University of Queensland

GWAS in Admixed Populations via Variable Selection in Finite Mixture of Regression Models

Luke Lloyd-Jones[1,☯] and Hien Nguyen[1,☯]

[1]Centre for Neurogenetics and Statistical Genomics, Queensland Brain Institute, University of Queensland, Brisbane, QLD
[2]Mathematics and Physices, University of Queensland, St Lucia, Brisbane, QLD, AUS

Last updated April 27, 2015

Model and method for variable selection

Let Y be a response phenotype of interest and let $\mathbf{x} = (x_1, x_2, \ldots, x_p)'$ be the vector of covariates believed to have an effect on Y. The finite mixture model is defined as follows:

Let $\mathcal{G} = \{f(y; \theta, \phi); (\theta, \phi) \in \Theta \times (0, \infty)\}$ be a family of parametric density functions of Y with respect to a σ-finite measure ν, where $\theta \subset R$ and ϕ is a dispersion parameter. We say that $(\mathbf{x}, Y)$ follows a FMR model of order K if the conditional density function Y given $\mathbf{x}$ has the form

$$f(y; \mathbf{x}, \mathbf{\Psi}) = \sum_{k=1}^{K} \pi_k f(y; \theta_k(\mathbf{x}), \phi_k)$$

with $\theta_k(\mathbf{x}) = h(\mathbf{x}'\boldsymbol{\beta}_k)$ (some mean function), $k = 1, 2, \ldots, K$, for a given link function $h(.)$, and for some $\mathbf{\Psi} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_k, \boldsymbol{\phi}, \boldsymbol{\pi})$ with $\boldsymbol{\beta}_k = (\beta_{k1}, \beta_{k1}, \ldots, \beta_{kP})'$, $\boldsymbol{\phi} = (\phi_1, \phi_2, \ldots, \phi_K)'$ $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_{K-1})'$ such that $\pi_k > 0$ and $\sum_{k=1}^{K} \pi_k = 1$.

In the case where $\mathbf{x}$ is random, we assume that its density $f(\mathbf{x})$ is functionally independent of the parameters in the FMR model. Thus the statistical inference can be done based purely on the conditional density.

Let $(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_n, y_n)$ $(i \in 1, \ldots, n)$ be a sample of observations from the FMR model. The conditional log-likelihood function of $\mathbf{\Psi}$ is given by

$$l_n(\mathbf{\Psi}) = \sum_{i=1}^{n} \log \left\{ \sum_{k=1}^{K} \pi_k f(y_i; \theta_k(\mathbf{x}_i), \phi_k) \right\}.$$

When the effect of a component of $\mathbf{x}$ is not significant the corresponding ordinary maximum likelihood estimate is often close to but not equal to zero. Thus this covariate is not excluded from the model. We define a penalised log-likelihood function as

$$\tilde{l}_n(\mathbf{\Psi}) = l_n(\psi) - \mathbf{p}_n(\mathbf{\Psi})$$

with the penalty function

$$\mathbf{p}_n(\mathbf{\Psi}) = \sum_{k=1}^{K} \pi_k \left\{ \sum_{j=1}^{P} p_{nk}(\beta_{kj}) \right\},$$

where the $p_{nk}(\beta_{kj})$ values are nonnegative and nondecreasing functions in $|\beta_{kj}|$. By maximising $\tilde{l}_n(\mathbf{\Psi})$ that contains a penalty, there is a positive chance of having some estimated values of β that are equal to 0 and thus of selecting a model. In the penalty function we choose the penalty imposed by the regression coefficients within the kth component of the FMR model to be proportional to π_k. Similar to relating the penalty to the sample size. We will focus on the SCAD penalty, where we let $(.)_+$ be the positive part of a quantity

$$p'_{nk}(\beta) = \gamma_{nk}\sqrt{n} I\{\sqrt{n}|\beta| \leq \gamma_{nk}\} + \frac{\sqrt{n}(a\gamma_{nk} - \sqrt{n}|\beta|)_+}{a-1} I\{\sqrt{n}|\beta| > \gamma_{nk}\}$$

Numerical solution

Let $(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_n, y_n)$ be a random sample of observations from the FMR model. In the context of finite mixture models, the EM algorithm can be used. However, due to the condition that for all n and k, $p_{nk}(0) = 0$ and $p_{nk}(\beta)$ is symmetric, non-negative, non-decreasing and twice differentiable for all β in $(0, \infty)$ with at most a few exceptions, which is essential for sparsity, the $p_{nk}(\beta)$'s are not differentiable at $\beta = 0$. The Newton-Raphson algorithm can not be used directly in the M step of the EM unless it is properly adapted to deal with the single non-smooth point at $\beta = 0$. To deal with this problem we can replace $p_{nk}(\beta)$ by a local quadratic approximation.

$$p_{nk}(\beta) \approx p_{nk}(\beta_0) + \frac{p'_n(\beta_0)}{2\beta_0}(\beta^2 - \beta_0^2),$$

in a neighbourhood of β_0. This function increases to infinity whenever $|\beta| \to \infty$, which is more suitable for the application that the simple Taylor's expansion. We replace $\mathbf{p}_n(\mathbf{\Psi})$ in the penalised log-likelihood function by the following function:

$$\tilde{\mathbf{p}}_n(\mathbf{\Psi}; \mathbf{\Psi}^{(m)}) = \sum_{k=1}^{K} \pi_k \sum_{j=1}^{P} \left\{ p_{nk}(\beta_{jk}^{(m)}) + \frac{p'_n(\beta_{jk}^{(m)})}{2\beta_{jk}^{(m)}} (\beta_{jk}^2 - \beta_{jk}^{(m)^2}) \right\}$$

The revised EM algorithm is as follows: Let the complete log-likelihood function be

$$l_n^c(\mathbf{\Psi}) = \sum_{i=1}^{n} \sum_{k=1}^{K} z_{ik} [\log \pi_k + \log\{f(y_i; \theta_k(\mathbf{x}_i), \phi_k)\}]$$

where the z_{ik}s are indicator variables showing the component membership of the ith observation in the FMR model and are unobserved imaginary variables. The penalised complete log-likelihood function is then given by $\tilde{l}_n^c(\mathbf{\Psi}) = l_n^c(\mathbf{\Psi}) - \mathbf{p}_n(\mathbf{\Psi})$. The EM algorithm maximises $\tilde{l}_n^c(\mathbf{\Psi})$ in the following two steps:

E step: The E step computes the conditional expectation of the function $\tilde{l}_n^c(\mathbf{\Psi})$ with respect to z_{ik}, given the data $(\mathbf{x}_1, y_1)$ and assuming that the current estimate $\mathbf{\Psi}^{(m)}$ gives the true parameters of the model. The conditional expectation is

$$Q(\mathbf{\Psi}, \mathbf{\Psi}^{(m)}) = \sum_{i=1}^{n} \sum_{k=1}^{K} w_{ik}^{(m)} \log \pi + \sum_{i=1}^{n} \sum_{k=1}^{K} w_{ik}^{(m)} \log\{f(y_i; \theta_k(\mathbf{x}_i), \phi_k)\} - \mathbf{p}_n(\mathbf{\Psi})$$

where the weights

$$w_{ik}^{(m)} = \frac{\pi_k^{(m)} f(y_i; \theta_k^{(m)}(\mathbf{x}_i), \phi_k^{(m)})}{\sum_{l=1}^{K} \pi_l^{(m)} f(y_i; \theta_l^{(m)}(\mathbf{x}_i), \phi_l^{(m)})}$$

are the conditional expectation of the unobserved z_{ik}.

M step: The M step on the $(m+1)$th iteration maximises the function $Q(\mathbf{\Psi}, \mathbf{\Psi}^{(m)})$ with respect to $\mathbf{\Psi}^{(m)}$. In the usual EM algorithm, the mixing proportions are updated by

$$\pi_k^{(m+1)} = \frac{1}{n} \sum_{i=1}^{n} w_{ik}^{(m)}$$

which maximise the leading term of $Q(\mathbf{\Psi}, \mathbf{\Psi}^{(m)})$.

Given this, we consider that the π_k are constant in $Q(\mathbf{\Psi}, \mathbf{\Psi}^{(m)})$, and maximise $Q(\mathbf{\Psi}, \mathbf{\Psi}^{(m)})$ with respect to the other parameters in $\mathbf{\Psi}$. By replacing $\mathbf{p}_n(\mathbf{\Psi})$ by $\tilde{\mathbf{p}}_n(\mathbf{\Psi}; \mathbf{\Psi}^{(m)})$ in $Q(\mathbf{\Psi}, \mathbf{\Psi}^{(m)})$, the regression coefficients are updated by solving

$$\sum_{i=1}^{n} w_{ik}^{(m)} \frac{\partial}{\partial \beta_{kj}} \{\log f(y_i; \theta_k(\mathbf{x}_i), \phi_k^{(m)})\} - \pi_k \frac{\partial}{\partial \beta_{kj}} \tilde{p}_{nk}(\beta_{kj})$$

where $\tilde{p}_{nk}(\beta_{kj})$ is the corresponding term in $\tilde{\mathbf{p}}_n(\mathbf{\Psi}, \mathbf{\Psi}^{(m)})$, for $k = 1, 2, \ldots, K; j = 1, 2, \ldots, P$. The updated estimates $\phi_k^{(m+1)}$ of the dispersion parameters are obtained by solving

$$\sum_{i=1}^{n} w_{ik}^{(m)} \frac{\partial}{\partial \phi_k} \{\log f(y_i; \theta_k(\mathbf{x}_i), \phi_k)\} = 0$$

Algorithm components

To begin, we start with a simple mixture of two normals i.e., $k = 2$, variances unknown, and we will use the SCAD penalty function. Given this, we need to find the following quantities

$$\frac{\partial}{\partial \beta_{kj}} \{\log f(y_i; \mathbf{x}_i' \boldsymbol{\beta}_k, \sigma_k^2)\}$$

where

$$f(y_i; \mathbf{x}_i' \boldsymbol{\beta}_k, \sigma_k^2) = \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left[-\frac{(y_i - \sum_{j=1}^P x_{ij}\beta_{kj})^2}{2\sigma_k^2} \right]$$

$$\log\{f(y_i; \mathbf{x}_i' \boldsymbol{\beta}_k, \sigma_k^2)\} = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma_k^2) - \frac{1}{2}\frac{(y_i - \sum_{j=1}^P x_{ij}\beta_{kj})^2}{\sigma_k^2}$$

$$\begin{aligned}\frac{\partial}{\partial \beta_{k\bar{j}}} \log\{f(y_i; \mathbf{x}_i' \boldsymbol{\beta}_k, \sigma_k^2)\} &= \frac{\partial}{\partial \beta_{k\bar{j}}} \left[-\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma_k^2) - \frac{1}{2}\frac{(y_i - \sum_{j=1}^P x_{ij}\beta_{kj})^2}{\sigma_k^2} \right] \\ &= \frac{(y_i - \sum_{j=1}^P x_{ij}\beta_{kj})}{\sigma_k^2} x_{i\bar{j}}\end{aligned}$$

where $\bar{j}$ denotes the jth value that we are taking the derivative with respect to. We also need

$$\frac{\partial}{\partial \beta_{k\bar{j}}} \tilde{p}_{nk}(\beta_{kj})$$

where

$$\tilde{p}_{nk}(\beta_{kj}) = \pi_k \sum_{j=1}^P \left\{ p_{nk}(\beta_{jk}^{(m)}) + \frac{p_{nk}'(\beta_{jk}^{(m)})}{2\beta_{jk}^{(m)}} (\beta_{jk}^2 - \beta_{jk}^{(m)^2}) \right\}$$

the $\beta_{jk}^{(m)}$ are treated as not dependent on β_{jk} and thus

$$\frac{\partial}{\partial \beta_{k\bar{j}}} \tilde{p}_{nk}(\beta_{kj}) = \frac{p_{nk}'(\beta_{\bar{j}k}^{(m)})}{\beta_{\bar{j}k}^{(m)}} \beta_{\bar{j}k}\pi_k$$

We can now attempt to solve

$$\sum_{i=1}^n w_{ik}^{(m)} x_{i\bar{j}} \frac{(y_i - \sum_{j=1}^P \beta_{kj}x_{ij})}{\sigma_k^2} - \frac{p_{nk}'(\beta_{\bar{j}k}^{(m)})}{\beta_{\bar{j}k}^{(m)}} \beta_{\bar{j}k}\pi_k = 0$$

for $\beta_{\bar{j}k}$. Express the first term as a split with the variable of interest

$$\sum_{i=1}^n w_{ik}^{(m)} x_{i\bar{j}} \frac{(y_i - \sum_{j\neq\bar{j}}^P \beta_{kj}x_{ij} - \beta_{k\bar{j}}x_{i\bar{j}})}{\sigma_k^2} - \frac{p_{nk}'(\beta_{\bar{j}k}^{(m)})}{\beta_{\bar{j}k}^{(m)}} \beta_{\bar{j}k}\pi_k = 0$$

$$\frac{\sum_{i=1}^n w_{ik}^{(m)} x_{i\bar{j}}(y_i - \sum_{j\neq\bar{j}}^P \beta_{kj}x_{ij})}{\sigma_k^2} - \frac{\sum_{i=1}^n w_{ik}^{(m)} \beta_{k\bar{j}} x_{i\bar{j}}^2}{\sigma_k^2} - \frac{p_{nk}'(\beta_{\bar{j}k}^{(m)})}{\beta_{\bar{j}k}^{(m)}} \beta_{\bar{j}k}\pi_k = 0$$

$$\frac{\sum_{i=1}^n w_{ik}^{(m)} x_{i\bar{j}}(y_i - \sum_{j\neq\bar{j}}^P \beta_{kj}x_{ij})}{\sigma_k^2} = \beta_{\bar{j}k} \left[\frac{\sum_{i=1}^n w_{ik}^{(m)} x_{i\bar{j}}^2}{\sigma_k^2} + \frac{p_{nk}'(\beta_{\bar{j}k}^{(m)})}{\beta_{\bar{j}k}^{(m)}} \pi_k \right]$$

$$\frac{\sum_{i=1}^n w_{ik}^{(m)} x_{i\bar{j}}(y_i - \sum_{j\neq\bar{j}}^P \beta_{kj}x_{ij})}{\sigma_k^2} = \beta_{\bar{j}k} \left[\frac{\beta_{\bar{j}k}^{(m)} \sum_{i=1}^n w_{ik}^{(m)} x_{i\bar{j}}^2}{\beta_{\bar{j}k}^{(m)}\sigma_k^2} + \frac{\pi_k\sigma_k^2 p_{nk}'(\beta_{\bar{j}k}^{(m)})}{\sigma_k^2\beta_{\bar{j}k}^{(m)}} \right]$$

$$\frac{\sum_{i=1}^n w_{ik}^{(m)} x_{i\bar{j}}(y_i - \sum_{j\neq\bar{j}}^P \beta_{kj}x_{ij})}{\sigma_k^2} = \beta_{\bar{j}k}\left[\frac{\beta_{\bar{j}k}^{(m)}\sum_{i=1}^n w_{ik}^{(m)}x_{i\bar{j}}^2 + \pi_k\sigma_k^2 p'_{nk}(\beta_{\bar{j}k}^{(m)})}{\sigma_k^2\beta_{\bar{j}k}^{(m)}}\right]$$

$$\frac{\beta_{\bar{j}k}^{(m)}\sum_{i=1}^n w_{ik}^{(m)}x_{i\bar{j}}(y_i - \sum_{j\neq\bar{j}}^P \beta_{kj}x_{ij})}{\beta_{\bar{j}k}^{(m)}\sum_{i=1}^n w_{ik}^{(m)}x_{i\bar{j}}^2 + \pi_k\sigma_k^2 p'_{nk}(\beta_{\bar{j}k}^{(m)})} = \beta_{\bar{j}k}$$

1 Variable selection using MM algorithms Hunter and Li

In the case of SCAD, Lasso, or bridge regression, the penalised log-likelihood is non differentiable; with SCAD or bridge regression the function is also non concave. They explore the connection with local quadratic approximation and MM algorithm.

The penalty functions $p_j(.)$ and tuning parameters λ_j are not necessarily the same for all j. They use the same penalisation for every component of $\boldsymbol{\beta}$ and write $p_\lambda(|\beta_j|)$. The AIC and BIC can be viewed as penalised likelihoods with $\lambda = \sqrt{(}2/n)$ and $\sqrt{(}\log(n)/n)$. For fixed $a > 2$, the SCAD penalty is the continuous function $p_\lambda(.)$ defined by $p_\lambda(0) = 0$ and for $\beta \neq 0$ and $p'_\lambda(|\beta|)$ defined above. The penalty functions $p_\lambda(\beta)$ are nondecreasing and concave (they say that SCAD is non concave). They go on to say that SCAD is ok.

They show that the local quadratic approximation is an instance of an MM algorithm. They say that their derivation applies to any of the penalty functions aforementioned and SCAD is included.

2 Hastie and Tibshirani on penalised regression Hastie et al. (2009)

3 Optimization - Penalty and Barrier Methods Lange et al. (2000)

It is profitable to view penalties and barriers from the perspective of the MM algorithm. We can engineer barrier tuning constants in a constrained minimisation problem so that the objective function is forced steadily downhill.

Lange writes $g_+(\mathbf{x})$ to be $g_+(\mathbf{x}) = \max\{g(\mathbf{x}), 0\}$.

References

Trevor Hastie, Robert Tibshirani, Jerome Friedman, T Hastie, J Friedman, and R Tibshirani. *The elements of statistical learning*, volume 2. Springer, 2009.

Kenneth Lange, David R Hunter, and Ilsoon Yang. Optimization transfer using surrogate objective functions. *Journal of computational and graphical statistics*, 9(1):1–20, 2000.