STAT3016/4116/7016
Introduction to Bayesian Data Analysis

RSFAS, College of Business and Economics, ANU

Linear and generalized linear mixed effects models

A hierarchical regression model

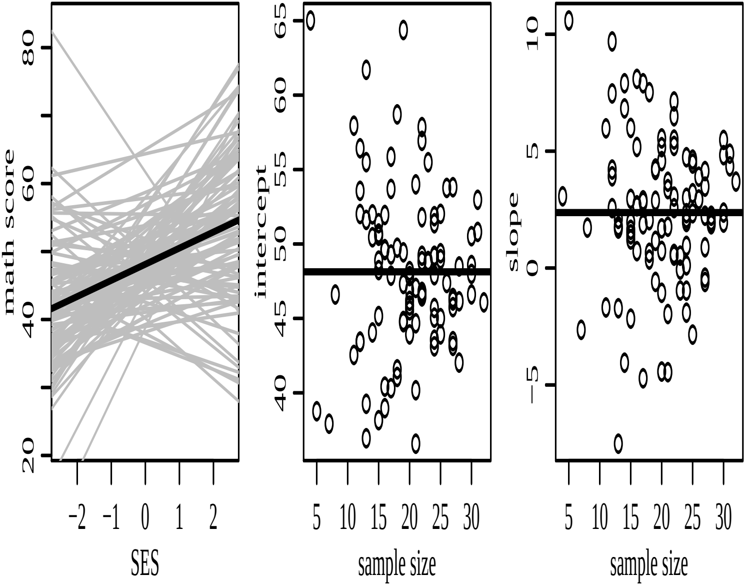
Example: Recall the school data from Chapter 8. Let's now examine the relationship between math scores and another variable, socioeconomic status (SES). Perhaps the relationship between mathscore and SES varies from school to school.

- Ordinary least squares model?
- Hierarchical Model? hierarchical model: y_{ij} ~ iid normal (theta_j, sigma^2) within group model
  theta_j ~ iid normal (mu, tau^2) between group model
- Bayesian Hierarchical Model?

now for hierarchical linear regression model, use betaTx
to replace y_{ij} and beta to replace theta_j

Least squares analysis of mathscore data

when the sample size is small,
the variability is large



A hierarchical regression model

Priors

$$\boldsymbol{\theta} \sim MVN(\boldsymbol{\mu}_0, \Lambda_0)$$

$$\boldsymbol{\Sigma} \sim inverse - Wishart(\eta_0, S_0^{-1})$$

$$\sigma^2 \sim inverse - gamma(\nu_0/2, \nu_0\sigma_0^2/2)$$

Sampling Model

$$\mathbf{Y}_j \sim MVN(\mathbf{X}_j\beta_j, \sigma^2\mathrm{I})$$

$$\boldsymbol{\beta}_j \overset{\text{iid}}{\sim} MVN(\theta, \Sigma)$$

(j=1,...,m; m groups)

A hierarchical regression model

Full conditional distributions of $\boldsymbol{\beta}_j$

$$Var[\boldsymbol{\beta}_j|\mathbf{y}_j, \mathbf{X}_j, \sigma^2, \boldsymbol{\theta}, \Sigma] = (\Sigma^{-1} + \mathbf{X}_j^T\mathbf{X}_j/\sigma^2)^{-1}$$

$$E[\boldsymbol{\beta}_j|\mathbf{y}_j, \mathbf{X}_j, \sigma^2, \boldsymbol{\theta}, \Sigma] = (\Sigma^{-1} + \mathbf{X}_j^T\mathbf{X}_j/\sigma^2)^{-1}(\Sigma^{-1}\boldsymbol{\theta} + \mathbf{X}_j^T\mathbf{y}_j/\sigma^2)$$

Full conditional distributions of $\boldsymbol{\theta}$

$$Var[\boldsymbol{\theta}|\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_m] = (\Lambda_0^{-1} + m\Sigma^{-1})^{-1}$$

$$E[\boldsymbol{\theta}|\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_m] = (\Lambda_0^{-1} + m\Sigma^{-1})^{-1}(\Lambda_0^{-1}\boldsymbol{\mu}_0 + m\Sigma^{-1}\bar{\boldsymbol{\beta}})$$

$(\bar{\boldsymbol{\beta}} = \frac{1}{m}\sum_j \boldsymbol{\beta}_j)$

A hierarchical regression model

Full conditional distributions of $\mathbf{\Sigma}$

$$\Sigma|\boldsymbol{\theta}, \boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_m \sim Inverse - Wishart(\eta_0 + m, [\mathbf{S}_0 + \mathbf{S}_\theta]^{-1})$$

where $\mathbf{S}_\theta = \sum_{j=1}^m (\boldsymbol{\beta}_j - \boldsymbol{\theta})(\boldsymbol{\beta}_j - \boldsymbol{\theta})^T$

Full conditional distributions of σ^2

$$\sigma^2 \sim Inverse - Gamma([\nu_0 + \sum_j n_j]/2, [\nu_0\sigma_0^2 + SSR]/2)$$

where $SSR = \sum_{j=1}^m \sum_{i=1}^{n_j} (y_{i,j} - \boldsymbol{\beta}_j^T \mathbf{x}_{i,j})^2$.

A hierarchical regression model - example

Mathscore data example

$\boldsymbol{\mu}_0 = \frac{1}{m}\sum_{j=1}^{m}\hat{\boldsymbol{\beta}}_{j,OLS}$;

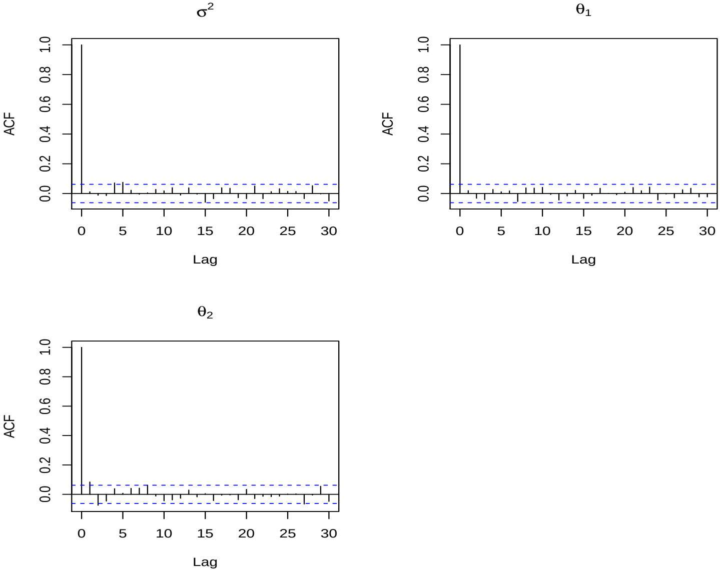
$\Lambda_0 = \mathbf{S}_0 = Cov(\hat{\boldsymbol{\beta}}_{j,OLS})$;

$\eta_0 = p + 2 = 4$;

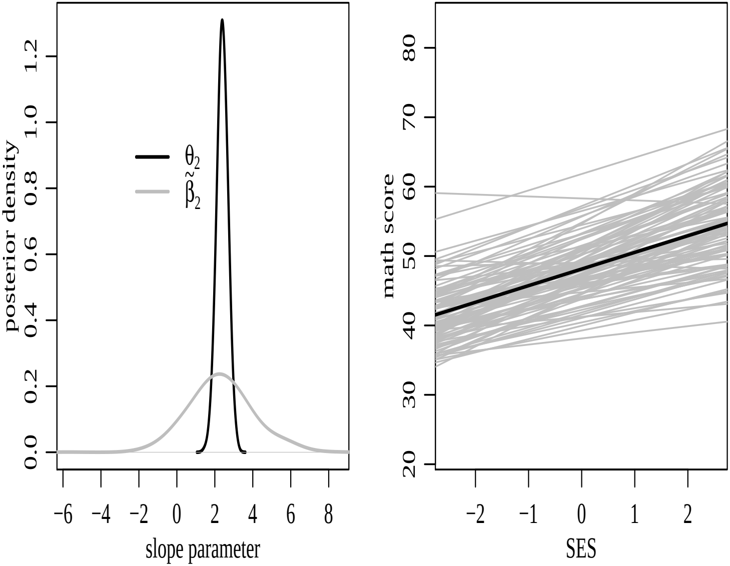
$\sigma_0^2 = \frac{1}{m}\sum_{j=1}^{m}\hat{\sigma}_{j,OLS}^2$; $\nu_0 = 1$

S=10000, thin=10

A hierarchical regression model - example



A hierarchical regression model - example



Generalised Linear Mixed effects models

$$\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_m \overset{\text{iid}}{\sim} \text{MVN}(\boldsymbol{\theta}, \Sigma)$$

$$p(\mathbf{y}_j | X_j \boldsymbol{\beta}_j, \gamma) = \prod_{i=1}^{n_j} p(y_{i,j} | \boldsymbol{\beta}_j^T \mathbf{x}_{i,j}, \gamma)$$

γ: scale parameter; mean of y is a function of $\boldsymbol{\beta}_j^T \mathbf{x}_{i,j}$

Generalised Linear Mixed effects models

A Metropolis-Gibbs algorithm

- Gibbs steps for $\boldsymbol{\theta}, \Sigma$
- Metropolis step for $\boldsymbol{\beta}_j$

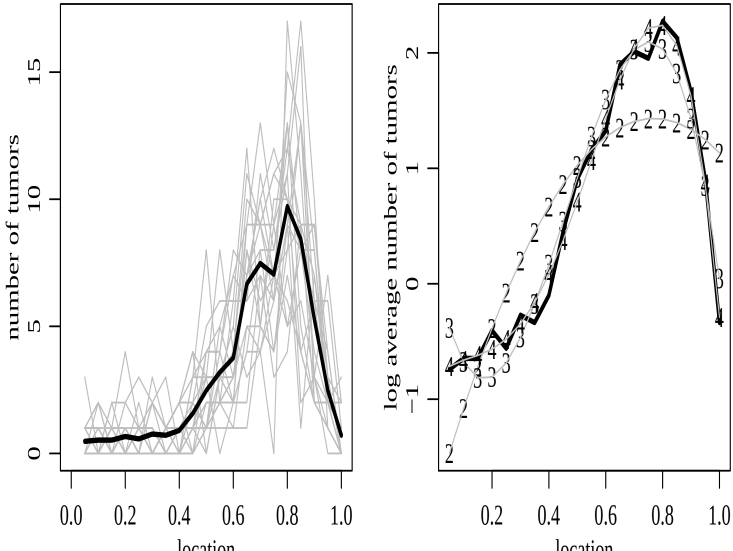
1. Sample $\boldsymbol{\beta}_j^* \sim \mathrm{MVN}(\boldsymbol{\beta}_j^{(s)}, V_j^{(s)})$
2. Compute
$$r = \frac{p(\mathbf{y}_j|X_j, \boldsymbol{\beta}_j^*)p(\boldsymbol{\beta}_j^*|\boldsymbol{\theta}^{(s)}, \Sigma^{(s)})}{p(\mathbf{y}_j|X_j, \boldsymbol{\beta}_j^{(s)})p(\boldsymbol{\beta}_j^{(s)}|\boldsymbol{\theta}^{(s)}, \Sigma^{(s)})}$$
3. Sample $u \sim \mathrm{uniform}(0,1)$. Set $\boldsymbol{\beta}_j^{(s+1)}$ to $\boldsymbol{\beta}_j^*$ if $u < r$ and to $\boldsymbol{\beta}_j^{(s)}$ if $u > r$.

$V_j^{(s)}$ is the proposal variance; perhaps set equal to a scaled version of $\Sigma^{(s)}$

Generalised Linear Mixed effects models - example

Analysis of tumour location data

($m = 21$; $n_j = 20$)



Generalised Linear Mixed effects models - example

Analysis of tumour location data

Model: $Y_{x,j}$: tumor count of mouse j at location x, where $Y_{x,j} \sim \mathrm{Pois}(\exp(f_j(x)))$, and $f_j(x) = \beta_{0,j} + \beta_{1,j}x + \beta_{2,j}x^2 + ... + \beta_{p-1,j}x^{p-1}$ $(x \in [0,1])$

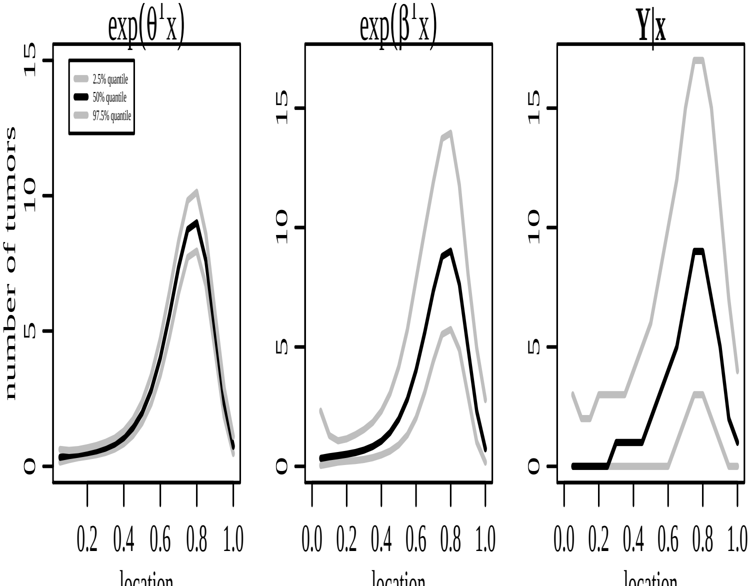
A fourth-degree polynomial was fit for f_j based on visual inspection of graphs of the data.

Unit information priors

- Prior guess for β_j's: regress $\log(y_{1,j} + 1/n) \ldots \log(y_{n,j} + 1/n)$ on $(\mathbf{x}_1, ..., \mathbf{x}_{20})$ (where $\mathbf{x}_i = (1, x_i, x_i^2, x_i^3, x_i^4)$ for $x \in (0.05, 0.10, ..., 0.95)$) → obtain estimates $\tilde{\boldsymbol{\beta}}_j$'s.
- Set $\boldsymbol{\mu} = \frac{1}{m}\sum_{j=1} \tilde{\boldsymbol{\beta}}_j$, and a prior covariance matrix equal to the sample covariance matrix of the $\tilde{\boldsymbol{\beta}}_j$'s. Set S_0 equal to this sample covariance matrix and $\eta_0 = p + 2 = 7$ (diffuse prior on variance of β's).

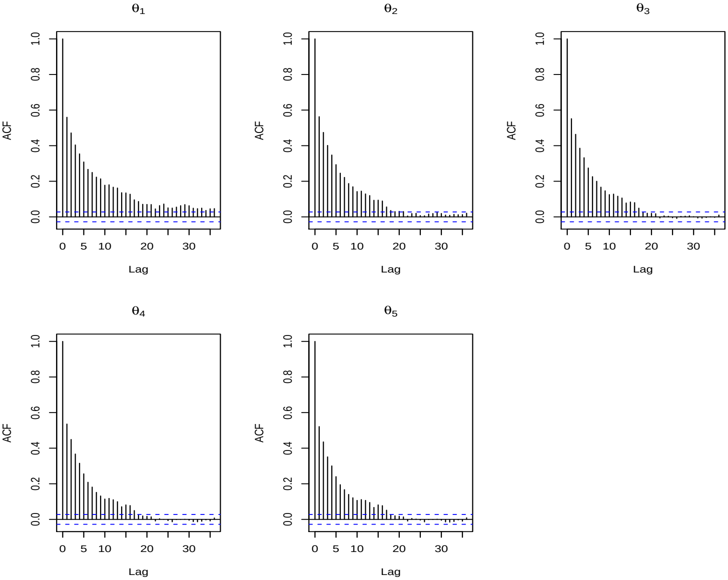
Generalised Linear Mixed effects models - example

Analysis of tumour location data



Generalised Linear Mixed effects models - example

Analysis of tumour location data - some MCMC diagnostics



Parameter expansion

Suppose a group level variance parameter happens to be estimated near zero. Then in the updating step for the group level coefficients, $\beta_1, ..., \beta_J$, these parameters will be pooled towards their common mean.

But the updating step for the group level variance parameter is based on the coefficients $\beta_1, ..., \beta_J$, and so will be estimated close to zero again, and the algorithm can get stuck.

Solution: Rescale the β_j's by multiplying the entire vector of group level coefficients $\boldsymbol{\beta}$ by a constant.

Parameter expansion - tumor location example

The parameter expanded version of the tumor location model is:

$Y_{x,j}$: tumor count of mouse j at location x, where
$Y_{x,j} \sim \mathrm{Pois}(\exp(f_j(x))$, and
$f_j(x) = \alpha\omega_{0,j} + \alpha\omega_{1,j}x + \alpha\omega_{2,j}x^2 + \alpha\omega_{3,j}x^3 + \alpha\omega_{4,j}x^4$ $(x \in [0,1])$

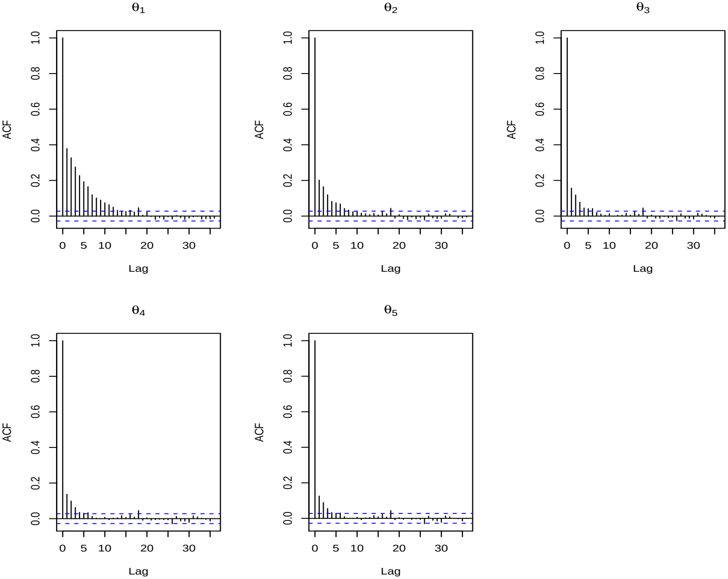
Here, α is a redundant multiplicative working parameter. The regression parameters of interest are now $\beta_j = \alpha\omega_j$, with group level standard deviation of $\sigma_\beta = |\alpha|\sigma_\omega$.

Parameter expansion

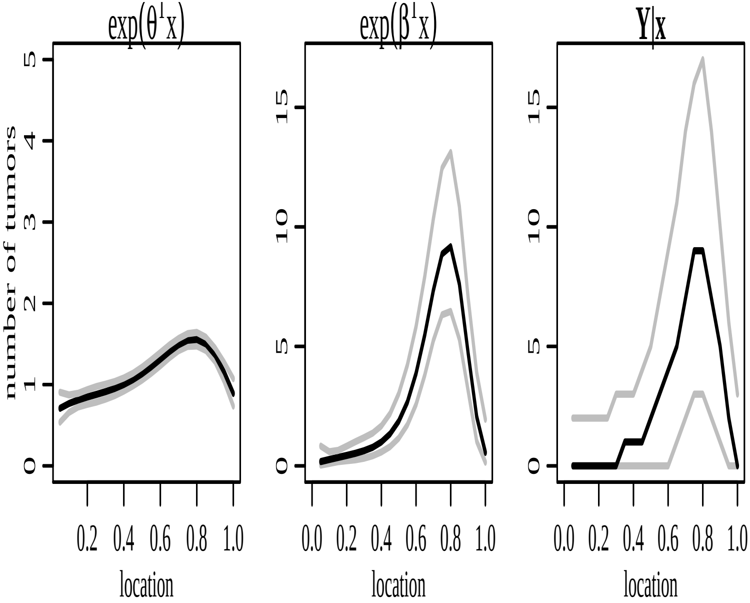
Why does it work?

- We transform from β to ω via multiplaction by α that rescales and potentially decorrelates the components of β (equivalent to rotating the parameter space)
- See Gelman et al. (2004) for theoretical derivations to show the increase in speed of convergence under parameter expansion.
- The likelihood function $p(y_j|\theta, \sigma_\beta)$ obtained by integrating out the ω_j's in the parameter expanded model is the same as that obtained by integrating out the β_j's in the original model. So we can fit either model to obtain a Monte Carlo sample from the same posterior distribution $p(y_j|\theta, \sigma_\beta)$.
- The added parameter α is not necessarily of any statistical interest.

Parameter expansion - Tumor location example - MCMC diagnostics



Parameter expansion - Tumor location example - results



Exercise - a nonlinear hierarchical growth curve model

The following table displays data on the growth of five orange trees over time. The response y_{ij} is the trunk circumference recorded at time x_j, $j = 1, .., 7$ and $i = 1, ..., 5$.

	Reponse for Tree Number				
x	1	2	3	4	5
118	30	33	30	32	30
484	58	69	51	62	49
664	87	111	75	112	81
1004	115	156	108	167	125
1231	120	172	115	179	142
1372	142	203	138	209	174
1582	145	203	140	214	177

Exercise - a nonlinear hierarchical growth curve model

One assumes y_{ij} is normally distributed with mean η_{ij} and variance σ^2, where the means satisfy the non-linear growth model

$$\eta_{ij} = \frac{\phi_{1i}}{1 + \phi_{i2}\exp(\phi_{i3}x_j)}$$

Suppose one reexpresses the parameters as the real-valued parameters:

$$\theta_{i1} = \log\phi_{i1};\ \ \theta_{i2} = \log(\phi_{i2}+1); \theta_{i3} = \log(-\phi_{i3}),\ \ i = 1, ..., 5$$

Let θ_i represent the vector of growth parameters for the ith tree.

Exercise - a nonlinear hierarchical growth curve model

To reflect a prior belief in similarity in the growth patterns of the five trees, one assumes that $\{\theta_i, i = 1, ..., 5\}$ are a random sample from a multivariate normal distribution with mean vector μ and variance covariance matrix Ω.

For the hyperprior distribution, assume Ω^{-1} follows a Wishart distribution with parameters R and 3, where R is a diagonal matrix with diagonal elements 0.1. Assume μ is multivariate normal with mean vector μ_0 (equal to the zero vector) and variance-covariance matrix M. Let M^{-1} be a diagonal matrix with diagonal elements 1.0×10^{-6}.

Implement a MCMC algorithm to approximate the posterior distribution of all parameters. Obtain 90% posterior interval estimates for all parameters.