Hierarchical Joint Effects Selection in Mixed Models

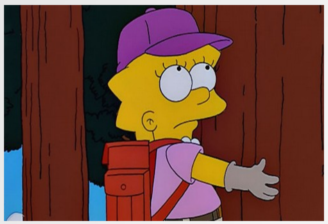
Francis Hui (ANU)
Samuel Muller (USYD)
Alan Welsh (ANU)

Aims of this talk:

- Generalized Linear Mixed Models (GLMMs)
- Penalized Likelihood Methods
  - What's been done for penalized variable selection in GLMMs?
- The CREPE Estimator for Joint Selection in GLMMs
  - Hui et al. *Composite Effects Selection in Mixed Models using CREPE.* Stat Sinica: In review.

Some images courtesy of Google images

Generalized Linear Mixed Models (GLMMs)



Over many years...

Has tree experienced defoliation?
1 = yes; 0 = no

Physical characteristics, soil chemistry, weather etc...

- Longitudinal dataset

Response matrix

Year 1 Year 2 … Year 20

Tree 1, Tree 2, . . . Tree 100

$$\begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1m} \\ y_{21} & y_{22} & \cdots & y_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{nm} \end{pmatrix}$$

Covariates for tree i = 1,....,n

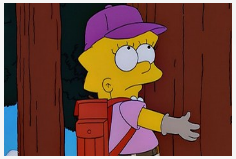
Year 1 Year 2 … Year 20

Rainfall, Fertilization, . . . Inclination

$$\begin{pmatrix} 1.34 & 1.48 & \ldots & 0.79 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 62 & 62 & \ldots & 62 \end{pmatrix}$$

Generalized Linear Mixed Models (GLMMs)

- Longitudinal dataset
  - Has there been a change in forest health over time?
  - What are the important predictors of forest health?



- Generalized Linear Mixed models

$$g(\mu_{ij}) = \boldsymbol{x}_{ij}^T \boldsymbol{\beta} + \boldsymbol{z}_{ij}^T \boldsymbol{b}_i; \quad i = 1, \ldots, n; \; j = 1, \ldots, m$$

Population averaged response

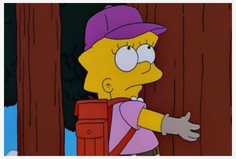
Between-cluster variability

$$\boldsymbol{b}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Gamma}\boldsymbol{\Gamma}^T)$$

Cholesky decomposition or eigendecomposition

Generalized Linear Mixed Models (GLMMs)

- Longitudinal dataset
  - What are the important predictors of forest health?



- Joint variable selection of fixed and random effects

$$g(\mu_{ij}) = \boldsymbol{x}_{ij}^T \boldsymbol{\beta} + \boldsymbol{z}_{ij}^T \boldsymbol{b}_i; \quad i = 1, \ldots, n; \ j = 1, \ldots, m$$

Select these two things!

$$\boldsymbol{b}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}\boldsymbol{\Gamma}^T)$$

- Some complications...
  - Lots of candidate models 😱
  - How to select the elements of the covariance matrix 😕
  - There's a hierarchical structure there: **"we usually only consider time-varying covariates that have been included in the fixed effects"** (Cheng et al., 2010) 😀

The CREPE estimator is designed to resolve the three problems above!

Penalized likelihood methods

- "Lots" of candidate models
  - Well, at least more than the $2^p - 1$ in GLMs

- One solution: Add a penalty to the likelihood

$$\hat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}} \; \ell(\boldsymbol{\Psi}) - p_\lambda(\boldsymbol{\beta}),$$

Tuning parameter

- Choose a penalty that is non-differentiable at zero => induces sparsity
  - lasso; adaptive lasso; SCAD etc...

Penalized likelihood methods

- What's been done for penalized likelihood in GLMMs?

$$g(\mu_{ij}) = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_{ij}^T\boldsymbol{b}_i; \quad i = 1, \ldots, n; \; j = 1, \ldots, m$$

Penalize these two things!

$$\boldsymbol{b}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}\boldsymbol{\Gamma}^T)$$

- LMMs:
  - M-ALASSO (Bondell et al., 2010)

$$p_\lambda(\boldsymbol{\Psi}) = \lambda \sum_k \tilde{w}_k|\beta_k| + \lambda \sum_k \tilde{v}_k|[\boldsymbol{\Gamma}]_{kk}|$$

"additive" penalty

Penalized likelihood methods

- What's been done for penalized likelihood in GLMMs?

$$g(\mu_{ij}) = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_{ij}^T\boldsymbol{b}_i; \quad i = 1, \ldots, n;\ j = 1, \ldots, m$$

Penalize these two things!

$$\boldsymbol{b}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}\boldsymbol{\Gamma}^T)$$

- LMMs:
  - M-ALASSO (Bondell et al., 2010)
  - ALASSO (Lin et al., 2013)

$$p_\lambda(\boldsymbol{\Psi}) = \lambda \sum_k \tilde{w}_k |\beta_k| + \lambda \sum_k \tilde{v}_k |[\boldsymbol{\Gamma}]_{kk}|$$

$$\text{Stage 1:} \quad p_\lambda(\boldsymbol{\Psi}) = \lambda_1 \sum_k \tilde{v}_k |[\boldsymbol{\Gamma}\boldsymbol{\Gamma}^T]_{kk}|; \quad \text{Stage 2:} \quad p_\lambda(\boldsymbol{\Psi}) = \lambda_2 \sum_k \tilde{w}_k |\beta_k|$$

Do random effects, then fixed effects

Penalized likelihood methods

- What's been done for penalized likelihood in GLMMs?

$$g(\mu_{ij}) = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_{ij}^T\boldsymbol{b}_i; \quad i = 1, \dots, n;\ j = 1, \dots, m$$

Penalize these two things!

$$\boldsymbol{b}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}\boldsymbol{\Gamma}^T)$$

- LMMs:
  - M-ALASSO (Bondell et al., 2010)

$$p_\lambda(\boldsymbol{\Psi}) = \lambda \sum_k \tilde{w}_k|\beta_k| + \lambda \sum_k \tilde{v}_k|[\boldsymbol{\Gamma}]_{kk}|$$

  - ALASSO (Lin et al., 2013)

$$\text{Stage 1:} \quad p_\lambda(\boldsymbol{\Psi}) = \lambda_1 \sum_k \tilde{v}_k|[\boldsymbol{\Gamma}\boldsymbol{\Gamma}^T]_{kk}|; \quad \text{Stage 2:} \quad p_\lambda(\boldsymbol{\Psi}) = \lambda_2 \sum_k \tilde{w}_k|\beta_k|$$

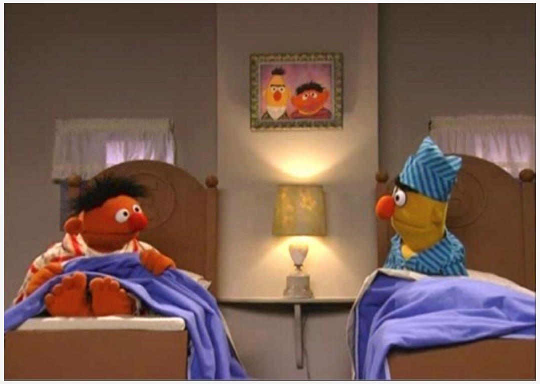
  - SCAD-P (Fan and Li, 2012); Iterative (Peng and Lu, 2012)
    - Both two stage process like ALASSO

- GLMMs:
  - Tweak the M-ALASSO for GLMMs (Ibrahim et al., 2011)

Penalized likelihood methods

- A basic problem:
  - M-ALASSO (Bondell et al., 2010) $p_\lambda(\boldsymbol{\Psi}) = \lambda \sum_k \tilde{w}_k |\beta_k| + \lambda \sum_k \tilde{v}_k |[\boldsymbol{\Gamma}]_{kk}|$
  - ALASSO (Lin et al., 2013) Stage 1: $p_\lambda(\boldsymbol{\Psi}) = \lambda_1 \sum_k \tilde{v}_k |[\boldsymbol{\Gamma}\boldsymbol{\Gamma}^T]_{kk}|$; Stage 2: $p_\lambda(\boldsymbol{\Psi}) = \lambda_2 \sum_k \tilde{w}_k |\beta_k|$
  - SCAD-P (Fan and Li, 2012); Iterative (Peng and Lu, 2012)
  - Tweak the M-ALASSO for GLMMs (Ibrahim et al., 2011)

All the penalties above treat the selection of fixed and random effects as **separate** processes.



Penalized likelihood methods

- A basic problem:
  - M-ALASSO (Bondell et al., 2010)

$$p_{\lambda}(\boldsymbol{\Psi}) = \lambda \sum_{k} \tilde{w}_k |\beta_k| + \lambda \sum_{k} \tilde{v}_k |[\boldsymbol{\Gamma}]_{kk}|$$

  - ALASSO (Lin et al., 2013)

$$\text{Stage 1:} \quad p_{\lambda}(\boldsymbol{\Psi}) = \lambda_1 \sum_{k} \tilde{v}_k |[\boldsymbol{\Gamma}\boldsymbol{\Gamma}^T]_{kk}|; \quad \text{Stage 2:} \quad p_{\lambda}(\boldsymbol{\Psi}) = \lambda_2 \sum_{k} \tilde{w}_k |\beta_k|$$

  - SCAD-P (Fan and Li, 2012); Iterative (Peng and Lu, 2012)
  - Tweak the M-ALASSO for GLMMs (Ibrahim et al., 2011)

All the penalties above treat the selection of fixed and random effects as **separate** processes.

There's a hierarchical structure for longitudinal GLMMs!

"We usually only consider time-varying covariates that have been included in the fixed effects" (Cheng et al., 2010)

Covariates

Fixed effect only

Composite (fixed and random) effect

Penalized likelihood methods

- A basic problem:
  - M-ALASSO (Bondell et al., 2010) $$p_\lambda(\Psi) = \lambda \sum_k \tilde{w}_k|\beta_k| + \lambda \sum_k \tilde{v}_k|[\Gamma]_{kk}|$$
  - ALASSO (Lin et al., 2013) $$\text{Stage 1:} \quad p_\lambda(\Psi) = \lambda_1 \sum_k \tilde{v}_k|[\Gamma\Gamma^T]_{kk}|; \quad \text{Stage 2:} \quad p_\lambda(\Psi) = \lambda_2 \sum_k \tilde{w}_k|\beta_k|$$
  - SCAD-P (Fan and Li, 2012); Iterative (Peng and Lu, 2012)
  - Tweak the M-ALASSO for GLMMs (Ibrahim et al., 2011)

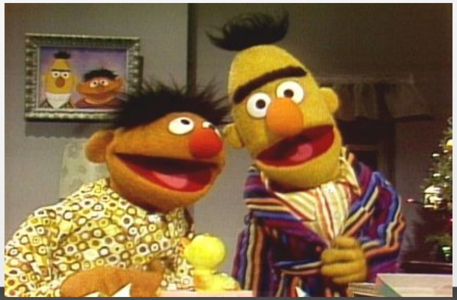
All the penalties above treat the selection of fixed and random effects as **separate** processes.

There's a hierarchical structure for longitudinal GLMMs!

Covariates → Fixed effect only

Covariates → Composite (fixed and random) effect

- Design and use a penalty that automatically incorporates this structure!



The CREPE Estimator...ingredients

- GLMMs

$$g(\mu_{ij}) = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_{ij}^T\boldsymbol{b}_i; \quad i = 1, \ldots, n; \; j = 1, \ldots, m$$

$$\boldsymbol{b}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Gamma}\boldsymbol{\Gamma}^T)$$

- Fixed effects
  - Adaptive lasso

$$p_\lambda(\boldsymbol{\beta}) = \sum_k \tilde{w}_k|\beta_k| \quad \text{where} \quad \tilde{w}_k = \tilde{\beta}_k^{-\gamma},$$

- Random effects ☹
  - Adaptive group lasso

$$p_\lambda(\boldsymbol{\gamma}_k) = \sum_k \tilde{v}_k||\boldsymbol{\gamma}_k|| \quad \text{where} \quad \tilde{v}_k = ||\tilde{\boldsymbol{\gamma}}_k||^{-\gamma},$$

Shrinking all elements in row k of the eigendecomposition to zero simultaneously

The CREPE Estimator...ingredients

- GLMMs

$$g(\mu_{ij}) = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_{ij}^T\boldsymbol{b}_i; \quad i = 1, \ldots, n;\ j = 1, \ldots, m$$

$$\boldsymbol{b}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}\boldsymbol{\Gamma}^T)$$

- Fixed effects
  - Adaptive lasso

$$p_\lambda(\boldsymbol{\beta}) = \sum_k \tilde{w}_k|\beta_k| \quad \text{where} \quad \tilde{w}_k = \tilde{\beta}_k^{-\gamma},$$

- Random effects 😕
  - Adaptive group lasso

$$p_\lambda(\boldsymbol{\gamma}_k) = \sum_k \tilde{v}_k\|\boldsymbol{\gamma}_k\| \quad \text{where} \quad \tilde{v}_k = \|\tilde{\boldsymbol{\gamma}}_k\|^{-\gamma},$$

- One more thing… 😀

Covariates → Fixed effect only

Covariates → Composite (fixed and random) effect

The CREPE Estimator

- GLMMs $g(\mu_{ij}) = \boldsymbol{x}_{ij}^T \boldsymbol{\beta} + \boldsymbol{z}_{ij}^T \boldsymbol{b}_i; \quad i = 1, \ldots, n; \; j = 1, \ldots, m$

$$\boldsymbol{b}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}\boldsymbol{\Gamma}^T)$$

- CREPE (Composite Random Effects PEnalty)

Was the covariate included as a composite effect?

$$\ell_{pen}(\boldsymbol{\Psi}) = \ell(\boldsymbol{\Psi}) - \lambda \sum_{k=1}^{p} \tilde{w}_k \sqrt{\beta_k^2 + \mathbb{1}_{\{k \in \alpha_c\}} \tilde{v}_k ||\boldsymbol{\gamma}_k||},$$

The CREPE Estimator

- GLMMs $$g(\mu_{ij}) = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_{ij}^T\boldsymbol{b}_i; \quad i = 1, \ldots, n; \; j = 1, \ldots, m$$
$$\boldsymbol{b}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}\boldsymbol{\Gamma}^T)$$
- CREPE (Composite Random Effects PEnalty)

Was the covariate included as a composite effect?

$$\ell_{pen}(\boldsymbol{\Psi}) = \ell(\boldsymbol{\Psi}) - \lambda \sum_{k=1}^{p} \tilde{w}_k \sqrt{\beta_k^2 + \mathbb{1}_{\{k \in \alpha_c\}} \tilde{v}_k ||\boldsymbol{\gamma}_k||},$$

- If the covariate is included as a purely fixed effect, CREPE => Adaptive lasso
- CREPE incorporates the hierarchical nature of the covariates:
  - By design, you're either a fixed effect or composite effect...can't be purely a random effect!!!

CREPE Sims

- Linear Mixed Models (Gaussian responses)
- Methods to compare: 1) CREPE, 2) M-ALASSO (Bondell et al., 2010), 3) ALASSO (Lin et al., 2013)
- $$p = \dim(\boldsymbol{x}_{ij}) = \lceil 7n^{1/4} \rceil$$

$$\boldsymbol{z}_{ij} = \text{equals first 8 elements of } \boldsymbol{x}_{ij}$$

$$y_{ij} = \boldsymbol{x}_{ij}^T \boldsymbol{\beta}_0 + \boldsymbol{z}_{ij}^T \boldsymbol{b}_i + \epsilon_{ij}$$

$$\boldsymbol{\beta}_0 = (-1, 3, 1.5, 0, 0, 2, 1, 0, 0, 1, 0, 0, -1, \ldots)$$

$$\boldsymbol{\Gamma}_0 \boldsymbol{\Gamma}_0^T = \begin{pmatrix} 9 & 4.8 & 0 & 0 & 0 & 0 & 0 & 0 \\ 4.8 & 4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

CREPE Sims

- Linear Mixed Models (Gaussian responses)
- Methods to compare: 1) CREPE, 2) M-ALASSO (Bondell et al., 2010), 3) ALASSO (Lin et al., 2013)
- $$p = \dim(\boldsymbol{x}_{ij}) = \lceil 7n^{1/4} \rceil$$

$$\boldsymbol{z}_{ij} = \text{equals first 8 elements of } \boldsymbol{x}_{ij}$$

$$y_{ij} = \boldsymbol{x}_{ij}^T \boldsymbol{\beta}_0 + \boldsymbol{z}_{ij}^T \boldsymbol{b}_i + \epsilon_{ij}$$

$$\boldsymbol{\beta}_0 = (-1, 3, 1.5, 0, 0, 2, 1, 0, 0, 1, 0, 0, -1, \ldots)$$

$$\boldsymbol{\Gamma}_0 \boldsymbol{\Gamma}_0^T = \begin{pmatrix} 9 & 4.8 & 0 & 0 & 0 & 0 & 0 & 0 \\ 4.8 & 4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

FP = # of false positives for fixef (overfitting)

FN = # of false negatives for fixef (underfitting)

%RE = percentage of datasets with correct ranef structure

%S = percentage of datasets where non-hierarchical shrinkage occurred

CREPE Sims

- Linear Mixed Models (Gaussian responses)

- Methods to compare: 1) CREPE, 2) M-ALASSO (Bondell et al., 2010), 3) ALASSO (Lin et al., 2013)

- $p = \dim(\boldsymbol{x}_{ij}) = \lceil 7n^{1/4} \rceil$

  $\boldsymbol{z}_{ij}$ = equals first 8 elements of $\boldsymbol{x}_{ij}$

  $y_{ij} = \boldsymbol{x}_{ij}^T \boldsymbol{\beta}_0 + \boldsymbol{z}_{ij}^T \boldsymbol{b}_i + \epsilon_{ij}$

FP = # of false positives for fixef (overfitting)

FN = # of false negatives for fixef (underfitting)

%RE = percentage of datasets with correct ranef structure

%S = percentage of datasets where non-hierarchical shrinkage occurred

n	m	CREPE			M-ALASSO				ALASSO			
		FP	FN	%RE	FP	FN	%RE	%S	FP	FN	%RE	%S
30	10	0.65	0	54	1.45	0.45	1	25	1.86	5.31	2	85
$(p_f = 17)$	20	0.31	0.01	94	0.98	0.16	2	2	1.04	4.13	8	62
60	10	0.46	0	64	1.05	0.05	0	11	0.25	8.12	1	70
$(p_f = 20)$	20	0.12	0	98	0.71	0.03	0	13	0.04	7.79	11	66

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Thanks that was delicious!

- For longitudinal GLMMs, CREPE builds in the hierarchical structure that covariates should end up as either fixed or composite effects.
- Outperform the limited stuff that is currently out there

Thanks to everyone for listening 😋