Facebook

Mixed effects models: Inference and Maximum Likelihood

CMAP, Ecole Polytechnique

September 27th, 2017

Plan

1. Problem
2. Parameter Estimation: SAEM algorithm
3. Inference: MCMC
4. Applications

Problem

Notations

- $Z = (Z_1, \ldots, Z_N) \in \mathcal{Z} \subset \mathbb{R}^d$, $Y = (Y_1, \ldots, Y_N) \in \mathcal{Y} \subset \mathbb{R}^d$ Random Variables
- Measurable spaces with the associated incomplete and complete densities $p(y_i, \theta)$ and $p(y_i, z_i, \theta)$ where:
  - y_i is observed
  - z_i is the individual parameter
  - θ is the population parameter
- Continuous, non linear and mixed effects models:

$$y_i = f(z_i) + \epsilon_i \tag{1}$$

  Where:
  - The structural model $f : \Theta \rightarrow \mathbb{R}$ is non linear and twice differentiable
  - $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ and $\sigma \in \mathbb{R}$
  - $z_i \sim \mathcal{N}(z_{pop}, w^2)$ such that $z_i = z_{pop} + \eta_i$ with $\eta_i \sim \mathcal{N}(0, \omega^2)$
  - $y_i | z_i \sim \mathcal{N}(f(z_i), \sigma^2)$
- In our case $\theta = (z_{pop}, \omega, \sigma)$

Maximum likelihood

The goal is to compute the maximum likelihood estimate

$$\theta^{ML} = \arg\max_{\theta \in \Theta} p(y, \theta) \tag{2}$$

The EM algorithm (Dempster, Laird and Rubin) is an iterative algorithm that computes this quantity by maximizing an auxialiary quandity $Q : \Theta^2 \to \mathbb{R}$ at a given parameter estimate θ':

$$Q(\theta, \theta) = \mathbb{E}_{p(z|y,\theta')} [\log p(y, z, \theta)] \tag{3}$$

Unfortunately, in the framework of nonlinear mixed effects models, there is no explicit expression for the E-step since the relationship between observations y and individual parameters z is nonlinear so the expectation cannot be computed in closed form. Thus, we use a stochastic version of the EM.

Parameter Estimation: SAEM algorithm

SAEM

At a given θ^{k-1}:

1. $z_i^k \sim p(z_i|y_i, \theta^{k-1})$
2. $Q^k(\theta) = Q^{k-1}(\theta) + \gamma_k(\log p(y_i, z_i^k, \theta) - Q^{k-1}(\theta))$
3. $\theta^k = \arg\max_{\theta \in \Theta} Q^k(\theta)$

My goal being to accelerate this algorithm, three approaches have been taken yet:

- Incremental/mini-batch strategies (see Neal & Hinton on the EM or F.Bach on the SGD)
- Faster MCMC dynamics (SG-MCMC based on an ITO diffusion, see for instance MALA based on a Langevin diffusion)
- Efficient proposal for a Metropolis Hastings algorithm based on a Laplace Approximation of the incomplete log likelihood $\log p(y, \theta)$

Inference: MCMC

Inference: MCMC

- First Order Conditional Estimation Method
  1. Laplace Approximation of the incomplete likelihood around the MAP ($z_{i,MAP} = \arg\max_z p(z_i|y_i,\theta)$)

$$p(y_i,\theta) = \int p(y_i,z_i,\theta)dz_i = \int e^{\log p(y_i,z_i,\theta)}dz_i \tag{4}$$

$$\approx e^{\log p(y_i,z_{i,MAP},\theta)}\sqrt{\frac{(2\pi)^p}{|-\nabla^2 \log p(y_i,z_{i,MAP},\theta)|}} \tag{5}$$

  2. Approximation of the posterior

$$-2\log p(y_i,\theta) = \underbrace{-p\log(2\pi) + \log(|-\nabla^2 \log p(y_i,z_{i,MAP},\theta)|)}_{\approx -2\log p(z_{i,MAP}|y_i,\theta)} \tag{6}$$

$$-2\log p(y_i,z_{i,MAP},\theta) \tag{7}$$

- Equivalent to linearizing the structural model f around the MAP:

$$y_i \approx z_{i,MAP} + (z_i - z_{i,MAP})\nabla_{z_i} f(z_{i,MAP}) + \epsilon_i \implies z_i|y_i \sim \mathcal{N}(\mu,\Gamma) \tag{8}$$

Comparison

- New independent proposal. At a given iteration k:

$$z_i^k \sim \mathcal{N}(z_{i,MAP}, |-\nabla^2 \log p(y_i, z_{i,MAP}, \theta)|) \tag{9}$$

- Comparison of the MCMC properties with:
  - Random Walk Metropolis (proposal $\mathcal{N}(z_i^k, \Omega)$)
  - Metropolis Adjusted Langevin Algorithm (proposal based on the Langevin dynamics $\mathcal{N}(z_i^k - \gamma_k \nabla \log \pi(z_i^k), \sqrt{2\gamma_k})$)

m	RWM	MALA	Laplace
JSS	16.128	8.872	16.128
AUTOCORR	0.641	-0.466	0.641
VAR	0.45	0.421	0.45

Applications

PK-PD datasets

- The data comes from 37 winter wheat experiments carried out between 1990 and 1996 on commercial farms near Paris, France. Each experiment was from a different site.
- Each experiment consisted of five to eight different nitrogen fertiliser rates, for a total of 224 nitrogen treatments. Nitrogen fertilizer was applied in two applications during the growing season. For each nitrogen treatment, grain yield (adjusted to 150 g.kg-1 grain moisture content) was measured.
- In this problem the sites are denoted by the index "i" and are the individuals in the dataset, the predictor is the dosage, the response is the grain yield and the covariate is the soil nitrogen

PK-PD datasets

We use a Linear Plateau model here and the the structural model is:

$$f(z_i) = \begin{cases} (Y_{max})_i + B_i * (t_i - (X_{max})_i), & \text{if } t \geq (X_{max})_i \\ (Y_{max})_i, & \text{otherwise} \end{cases} \tag{10}$$

Where $z_i = ((X_{max})_i, (Y_{max})_i, B_i)$.

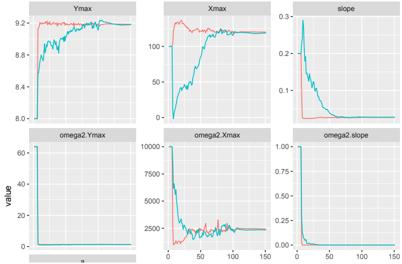

Figure: Estimate of X_{Max} for 100 replicates

Thank you