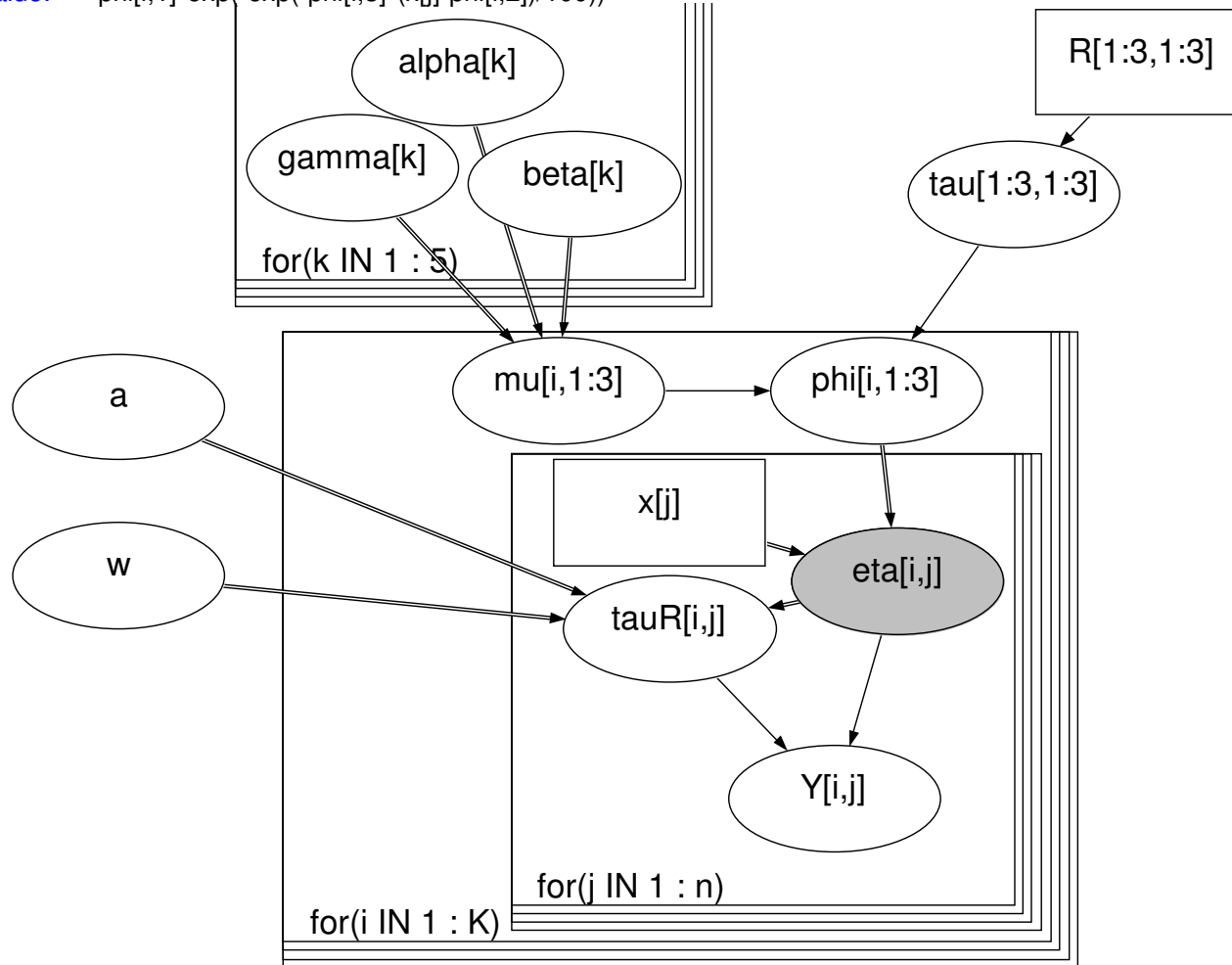
### Modèle de variance résiduelle par souche

$$\sigma_{e_{ht}} = \exp\left(\underline{a}_h + \underline{b}_h t^* + \underline{c}_h t^{*2}\right) \quad t^* = (t - 20) / 100 \quad \text{où } t = \text{age(sem)}$$

Ici  $\underline{a} = a\text{sigma}$ ,  $\underline{b} = b\text{sigma}$ ,  $\underline{c} = c\text{sigma}$

# NLMM: Ex croissance/graphe

name: eta[i,j]    type: logical    link: identity  
 value: phi[i,1]\*exp(-exp(-phi[i,3]\*(x[j]-phi[i,2])/100))





## NLMM: Ex croissance/graphe/code

```
model{
  for( i in 1 : K ) {
    for( j in 1 : n ) {
      Y[i , j] ~ dnorm(eta[i , j], tauR[i , j])
    }
  }
  for( l in 1 : 3 ) {
    for( k in 1 : 5 ) {
      alpha[k , l] ~ dnorm(0.0, 1.0E-6)
    }
  }
  for( i in 1 : K ) {
    for( j in 1 : n ) {
      eta[i , j] <- phi[i , 1] * exp( -exp(( -phi[i , 3]) * (x[j] - phi[i , 2]) / 100))
      tauR[i , j] <- (1 / exp(a)) * pow(eta[i , j], w)
    }
    phi[i , 1:3] ~ dmnorm(mu[i , 1:3], tau[1:3 , 1:3])
    mu[i , 1:3] <- inprod(X[1:5 , i], alpha[1:5 , 1:3])
  }
  a ~ dunif(-20, 20)
  tau[1:3 , 1:3] ~ dwish(R[1:3 , 1:3], 4)
  w ~ dunif(-5, 5)
}
```

## NLMM: Ex croissance/comparaison de modèles

Comparison of models of residual variance of NLMM for growth data in chicken

Model	#par	pD	DIC	Gain in DIC
Mixed/Himogeneous	1	502	36607	0
Exp(Age Quadra*Strain)	15	547	34227	2380
Power of mean	2	538	34431	2176
Constant +Power	3	552	34402	2205
Exp(Age Quadra*Strain+Rand Inter)	16	674	33987	2620(+240)

## Hétéroscédasticité dans les modèles à seuils

McCullach, 1982, Foulley & Gianola, 1996; Jaffrézic Robert Foulley, 1999

Soit  $\pi_{ij}$  probabilité de réponse  $j$  pour une observation  $r$  de la strate  $i$

$$\pi_{ij} = \Phi\left(\frac{\tau_j - \eta_i}{\sigma_i}\right) - \Phi\left(\frac{\tau_{j-1} - \eta_i}{\sigma_i}\right)$$

$\Phi(\cdot)$  CDF de  $N(0,1)$

$$\eta_i = \mathbf{x}_i' \boldsymbol{\beta} + \mathbf{z}_i' \mathbf{u}; \quad \ln(\sigma_i) = \mathbf{p}_i' \boldsymbol{\delta} + \mathbf{q}_i' \mathbf{v} \quad \text{ou} \quad \ln(\sigma_i) \sim \mathcal{N}(\mathbf{p}_i' \boldsymbol{\delta}, \mathbf{q}_i' \boldsymbol{\Gamma} \mathbf{q}_i)$$

$$l_{ir} = \mathbf{x}_i' \boldsymbol{\beta} + \mathbf{z}_i' \mathbf{u} + \sigma_i e_{ir}^* \quad \text{et} \quad y_{ijr} = 1 \Leftrightarrow \tau_{j-1} < l_{ir} \leq \tau_j$$

Inférence a posteriori **Foulley & Jaffrézic**, 2009, Computational Methods & Programs in Biomedicine

<http://w3.jouy.inra.fr/unites/miaj/public/matrisq/Contacts/applibugs.ressources.html>

Application par Whegang et al (2010) Essais randomisés antipaludiques

# Détection de loci sélectionnés

Modèle dit de Wright-Fisher intégrant **mutation, dérive et migration**  
Cf Crow-Kimura (1970), Crow (1986), Beaumont Balding (2004)  
Rousset (2007)

Modèle hiérarchique à plusieurs niveaux

$i$ =locus;  $j$ =(sous)population

$a_{ij}$ =Nombre de gènes d'un type allélique au locus  $i$  dans la pop  $j$

$p_{ij}$  = Fréquence de cet allèle au locus  $i$  dans la pop  $j$

$\lambda_{ij}$  = Proportion d'immigrants dans la population  $j$  (par génération)

$$0) \boxed{a_{ij} \mid p_{ij} \sim B(n_{ij}, p_{ij})} \quad 1) \boxed{p_{ij} \mid x_i, \lambda_{ij} \sim \text{Beta}(\lambda_{ij} x_i, \lambda_{ij} (1 - x_i))}$$

$$E(p_{ij}) = x_i \text{ (fréquence dans le gene-pool)}$$

$$\text{Var}(p_{ij}) = x_i (1 - x_i) \underbrace{(1 + \lambda_{ij})^{-1}}_{F_{ij}}$$

Marginale après intégration de  $p_{ij}$  :  $a_{ij} \mid x_i, \lambda_{ij} \sim \text{Beta - Binomiale}$

## Détection de loci sélectionnés

2) Modèle logistique sur  $F_{ij}$

$$\eta_{ij} = \ln \underbrace{F_{ij} / (1 - F_{ij})}_{1/\lambda_{ij}} = \mu + \underbrace{\alpha_i}_{\text{locus}} + \underbrace{\beta_j}_{\text{population}} + \underbrace{\gamma_{ij}}_{\text{écart additivité}}$$

$$\eta_{ij} \mid a_i, b_j, \sigma_\gamma^2 \sim_{id} \mathcal{N} \left( \underbrace{a_i + b_j}_{\mu + \alpha_i + \beta_j}, \sigma_\gamma^2 \right)$$

sous l'hypothèse de neutralité  $F_{ij}$  ne dépend pas de  $i$

a) approche de mélange: "Spike & slab priors"

$$a_i = \delta_i a_i^* \quad \delta_i = I(\text{locus } i \text{ sélectionné})$$

$$a_i^* \sim \mathcal{N}(0, \sigma_{\alpha^*}^2) \text{ Clyde (1999)}$$

$$\delta_i \sim \text{Ber}(P) \quad P \sim \text{Beta}(u_1, u_2)$$

Riebler, Held Stephan (2008), Gautier et al (2009)

## Détection de loci sélectionnés

*b*) approche du Lasso Bayésien (Park & Casella, 2008)

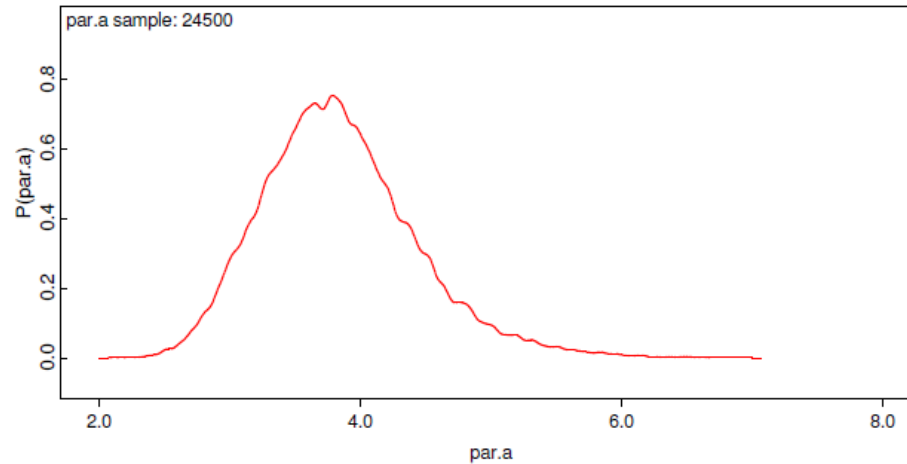
$a_i \sim_{iid} \mathcal{N}(0, \sigma_\alpha^2) \quad \sigma_\alpha^2 | \gamma \sim \text{Exp}(\gamma) \quad \gamma = \lambda^2 / 2, \quad \lambda = \text{paramètre de pénalité}$

la marginale  $a_i | \gamma \sim \text{DE}(\lambda)$  (Andrews & Mallows, 1974)

$\gamma \sim \mathcal{G}(u, v)$  par ex  $u = 1/2, v \approx 0$  (essai d'application en 2008)

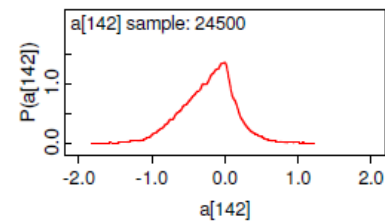
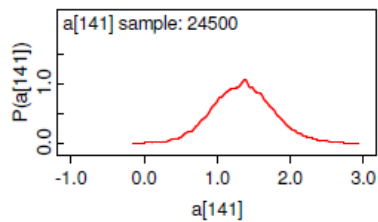
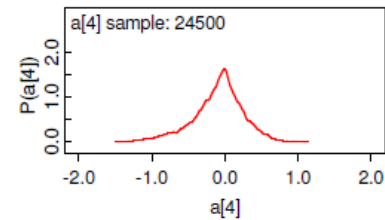
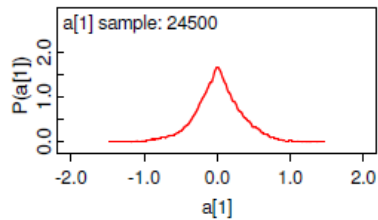
Ces modèles sont maintenant appliqués pour la détection des marqueurs dans la sélection QTL-MAS végétale (Fang et al, 2012, TAG) et animale (Legarra et al, 2011; Colombani et al, 2013)

# Détection de loci sélectionnés



par.a 3.844 0.5941 0.03861 2.838 3.794 5.208 501 24500

par.a=(lambda\*\*2)/2



## A priori sur la dispersion des effets aléatoires

MAP of a variance component in a random effect model  $\mathbf{u} \sim \mathcal{N}(0, \sigma^2 I_q)$  under different objective priors

	$\ln(\sigma)$	$\sigma$	$\sigma^2$
$\pi(\sigma^2) = \sigma^{2(\delta-1)}$	$\frac{1}{2} \ln \frac{E_c(u'u)}{q-2\delta}$	$\sqrt{\frac{E_c(u'u)}{q+1-2\delta}}$	$\frac{E_c(u'u)}{q+2(1-\delta)}$
$\delta = 1$	$q-2$	$q-1$	$q$
$\delta = 1/2$	$q-1$	$q$	$q+1$
$\delta = 0$	$q$	$q+1$	$q+2$
$\pi(\sigma) = \lambda \exp(-\lambda\sigma)$	$\lambda\sigma^3 + (q-1)\sigma^2 - E_c(u'u) = 0$	$\lambda\sigma^3 + q\sigma^2 - E_c(u'u) = 0$	$\lambda\sigma^3 + (q+1)\sigma^2 - E_c(u'u) = 0$

1)  $\pi(\sigma^2) = \sigma^{2(\delta-1)}$  soit  $\pi(\sigma^2) = 1$  (REML);  $\pi(\sigma) = 1$  (type Gelman),  $\pi(\sigma^2) = 1/\sigma^2$  (Jeffreys) pour  $\delta = 1, 1/2, 0$  respectivement

2)  $\pi(\sigma) = \lambda \exp(-\lambda\sigma)$ ;  $\Pr(\sigma > t) = \exp(-\lambda t) = \alpha \Rightarrow \lambda = -\log \alpha / t$  PC priors Martins et al (2014)



## A priori sur la dispersion des effets aléatoires

MAP of a variance component in a random effect model  $\mathbf{u} \sim \mathcal{N}(0, \sigma^2 I_q)$  under different objective priors

	$\ln(\sigma)$	$\sigma$	$\sigma^2$
$\pi(\sigma^2) = \sigma^{2(\delta-1)}$	$\frac{1}{2} \ln \frac{E_c(u'u)}{q-2\delta}$	$\sqrt{\frac{E_c(u'u)}{q+1-2\delta}}$	$\frac{E_c(u'u)}{q+2(1-\delta)}$
$\delta = 1$	$q-2$	$q-1$	$q$
$\delta = 1/2$	$q-1$	$q$	$q+1$
$\delta = 0$	$q$	$q+1$	$q+2$
$\pi(\sigma) = \lambda \exp(-\lambda\sigma)$	$\lambda\sigma^3 + (q-1)\sigma^2 - E_c(u'u) = 0$	$\lambda\sigma^3 + q\sigma^2 - E_c(u'u) = 0$	$\lambda\sigma^3 + (q+1)\sigma^2 - E_c(u'u) = 0$

1)  $\pi(\sigma^2) = \sigma^{2(\delta-1)}$  soit  $\pi(\sigma^2) = 1$  (REML);  $\pi(\sigma) = 1$  (type Gelman),  $\pi(\sigma^2) = 1/\sigma^2$  (Jeffreys) pour  $\delta = 1, 1/2, 0$  respectivement

2)  $\pi(\sigma) = \lambda \exp(-\lambda\sigma)$ ;  $\Pr(\sigma > t) = \exp(-\lambda t) = \alpha \Rightarrow \lambda = -\log \alpha / t$  PC priors Martins et al (2014)

## DIC vs Déviance marginale

$DIC = \bar{D} + p_D$  où  $p_D = \bar{D} - D(\bar{\theta})$  complexité

Quid de la pénalité à affecter à  $\bar{D}$  pour obtenir  $D_m$

$$KL(\pi(\theta | \mathbf{y}), \pi(\theta)) = \int \ln \frac{\pi(\theta | \mathbf{y})}{\pi(\theta)} \pi(\theta | \mathbf{y}) d\theta$$

$$KL = \int \ln \frac{f(\mathbf{y} | \theta) \cancel{\pi(\theta)}}{m(\mathbf{y}) \cancel{\pi(\theta)}} \pi(\theta | \mathbf{y}) d\theta$$

$$KL = E_{\theta | \mathbf{y}} [\ln f(\mathbf{y} | \theta)] - \ln m(\mathbf{y})$$

$$-2KL = \bar{D} - D_m \Rightarrow \boxed{D_m = \bar{D} + 2KL}$$

## Calcul de la vraisemblance marginale

$$\log m(\mathbf{y}) = \int_0^1 E_{\theta|\mathbf{y},t} [\log f(\mathbf{y}|\theta)] dt \quad 0 \leq t \leq 1$$

où  $\theta|\mathbf{y},t$  a pour densité:

$$\pi(\theta|\mathbf{y},t) = \frac{f(\mathbf{y}|\theta)^t \pi(\theta)}{\int f(\mathbf{y}|\theta)^t \pi(\theta) d\theta}$$

Thermodynamic integration (end of the 70's)

Ripley (1988), Ogata (1989), Neal (1993)

"Path sampling" (Gelman & Meng, 1998)

Lartillot & Philippe (2006) "Annealing-Melting"

Friel & Petit (2008) "Power Posteriors"

## Vraisemblance fractionnaire

$\pi(\boldsymbol{\theta}, b) \propto f(\mathbf{y} | \boldsymbol{\theta})^b \pi(\boldsymbol{\theta}) \quad 0 \leq b \leq 1$  (O'Hagan, 1995)

fraction  $b$  de la vraisemblance sert à régler l'a priori

$$m^F(\mathbf{y}, b) = \int f(\mathbf{y} | \boldsymbol{\theta})^{1-b} \pi(\boldsymbol{\theta}, b) d\boldsymbol{\theta}$$

$$m^F(\mathbf{y}, b) = \frac{\int f(\mathbf{y} | \boldsymbol{\theta}) \pi(\boldsymbol{\theta}) d\boldsymbol{\theta}}{\int f(\mathbf{y} | \boldsymbol{\theta})^b \pi(\boldsymbol{\theta}) d\boldsymbol{\theta}} = \frac{m(\mathbf{y}, 1)}{m(\mathbf{y}, b)}$$

PP directly provides

$-\pi(\boldsymbol{\theta}, b)$  via  $\pi(\boldsymbol{\theta} | \mathbf{y}, t = b)$

$$-\log m^F(\mathbf{y}, b) = \int_b^1 E_{\boldsymbol{\theta} | \mathbf{y}, t} [\log f(\mathbf{y} | \boldsymbol{\theta})] dt$$

## Vraisemblance marginale/ex Pothoff

$i$ : subscript for individual  $i = 1, \dots, I = 25$  (11 girls+16 boys)

$j$ : subscript for measurement at age  $t_j$  (8,10,12,14 yrs)

1) Purely Fixed Model

$$y_{ij} = \underbrace{(\alpha_0 + \alpha x_i)}_{\text{intercept}} + \underbrace{(\beta_0 + \beta x_i)}_{\text{pente}} (t_j - 8) + e_{ij}$$

2) Random intercept model

$$y_{ij} = (\alpha_0 + \alpha x_i + a_i) + (\beta_0 + \beta x_i)(t_j - 8) + e_{ij}$$

3) Random intercept & slope model assuming independent effects

$$y_{ij} = (\alpha_0 + \alpha x_i + a_i) + (\beta_0 + \beta x_i + b_i)(t_j - 8) + e_{ij}$$

or

$$y_{ij} = \phi_{i1} + \phi_{i2} (t_j - 8) + e_{ij}, \quad y_{ij} \sim_{\text{id}} \mathcal{N}(\eta_{ij}, \sigma_e^2)$$

$$\text{with } \phi_i = \begin{pmatrix} \phi_{i1} \\ \phi_{i2} \end{pmatrix} \sim \mathcal{N} \left[ \begin{pmatrix} \alpha_0 + \alpha x_i \\ \beta_0 + \beta x_i \end{pmatrix}, \begin{pmatrix} \sigma_a^2 & 0 \\ 0 & \sigma_b^2 \end{pmatrix} \right]$$

4) Random intercept & slope model assuming correlated effects

$$\phi_i = \begin{pmatrix} \phi_{i1} \\ \phi_{i2} \end{pmatrix} \sim \mathcal{N} \left[ \begin{pmatrix} \alpha_0 + \alpha x_i \\ \beta_0 + \beta x_i \end{pmatrix}, \begin{pmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{pmatrix} \right]$$

## V. marginale/Ex Pothoff/a priori

1<sup>st</sup> level:  $y_{ij} \sim_{\text{id}} \mathcal{N}(\eta_{ij}, \sigma_e^2)$  with  $\eta_{ij} = \phi_{i1} + \phi_{i2}(t_j - 8)$

2<sup>nd</sup> level:

$$2a) \phi_i = \begin{pmatrix} \phi_{i1} \\ \phi_{i2} \end{pmatrix} \sim \mathcal{N} \left[ \begin{pmatrix} \alpha_0 + \alpha x_i \\ \beta_0 + \beta x_i \end{pmatrix}, \underbrace{\begin{pmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{pmatrix}}_{\Sigma} \right]$$

$$2b) \sigma_e \sim U(0, \Delta_e) \text{ or } \sigma_e^2 \sim \text{InvG}(1, \underline{\sigma}_e^2)$$

3<sup>rd</sup> level:

Fixed effects:  $\alpha_0, \alpha, \beta_0, \beta \sim U(\text{inf}, \text{sup})$

Var (Covar) components:

–If  $\sigma_{ab} = 0$ , then i)  $\sigma_a \sim U(0, \Delta_a)$ , same for  $\sigma_b \sim U(0, \Delta_b)$

or ii)  $\sigma_a^2 \sim \text{InvG}(1, \underline{\sigma}_a^2)$ , same for  $\sigma_b^2 \sim \text{InvG}(1, \underline{\sigma}_b^2)$

–If  $\sigma_{ab} \neq 0$ , then i)  $\sigma_a \sim U(0, \Delta_a)$ ,  $\sigma_b \sim U(0, \Delta_b)$ ,  $\rho \sim U(-1, 1)$

or ii)  $\Omega \sim W((\nu \underline{\Sigma})^{-1}, \nu)$  for  $\Omega = \Sigma^{-1}$

with  $\nu = \dim(\Omega) + 1$  and  $\underline{\Sigma}$  known location parameter

\*Take care as Winbugs uses another notation ie  $W((\nu \underline{\Sigma}), \nu)$

## V. marginale/Ex Pothoff/comparaison de modèles

Table : Criteria of comparison of models applied to Pothoff & Roy's data set  
(missing data version of Little and Rubin, N=99)

Models	Residual Likelihood			Uniform priors		Inv-Gamma Wishart priors		Fractional priors	
	Do	AIC	BIC	Dm	DIC	Dm	DIC	Dm (0.05)	Dm (0.15)
Fixed	884.0	886.0	888.6	919.5	907.6	919.4	907.4	866.8	770.6
Random Intercept	843.6	847.6	850.2	884.8	837.6	880.5	837.5	828.9	731.6
Intercept+Slope Indep	842.7	848.7	852.6	886.6	833.8	880.1	833.7	823.8	725.8
Intercept+Slope Correl	842.4	850.4	855.5	887.2	835.1	880.1	834.6	830.2	731.7

Priors on fixed effects are:  $\alpha_0 \sim U(0, 500)$ ,  $\alpha \sim U(-50, 20)$ ,  $\beta_0 \sim U(4, 12)$ ,  $\beta \sim U(-10, 4)$

Priors on variance covariance components are:

-uniform on standard deviations and correlation:  $\sigma_e \sim U(0, 50)$ ,  $\sigma_1 \sim U(0, 100)$ ,  $\sigma_2 \sim U(0, 20)$ ,  $\rho \sim U(-1.0, 1.0)$

-inverted gamma or Wishart on variances:  $\sigma_e^{-2} \sim G(1, 200)$ ,  $\sigma_1^{-2} \sim G(1, 300)$ ,  $\sigma_2^{-2} \sim G(1, 2)$ ,  $\Sigma^{-1} \sim \mathcal{W}((3\underline{\Sigma})^{-1}, 3)$

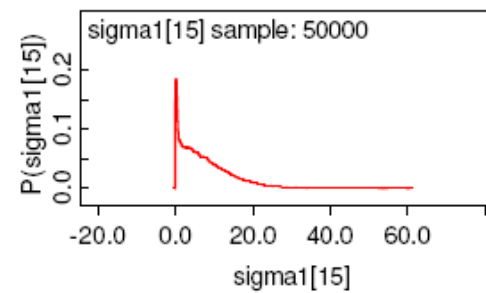
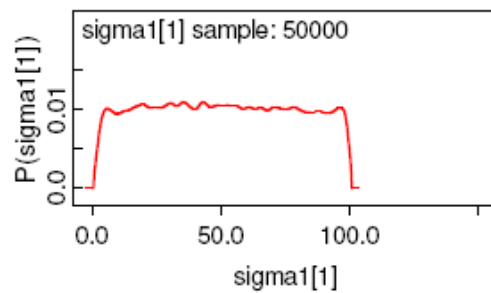
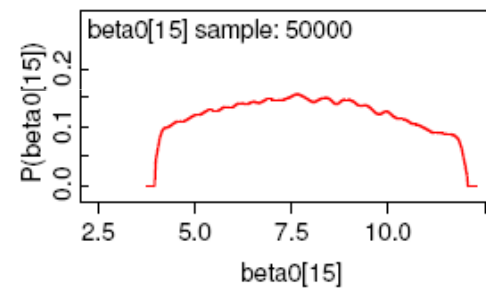
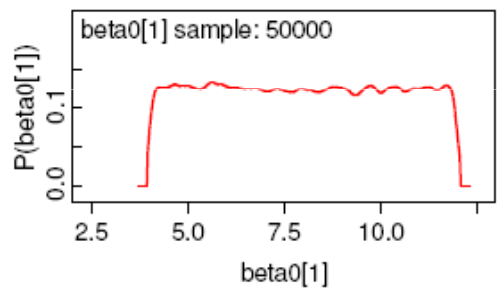
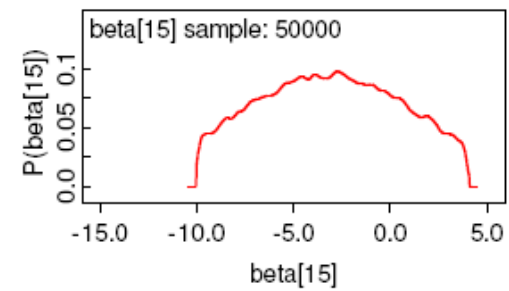
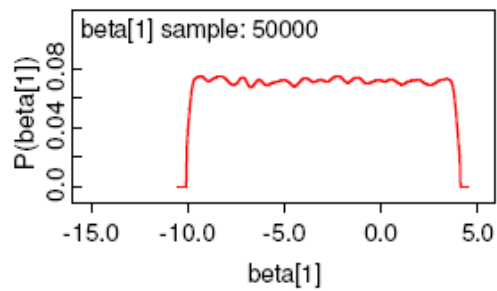
where  $\underline{\Sigma} = \text{Diag}(300, 2)$

Fractional "priors" (O'Hagan (1995)) were based on a fraction (b=0.05, 0.15) of the likelihood to turn initial uniform into data-based priors.

Do : minimum frequentist deviance : Dm : marginal (integrated) deviance, AIC, BIC computed as in SAS-proc mixed and DIC as in Winbugs.

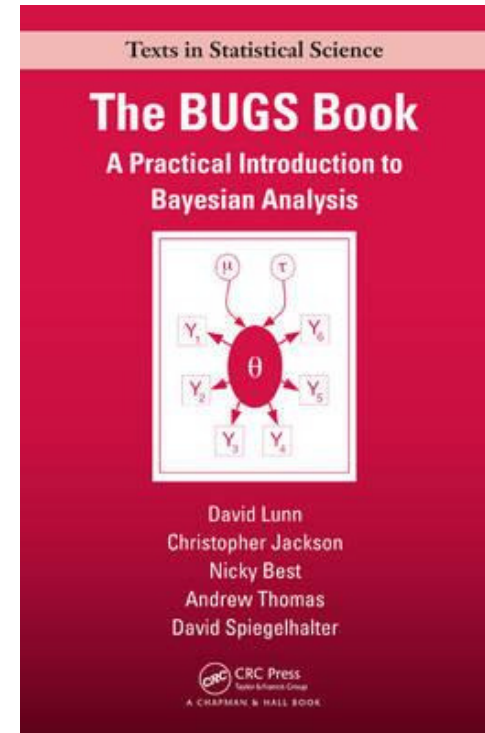
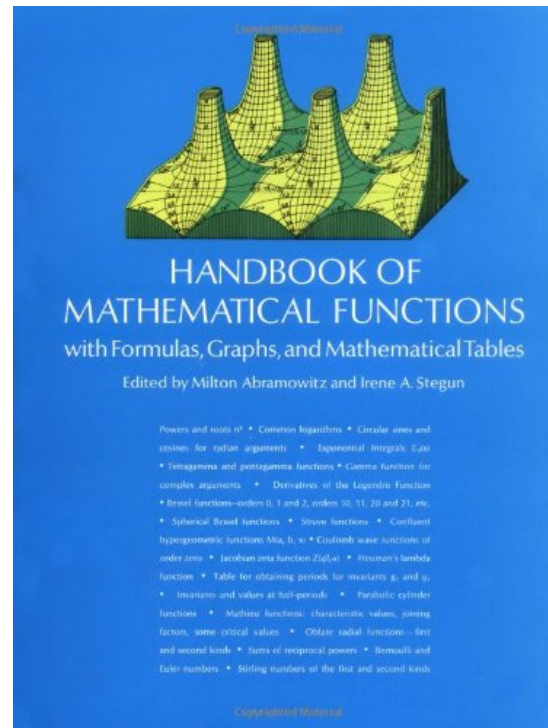
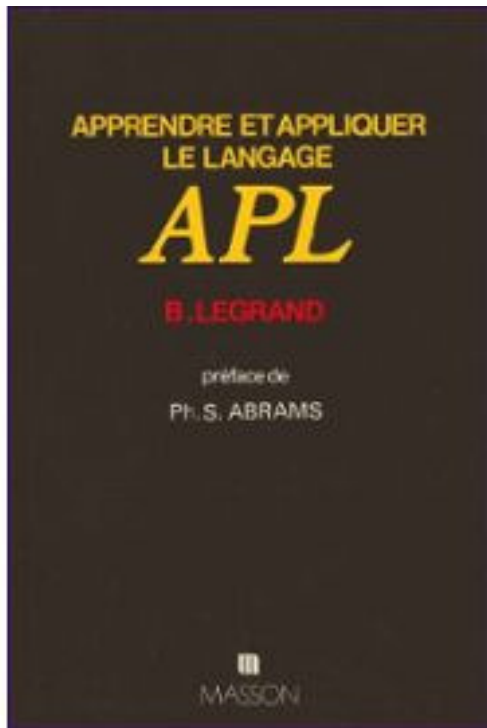
Computations involved a trapezoidal integration with a grid of 30 points on the (0,1) defined as  $t_i = (i/N)^3$  with  $i = 1, \dots, 30$  and  $N = 30$

## Résultats/a priori fractionnaire (b=0 vs 0.125)





## Mes ouvrages de référence



## Epilogue

- « Pour bien connaître les choses, il faut en savoir le détail; et comme il est presque toujours infini, nos connaissances sont toujours superficielles et imparfaites »  
La Rochefoucault (1678) Maximes, No 106
- « Recherche: partir de ce que l'on croit savoir et tirer sur le fil en souhaitant qu'il se brise. »  
J Rostand (1959) Carnet d'un biologiste, Stock

## Remerciements

### *-Etudiants Master & Thèse & Postdocs*

**Catherine Beaumont, Baddi Besbes, Laurianne Canario, Claire Chauvin**, Serguei Derban, Christine Dillmann, **Mylène Duval**, Françoise Clerget-Darpoux, **Eric Hanocq**, Barbara Heude, Toby Dylan Hocking, Luc Janss, Ina Hoeschele, Clara Israel, Florence Jaffrezic, **Eduardo Manfredi, Guillemette Marot, Cristian Meza**, Anna Maria Nebraska, Elie Maza, Miguel Perez-Enciso, **Christèle Robert-Granié, Andrea Rau, Magali San Cristobal**, Yoan Soussan, Caroline Thaon d'Arnoldi, **Etienne Verrier**.

### *-Collègues*

Christèle Robert-Granié, Céline Delmas, Florence Jaffrézic  
L Ollivier, M Gautier, JJ Colleau, B Langlois, A Neau  
JM Marin  
D Gianola, RL Quaas

### *-Groupes*

Applibugs, BaBayes, Monolix, Xian's Og, Jeu logique du Monde

## Quelques pronostics pour les 1ers matches de la coupe du monde

Groupe	No match	Equipe A	EquipeB	[1]	[X]	[2]
A	1	BRA	HRV	0.6883	0.1992	0.1124
A	2	MEX	CMR	0.5656	0.2504	0.184
B	3	ESP	NLD	0.4795	0.2754	0.2452
B	4	CHL	AUS	0.5486	0.2562	0.1953
C	5	COL	GRC	0.4725	0.277	0.2505
C	6	CIV	JPN	0.3911	0.2888	0.3201
D	7	URY	CRI	0.5498	0.2558	0.1944
D	8	GBR	ITA	0.3796	0.2894	0.331
E	9	CHE	ECU	0.3516	0.29	0.3585
E	10	FRA	HND	0.5622	0.2518	0.186
F	11	ARG	BIH	0.6118	0.2331	0.1551
F	12	IRN	NGA	0.3455	0.2898	0.3647
G	13	DEU	PRT	0.516	0.2661	0.2179
G	14	GHA	USA	0.2337	0.2719	0.4944
H	15	BEL	DZA	0.5718	0.2483	0.1799
H	16	RUS	KOR	0.493	0.2723	0.2346