Variable Screening and Parameter Estimation for High-Dimensional Generalized Linear Mixed Models Using ℓ_1-Penalization

Jürg Schelldorfer

joint work with Peter Bühlmann

ETH Zurich

useR! 2011

August 16, 2011

Riboflavin Production in Bacillus Subtilis

A data set provided by DSM Nutritional Products.

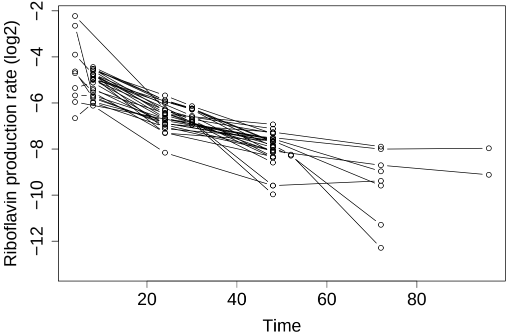
Goal: improve riboflavin production rate by genetic engineering

response variable $Y \in \mathbb{R}$: riboflavin (log-)production rate
covariates $X \in \mathbb{R}^p$: expressions from genes
$n = 111$ observations and $p = 4088$ variables

$\hookrightarrow$ "simple" high-dimensional regression problem, but...

Riboflavin Production in Bacillus Subtilis

...we know more about the data...



$n = 111$ observations and $p = 4088$ variables
$N = 28$ groups with $n_i \in \{2, ..., 6\}$ observations per group

$\hookrightarrow$ high-dimensional longitudinal data

General framework

$\hookrightarrow$ high-dimensional longitudinal data

when $n \ll p$

Goal 1: variable selection

Goal 2: parameter estimation

Table of Contents

Overview

	$n > p$	$n \ll p$
Generalized Linear Models (GLMs)	MLE [glm]	Lasso [glmnet]
Generalized Linear Mixed Models (GLMMs)	MLE [glmer]	?

n: number of observations
p: number of variables

Generalized Linear Model (GLM)

For n observations (y_i, x_i^T)

- (y_i, x_i^T) independent for $i = 1, \ldots, n$
- y_i has a density

$$\exp\left\{\phi^{-1}\Big(y_i\xi_i - b(\xi_i)\Big) + c(y_i, \phi)\right\} \text{with } \boldsymbol{\mu}_i = \mathbb{E}[y_i]$$

- $g(\boldsymbol{\mu}) = \boldsymbol{\eta}$ with $\boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta}$

Estimate $\boldsymbol{\beta}$ by

$$\hat{\boldsymbol{\beta}}_{MLE} = argmin_{\boldsymbol{\beta}} - \ell(\boldsymbol{\beta})$$

ℓ_1-regularized Generalized Linear Model

For $n \ll p$ we should not use the MLE. Use the Lasso (Tibshirani, 1996)

$$\hat{\boldsymbol{\beta}}(\lambda) = \textit{argmin}_{\boldsymbol{\beta}} - \ell(\boldsymbol{\beta}) + \lambda \|\boldsymbol{\beta}\|_1 \quad , \quad \lambda \geq 0$$

with the following properties:

- The Lasso does variable selection (i.e. some coefficients are set exactly to zero)
- Convex optimiziation problem, which can be solved efficiently

Generalized Linear Mixed Model (GLMM) I

Notation:

$g = 1, ..., N$ independent groups/clusters/subjects
$j = 1, .., n_g$ observations for group/cluster/subject g
$n = \sum_{g=1}^{N} n_g$ total number of observations

$\boldsymbol{y}$: n-dim response variable
$\boldsymbol{b}$: q-dim (correlated) random effects

$\boldsymbol{\beta} \in \mathbb{R}^p$ fixed-effects parameters
$\boldsymbol{\theta} \in \mathbb{R}^d$ covariance parameters
ϕ dispersion parameter

$\boldsymbol{X}$: $n \times p$ model matrix for $\boldsymbol{\beta}$
$\boldsymbol{Z}$: $n \times q$ model matrix for $\boldsymbol{b}$
$\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$: $q \times q$ covariance matrix, determined by $\boldsymbol{\theta}$

Generalized Linear Mixed Model (GLMM) II

Model Assumptions:

- $y_i|\boldsymbol{b}$ are independent for $i = 1, \ldots, n$
- $y_i|\boldsymbol{b}$ has a density of the form

$$\exp\left\{\phi^{-1}\Big(y_i\xi_i - b(\xi_i)\Big) + c(y_i,\phi)\right\} \text{ with } \boldsymbol{\mu}_i = \mathbb{E}[y_i|\boldsymbol{b}]$$

- $g(\boldsymbol{\mu}) = \boldsymbol{\eta}$ with $\boldsymbol{\eta} = \boldsymbol{X}\beta + \boldsymbol{Z}\boldsymbol{b}$
- $\boldsymbol{b} \sim \mathcal{N}_q(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}})$ with $\boldsymbol{\Sigma}_{\boldsymbol{\theta}} \geq 0$ for $\boldsymbol{\theta} \in \mathbb{R}^d$

$$(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}, \hat{\phi})_{MLE} = argmin_{\boldsymbol{\beta},\boldsymbol{\theta},\phi} - \log L(\boldsymbol{\beta}, \boldsymbol{\theta}, \phi)$$

Recap

	$n > p$	$n \ll p$
Generalized Linear Models (GLMs)	MLE [`glm`] ✔	Lasso [`glmnet`] ✔
Generalized Linear Mixed Models (GLMMs)	MLE [`glmer`] ✔	!

High-dimensional GLMM Set-up

Additionally to a GLMM, assume

- $n = \sum_{i=1}^{N} n_g \ll p$
- the true $\boldsymbol{\beta}_0$ is sparse
- $d = dim(\boldsymbol{\theta})$ small, say $d \leq 6$

Aim: Estimate $\boldsymbol{\beta}, \boldsymbol{\theta}$, ϕ and predict $\boldsymbol{b}$

The GLMMLasso estimator

Key Idea 1: Lasso-type penalty

objective function:

$$Q_\lambda(\boldsymbol{\beta}, \boldsymbol{\theta}, \phi) := -2 \log L(\boldsymbol{\beta}, \boldsymbol{\theta}, \phi) + \lambda \|\boldsymbol{\beta}\|_1,$$

Estimate $(\boldsymbol{\beta}, \theta, \phi)$ by

$$(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}, \hat{\phi}) := \mathit{argmin}_{\boldsymbol{\beta}, \boldsymbol{\theta}, \phi} Q_\lambda(\boldsymbol{\beta}, \boldsymbol{\theta}, \phi).$$

In general, $L(\boldsymbol{\beta}, \boldsymbol{\theta}, \phi)$ cannot be computed explicitly.

The GLMMLasso estimator

Key Idea 2: Laplace approximation to approximate the integrand of $L(\boldsymbol{\beta}, \boldsymbol{\theta}, \phi)$ by a quadratic function.

$$I = \int_{\mathbb{R}^q} e^{-S(\boldsymbol{b})} d\boldsymbol{b} \approx (2\pi)^{q/2} |S''(\tilde{\boldsymbol{b}})|^{-1/2} e^{-S(\tilde{\boldsymbol{b}})}$$

where $\tilde{\boldsymbol{b}} = argmin_{\boldsymbol{b}} S(\boldsymbol{b})$ is the mode of $-S(\boldsymbol{b})$.

Hence

$$Q_\lambda(\boldsymbol{\beta}, \boldsymbol{\theta}, \phi) \rightsquigarrow \tilde{Q}_\lambda^{LA}(\boldsymbol{\beta}, \boldsymbol{\theta}, \phi)$$

The GLMMLasso estimator

The GLMMLasso estimator is

$$(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}, \hat{\phi}) := \textit{argmin}_{\boldsymbol{\beta}, \boldsymbol{\theta}, \phi} \tilde{Q}_{\lambda}^{LA}(\boldsymbol{\beta}, \boldsymbol{\theta}, \phi)$$

This is a high-dimensional, non-convex optimization problem!

The GLMMLasso algorithm

How to calculate

$$(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}, \hat{\phi}) := \textit{argmin}_{\boldsymbol{\beta},\boldsymbol{\theta},\phi} \tilde{Q}_{\lambda}^{LA}(\boldsymbol{\beta}, \boldsymbol{\theta}, \phi)?$$

Key Idea 3: coordinatewise optimization with inexact line search

i.e. optimize $\tilde{Q}_{\lambda}^{LA}$ w.r.t. one coordinate keeping all other coordinates fixed (Tseng and Yun, 2009):

The GLMMLasso algorithm

$$
\begin{aligned}
(\boldsymbol{\beta}, \boldsymbol{\theta}, \phi) &= (\beta_1, \ldots, \beta_p, \theta_1, \ldots, \theta_d, \phi) \in \mathbb{R}^{p+d+1} \\
&= (\beta_1, \ldots, \beta_p, \theta_1, \ldots, \theta_d, \phi) \\
&= (\beta_1, \ldots, \beta_p, \theta_1, \ldots, \theta_d, \phi) \\
&= (\beta_1, \ldots, \beta_p, \theta_1, \ldots, \theta_d, \phi) \\
&= (\beta_1, \ldots, \beta_p, \theta_1, \ldots, \theta_d, \phi) \\
&= (\beta_1, \ldots, \beta_p, \theta_1, \ldots, \theta_d, \phi) \\
&\ldots \\
&= (\beta_1, \ldots, \beta_p, \theta_1, \ldots, \theta_d, \phi) \\
&= (\beta_1, \ldots, \beta_p, \theta_1, \ldots, \theta_d, \phi) \\
&\ldots
\end{aligned}
$$

What does the title mean...

1. Step: Variable Screening by GLMMLasso
Imposed by the Lasso (variable selection too restrictive)

2. Step: Parameter Estimation:
Key Idea 4: Refitting by ML with the selected (non-zero) variables to get accurate parameter estimates.

Riboflavin Production in Bacillus Subtilis

Gaussian linear mixed model:

$$y_{ij} = \boldsymbol{x}_{ij}^T \beta + b_{ik_1} z_{ijk_1} + b_{ik_2} z_{ijk_2} + \varepsilon_{ij} \quad i = 1, \ldots, N, \quad j = 1, \ldots, n_g$$

with $b_{ik_1} \sim \mathcal{N}(0, \theta_{k_1}^2)$, $b_{ik_2} \sim \mathcal{N}(0, \theta_{k_2}^2)$ and $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$.

Conclusions:

- variability between groups ($\hat{\sigma}^2 = 0.15$, $\hat{\theta}_{k_1}^2 = 0.03$, $\hat{\theta}_{k_2}^2 = 0.06$)
- one dominating gene

Take-home message

	$n > p$	$n \ll p$
Generalized Linear Models (GLMs)	MLE [glm]	Lasso [glmnet]
Generalized Linear Mixed Models (GLMMs)	MLE [glmer]	**GLMMLasso** [**glmmlasso**]

Thank you!

schelldorfer@stat.math.ethz.ch
http://stat.ethz.ch/people/schell

Contact:

schelldorfer@stat.math.ethz.ch
http://stat.ethz.ch/people/schell

R package:

http://stat.ethz.ch/people/schell

References

- P. Tseng and S. Yun ; A Coordinte Gradient Descent Method for Nonsmooth Separable Minimization ; Mathematical Programming (2009)
- R. Tibshirani ; Regression Shrinkage and Selection via the Lasso ; J. R. Stat. Soc. (1996)
- J. Schelldorfer, P. Bühlmann and S. van de Geer ; Estimation for High-Dimensional Linear Mixed-Effects Models Using ℓ_1-penalization ; The Scandinavian Journal of Statistics (2011)
- D. Bates ; Computational Methods for Mixed Models ; Vignette for lme4 (2011)