

Converting a population parameter into a random variable with small variance

Belhal Karimi

1 Convert the parameter into random variables with variance that tends to infinity

The initial model $p(y_i, \psi_i; \beta, \theta)$ becomes $p(y_i, \psi_i, \beta; \theta)$.

The new model can now be written as:

$$p(y_i|\psi_i, \beta; \theta)p(\psi_i; \theta)\pi(\beta) \quad (1)$$

where $\pi(\beta)$ is a prior on our new random variable β .

The challenge is to see if we keep the same convergence properties in the SAEM algorithm when we operate that change.

Let's take a simple example: a logistic regression:

$$y_{ij} = f(t_{ij}, \psi_i) + \epsilon_{ij} = \frac{1}{\alpha_i + e^{-\beta_i t_{ij}}} + \epsilon_{ij} \quad (2)$$

For all individuals $\beta_i = \beta$ and we have j observations per individual i . For the moment we will forget about the number of observations j . With $\psi_i = (\alpha_i, \beta)$ Hypothesis:

$$y_i \sim N\left(\frac{1}{\alpha_i + e^{-\beta t}}, a_i^2\right) \quad (3)$$

1.1 Fixed population parameter

The model is written as

$$p(y, \psi; \theta) = p(y|\psi; \theta)p(\psi; \theta) \quad (4)$$

With $\theta = (\theta_\alpha, \beta)$ In that case, we can not consider that the model belongs to the exponential family. Indeed we can not write $(y_i - \frac{1}{\eta_i + e^{-\psi_i t}})$ as a scalar product between the sufficient statistic and a vector of parameter.

The solution consists in putting variability on the parameter β so that the sufficient statistics becomes a function of y_i and ψ_i . In that case we belong to the exponential family.

Remark: Instead of putting variability, we can also put a prior on this parameter. We'll see in the sequel what differences it makes in practice.

In that part, we consider that the parameters are fixed $\theta = (\theta_\alpha, \beta)$. Running the SAEM gives us the maximum likelihood estimate $\theta^{ML} = (\theta_\alpha^{ML}, \beta^{ML})$

The issue now is to see if the estimates obtained while putting variability, or a prior, leads to the same maximum likelihood estimate when we tends the variability to zero.

1.2 Parameter with variability

In the second algorithm, β is a random variable such as $\beta \sim N(\beta^*, \sigma_\beta^2)$
The model becomes:

$$p(y, \psi, \beta; \theta) = p(y|\psi, \beta; \theta)p(\psi; \theta)\pi(\beta) \quad (5)$$

$$(6)$$

with $\theta = (\theta_\alpha, \beta^*, \sigma_\beta^2)$

If we then look at what is happening in the SAEM algorithm, nothing changes in the simulation phase where the latent data are drawn from same the conditional $p(\psi_i|y_i, \theta_\alpha, \beta, \theta_y)$.

Yet, the new algorithm requires to draw a new β at each iteration from the conditional $p(\beta|y_i, \eta, \theta_\beta, \theta_y)$.

$$(\eta_i)^{k+1} \sim p(\eta_i|y_i, \beta^k, \theta_\alpha, \theta_y) \quad (7)$$

$$\beta^{k+1} \sim p(\beta|y_i, \eta^{k+1}, \theta_\beta, \theta_y) = Cp(y_i|\beta, \eta^{k+1}, \theta_y)p(\beta, \theta_\beta) \quad (8)$$

This consists in a Gibbs sampling where we alternatively simulate from two posteriors using MCMC transistion kernels.

Question: Convergence of $\hat{\theta}(\sigma_\beta) = (\hat{\theta}_\alpha(\sigma_\beta), \hat{\beta}(\sigma_\beta))$ when $\sigma_\beta \mapsto 0$

We know that when the complete model belongs to the exponential family, the saem converges to a subset of the EM solution set. In this case we are linking our non-exponential model to an exponential one by putting variability on the fixed parameter.

Yet, here the algorithm implies a gibbs sampling (to sample η and β alternatively) and of course two MCMC. Let's first consider the case when we can directly sample from those two conditional distributions (no mcmc). The challenge is to see if the parameters estimates at convergence of this algo are the MLE.

So here we use only one MCMC method, known as Gibbs sampling, to sample from the conditionals of interest. Keep in mind that each conditional can be sampled via an MCMC method as well (MH for instance). We'll consider only one iteration of the Gibbs sampling to begin with.

In the current implementation of the SAEM, the sampling is done on η instead of ψ but we can shift from one to another by adding the fixed effect to the random effects η_i :

$$\psi_i = \eta_i + \beta \quad (9)$$

After the gibbs sampling the update of the individual parameter component ψ_i looks like:

$$\psi_i^k = \eta_i^k + \beta^k \quad (10)$$

The stochastic approximation is:

$$s_k = s_{k-1} + \gamma_k \left(\sum_{i=1}^N \tilde{S}(y_i, \eta_i^k, \beta^k) - s_{k-1} \right) \quad (11)$$

And finally the M-step:

$$\theta_k = \hat{\theta}(s_k) \quad (12)$$

We have to keep in mind that this θ_k depends on σ_β .

The problem is not around the Gibbs sampling. Indeed when the random effects are already multidimensional, gibbs sampling is already used.

Let's name all the models:

1. M1 is the model where β is not random
2. M2 is the model where β is a random variable following a gaussian $\mathcal{N}(\beta^*, \sigma_\beta)$
3. M3 is the model where we put a prior on β

In M2, we choose a variance for β that does not change throughout the iterations. The SAEM converges to the maximum likelihood of model M2: $\theta^{MLM2}(\sigma_\beta) = (\theta_\alpha^{MLM2}(\sigma_\beta), \beta_*^{MLM2}(\sigma_\beta))$. Two interesting questions arise:

1. What are the properties of this estimator and where does it converge when σ_β tends to 0? Does it converge to θ^{MLM1} ?
2. If the variance on β , σ_β is converging to zero throughout the iterations, will the estimator converge to θ^{MLM1} ?

In the latter case, the whole point is to see if we could converge to the right MLE while defining the population parameter as a random variable (wheres we want to solve the problem when the parameter is fixed) by making its variance tend to zero throughout the iterations.

1.2.1 Properties of the estimator when the population parameter is a random variable

First, let's focus on the properties of our estimator

$$\theta^{MLM2}(\sigma_\beta) = (\theta_\alpha^{MLM2}(\sigma_\beta), \beta_*^{MLM2}(\sigma_\beta)) \quad (13)$$

In this case the most important thing is to evaluate the expectation (bias), the variance, the efficiency and in the case of large samples, the consistency. Indeed the variability put on the population parameter induces some variability since at each iteration N simulations of the random variable β are sampled from their conditional distributions. The goal is to evaluate this variance and the

expectation of the estimator when the SAEM algorithm has converged. Those two quantities depends on the value of the variance of β . Checking that the parameter estimate of the random effects converges to the solution of the model 1 when the variance tends to zero will be crucial.

For exploratory purposes, we are going to study the case of the linear gaussian mixed effects model. In this case, we can compute all the posterior distributions and as a result explicitly write all of our estimators. First of all, the maximum likelihood estimators are computed with one of the individual parameters component being fixed. Then, variability is being introduced: this same component is considered as a random variable, gaussian with a non zero mean and a variance.

1.3 Putting a prior on the parameter

In that case we are only putting a prior on $\beta \sim \mathcal{N}(\beta^*, \sigma_\beta^2)$ where now the parameters of that prior does not change from an iteration to another.

So the vector of parameter is now $\theta = (\theta_\alpha, \beta)$. The SAEM gives us a estimate of the maximum likelihood that depends obviously on the value of the variance of the prior $\hat{\theta}(\sigma_\beta) = (\hat{\theta}_\alpha, \hat{\beta})$ with $\hat{\beta}^{MAP}(\sigma_\beta) = \operatorname{argmax}(p(\beta|y, \hat{\theta}_\alpha))$.

Question: convergence of $\hat{\beta}^{MAP}$ when $\sigma_\beta \mapsto 0$

In this case the implementation is kind of the same and implies a Gibbs sampling where we alternatively sample from two mcmc (on the individual parameters and the fixed param we put a prior on). The difference is that the parameters of that prior are not re-evaluated.