Mixed effects models: modelling and inference

Ecole de printemps de la chaire
Modélisation mathématique et biodiversité

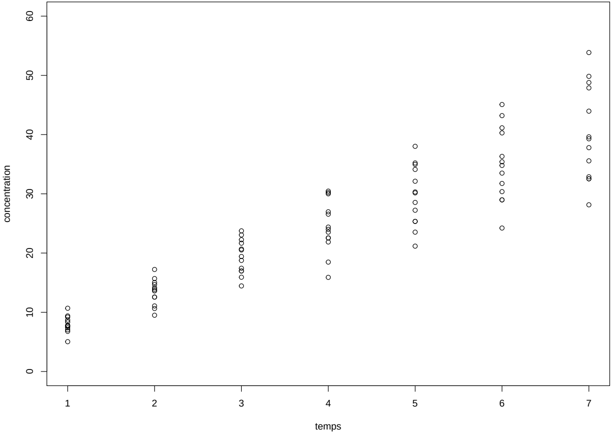
Mai 2019

Estelle Kuhn

INRA, MaIAGE

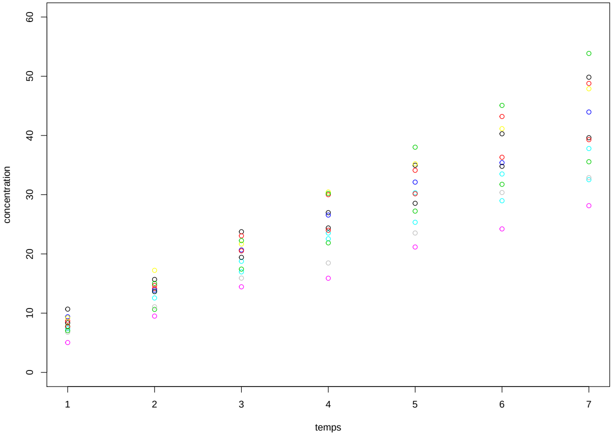
Outline

Measurements of concentration at several times



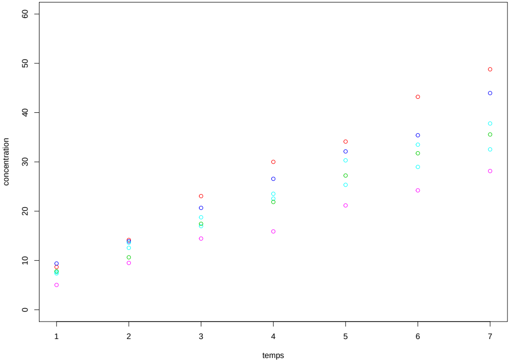
Repeated measurements of concentration at several times

$\Rightarrow$ concentrations for each individual of the population is measured at several times



Repeated measurements of concentration at several times

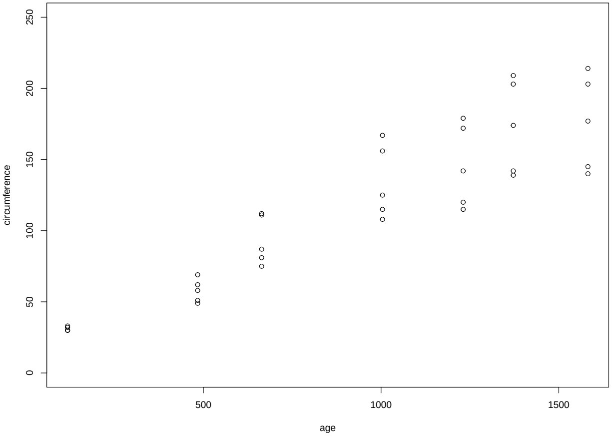
zoom on some individuals



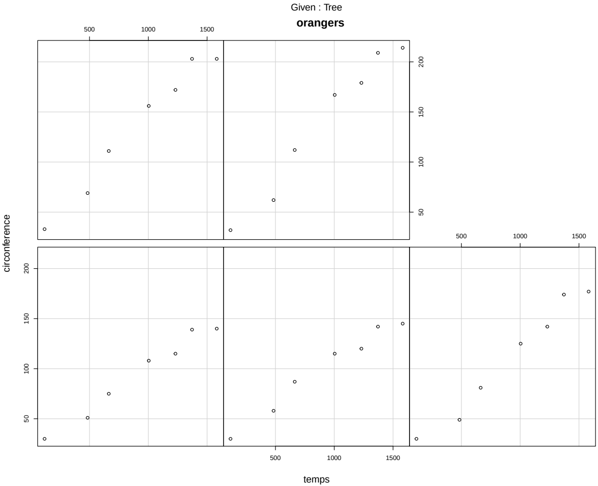
model suggestion for each individual i $y(t) = A_i + B_i t$ meaning that slope and intercept depend on individual i

Observations of growing process of orange trees

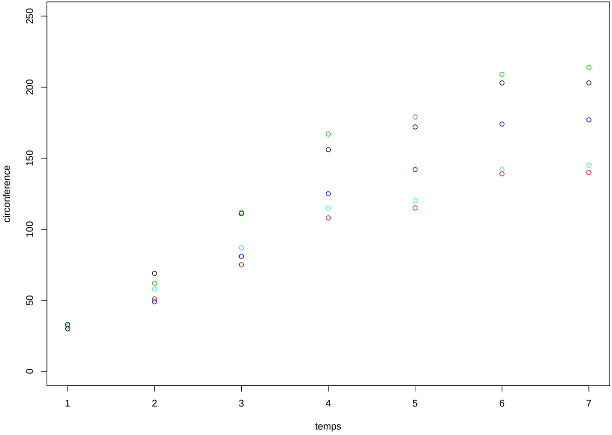
[Pinheiro and Bates (2000)]



Observations of growing process of five orange trees



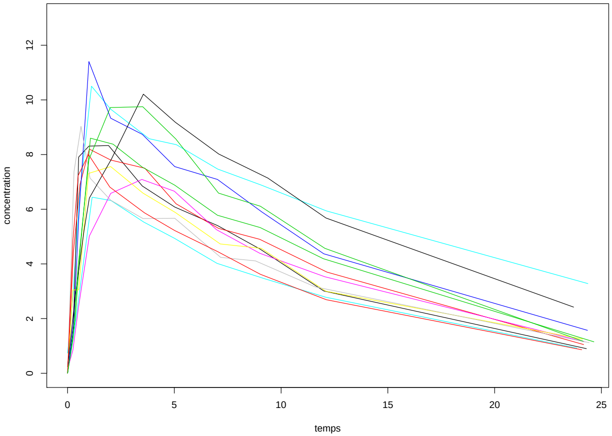
Observations of growing process of five orange trees



individual model suggestion $y(t) = \frac{\varphi_{i1}}{1+\exp\left(-\frac{t-\varphi_{i2}}{\varphi_{i3}}\right)}$

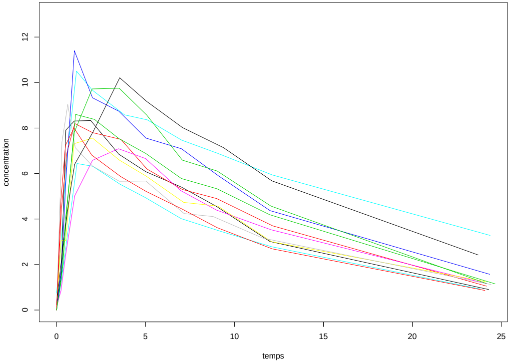
Theophylline concentration along time

[Davidian and Giltinian (1995)]



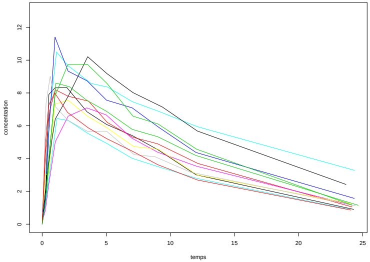
12 subjects, same oral dose (mg/kg) times in hours theophylline concentration in mg/L

Theophylline concentration along time



- Similarly-shaped concentration-time profiles across subjects
- Peak, rise, decay vary considerably
- Attributable to inter-subject variation in underlying PK processes (absorption, etc)

Some pharmacokinetic objectives



- Understanding intra-subject processes of drug absorption, distribution, and elimination governing achieved concentrations
  $\Rightarrow$ variabilities intra (within) subject and inter subject
- Understanding variations of these processes across subjects
  $\Rightarrow$ fundamental for developing dosing strategies and guidelines

Pharmacokinetic (PK) models

- One-compartment model for theophylline following oral dose d at time 0 describing the evolution of drug concentration over time.

$$y(t) = \frac{d_i ka_i}{V_i ka_i - Cl_i}\left[e^{-\frac{Cl_i}{V_i}t} - e^{-ka_i t}\right]$$

  where V_i, ka_i and Cl_i respectively denote the volume of the central compartment, the drug's absorption rate constant and the drug's clearance of individual i.
- ka_i, Cl_i and V_i summarize PK processes underlying observed concentration profiles for subject i

Some statistical objectives

- Determine mean/median values of ka_i, Cl_i and V_i and how they vary in the population of subjects
- Elucidate whether some of this variation is associated with subject characteristics (e.g. weight, age, renal function)
- Develop dosing strategies for subpopulations with certain characteristics (e.g. elderly, female)

General context

- Consider a response evolving over time (or other conditions) within individuals from a population of interest
- Inference focuses on mechanisms that underlie individual profiles of repeated measurements of the response and how these vary in the population
- A model for individual profiles with parameters that may be interpreted as representing such features or mechanisms is available

$\Rightarrow$ Common situations in agricultural, environmental, biomedical, economical applications

General setting of repeated measurements

- Measurements are repeated on each of N individuals
- Y_{ij} denotes the response at the jth measurement for individual i for $1 \leq j \leq J$
- X_i covariates of individual i

Example of Theophylline dataset

- Y_{ij} is drug concentration for subject i at time t_{ij}
- X_i contains subject characteristics such as weight, age, renal function, smoking status, etc for subject i

Individual-level model (Stage 1)

modelling the observations for $1 \leq i \leq N, 1 \leq j \leq J$

$$Y_{ij} = f(X_{ij}, \varphi_i) + \varepsilon_{ij}$$

- f function governing within-individual behavior
- $X_i = (X_{ij})_j$ covariates of individual i
- φ_i parameters specific to individual i
- ε_{ij} centered random error term

Example: Theophylline pharmacokinetic model

$$f(d_i, t_{ij}, \varphi_i) = \frac{d_i ka_i}{V_i ka_i - Cl_i}\left[e^{-\frac{Cl_i}{V_i} t_{ij}} - e^{-ka_i t_{ij}}\right]$$

where $\varphi_i = (ka_i, Cl_i, V_i)$ absorption rate, volume, and clearance for subject i

- $E(Y_{ij}|X_{ij}, \varphi_i) = f(X_{ij}, \varphi_i) \Rightarrow f$ represents an average profile
- f may not capture all within-individual variations

Population model (Stage 2)

Modeling the individual parameters for $1 \leq i \leq N$

$$\varphi_i = U_i \beta + V_i b_i$$

- U_i, V_i covariates of individual i
- β fixed effects of size d_f
- b_i centered random effects of individual i of size d_r

$\Rightarrow$ characterizes how elements of φ_i vary across individuals due to

- association with covariates modeled by β
- unexplained variation in the population represented by b_i

Example: Theophylline pharmacokinetic model
ka_i, Cl_i and V_i are individual random parameters such that
$\log ka_i = \log(ka) + b_{i,1},\ b_{i,1} \sim \mathcal{N}(0, \gamma_1)$
$\log Cl_i = \log(Cl) + \beta BW_i + b_{i,2},\ b_{i,2} \sim \mathcal{N}(0, \gamma_2)$
$\log V_i = \log(V) + b_{i,3},\ b_{i,3} \sim \mathcal{N}(0, \gamma_3)$
where BW_i is the body weight of individual i

Mixed effect model: art of modeling variabilities ?

- Modeling the observations for $1 \leq i \leq N, 1 \leq j \leq J$

$$Y_{ij} = f(X_{ij}, \varphi_i) + \varepsilon_{ij},$$

- Modeling the individual parameters for $1 \leq i \leq N$

$$\varphi_i = U_i \beta + V_i b_i$$

where

- X_i, U_i, V_i covariates of individual i
- β fixed effects
- b_i random effects of individual i
- φ_i parameters specific to individual i

Usual assumptions:

- $(b_i)_i$ are independent identically distributed
- $(Y_{ij} | b_i)_j$ are independent

Linear mixed effect models

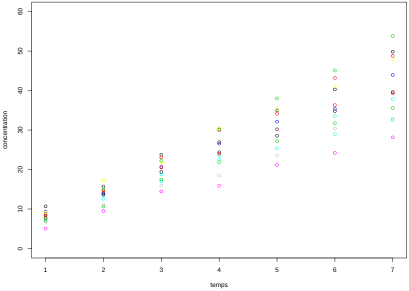
[Davidian and Giltinian (1995)]

$$Y_{ij} = X_{ij}\beta + Z_{ij}b_i + \varepsilon_{ij} \ , \ 1 \leq i \leq N, \ 1 \leq j \leq J$$

- $Y_i = (Y_{ij})_j$ is the observation vector for individual i
- X_i and Z_i are matrices of known covariates of individual i
- β is the vector of fixed effects
- $b_i \overset{iid}{\sim} \mathcal{N}(0, \Gamma)$
- ε_i is a random error vector, with $\varepsilon_i \overset{iid}{\sim} \mathcal{N}_J(0, \Sigma)$

$\Rightarrow$ Parameters of models: $\theta = (\beta, \Gamma, \Sigma)$

Example of concentrations with slope and intercept depending on the individual



$$Y_{ij} = (A + a_i) + (B + b_i)t_{ij} + \varepsilon_{ij} \ , \ 1 \leq i \leq N, \ \ 1 \leq j \leq J$$

with $a_i \overset{iid}{\sim} \mathcal{N}(0, \gamma_a^2)$ and $b_i \overset{iid}{\sim} \mathcal{N}(0, \gamma_b^2)$ and $\varepsilon_{ij} \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$

Nonlinear mixed effects model

[Davidian Giltinian (1995), PInheiro Bates (2000), Lavielle (2014)]
$\Rightarrow$ the function f is nonlinear in the individual parameter φ_i

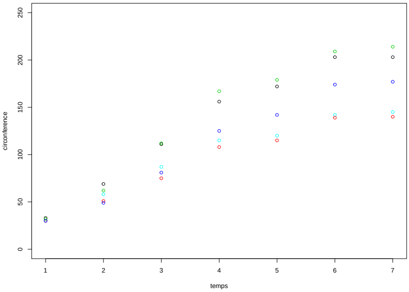
$$\begin{cases} Y_{ij} & = f(X_{ij}, \varphi_i) + \varepsilon_{ij}, \quad 1 \leq i \leq N, \ 1 \leq j \leq J \\ \varphi_i & = U_i \beta + V_i b_i, \quad 1 \leq i \leq N \end{cases}$$

where

- $Y_i = (Y_{ij})_j$ is the observation vector for individual i
- X_i and U_i, V_i are matrices of known covariates of individual i
- β is the vector of fixed effects
- b_i is a random effect of individual i, e.g. $b_i \overset{iid}{\sim} \mathcal{N}(0, \Gamma)$
- ε_i is a random error vector, e.g. $\varepsilon_i \overset{iid}{\sim} \mathcal{N}_J(0, \Sigma)$

$\Rightarrow$ Parameters of models: $\theta = (\beta, \Gamma, \Sigma)$

Example of the orange trees



$$Y_{ij} = \frac{\varphi_{i1}}{1+\exp\left(-\frac{t_j-\varphi_{i2}}{\varphi_{i3}}\right)} + \varepsilon_{ij}, \quad \text{with} \quad \varphi_i = \beta + b_i \ ,$$

where $\beta = (\beta_1, \beta_2, \beta_3) \in \mathbb{R}^3$, $b_i \overset{iid}{\sim} \mathcal{N}_3(0, \Gamma)$ and $\varepsilon_{ij} \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$.

Representation as hierarchical model

? link between mixed effects models and hierarchical models

$\Rightarrow$ differents representations of the same model [Lavielle (2014)]

$$\begin{cases} \varphi_i &= & U_i\beta + V_i b_i & \text{with} & b_i \sim q(.;\Gamma) & \text{(stage 2)} \\ Y_i &= & f(X_i,\varphi_i) + \varepsilon_i & \text{with} & \varepsilon_i \sim q(.;\Sigma) & \text{(stage 1)} \end{cases}$$

more generaly :

$$\begin{cases} b_i & \sim & q(.;\Gamma) \\ Y_i | b_i; X_I, U_i, V_i & \sim & q(.;\beta,\Sigma) \end{cases}$$

$\Rightarrow$ latent variables model structure

Summary of the day

$$\begin{cases} Y_{ij} &= f(X_{ij}, \varphi_i) + \varepsilon_{ij}, \quad 1 \le i \le N,\ 1 \le j \le J \\ \varphi_i &= U_i \beta + V_i b_i, \quad\quad 1 \le i \le N \end{cases}$$

with $b_i \overset{iid}{\sim} q(.;\Gamma)$ and $\varepsilon_{ij} \overset{iid}{\sim} q(.;\Sigma)$
wth parameters of models: $\theta = (\beta, \Gamma, \Sigma)$

- modeling observation level and individual level
- combining fixed effects and random effects
- possibly linear or nonlinear dependency of the response
- possible with heteroscedastic error model:

$$\begin{cases} Y_i &= f(X_i, \varphi_i) + g(X_i, \varphi_i)\varepsilon_i \quad \text{with} \quad \varepsilon_i \sim q(.;\Sigma) \\ \varphi_i &= U_i \beta + V_i b_i \quad \text{with} \quad b_i \sim q(.;\Gamma) \end{cases}$$

- representation as hierarchical modeling
- latent variables model structure

Context of plant breeding

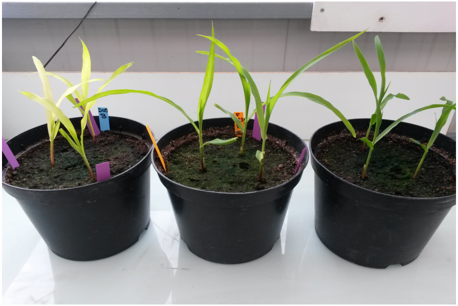


Figure: Maïs en stress froid (INRA Mons)

- genotype by environment interaction
- Challenge : find the "best" variety for a given environment
- Opportunity : adaption to climate change

Data acquisition

phenotyping platform in controled condition
$\Longrightarrow$ measurement of biomass, height, yield

Figure: Phenoarch INRA Montpellier

Data acquisition

phenotyping platform in open field

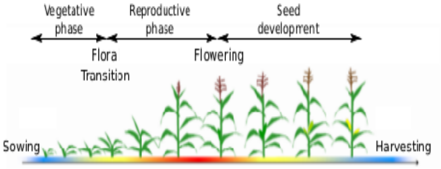
$\Longrightarrow$ under semi controlled condition

Figure: Pheno3C INRA Clermont-Ferrand

$\Longrightarrow$ Using data to calibrate crop model

$\Longrightarrow$ Compute "good" values for parameters as root emergence rate, leaf emergence rate

Modeling plant growth process

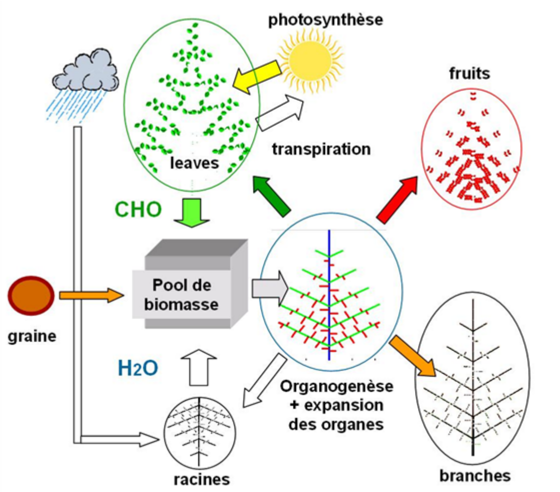


$\Longrightarrow$ Many questions:

- times of interest: floral transition, flowering time, leaf appearance, root appearance
- covariables of interest
- genotypic effect

$\Longrightarrow$ Describe the growth process by ecophysiological model

Crop growth modeling



$\Longrightarrow$ many unknown mecanistic parameters

Ecophysiological modeling: Greenlab model

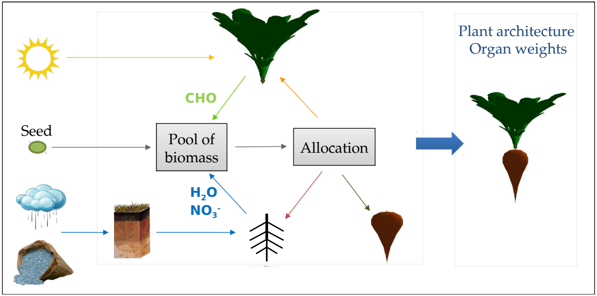


Figure: Overview of the Greenlab model

$\Longrightarrow$ estimate many unknown mechanistic parameters

$\Longrightarrow$ modeling the different levels of variability

$\Longrightarrow$ identify which parameters depend on the genotype

$\Longrightarrow$ reduce the number of parameters to estimate

Mixed effects model for crop model analysis

$\Longrightarrow$ modeling observations conditionaly to individual paramater

$$y_{ijk} = f(\varphi_i, e_j) + \varepsilon_{ijk}, \quad 1 \leq i \leq N, \quad 1 \leq j \leq J, \quad 1 \leq k \leq K$$

with y_{ijk} measurement of plant kth of genotype i in environnemental condition j
φ_i parameter of genotype i
e_j environnemental covariates
Σ population parameters vector

$\Longrightarrow$ modeling genotypic variability of crop model parameter using individual parameter meaning that for genotype i model parameter are modeled by:

$$\varphi_i = \beta + b_i \text{ with } b_i \sim \mathcal{N}(0; \Gamma), \quad 1 \leq i \leq N,$$

$\Longrightarrow$ model parameters $\theta = (\beta, \Gamma, \Sigma) \in \Theta$

Inference in mixed effects models

- Linear mixed effects models

$$Y_{ij} = X_{ij}\beta + Z_{ij}b_i + \varepsilon_{ij} \;\; 1 \leq i \leq N, \; 1 \leq j \leq J$$

- Nonlinear mixed effects models

$$\begin{cases} Y_{ij} & = & f(X_{ij}, \varphi_i) + \varepsilon_{ij} & 1 \leq i \leq N, \; 1 \leq j \leq J \\ \varphi_i & = & U_i\beta + V_i b_i, & 1 \leq i \leq N \end{cases}$$

with U_i, V_i and Z_i design matrices,
β population parameters also called fixed effects,
$b_i \overset{iid}{\sim} q(.;\Gamma)$ random effects
$\varepsilon_i \overset{iid}{\sim} q(.;\Sigma)$ noise term independent of b_i
f a nonlinear function of φ_i.

Parameters of models: $\theta = (\beta, \Gamma, \Sigma)$

Statistical issues

Consider the following mixed effects model:

$$\left\{ \begin{array}{lcll} Y_{ij} & = & f(X_{ij}, \varphi_i) + \varepsilon_{ij} & 1 \leq i \leq N,\ 1 \leq j \leq J \\ \varphi_i & = & U_i \beta + V_i b_i, & 1 \leq i \leq N \end{array} \right.$$

with $b_i \overset{iid}{\sim} q(.; \Gamma)$ random effects
and $\varepsilon_{ij} \overset{iid}{\sim} q(.; \Sigma)$ noise term independent of (b_i)

Objectives:

- estimate model parameters $\theta = (\beta, \Gamma, \Sigma) \in \Theta$
- predict individual output as $\hat{\varphi}_i$ or $\hat{Y}_i$
- test if some fixed effects β are significant
- test if some random effects (b_i) are fixed
- ...

Likelihoods in mixed effects model

Consider random variables $(Y_i, b_i)_i$ following the model given by:

$$\left\{ \begin{array}{lcll} Y_{ij} & = & f(X_{ij}, \varphi_i) + \varepsilon_{ij} & 1 \leq i \leq N,\ 1 \leq j \leq J \\ \varphi_i & = & U_i \beta + V_i b_i, & 1 \leq i \leq N \end{array} \right.$$

with $b_i \overset{iid}{\sim} q(.;\Gamma)$ and $\varepsilon_{ij} \overset{iid}{\sim} q(.;\Sigma)$ with $\theta = (\beta, \Gamma, \Sigma) \in \Theta$

Define the complete likelihood:

$$\begin{aligned} L_{comp}(\theta; Y_1^N, b_1^N) &= \prod_1^N L_{comp}(\theta; Y_i, b_i) \\ &= \prod_1^N \left(p(Y_i | b_i; \beta, \Sigma) p(b_i; \Gamma)\right) \end{aligned}$$

$\Rightarrow$ the random effects (b_i) are non observed

$\Rightarrow$ integrate over the random effects b_i

Likelihoods in mixed effects model

Define the observed (or marginal) likelihood:

$$
\begin{aligned}
L_{marg}(\theta; Y_1^N) &= \prod_1^N L_{marg}(\theta; Y_i) \\
&= \prod_1^N \int L_{comp}(\theta; Y_i, b_i) db_i \\
&= \prod_i \int p(Y_i | b_i; \beta, \Sigma) p(b_i; \Gamma) db_i
\end{aligned}
$$

Define the maximum likelihood estimate (MLE) by:

$$\hat{\theta}_N = \arg\max_{\theta \in \Theta} L_{marg}(\theta; Y_1^N)$$

- ? theoretical properties of MLE? as N goes to infinity? consistency ? asymptotic normality ?
- computational aspects

Maximum likelihood estimator: consistency

[Nie, Metrika (2006)]

$$\hat{\theta}_N = \arg\max_{\theta \in \Theta} L_{marg}(\theta; Y_1^N)$$

Under regularity and moment conditions on the model, the MLE estimator $\hat{\theta}_N$ exists almost surely and

$$\lim_{N \Rightarrow +\infty} \hat{\theta}_N = \theta_0 \ P_{\theta_0} - p.s.$$

$\Rightarrow$ Example of logistic model for orange trees satisfayes these conditions.

Maximum likelihood estimator: convergence rates

- in general regular parametric models MLE is $\sqrt{N}$ consistent
- what is the role of J in mixed effects model?

Example of balanced ANOVA model with one way:

$$Y_{ij} = \alpha + b_i + \varepsilon_{ij}\ ,\ 1 \leq i \leq N,\ 1 \leq j \leq J$$

with $b_i \overset{iid}{\sim} \mathcal{N}(0, \gamma^2)$, $\varepsilon_i \overset{iid}{\sim} \mathcal{N}_J(0, \sigma^2)$, ε_i independent of (b_i)

$\Rightarrow \hat{\alpha} - \alpha_0 = O_p(N^{-1/2})$

$\Rightarrow \hat{\gamma}^2_{MLE} - \gamma^2_0 = O_p((N)^{-1/2})$

$\Rightarrow \hat{\sigma}^2_{MLE} - \sigma^2_0 = O_p((NJ)^{-1/2})$

Maximum likelihood estimator: convergence rates

- in general regular parametric models MLE is $\sqrt{N}$ consistency
- what is the role of J in mixed effects model?

Example of with "intercept and slope":

$$Y_{ij} = \alpha + \beta X_j + b_i + \varepsilon_{ij} \; , \; 1 \leq i \leq N, \; 1 \leq j \leq J$$

with $b_i \overset{iid}{\sim} \mathcal{N}(0, \gamma^2)$, $\varepsilon_i \overset{iid}{\sim} \mathcal{N}_J(0, \sigma^2)$, ε_i independent of (b_i)

$\Rightarrow \hat{\alpha} - \alpha_0 = O_p(N^{-1/2})$

$\Rightarrow \hat{\beta} - \beta_0 = O_p((NJ)^{-1/2})$

Maximum likelihood estimator: convergence rates

[Nie, JSPI (2007)]

$$\begin{cases} Y_{ij} = f(X_{ij}, \varphi_i) + \varepsilon_{ij} & 1 \le i \le N, \ 1 \le j \le J \\ \varphi_i = U_i \beta + V_i b_i, & 1 \le i \le N \end{cases}$$

with $b_i \overset{iid}{\sim} q(.; \Gamma)$ and $\varepsilon_{ij} \overset{iid}{\sim} q(.; \Sigma)$ with parameters $\theta = (\beta, \Gamma, \Sigma)$

$$\hat{\theta}_N = \arg\max_{\theta \in \Theta} L_{marg}(\theta; Y_1^N)$$

Under regularity assumptions and moment conditions on the model

- For fixed J, $\hat{\theta}_N$ is $\sqrt{N}$ consistent when N tends to infinity.
- the MLE $\hat{\beta}_N$ for β is $\sqrt{NJ}$ consistent and
  the MLE $\hat{\Gamma}_N$ for Γ is $\sqrt{N}$ consistent when N and J tend to infinity

Moreover the asymptotic covariance matrix is equal to the inverse of the Fisher matrix information.

Maximum likelihood estimation: computational aspect

Recall the definition of observed (or marginal) likelihood:

$$L_{marg}(\theta; Y_1^N) \;=\; = \prod_i \int \underbrace{p(Y_i|b_i;\beta,\Sigma)}_{stage1}\underbrace{p(b_i;\Gamma)}_{stage2} db_i$$

and of the maximum likelihood estimator (MLE):

$$\hat{\theta}_n = \arg\max_{\theta\in\Theta} L_{marg}(\theta; Y_1^N)$$

Example of gaussian linear mixed effects model:

$$Y_i = X_i\beta + Z_i b_i + \varepsilon_i \;,\; 1 \le i \le N,$$

with $b_i \overset{iid}{\sim} \mathcal{N}(0,\Gamma)$, $\varepsilon_i \overset{iid}{\sim} \mathcal{N}_J(0,\Sigma)$, ε_i independent of (b_i)
$\Rightarrow Y_i \sim \mathcal{N}(X_i\beta, Z_i'\Gamma Z_i + \Sigma)$

Maximum likelihood estimation: computational aspect

- Exact likelihood methods: Maximize likelihood "directly" using deterministic or stochastic approximation to the integrals
  - Deterministic approximation (Quadrature, Adaptive Gaussian quadrature)
  - Stochastic approximation (Importance sampling, brute-force Monte Carlo integration)

  $\Rightarrow$ computationaly expensive in particular in high-dimensional setting
- inference based on linearization of the likelihood
  $\Rightarrow$ no guarantee of convergence
- iterative procedure based on individual estimates
  $\Rightarrow$ no guarantee of convergence
- tools for maximum likelihood estimation in latent variables model

Some existing approximate methods (non exhaustive)

- Methods based on approximations of the likelihood
  - First order methods (FO, Beal and Sheiner, 1982)
  - First order conditional methods (FOCE, Lindstrom and Bates, 1990)
  - Laplace-EM (Vonesh, 1996) also called mode approximation

  No convergence property or with non realistic assumptions, default of convergence.

- Methods based on the exact likelihood
  - MCEM algorithm (Walker, 1996 ; Fort and Moulines, 2004)
  - SAEM algorithm (Delyon, Lavielle and Moulines, 1999)

  Convergence property

Estimation in latent variables model

Heuristic of approach in latent variables models

$$\begin{cases} Y_{ij} & = & f(X_{ij}, \varphi_i) + \varepsilon_{ij} & 1 \leq i \leq N, \ 1 \leq j \leq J \\ \varphi_i & = & U_i \beta + V_i b_i, & 1 \leq i \leq N \end{cases}$$

with $b_i \overset{iid}{\sim} q(.;\Gamma)$ and $\varepsilon_{ij} \overset{iid}{\sim} q(.;\Sigma)$

- Observed data (Y_i) $\Rightarrow$ observed vectors
- Random effects (b_i) $\Rightarrow$ latent variables

$\Rightarrow$ if (b_i) were observed,
then consider as objective function $\log L_{comp}(\theta; Y_1^N, b_1^N)$

$\Rightarrow$ instead consider the quantity $E[\log L_{comp}(\theta; Y_1^N, b_1^N))|Y_1^N; \theta]$.

$\Rightarrow$ iterative approach: maximize in θ the quantity
$Q(\theta|\theta_{current}) = E[\log L_{comp}(\theta; Y_1^N, b_1^N))|Y_1^N; \theta_{current}]$.

The EM algorithm [Dempster et al. (1977), Wu (1983), Vaida (2005)]

Iteration k of the algorithm:

- Expectation step :

$$Q(\theta|\theta_{k-1}) = E[\log L_{comp}(Y, b; \theta)|Y; \theta_{k-1}]$$

- Maximization step :

$$\theta_k = \arg\max_{\theta \in \Theta} \; Q(\theta|\theta_{k-1})$$

Proposition

If $Q(\theta_{k-1}|\theta_{k-1}) \leq Q(\theta_k|\theta_{k-1})$,
then $\log L_{marg}(\theta_{k-1}; Y_1^N) \leq \log L_{marg}(\theta_k; Y_1^N)$

Proposition

Under regularity condition on the model, the sequence (θ_k) *converges toward a critical point of the observed likelihood* L_{marg}.

Limits of EM algorithm

Iteration k of the algorithm:

- Expectation step :

$$Q(\theta|\theta_{k-1}) = E[\log L_{comp}(Y, b; \theta)|Y; \theta_{k-1}]$$

- Maximization step :

$$\theta_k = \arg\max_{\theta \in \Theta} \; Q(\theta|\theta_{k-1})$$

$\Rightarrow$ Limits of EM algorithm:

- theory in exponential model
- nature of the limit point
- convergence depends on the initial guess
- expression of $Q(\theta|\theta')$ often analytically intractable
  $\Rightarrow$ approximate the quantity $Q(\theta|\theta')$?

Heuristics of the stochastic approximation

Quantity of interest in the EM algorithm:

$$Q(\theta|\theta') = E(\log L_{comp}(y, b; \theta)|y; \theta')$$

$\Rightarrow$ build a sequential approximation of this quantity: at iteration k

- simulate a realization b_k of the random effects
- compute $Q_k(\theta) = Q_{k-1}(\theta) + \gamma_k \left(\log L_{comp}(y, b_k; \theta) - Q_{k-1}(\theta)\right)$ where (γ_k) is a positive decreasing step size sequence.

Then, we have:

$$\frac{Q_k(\theta) - Q_{k-1}(\theta)}{\gamma_k} = E[\log L_{comp}(y, b; \theta)|y; \theta] - Q_{k-1}(\theta)$$
$$+ \log f(y, b_k; \theta) - E[\log L_{comp}(y, b; \theta)|y; \theta]$$

$$\frac{Q_k(\theta) - Q_{k-1}(\theta)}{\gamma_k} \approx E[\log L_{comp}(y, b; \theta)|y; \theta] - Q_{k-1}(\theta) + e_k$$

If $b_k \sim p(\cdot|y, \theta)$ then $e_k \approx 0$

Stochastic Approximation of the EM algorithm

[Delyon, Lavielle, Moulines (1999) AS]

Iteration k of the algorithm:

- Simulation step : $b^k \sim \pi_{\theta_{k-1}}(.|y)$
  where π_θ is the distribution of b conditionaly to y
- Stochastic approximation :
  $Q_k(\theta) = Q_{k-1}(\theta) + \gamma_k[\log L_{comp}(y, b^k, \theta) - Q_{k-1}(\theta)]$
  where (γ_k) is a decreasing sequence of positive step-sizes.
- Maximisation step : $\theta_k = \arg\max_{\theta\in\Theta} Q_k(\theta)$

$+$ converges almost surely toward a stationary point $\widehat{\theta}$ of L_{marg}
– theory in exponential model
– nature of the limit point
– convergence depends on the initial guess

Extension of SAEM algorithm using MCMC procedure

[K. Lavielle (2004) , Allassonnière, K., Trouvé (2010)]

- Simulation step : $b^k \sim \Pi_{\theta_{k-1}}(b^{k-1}, \cdot)$
  where Π_θ is a transition probability of an ergodic Markov Chain having the posterior distribution $p(\cdot|y,\theta)$ as stationary distribution,
- Stochastic approximation :
  $Q_k(\theta) = Q_{k-1}(\theta) + \gamma_k \left(\log L_{comp}(y, b^k, \theta) - Q_{k-1}(\theta)\right)$
- Maximisation step : $\theta_k = \arg\max_{\theta\in\Theta} Q_k(\theta)$

Simulation step : one step of a Metropolis Hastings algorithm

- simulate a candidate from a proposal distribution
  $b^c \sim q_{\theta_{k-1}}(.|b^{k-1})$
- accept or reject this candidate with probability

$$\alpha(b^{k-1}, b^c) = \min\left(1, \frac{p(b^c|y,\theta)q_{\theta_{k-1}}(b^{k-1}|b^c)}{p(b^{k-1}|y,\theta)q_{\theta_{k-1}}(b^c|b^{k-1})}\right)$$

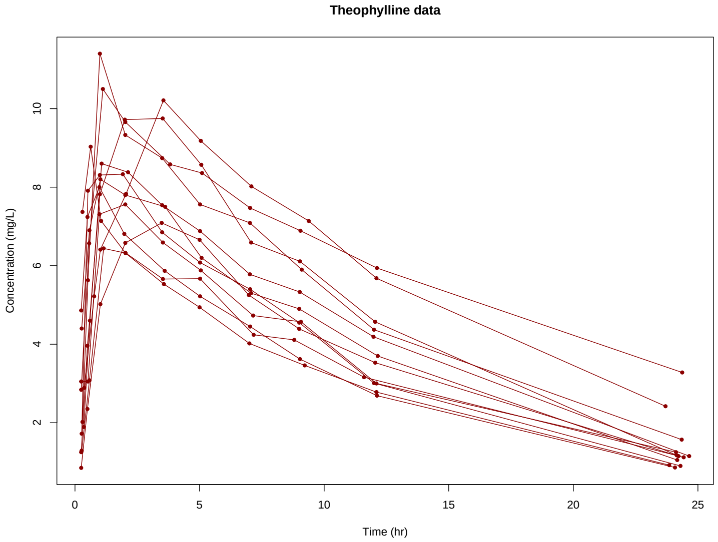
Estimation of parameters of theophylline model

⇒ use saemix R package [Comets et al (2017)]

Example of R code

```
library(saemix)
#data creation
data("theo.saemix")
theo.data <- saemixData(name.data = theo.saemix,
 header = TRUE,sep = " ", na = NA,
name.group = c("Id"),
 name.predictors = c("Dose", "Time"),
name.response = c("Concentration"),
name.covariates = c("Weight", "Sex"),
units = list(x = "hr",y = "mg/L", covariates =
c("kg", "-")),
name.X = "Time")
plot(theo.data,type = "b", col = "DarkRed", main =
"Theophylline data")
```

Estimation of parameters of theophylline model



Estimation of parameters of theophylline model

```
#model definition
model1cpt <- function(psi, id, xidep) {
dose <- xidep[, 1]
tim <- xidep[, 2]
ka <- psi[id, 1]
V <- psi[id, 2]
CL <- psi[id, 3]
k <- CL / V
ypred <- dose * ka / (V * (ka - k)) * (exp( -k * tim)
-+ exp( - ka * tim))
return(ypred)
}
```

$\Rightarrow$ correspond to model equation defined above:

$$f(d_i, \varphi_i, t) = \frac{d_i ka_i}{V_i ka_i - Cl_i}\left[e^{-\frac{Cl_i}{V_i}t} - e^{-ka_i t}\right]$$

Estimation of parameters of theophylline model

```
# model structure definition
theo.model <- saemixModel(model = model1cpt,
description = "One-compartment model with first-order
absorption",
psi0 = matrix(c(1, 20, 0.5), ncol = 3,
byrow = TRUE, dimnames = list(NULL, c("ka", "V",
"CL"))), transform.par = c(1, 1, 1),
covariate.model = matrix(c(0, 0, 1, 0, 0, 0),
ncol = 3, byrow = TRUE))
#option definition
opt <- list(save = FALSE, save.graphs = FALSE)
#fitting model with data
theo.fit <- saemix(theo.model, theo.data, opt)
```

Estimation of parameters of theophylline model: results

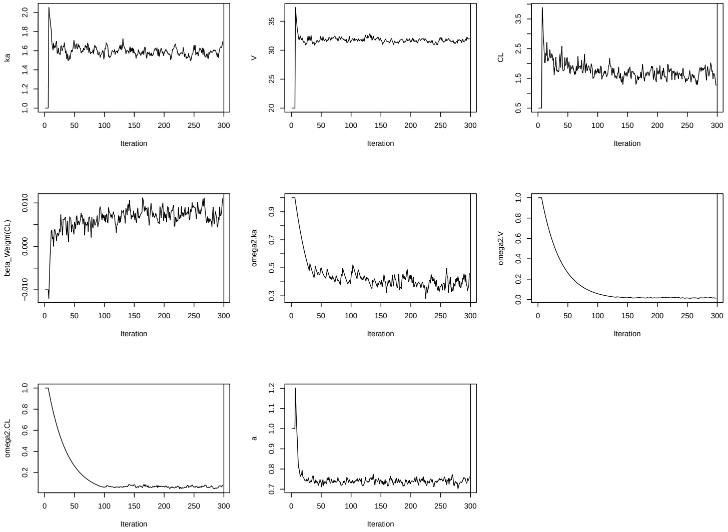
Fixed effects

Parameter	Estimate	SE	CV(%)	p-value
[1,] ka	1.5786	0.2947	18.7	-
[2,] V	31.6605	1.4322	4.5	-
[3,] CL	1.5521	0.9683	62.4	-
[4,] β_W(CL)	0.0082	0.0089	108.3	0.18
[5,] a	0.7429	0.0569	7.7	-

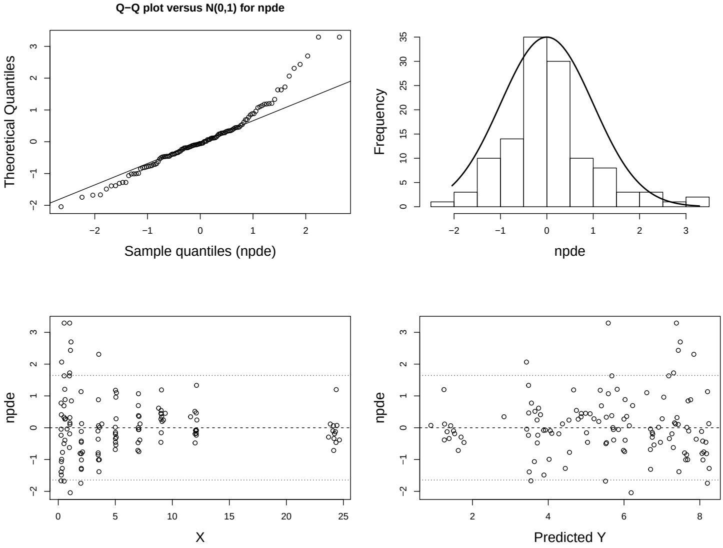
Variance of random effects

Parameter	Estimate	SE	CV(%)
ka omega2.ka	0.368	0.1668	45
V omega2.V	0.017	0.0096	57
CL omega2.CL	0.065	0.0324	50

Estimation of parameters of theophylline model



Estimation of parameters of theophylline model



Estimation of parameters of theophylline model

```
Statistical criteria
--------------------------------------------------------
Likelihood computed by linearisation
-2LL= 343.427
AIC = 359.427
BIC = 363.3063
Likelihood computed by importance sampling
-2LL= 344.8205
AIC = 360.8205
BIC = 364.6997
```

Prediction in mixed effects model

Consider a mixed effects model:

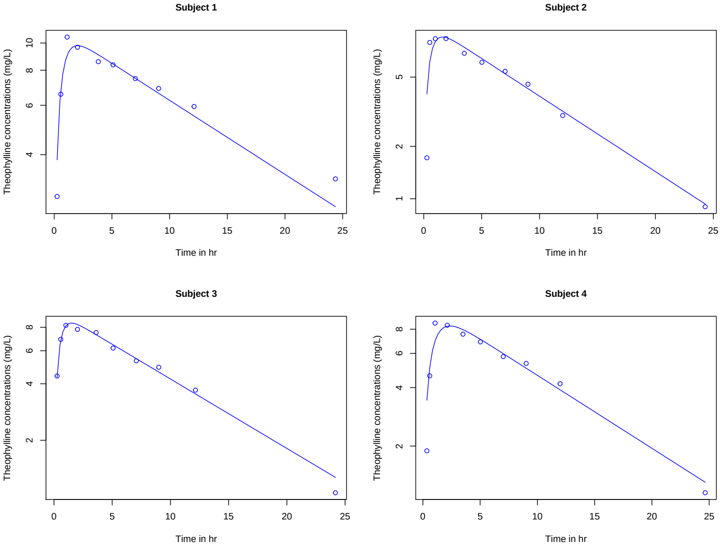
$$\begin{cases} Y_{ij} &= f(X_{ij}, \varphi_i) + \varepsilon_{ij} \quad 1 \leq i \leq N,\ 1 \leq j \leq J \\ \varphi_i &= U_i\beta + V_i b_i, \quad 1 \leq i \leq N \end{cases}$$

with $b_i \overset{iid}{\sim} q(.;\Gamma)$ and $\varepsilon_{ij} \overset{iid}{\sim} q(.;\Sigma)$

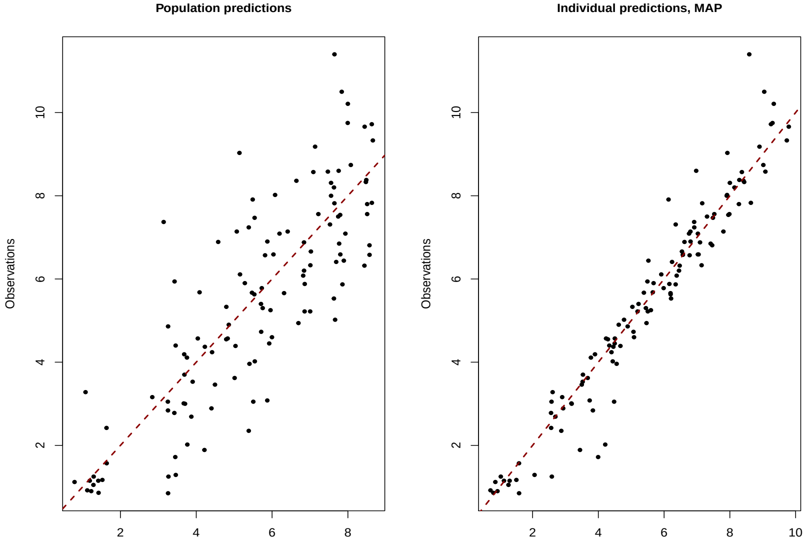
$\Rightarrow$ predicted values for random effects for $1 \leq i \leq N$:
$\hat{b}_i = E(b_i|Y_i)$ or $\hat{b}_i = \arg\max q(b_i|Y_i)$

$\Rightarrow \hat{\varphi}_i = U_i\hat{\beta} + V_i\hat{b}_i$ and $\hat{Y}_i = f(X_i, \hat{\varphi}_i)$

Prediction of individual profiles of theophylline model



Comparision between population predictions and individual predictions in theophylline model



List of toolboxs (non exhaustive)

- R package nlme [Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., and R Core Team (2019)].
- R package lme4 [Bates et al. (2019)]
- R package saemix [Comets, E., Lavenu, A., and Lavielle, M. (2017)]
- SPSS (2002). Linear mixed-effects modeling in SPSS. An introduction to the MIXED procedure.
- SAS Proc NLMIXED
- MONOLIX (2013)

Summary of the day

$$\begin{cases} Y_{ij} &= f(X_{ij}, \varphi_i) + \varepsilon_{ij}, \quad 1 \leq i \leq N,\ 1 \leq j \leq J \\ \varphi_i &= U_i \beta + V_i b_i, \quad\quad 1 \leq i \leq N \end{cases}$$

with $b_i \overset{iid}{\sim} q(.;\Gamma)$ and $\varepsilon_{ij} \overset{iid}{\sim} q(.;\Sigma)$ with $\theta = (\beta, \Gamma, \Sigma)$

- Define the complete likelihood: $L_{comp}(\theta; Y_1^N, b_1^N)$
- Define the observed likelihood: $L_{marg}(\theta; Y_1^N) = \int L_{comp}(\theta; Y_1^N, b_1^N) db_1^N$
- Define the maximum likelihood estimate (MLE) by:

$$\hat{\theta}_N = \arg\max_{\theta \in \Theta} L_{marg}(\theta; Y_1^N)$$

- good properties for $\hat{\theta}_N$
- efficient convergent stochastic algorithm to evaluate $\hat{\theta}_N$
- corresponding toolbox and R packages

Extension of SAEM algorithm using MCMC procedure

[K. et al. (2004), Allassonniere et al. (2010)]

- Simulation step : $b^k \sim \Pi_{\theta_{k-1}}(b^{k-1}, \cdot)$
  where Π_θ is a transition probability of an ergodic Markov Chain having the posterior distribution $p(\cdot|y,\theta)$ as stationary distribution,
- Stochastic approximation :
  $Q_k(\theta) = Q_{k-1}(\theta) + \gamma_k \left(\log L_{comp}(y, b^k, \theta) - Q_{k-1}(\theta)\right)$
- Maximisation step : $\theta_k = \arg\max_{\theta \in \Theta} Q_k(\theta)$

Simulation step : one step of a Metropolis Hastings algorithm

- simulate a candidate from a proposal distribution
  $b^c \sim q_{\theta_{k-1}}(.|b^{k-1})$
- accept or reject this candidate with probability

$$\alpha(b^{k-1}, b^c) = \min\left(1, \frac{p(b^c|y,\theta)q_{\theta_{k-1}}(b^{k-1}|b^c)}{p(b^{k-1}|y,\theta)q_{\theta_{k-1}}(b^c|b^{k-1})}\right)$$

Additional comments and discussions on maximum likelihood estimation in mixed effects models

- tuning of the parameters in stochastic algorithms
- tuning of the MCMC procedure
- computation of the likelihood
- computation of the Fisher information matrix
- identifiability of the model

Alternative approach: bayesian inference

- consider θ as a random variable
- choose a prior distribution for θ denoted by π

$$\begin{cases} \theta & \sim & \pi \\ b_i & \overset{iid}{\sim} & q(.;\Gamma) \\ Y_i|b_i;X_i,U_i,V_i & \overset{i}{\sim} & q(.;\beta,\Sigma) \end{cases}$$

- simulate a (quasi) sample of the distribution of (θ, b) conditionaly to the observation Y

$\Rightarrow$ use intensive computational tools as MCMC, importance sampling, ABC

Testing fixed effects in mixed effects model

$$\begin{cases} Y_{ij} &= f(X_{ij}, \varphi_i) + \varepsilon_{ij}, \quad 1 \leq i \leq N,\ 1 \leq j \leq J \\ \varphi_i &= U_i \beta + V_i b_i, \quad 1 \leq i \leq N \end{cases}$$

with $b_i \overset{iid}{\sim} q(.;\Gamma)$ and $\varepsilon_{ij} \overset{iid}{\sim} q(.;\Sigma)$ with $\theta = (\beta, \Gamma, \Sigma)$

- test whether the covariate effect β is significant or not

Example: Theophylline pharmacokinetic model
ka_i, Cl_i and V_i are individual random parameters such that
$\log ka_i = \log(ka) + b_{i,1},\ b_{i,1} \sim \mathcal{N}(0, \gamma_1)$
$\log Cl_i = \log(Cl) + \beta BW_i + b_{i,2},\ b_{i,2} \sim \mathcal{N}(0, \gamma_2)$
$\log V_i = \log(V) + b_{i,3},\ b_{i,3} \sim \mathcal{N}(0, \gamma_3)$
where BW_i is the body weight of individual i

Likelihood ratio test statistic

Let $(Y_1, ..., Y_N)$ be a sample having density f_θ, $\theta \in \Theta \subset \mathbb{R}^q$

Consider the test defined by
$H_0 :$ "$\theta \in \Theta_0$" against $H_1 :$ "$\theta \in \Theta_1$"

Then the likelihood ratio test statistic equals to

$$LRT_N = -2 \log \left(\frac{\sup_{\theta \in \Theta_0} L_N(\theta)}{\sup_{\theta \in \Theta_1} L_N(\theta)} \right) = 2(\ell_N(\hat{\theta}_{H_1}) - \ell_N(\hat{\theta}_{H_0}))$$

with $L_N(\theta) = \prod_1^N f_\theta(Y_i)$

Asymptotic distribution of the LRT statistic for linear hypotheses defined by equalities when Θ is open

Consider the test defined by
H_0 : "$R\theta = 0$" against H_1 : "$R\theta \neq 0$"
where R is a full rank matrix of size $r x p$.

Then, assuming regularity conditions, under H_0:

$$LRT_N = -2\log\left(\frac{\sup_{\theta\in\Theta_0} L_N(\theta)}{\sup_{\theta\in\Theta_1} L_N(\theta)}\right) = 2(\ell_N(\hat{\theta}_{H_1}) - \ell_N(\hat{\theta}_{H_0})) \xrightarrow{\mathcal{L}} \chi^2(r)$$

Application to testing the effect of one covariate

Consider the test defined by
H_0 : "$\beta = 0$" against H_1 : "$\beta \neq 0$"

Then, assuming regularity conditions, under H_0:

$$LRT_N = -2\log\left(\frac{\sup_{\theta\in\Theta_0} L_N(\theta)}{\sup_{\theta\in\Theta_1} L_N(\theta)}\right) = 2(\ell_N(\hat{\theta}_{H_1}) - \ell_N(\hat{\theta}_{H_0})) \xrightarrow{\mathcal{L}} \chi^2(1)$$

$\Rightarrow$ require to evaluate numerically the likelihood

$\Rightarrow$ asymptotic distribution

Test for variance components in mixed effects model

Objective: test that r random effects among p have null variances.

$$\begin{cases} Y_{ij} = f(X_{ij}, \varphi_i) + \varepsilon_{ij}, & 1 \leq i \leq N,\ 1 \leq j \leq J \\ \varphi_i = U_i\beta + V_i b_i, & 1 \leq i \leq N \end{cases}$$

with $b_i \overset{iid}{\sim} \mathcal{N}_p(0; \Gamma)$ and $\varepsilon_{ij} \overset{iid}{\sim} q(.; \Sigma)$

Let $\Gamma = \left(\begin{array}{c|c} \Gamma_1 & \Gamma_{12} \\ \hline \Gamma_{12}^t & \Gamma_2 \end{array}\right)$ where $\Gamma_1 \in \mathcal{S}_{p-r}^+$ and $\Gamma_2 \in \mathcal{S}_r^+$

$\Theta_0 = \{\theta \in \mathbb{R}^q | \beta \in \mathbb{R}^p, \Gamma_1 \in \mathcal{S}_{p-r}^+, \Gamma_2 = 0, \Gamma_{12} = 0, \Sigma \in \mathcal{S}_J^+\}$

$\Theta_1 = \{\theta \in \mathbb{R}^q | \beta \in \mathbb{R}^p, \Gamma \in \mathcal{S}_p^+, \Sigma \in \mathcal{S}_J^+\}$

$\Longrightarrow$ test $H_0 : \theta \in \Theta_0$ against $H_1 : \theta \in \Theta_1$

Asymptotic distribution of the LRT statistic for testing that one variance equal zero in mixed effects model with one single random effect

[Self and Liang (1987) Annals of Statistics]

$$Y_{ij} = X_j \beta + b_i + \varepsilon_{ij} \ , \ 1 \leq i \leq N, \ 1 \leq j \leq J$$

with $b_i \overset{iid}{\sim} \mathcal{N}(0, \gamma^2)$, $\varepsilon_i \overset{iid}{\sim} \mathcal{N}_J(0, \sigma^2)$, ε_i independent of (b_i)

Consider the test defined by
H_0 : "$\gamma^2 = 0$" against H_1 : "$\gamma^2 \neq 0$"

Then, assuming regularity conditions, under H_0:

$$LRT_n = 2(\ell_n(\hat{\theta}_{H_1}) - \ell_n(\hat{\theta}_{H_0})) \xrightarrow{\mathcal{L}} \frac{1}{2}\delta_0 + \frac{1}{2}\chi^2(1)$$

Asymptotic distribution of the LRT statistic for linear hypotheses defined by inequalities when Θ is open

Consider the test defined by
H_0 : "$R\theta = 0$" against H_1 : "$R\theta \geq 0$"
where R is a full rank matrix

Denote by θ_0 the true value being in H_0
and I_0 the corresponding Fisher information matrix.

Then, assuming regularity conditions, under H_0:

$$LRT_n \xrightarrow{\mathcal{L}} \min_{R\theta=0} (Z-\theta)^t I_0 (Z-\theta) - \min_{R\theta \geq 0} (Z-\theta)^t I_0 (Z-\theta)$$

where $Z \sim \mathcal{N}(0, I_0^{-1})$
$\Longrightarrow$ reduce to test the mean of a multivariate normal distribution
$\Longrightarrow$ identify the limit distribution

Example of testing one single variance is zero

[Self & Liang, 1987]
Let $\theta = (\beta, \gamma^2, \Sigma)$ and $\Theta = \mathbb{R} \times \mathbb{R}^+ \times \mathcal{S}_J^+$.
Consider H_0 : "$\gamma^2 = 0$" against H_1 : "$\gamma^2 \geq 0$"
Let $Z \sim \mathcal{N}(0, I_0^{-1})$

$$\begin{aligned} D(Z) &= Z' I(\theta_0) Z - \inf_{\theta \geq 0} (Z - \theta)^t I_0 (Z - \theta) \\ &= \|Z\|_{I_0}^2 - \inf_{\theta \geq 0} \|Z - \theta\|_{I_0}^2 \\ &= \|\tilde{Z}\|^2 - \inf_{\theta \geq 0} \|\tilde{Z} - \theta\|^2 \\ &= \|\tilde{Z}\|^2 1_{\tilde{Z} > 0} \\ &= \frac{1}{2}\chi^2(0) + \frac{1}{2}\chi^2(1) \end{aligned}$$

where $\tilde{Z} \sim \mathcal{N}(0, 1)$

Sketch of proof

- Using Taylor series expansion

$$\begin{aligned}\ell_N(\theta) &= \ell_N(\theta_0) + \frac{1}{2}n^{-1}S_n(\theta_0)I^{-1}(\theta_0)S_n(\theta_0) \\ &- \frac{1}{2}[Z_n - n^{1/2}(\theta-\theta_0)]^t I(\theta_0)[Z_n - n^{1/2}(\theta-\theta_0)] \\ &+ O_P(1)||\theta-\theta_0||^3\end{aligned}$$

  where $Z_n = n^{-1/2}I(\theta_0)^{-1}S_n(\theta_0)$.

- Define $u = n^{1/2}(\theta-\theta_0)$ and rewrite the likelihood ratio test statistics as:

$$\begin{aligned}LRT_n &= -2[\sup_{\theta\in\Theta_0}\ell_n(\theta) - \sup_{\theta\in\Theta_1}\ell_n(\theta)] \\ &= \inf_{Ru=0}||Z_n-u||_{I(\theta_0)} - \inf_{Ru\geq 0}||Z_n-u||_{I(\theta_0)}.\end{aligned}$$

$\Longrightarrow$ establish the limit distribution

Asymptotic distribution of the LRT statistic for general hypotheses when Θ is open

[Self and Liang (1987) Annals of statistics]

Consider the test defined by
H_0 : "$\theta \in \Theta_0$" against H_1 : "$\theta \in \Theta_1$"
Then, assuming regularity conditions, under H_0:

$$LRT_n = 2(\ell_n(\hat{\theta}_{H_1}) - \ell_n(\hat{\theta}_{H_0})) \xrightarrow{\mathcal{L}} D_T(Z),$$

where $Z \sim \mathcal{N}(0, I_0^{-1})$ and

$$D_T(z) = \|z - T(\Theta_0, \theta_0)\|_{I_0}^2 - \|z - T(\Theta_1, \theta_0)\|_{I_0}^2.$$

where $T(\Theta, \theta)$ is the tangent cone of Θ at θ

$\Longrightarrow$ using tangent cones to approximate Θ_0 and Θ_1

Limits of the existing results

Example of testing one variance equals to zero considering two correlated random effects:

Let $\theta = (\beta, \Gamma, \Sigma)$ with $\Gamma = \begin{pmatrix} \gamma_1^2 & \gamma_{12} \\ \gamma_{12} & \gamma_2^2 \end{pmatrix}$ and $\Theta = \mathbb{R}^2 \times \mathcal{S}_2^+ \times \mathcal{S}_J^+$.

Consider $H_0 : \theta \in \Theta_0$ against $H_1 : \theta \in \Theta_1$ with

$$\Theta_0 = \{\theta, \beta \in \mathbb{R}^2, \gamma_1^2 = \gamma_{12} = 0, \gamma_2^2 \geq 0, \Sigma \in \mathcal{S}_J^+\}$$
$$\Theta_1 = \{\theta, \beta \in \mathbb{R}^2, \gamma_1^2 \geq 0, \gamma_1^2\gamma_2^2 - \gamma_{12}^2 \geq 0, \gamma_2^2 \geq 0, \Sigma \in \mathcal{S}_J^+\}$$

$\Longrightarrow$ Θ is not open

$\Longrightarrow$ approxmation with cones for Θ_1 and Θ_0

$\Longrightarrow$ identify the limit distribution

Identifying the asymptotic distribution of the LRT statistics for testing variance components in nonlinear mixed effects model

[Baey, Cournède, K. (2019) CSDA]

Consider the test defined by
$H_0 : \theta \in \Theta_0$ against $H_1 : \theta \in \Theta_1$ where
$\Theta_0 = \{\theta \in \mathbb{R}^q | \beta \in \mathbb{R}^p, \Gamma_1 \in \mathcal{S}^+_{p-r}, \Gamma_2 = 0, \Gamma_{12} = 0, \Sigma \in \mathcal{S}^+_J\}$
$\Theta_1 = \{\theta \in \mathbb{R}^q | \beta \in \mathbb{R}^p, \Gamma \in \mathcal{S}^+_p, \Sigma \in \mathcal{S}^+_J\}$
Then, assuming regularity assumptions, under H_0:

$$LRT_n \xrightarrow{\mathcal{L}} \bar{\chi}^2(I_0^{-1}, T(\Theta_0, \theta_0)^{\perp} \cap T(\Theta_1, \theta_0)),$$

where $T(\Theta, \theta)$ is the tangent cone of Θ at θ and $\bar{\chi}^2(V, \mathcal{C})$ has a χ-bar square distribution (mixture of chi square distributions) with $\mathcal{C}$ a closed convex cone and V a positive definite matrix

The Chi-bar Square distribution

Let $\mathcal{C}$ be a closed convex cone of $\mathbb{R}^q$ and
V a positive definite matrix of size qxq.
Let $Z \sim \mathcal{N}(0, V)$
Then $\bar{\chi}^2(V, \mathcal{C}) = Z'V^{-1}Z - \inf_{\theta \in \mathcal{C}}(Z - \theta)'V^{-1}(Z - \theta)$
has a χ-bar square distribution and

$$\forall t \geq 0 \; P(\bar{\chi}^2(V, \mathcal{C}) \leq t) = \sum_{i=0}^{q} w_i(p, V, \mathcal{C}) P(\chi_i^2 \leq t)$$

where the weights $w_i(q, V, \mathcal{C})$ are some non-negative numbers summing up to one

Example of testing one variance equals to zero considering two independent random effects

Let $\theta = (\beta, \Gamma, \Sigma)$ with $\Gamma = \begin{pmatrix} \gamma_1^2 & 0 \\ 0 & \gamma_2^2 \end{pmatrix}$ and
$\Theta = \mathbb{R}^2 \times \mathbb{R}^+ \times \mathbb{R}^{+*} \times \mathcal{S}_J^+$.
Consider $H_0 : \gamma_1^2 = 0$ against $H_1 : \gamma_1^2 \geq 0$
Let $Z \sim \mathcal{N}(0, I_0^{-1})$

$$\begin{aligned} D(Z) &= \inf_{\theta_1 = 0} (Z - \theta)' I_0 (Z - \theta) - \inf_{\theta_1 \geq 0} (Z - \theta)^t I_0 (Z - \theta) \\ &= \tilde{Z}_1{}^2 - \inf_{\theta_1 \geq 0} (\tilde{Z}_1 - \theta_1)^2 \\ &= \tilde{Z}_1{}^2 1_{\tilde{Z}_1 > 0} \\ &= \frac{1}{2}\chi^2(0) + \frac{1}{2}\chi^2(1) \end{aligned}$$

where $\tilde{Z} \sim \mathcal{N}(0, 1)$

Evaluation of the empirical level of the test for one effect when two effects are non correlated in the linear model

$$Y_{ij} = \varphi_{1i} + \varphi_{2i} t_{ij} + \varepsilon_{ij} \ ,$$

Let $\Gamma = \begin{pmatrix} \gamma_1^2 & 0 \\ 0 & \gamma_2^2 \end{pmatrix}$

Consider $H_0 : \gamma_1 = 0$ against $H_1 : \gamma_1 \geq 0$

Table: Percentages of rejection for the LRT procedure for $n = 500$ for the nominal level of the test α on 300 repetitions.

α	$\hat{\alpha}_{0.5\chi_0^2+0.5\chi_1^2}$
0.01	0.010
0.05	0.046
0.10	0.093

Example of testing one variance equals to zero considering two correlated random effects

Let $\theta = (\beta, \Gamma, \Sigma)$ with $\Gamma = \begin{pmatrix} \gamma_1^2 & \gamma_{12} \\ \gamma_{12} & \gamma_2^2 \end{pmatrix}$ and $\Theta = \mathbb{R}^2 \times \mathcal{S}_p^+ \times \mathcal{S}_J^+$.
Consider $H_0 : \theta \in \Theta_0$ against $H_1 : \theta \in \Theta_1$

$\Theta_0 = \{\theta, \beta \in \mathbb{R}^2, \gamma_1^2 = \gamma_{12} = 0, \gamma_2^2 \geq 0, \Sigma \in \mathcal{S}_J^+\}$
$\Theta_1 = \{\theta, \beta \in \mathbb{R}^2, \gamma_1^2 \geq 0, \gamma_1^2\gamma_2^2 - \gamma_{12}^2 \geq 0, \gamma_2^2 \geq 0, \Sigma \in \mathcal{S}_J^+\}$

$$LRT_n \xrightarrow{d} \quad \frac{1}{2}\chi^2(1) + \frac{1}{2}\chi^2(2)$$

Evaluation of the empirical level of the test for one effect when two effects are correlated in the linear model

$$Y_{ij} = \varphi_{1i} + \varphi_{2i} t_{ij} + \varepsilon_{ij} \ ,$$

Let $\Gamma = \begin{pmatrix} \gamma_1^2 & \gamma_{12} \\ \gamma_{12} & \gamma_2^2 \end{pmatrix}$
Consider $H_0 : \theta \in \Theta_0$ against $H_1 : \theta \in \Theta_1$

Table: Percentages of rejection for the LRT procedure for $n = 500$ for the nominal level of the test α on 300 repetitions.

α	$\hat{\alpha}_{0.5\chi_1^2+0.5\chi_2^2}$	$\hat{\alpha}_{0.5\chi_0^2+0.5\chi_1^2}$
0.01	0.016	0.049
0.05	0.055	0.174
0.10	0.103	0.311

Perspectives

- need for efficient numerical evaluation of likelihood
- need for efficient numerical evaluation of Fisher information matrix
- limits of non asymptotic test procedure ...
  $\Rightarrow$ Likelihood ratio tests in linear mixed models with one variance component, Crainiceanu and Ruppert, JRSS B (2004)

Comments on the distribution of random effect

- centered distribution
- usual choice Gaussian distribution
- possible to choose other ones: Student, mixture ...
- test for the adequation of Gaussian distribution for random effects
  $\Rightarrow$ Diagnosing misspecification of the random-effects distribution in mixed models Drikvandi et al. Biometrics (2016)
- Nonparametric estimation of random effects densities in linear mixed-effects model. Comte F, Samson A, Journal of Nonparametric Statistics, (2012)

Summary of the day

$$\begin{cases} Y_{ij} & = & f(X_{ij}, \varphi_i) + \varepsilon_{ij}, & 1 \leq i \leq N,\ 1 \leq j \leq J \\ \varphi_i & = & U_i \beta + V_i b_i, & 1 \leq i \leq N \end{cases}$$

with $b_i \overset{iid}{\sim} q(.;\Gamma)$ and $\varepsilon_{ij} \overset{iid}{\sim} q(.;\Sigma)$ with $\theta = (\beta, \Gamma, \Sigma)$

- Testing procedure for fixed effects β via LRT
- Testing procedure for variance components Γ via LRT
- alternatives: Wald test, score test

Model choice criteria

Consider the mixed effects model

$$\begin{cases} Y_{ij} &= f(X_i, \varphi_i) + \varepsilon_{ij}, \quad 1 \leq i \leq N,\ 1 \leq j \leq J \\ \varphi_i &= U_i \beta + b_i, \quad\quad 1 \leq i \leq N \end{cases}$$

with $b_i \overset{iid}{\sim} \mathcal{N}(0, \Gamma)$ and $\varepsilon_{ij} \overset{iid}{\sim} q$ and $\theta = (\beta, \Gamma)$

Recall the
Bayesian information criterion defined as:

$$BIC = -2 \log L_{marg}(\widehat{\theta}; Y_1^N) + dim(\theta) \log(n_{obs})$$

$\Rightarrow$ what is the "real" sample size in mixed effects model? NJ? N?
From a practical point of view, the $\log(NJ)$ penalty is implemented in the R package nlme and in the SPSS procedure MIXED while the $\log(N)$ penalty is used in Monolix, saemix or in the SASproc NLMIXED.

Model choice criteria

[A note on BIC in mixed-effects models, Delattre, Lavielle, Poursat, EJS 2014]

Consider the following mixed effects model:

$$\left\{ \begin{array}{ccc} \varphi_i & \sim & q(.|U_i, \theta) \\ Y_i|\varphi_i; X_i & \sim & q(.|\varphi_i; X_i) \end{array} \right.$$

where $\varphi_i = U_i \beta + b_i$ with U_i block diagonal, $b_i \sim \mathcal{N}(0, \Gamma)$ and Γ is potentially degenerated.

$$\Gamma = \left(\begin{array}{cc} 0 & 0 \\ 0 & \Gamma_R \end{array} \right)$$

Denote the parameter $\theta = (\theta_F, \theta_R)$ where $\theta_F = \beta_F$ and $\theta_R = (\beta_R, \Gamma_R)$.

Consider the hybrid Bayesian information criterion defined as:

$$BIC_{hyb} = -2 \log L_{marg}(\widehat{\theta}; Y_1^N) + \dim(\theta_R) \log(N) + \dim(\theta_F) \log(NJ)$$

$\Rightarrow$ intensive simulation study to highlight the good statistical properties of this criterion

Model choice criteria

[A note on BIC in mixed-effects models, Delattre, Lavielle, Poursat, EJS 2014]
Consider the hybrid Bayesian information criterion defined as:

$$BIC_{hyb} = -2 \log L_{marg}(\widehat{\theta}; Y_1^N) + \dim(\theta_R) \log(N) + \dim(\theta_F) \log(NJ)$$

- In a pure fixed-effects model,
  $\Rightarrow \theta = \theta_F$
  $\Rightarrow$ penalty $\dim(\theta) \log(NJ)$
- if all the individual parameters are random,
  $\Rightarrow \theta = \theta_R$
  $\Rightarrow$ penalty $\dim(\theta) \log(N)$
- the criterion proposed appears to be an hybrid BIC version that automatically adapts to the random-effects structure of a mixed model

Variable selection in linear mixed effects model

Consider the linear mixed effect model:

$$Y_i = X_i\beta + Z_i b_i + \varepsilon_i \ , \ 1 \leq i \leq N,$$

- Y_i is the observation vector for individual i of size J
- X_i and Z_i are matrices of known covariates of individual i
- β is the vector of fixed effects of size p
- $b_i \overset{iid}{\sim} \mathcal{N}_q(0, \Gamma)$
- ε_i is a random error vector, with $\varepsilon_i \overset{iid}{\sim} \mathcal{N}_J(0, \Sigma)$

$\Rightarrow$ In case where $NJ << p$ not possible to use classical maximum likelihood approach

$\Rightarrow$ penalize the estimation criterion

Variable selection in linear mixed effects model

Consider the linear mixed effect model:

$$Y_i = X_i\beta + Z_i b_i + \varepsilon_i \ , \ 1 \leq i \leq N,$$

- Y_i is the observation vector for individual i of size J
- X_i and Z_i are matrices of known covariates of individual i
- β is the vector of fixed effects of size p
- $b_i \overset{iid}{\sim} \mathcal{N}_q(0, \Gamma(\gamma))$
- $\varepsilon_i \overset{iid}{\sim} \mathcal{N}_J(0, \sigma^2 I)$

$\Rightarrow Y_i \sim \mathcal{N}(X_i\beta, Z_i'\Gamma(\gamma)Z_i + \sigma^2 I)$

Variable selection in linear mixed effects model

[Estimation for High-Dimensional Linear Mixed-Effects Models Using L_1 Penalization, Schelldorfer et al., SJS (2011)]

Consider the linear mixed effect model:

$$Y_i = X_i\beta + Z_i b_i + \varepsilon_i \ , \ 1 \leq i \leq N,$$

where β is of size p, $b_i \overset{iid}{\sim} \mathcal{N}_q(0, \Gamma(\gamma))$ and $\varepsilon_i \overset{iid}{\sim} \mathcal{N}_J(0, \sigma^2 I)$

$\Rightarrow Y_i \sim \mathcal{N}(X_i\beta, Z_i'\Gamma(\gamma)Z_i + \sigma^2 I)$

Consider the setting where $NJ << p$, $dim(\gamma) << p$,
q might as high as p

Consider the following objective function

$$C(\beta, \gamma, \sigma^2) = \frac{1}{2}\log|V| + \frac{1}{2}(Y - X\beta)'V^{-1}(Y - X\beta) + \lambda\sum_{k=1}^{p}|\beta_k|$$

with $V = diag(V_1, \ldots, V_N)$ and $V_i = Z_i\Gamma(\gamma)Z_i + \sigma^2 I$
and define the penalized estimator $(\hat{\beta}, \hat{\gamma}, \hat{\sigma^2}) = \arg\max Q(\beta, \gamma, \sigma^2)$

Variable selection in linear mixed effects model

$$Y_i = X_i\beta + Z_i b_i + \varepsilon_i \ , \ 1 \leq i \leq N,$$

where β is of size p, $b_i \overset{iid}{\sim} \mathcal{N}_q(0, \Gamma(\gamma))$ and $\varepsilon_i \overset{iid}{\sim} \mathcal{N}_J(0, \sigma^2 I)$

$$C(\beta, \gamma, \sigma^2) = \frac{1}{2}\log|V| + \frac{1}{2}(Y - X\beta)' V^{-1}(Y - X\beta) + \lambda \sum_{k=1}^{p} |\beta_k|$$

$(\hat{\beta}, \hat{\gamma}, \hat{\sigma}^2) = \arg\max Q(\beta, \gamma, \sigma^2)$

- theoretical properties of consistency for the penalized estimates and for the support of β
- implemented in R package lmmlasso and glmmlasso

$\Rightarrow$ other approach [An iterative algorithm for joint covariate and random effect selection in nonlinear mixed effects models, Delattre et al. (2019)]

$\Rightarrow$ further work needed for variable selection in nonlinear mixed effects models

Short global summary

- fixed and random effects
- maximum likelihood estimator with good properties
- convergent stochastic algorithm to evaluate its value
- testing procedures for fixed effects and variance components
- model choice criteria
- variable selection in linear mixed effects model

Somes extensions of mixed models

- Modeling the observation level through a function defined by an Ordinary Differential Equation [Donnet S., Samson A., JSPI (2007)]
- Parametric inference for mixed models defined by stochastic differential equations, [Donnet S., Samson A. (2008) ESAIM PS (2008)]
- Parametric estimation of complex mixed models based on meta-model approach, [Barbillon P, Barthelemy C, Samson A, Statistics and Computing, (2017)]
- ...

Others models with random effects

- Maximum likelihood estimation in frailty models [K., El Nouty Stat Compu (2013)]
- Maximum likelihood estimation for stochastic differential equations with random effects [Delattre M., Genon-Catalot V., Samson A., SJS (2013)]
  $\Rightarrow$ Mixedsde: a R package to fit mixed stochastic differential equations [Dion C., Hermann S., Samson A. (2018)]
- ...

Books bibliography

- Mixed effects models in S and S-PLUS, J.C. Pinheiro D.M. Bates (2000)
- Linear mixed models for longitudinal data, G. Verbeke and G. Molenberghs (2000)
- Mixed effects models for the population approach, M. Lavielle (2014)
- Nonlinear Models for Repeated Measurement Data, M. Davidian and D.M. Giltinian (1995)