Title of Your Thesis

By
Bin Zhuo

A THESIS

submitted to

Oregon State University

University Honors College

in partial fulfillment of
the requirements for the
degree of

Honors Baccalaureate of Science in Statistics
(Honors Scholar)

Presented Month dd, yyyy
Commencement June 2016

AN ABSTRACT OF THE THESIS OF

Bin Zhuo for the degree of Honors Baccalaureate of Science in Statistics presented on Month dd, yyyy. Title: Title of Your Thesis

Abstract approved:

Yanming Di

This is the abstract for my honors thesis. I'm going to start here.

Key Words: keyword1, keyword2, keyword3

Corresponding e-mail address: zhuob@oregonstate.edu

Title of Your Thesis

By
Bin Zhuo

A THESIS

submitted to

Oregon State University

University Honors College

in partial fulfillment of
the requirements for the
degree of

Honors Baccalaureate of Science in Statistics
(Honors Scholar)

Presented Month dd, yyyy
Commencement June 2016

Honors Baccalaureate of Science in Statistics project of Bin Zhuo presented on Month dd, yyyy

APPROVED:

Yanming Di, Mentor, representing Department of Statistics

Committee Member Name, Committee Member, representing Committee Member Department

---

Committee Member Name, Committee Member, representing Committee Member Department

---

Toni Doolen, Dean, Oregon State University Honors College

I understand that my project will become part of the permanent collection of Oregon State University Honors College. My signature below authorizes release of my project to any reader upon request.

Bin Zhuo, Author

Contents

1 Introduction

1.1 Biological question of interest

Some biology here.

1.2 Generalized Linear Mixed Models

1.2.1 Classical Linear Models

In a classical linear model, a vector $\boldsymbol{y}$ of n observations is assumed to be a realization of random variable $\boldsymbol{Y}$ whose components are identically distributed with mean $\boldsymbol{\mu}$. The systematic part of this model is a specification of the mean μ over a few unknown parameters (McCullagh and Nelder, 1989). In the context of classical linear model, the mean is a function of p covariates $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_p$

$$\boldsymbol{\mu} = \beta_0 + \sum_{i=1}^{p} \beta_i \boldsymbol{X}_i \tag{1}$$

where β's are unknown parameters and need to be estimated from data. For jth[1] component Y_j with a random term allowing for measurement error ϵ , the model can be expressed as

$$Y_j = \beta_0 + \beta_1 x_{1j} + \ldots + \beta_p x_{pj} + \epsilon_j \tag{2}$$

It is often required that ϵ_i's meet Gauss-Markov assumption, i.e, $E(\epsilon_i) = 0$, $\text{Var}[\epsilon_i] = \sigma^2 < \infty$ and $\text{Cov}[\epsilon_i, \epsilon_j] = 0, \forall i \neq j$. In practice, the error term is frequently, if not always, assumed to be normally distributed, i.e. $\boldsymbol{\epsilon} \sim N(0, \sigma^2 \boldsymbol{I})$.

1.2.2 Linear Mixed Models

The Gauss-Markov assumption is vulnerable in real-world problems, for example, nonconstant variance, or correlated data where $\text{Cov}[\epsilon_i, \epsilon_j] \neq 0$. Without loss of generality, 2 in either case can be expressed as

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, E[\boldsymbol{\epsilon}] = \boldsymbol{0}, \text{Cov}[\boldsymbol{\epsilon}] = \sigma^2 \boldsymbol{V} \tag{3}$$

where $\boldsymbol{V}$ is a known positive definite matrix. Let $\boldsymbol{Y}^* = \boldsymbol{V}^{-1/2}\boldsymbol{Y} = \boldsymbol{V}^{-1/2}\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{V}^{-1/2}\boldsymbol{\epsilon}$. It follows that $\text{Cov}(\boldsymbol{Y}^*) = \sigma^2 \boldsymbol{I}$ and the techniques in classical linear models are readily used to estimate $\boldsymbol{\beta}$. However, this method relies on the assumption that $\boldsymbol{V}$ is known which is rarely, if ever, given. On the other hand, the strucure of $\boldsymbol{V}$, which depends on experiment setup, can often be specified by a few unknown parameters.

[1] Unless specified otherwise, we assume there are n observations, i.e. $j = 1, \ldots, n$

Nonindependence can occur in the form of serial correlation or cluster correlation (Rencher and Schaalje, 2008, chapter 17). Serial correlation is usually seen in experiments where multiple measurements are taken from a response variable on the same experimental unit (a.k.a. repeated measurements). Several covariance structures are available for implementation, and interested reader is referred to (Littell et al., 2006, chapter 5) for further information. Cluster correlation is present when measurements of a response variable are grouped in various ways. In many situations, the covariance of cluster correlated data can be specified using an extension of standard linear model by

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{Z_1 u_1} + \cdots + \boldsymbol{Z_q u_q} + \boldsymbol{\epsilon} \tag{4}$$

Equation (4) differs from the matrix form of (2) only in the $\boldsymbol{Z_i u_i}$ terms, which is the key part of linear mixed models. The $\boldsymbol{Z}_i$ are known $n \times p_i$ full rank matrices, usually used to specify membership of predictors in various subgroups. The most important innovation in this model is that instead of estimating $\boldsymbol{u}_i$'s as fixed parameters, they are assumed to be unknown random quantities. Similar to the property of $\boldsymbol{\epsilon}$, we assume $E[\boldsymbol{u}_i] = 0$, $\text{Cov}[\boldsymbol{u}_i] = \sigma_i^2 \boldsymbol{I}_{p_i}$ for $i = 1, \ldots, q$. It is in many cases reasonable to require that $\boldsymbol{u}_i$ are mutually independent, and that $\boldsymbol{u}_i$ is independent of $\boldsymbol{\epsilon}$ for $i = 1, \ldots, q$. If we further impose normal distribution on the random terms and errors, then (4) can be casted in a Bayesian framework.

$$\begin{aligned} \boldsymbol{y}|\boldsymbol{u}_1, \ldots, \boldsymbol{u}_q &\sim N_n(\boldsymbol{X\beta} + \sum_{i=1}^{q} \boldsymbol{Z_i u_i}, \sigma^2 \boldsymbol{I}_n), \\ \boldsymbol{u}_i &\sim N_{p_i}(0, \sigma_i^2 \boldsymbol{I}_{p_i}) \end{aligned} \tag{5}$$

The modeling issues are: (a) estimation of variance components σ_i^2 and σ^2; (b) estimation of random effects u_i if needed. For the variance component estimation, there are primarily three approaches: (i) procedures based on expected mean squares from analysis of variance (ANOVA); (ii) maximum likelihood (ML); and (iii) restricted/residual maximum likelihood (REML). For more details, see (Littell et al., 2006, Chapter 1)

1.2.3 Generalized Linear Mixed Models

Another perspective of classical linear models is that they can be arranged into three parts (McCullagh and Nelder, 1989, Chapter 2) .

1. the *random component* $\boldsymbol{Y}$ have certain distribution (usually Gaussian) with $E[\boldsymbol{Y}] = \boldsymbol{\mu}$.

2. the *systematic component*. The linear predictor $\boldsymbol{\eta}$ is formed by covariates $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p$

$$\boldsymbol{\eta} = \sum_{i=1}^{p} \beta_i \boldsymbol{x}_i = \boldsymbol{X\beta} \tag{6}$$

3. the *link* relates the random components and the systematic components by

$$\boldsymbol{\eta} = \boldsymbol{\mu} \tag{7}$$

A generalized linear model (GLM) allows two extensions. Firstly, the distribution in part 1 may come from another distribution (for example, Poisson, Gamma or Binomial). Secondly, in (7) $\boldsymbol{\eta}$ can relate to $\boldsymbol{\mu}$ by a monotonic function $\boldsymbol{\eta} = g(\boldsymbol{\mu})$. In this setting, classical linear model is a special case since it has normal random variables in part 1 and identity link in (7). Generalized linear mixed model (GLMM) is a natural generalization of GLM that further extends (6) to allow random effect, casted in a matrix notation

$$\boldsymbol{\eta} = \boldsymbol{X\beta} + \sum_{i=1}^{q} \boldsymbol{Z_i u_i} + \boldsymbol{\epsilon} \tag{8}$$

where $\boldsymbol{Z}_i, \boldsymbol{u}_i$ and $\boldsymbol{\epsilon}$ are specified in (4).

1.3 An example of Poisson Regression with Random Effects

As an example, suppose now we have RNA-Seq gene expression data from 3 randomly selected experiments, with 2 random treatments nested in each experiment and 2 replicates in each treatment. For a single gene, let $Y_{jkl} \sim \text{Poisson}(\mu_{jkl})$ be the read count for jth lab kth treatment and lth observation unit. The linear predictor η_{jkl} relates mean μ_{jkl} by (8).

$$\eta_{jkl} = \beta_0 + a_j + b_{k(j)} + \epsilon_{jkl} \tag{9}$$

where $j = 1, \ldots, 3$, $k = 1, 2$ and $l = 1, 2$; $a_j \sim N(0, \sigma_1^2), b_{k(j)} \sim N(0, \sigma_2^2)$ and $\epsilon_{jkl} \sim N(0, \sigma_0^2)$ are mutually independent random effects. If the data are sorted by experiment and treatment nested in experiment, then the model can be casted in the

form of Equation 8 with

$$q = 2, \boldsymbol{X} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \boldsymbol{Z}_1 = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \boldsymbol{Z}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Then

$$\boldsymbol{\Sigma} = \sigma_1^2 \boldsymbol{Z_1} \boldsymbol{Z_1'} + \sigma_2^2 \boldsymbol{Z_2} \boldsymbol{Z_2'} + \sigma_0^2 \boldsymbol{I}_{12} = \begin{bmatrix} \boldsymbol{\Sigma}_d & \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{\Sigma}_d & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{\Sigma}_d \end{bmatrix}$$

where $\boldsymbol{O}$ is a 4×4 matrix of 0 and

$$\boldsymbol{\Sigma}_d = \begin{bmatrix} \sigma_1^2 + \sigma_2^2 + \sigma_0^2 & \sigma_1^2 + \sigma_2^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 + \sigma_2^2 & \sigma_1^2 + \sigma_2^2 + \sigma_0^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2 + \sigma_2^2 + \sigma_3^2 & \sigma_1^2 + \sigma_2^2 \\ \sigma_1^2 & \sigma_2^2 & \sigma_1^2 + \sigma_2^2 & \sigma_1^2 + \sigma_2^2 + \sigma_0^2 \end{bmatrix}$$

The challenge due to the complexity of GLMM is the estimation of parameters. Next section summarizes current available methods in estimating variance components.

1.4 Estimation

There are three general approaches for estimating parameters under GLMM settings (Myers et al., 2012, Chapter 7): (i) using numerical method to approximate the integrals for the likelihood functions and obtaining the estimating equations; (ii) linearization of the conditional mean and then iteratively applying linear mixed model techniques to the approximated model; (iii) Bayesian approach.

In the following discussion, we assume conditional distribution of $\boldsymbol{Y}$ given $\boldsymbol{u}$ is $f_Y(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{u})$, the link function is $\boldsymbol{\eta} = g(\boldsymbol{\mu})$, and $\boldsymbol{\eta}$ relates the covariates by (8). We also assume random effect $\boldsymbol{u}$ to have some distribution $\boldsymbol{u} \sim \phi(\boldsymbol{u}|\boldsymbol{\Sigma})$.

1.4.1 Likelihood Function Approach

It is straightforward to write down the likelihood function of $\boldsymbol{Y}$.

$$L(\boldsymbol{Y}|\boldsymbol{\beta}, \boldsymbol{\Sigma}) = \int f(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{u})\phi(\boldsymbol{u}|\boldsymbol{\Sigma})d\boldsymbol{u} \tag{10}$$

A major challenge in estimating GLMM models is the integration of (10) over the n-dimensional distribution of $\boldsymbol{u}$. Numerical approximation are usually used in evaluating the integral. In this part we will only discuss the *Gauss-Hermite* (GH) quadrature that is recognized as a higher order Laplace approximation (Liu and Pierce, 1994).

Gauss-Hermite quadrature is defined in terms of integral of the form

$$\int_{-\infty}^{\infty} f(x)e^{-x^2}dx \tag{11}$$

The integral (11) is approximated by a weighted sum of $f(x)$

$$\int_{-\infty}^{\infty} f(x)e^{-x^2}dx \approx \sum_{i=1}^{m} w_i f(x_i) \tag{12}$$

where x_i are the zeros of mth order Hermite polynomial and w_i are corresponding weights. (12) gives the exact numerical value for all polynomials up to degree of $2m-1$. For a Hermite polynomial of degree n, x_i and w_i can be calculated as

$$x_i = i\text{th zero of } H_n(x), \quad w_i = \frac{2^{n-1}n!\sqrt{\pi}}{n^2[H_{n-1}(x_i)]^2} \tag{13}$$

An improved version of the regular Gauss-Hermite quadrature is to center and scale the quadrature points by the empirical Bayes estimate of the random effects and the Hessian matrix from the Bayes estimate suboptimization (Liu and Pierce, 1994). This procedure is called *Adaptive Gauss-Hermite* (AGH) quadrature (Pinheiro and Bates, 1995).

The AGH quadrature starts by maximizing the integrand $h(\boldsymbol{u}|\boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\Sigma}) = f(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{u})\phi(\boldsymbol{u}|\boldsymbol{\Sigma})$ in equation (10) with respect to random effects $\boldsymbol{u}$. The resulting estimate $\hat{\boldsymbol{u}}^{(n)}$ is the joint posterior modes for the random effects. Because $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ are unknown, they are replaced by the current estimates $\hat{\boldsymbol{\beta}}^{(n)}$ and $\hat{\boldsymbol{\Sigma}}^{(n)}$ at iteration n. The Hessian matrix $\hat{\boldsymbol{H}}^{(n)}$ can be obtained by evaluating the second order partial derivatives of $\log(h(\boldsymbol{u}|\boldsymbol{y}, \hat{\boldsymbol{\beta}}^{(n)}, \hat{\boldsymbol{\Sigma}}^{(n)}))$ at $\hat{\boldsymbol{u}}^{(n)}$. Consequently, $\hat{\boldsymbol{\Omega}}^{(n)} = -\hat{\boldsymbol{H}}^{(n)}$ is the estimated covariance matrix for the random effects posterior modes. It follows from equation (10) that for the ith cluster

$$L(\boldsymbol{Y}_i|\boldsymbol{\beta}, \boldsymbol{\Sigma}) = \int f(\boldsymbol{y}_i|\boldsymbol{\beta}, \boldsymbol{u})\phi(\boldsymbol{u}|\boldsymbol{\Sigma})d\boldsymbol{u} = \int \frac{f(\boldsymbol{y}_i|\boldsymbol{\beta}, \boldsymbol{u})\phi(\boldsymbol{u}|\boldsymbol{\Sigma})}{\phi(\boldsymbol{u}|\hat{\boldsymbol{u}}^{(n)}, \hat{\boldsymbol{\Omega}}^{(n)})}\phi(\boldsymbol{u}|\hat{\boldsymbol{u}}^{(n)}, \hat{\boldsymbol{\Omega}}^{(n)})d\boldsymbol{u} \tag{14}$$

[copied from SAS help] Let m be the number of quadrature points in each dimension (for each random effect) and Q the number of random effects. If $\boldsymbol{x} = (x_1, \ldots, x_m)$ are the nodes for standard Gauss-Hermite quadrature, and $\boldsymbol{x}_j^* = (x_{j_1}, \ldots, x_{j_Q})$ is a point on the Q dimensional quadrature grid, then the centered and scaled nodes are

$$\boldsymbol{a}_j^* = \hat{\boldsymbol{u}}^{(n)} + \sqrt{2}[\hat{\boldsymbol{\Omega}}^{(n)}]^{1/2}\boldsymbol{x}_j^* \tag{15}$$

The centered and scaled nodes, along with the Gauss-Hermite quadrature weights $\boldsymbol{w} = (w_1, \ldots, w_m)$ are used to construct the Q dimensional integral (14), approximated by

$$\begin{aligned} L(\boldsymbol{y}_i|\boldsymbol{\beta}, \boldsymbol{\Sigma}) &\approx \sum_{j_1=1}^{m} \cdots \sum_{j_Q=1}^{m} \frac{f(\boldsymbol{y}_i|\boldsymbol{\beta}, \boldsymbol{a}_j^*)\phi(\boldsymbol{a}_j^*|\boldsymbol{\Sigma})}{\phi(\boldsymbol{a}_j^*|\hat{\boldsymbol{u}}^{(n)}, \hat{\boldsymbol{\Omega}}^{(n)})} w_{j_1} \cdots w_{j_Q} \\ &= (2)^{Q/2}|\hat{\boldsymbol{\Omega}}^{(n)}|^{1/2} \sum_{j_1=1}^{m} \cdots \sum_{j_Q=1}^{m} \left[f(\boldsymbol{y}_i|\boldsymbol{\beta}, \boldsymbol{a}_j^*)\phi(\boldsymbol{a}_j^*|\boldsymbol{\Sigma}) \prod_{k=1}^{Q} w_{jk} \exp(x_{jk}^2) \right] \end{aligned} \tag{16}$$

Thus the multidimensional unbounded integrals are approximated by a finite summations. Now that the likelihood has the form of (16), a number of methods (e.g. Newton-Raphson or Fisher's scoring) can be used to estimate $(\boldsymbol{\beta}, \boldsymbol{\Sigma})$.

It should be noted, however, as the number of dimension Q increases, the computation for (16) grows exponentially since the total number of nodes is m^Q. Therefore it is difficult to implement AGH procedure with more than three random effects (Bolker et al., 2009).

1.4.2 Estimation based on Linearization

Linearization methods employ expansions to approximate the model by one based on pseudo-data with fewer nonlinear components. The generalized linear mixed model is approximated by a linear mixed model based on current values of the covariance parameter estimates. The resulting linear mixed model is then fit, which is itself an iterative process. The process of computing the linear approximation must be repeated several times until some criterion stabilizes. On convergence, the new parameter estimates are used to update the linearization, which results in a new linear mixed model.

Under GLMM framework, we have some conditional distribution of $\boldsymbol{Y}$ given $\boldsymbol{u}$. Without loss of generality, we assume

$$\begin{aligned} E[\boldsymbol{Y}|\boldsymbol{u}] = \boldsymbol{\mu} = g^{-1}(\boldsymbol{\eta}) = g^{-1}(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u}), \\ \text{Var}[\boldsymbol{Y}|\boldsymbol{u}] = \boldsymbol{S} \end{aligned} \tag{17}$$

where $\boldsymbol{u} \sim N(\boldsymbol{0}, \boldsymbol{D})$. The linearization is done by Taylor expansion of (17) about estimates $\boldsymbol{\eta}$. The *Penalized Quasi-likelihood* (PQL) or *Marginal Quasi-likelihood*

(MQL) estimate procedure developed by Breslow and Clayton (1993) may be used for this purpose.

1.4.3 Penalized Quasi-likelihood

The PQL procedure uses a first order Taylor expansion of $\boldsymbol{\beta}$ and $\boldsymbol{u}$, at $\tilde{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{u}}$, respectively

$$g^{-1}(\boldsymbol{\eta}) \approx g^{-1}(\hat{\boldsymbol{\eta}}) + \tilde{\boldsymbol{\Omega}}_P(\boldsymbol{\eta} - \tilde{\boldsymbol{\eta}}) \tag{18}$$

where $\tilde{\boldsymbol{\Omega}}_P$ is an $n \times n$ diagonal matrix whose (i,i) entry is $\partial g^{-1}(\boldsymbol{\eta}_i)/\partial \boldsymbol{\eta}_i$ evaluated at $\tilde{\boldsymbol{\eta}} = \boldsymbol{X}\tilde{\boldsymbol{\beta}} + \boldsymbol{Z}\tilde{\boldsymbol{u}}$. Multiplying both sides by $\tilde{\Omega}_P^{-1}$ and (18) can be rearranged as

$$\boldsymbol{X\beta} + \boldsymbol{Zu} \approx \tilde{\boldsymbol{\Omega}}_P^{-1}[g^{-1}(\boldsymbol{\eta}) - g^{-1}(\tilde{\boldsymbol{\eta}})] + \boldsymbol{X}\tilde{\boldsymbol{\beta}} + \boldsymbol{Z}\tilde{\boldsymbol{u}} \tag{19}$$

Note that the right hand side of (19) is just the expected value, given $\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{u}}$, of pseudo-response

$$\tilde{\boldsymbol{Y}} = \tilde{\boldsymbol{\Omega}}_P^{-1}[\boldsymbol{Y} - g^{-1}(\tilde{\boldsymbol{\eta}})] + \boldsymbol{X}\tilde{\boldsymbol{\beta}} + \boldsymbol{Z}\tilde{\boldsymbol{u}} \tag{20}$$

whose variance-covariance matrix given $\boldsymbol{u}$ is

$$\mathrm{Var}[\tilde{\boldsymbol{Y}}|\boldsymbol{u}] = \tilde{\boldsymbol{\Omega}}_P^{-1}\mathrm{Var}[\boldsymbol{Y}|\boldsymbol{u}]\tilde{\boldsymbol{\Omega}}_P^{-1} = \tilde{\boldsymbol{\Omega}}_P^{-1}\boldsymbol{S}\tilde{\boldsymbol{\Omega}}_P^{-1} \tag{21}$$

Then we can consider the model

$$\tilde{\boldsymbol{Y}} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{\epsilon} \tag{22}$$

which is a linear mixed model with pseudo response $\tilde{\boldsymbol{Y}}$ with covariance matrix

$$\boldsymbol{W} = \mathrm{Var}[\tilde{\boldsymbol{Y}}|\boldsymbol{u}] = \boldsymbol{ZDZ'} + \tilde{\boldsymbol{\Omega}}_P^{-1}\boldsymbol{S}\tilde{\boldsymbol{\Omega}}_P^{-1}. \tag{23}$$

Model (22) has exactly the same form as linear mixed model, except that an estimate of $(\boldsymbol{\beta}, \boldsymbol{u})$ is needed for calculating pseudo-response $\tilde{\boldsymbol{Y}}$. An iterative procedure can be used to estimate(22) by substituting raw data $\boldsymbol{y}$ for $\tilde{\boldsymbol{y}}$ and identity matrix $\boldsymbol{I}$ for $\boldsymbol{S}$ as starting values. Techniques for fitting LMM such as *restricted maximum likelihood* (REML) can be readily applied to estimate variance components $\boldsymbol{D}$, upon which $\hat{W}$ is calculated. The estimate for $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\hat{\boldsymbol{W}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\hat{\boldsymbol{W}}^{-1}\boldsymbol{X}\tilde{\boldsymbol{y}}, \tag{24}$$

and the estimate for random effect is

$$\hat{\boldsymbol{u}} = \hat{\boldsymbol{D}}Z\hat{\boldsymbol{W}}^{-1}(\tilde{\boldsymbol{y}} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) \tag{25}$$

Then the pseudo-response is updated and the procedure is repeated until convergence is reached for fixed effects and variance components. Note that (25) estimates a vector of random effect. For this reason, PQL is also referred to as *subject-specific* estimate

procedure.

1.4.4 Marginal Quasi-likelihood

One of the motivation for MQL is that usually one is more interested in estimating the marginal mean of the response than estimating the conditional mean as was done by (25) in PQL. Since $E[\boldsymbol{\eta}|\boldsymbol{u}] = \boldsymbol{X\beta} + \boldsymbol{Zu}$, the unconditional mean is $E[\boldsymbol{\eta}] = E[E(\boldsymbol{\eta}|\boldsymbol{u})] = \boldsymbol{X\beta}$. A first-order Taylor expansion of $E[\boldsymbol{Y}|\boldsymbol{u}]$ about $\boldsymbol{X\beta}$ is given by

$$E[\boldsymbol{Y}|\boldsymbol{u}] = g^{-1}(\boldsymbol{\eta}) \approx g^{-1}(\boldsymbol{X\beta}) + \tilde{\Omega}_M(\boldsymbol{\eta} - \boldsymbol{X\beta}) \tag{26}$$

where $\tilde{\Omega}_M$ is evaluated at $\boldsymbol{X\beta}$ (recall that for PQL, $\tilde{\Omega}_P$ is evaluated at $\boldsymbol{X\beta}+\boldsymbol{Zu}$). The unconditional expected value of $\boldsymbol{Y}$ is approximately $g^{-1}(\boldsymbol{X\beta})$ by (26). The variance of $\boldsymbol{Y}$ can then be derived from the relation $\mathrm{Var}(\boldsymbol{Y}) = E[\mathrm{Var}(\boldsymbol{Y}|\boldsymbol{u})] + \mathrm{Var}[E(\boldsymbol{Y}|\boldsymbol{u})]$, which yields

$$\mathrm{Var}[\boldsymbol{Y}] = \tilde{\Omega}_P \boldsymbol{ZDZ'}\tilde{\Omega}'_P + S_{\boldsymbol{\eta}_0} \tag{27}$$

A linearization is performed at $\boldsymbol{\eta}_0 = \boldsymbol{X\beta}_0$,

$$g^{-1}(\boldsymbol{\eta}) \approx g^{-1}(\boldsymbol{X\beta_0}) + \tilde{\Omega}_M(\boldsymbol{\eta} - \boldsymbol{X\beta}_0)$$

Multiplying both sides by $\tilde{\Omega}_M^{-1}$, it then can be arranged to

$$\boldsymbol{X\beta} + \boldsymbol{Zu} \approx \tilde{\Omega}_M^{-1}[g^{-1}(\boldsymbol{\eta}) - g^{-1}(\boldsymbol{\eta}_0)] + \boldsymbol{X\beta}_0$$

The pseudo-response is defined as

$$\tilde{\boldsymbol{Y}}_M = \tilde{\Omega}_M^{-1}[\boldsymbol{Y} - g^{-1}(\boldsymbol{\eta}_0)] + \boldsymbol{X\beta}_0 \tag{28}$$

Next we consider the linear mixed model

$$\tilde{\boldsymbol{Y}}_M = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{\epsilon}$$

where $\mathrm{Var}(\boldsymbol{\epsilon})$ is given by (27). The estimate for fixed effect parameters $\boldsymbol{\beta}$ and variance components is the same as those in PQL.

Note that the pseudo-response is not a function of $\boldsymbol{u}$ any more, so updating this quantity does not require calculating the random effects $\boldsymbol{u}$. MQL is also referred to as *population-averaged* estimate approach.

Pinheiro and Chao (2006) and Breslow and Lin (1995) showed that PQL approach may lead to asymptotically biased estimates and hence to inconsistency. It is not recommended to use simple PQL method in practice.

1.4.5 Bayes Approach

As discussed earlier, for models with higher dimensional integrals, it is not practical to evaluate the likelihood function by AGH procedure. Here we will describe the MCEM algorithm for For mixed models, a typical strategy is to treat the random effects to be missing data. Following this rationale, the the problem of estimating variance components associated with random effects can be simplified. Denote the *complete data* as $\boldsymbol{v} = (\boldsymbol{y}, \boldsymbol{u})$, the log-likelihood of $\boldsymbol{v}$ can be expressed as

$$\log \pi(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\boldsymbol{v}) = \log f(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{u}) + \log \phi(\boldsymbol{u}|\boldsymbol{\Sigma}) \tag{29}$$

The optimal solution in (29) can be obtained by *Expectation-Maximization* (EM) algorithm that can be readily implemented as follows:
E-Step. At $(k+1)$th iteration with $\boldsymbol{\beta}^{(k)}$ and $\boldsymbol{\Sigma}^{(k)}$ calculate

$$\begin{aligned} E_{\boldsymbol{\beta}^{(k)}}[\log f(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\boldsymbol{v})|\boldsymbol{y}] &= Q_1(\boldsymbol{\beta}, \boldsymbol{\beta}^{(k)}), \\ E_{\boldsymbol{\Sigma}^{(k)}}[\log \phi(\boldsymbol{\Sigma}|\boldsymbol{v})|\boldsymbol{y}] &= Q_2(\boldsymbol{\Sigma}, \boldsymbol{\Sigma}^{(k)}) \end{aligned} \tag{30}$$

M-Step. Maximize Q_1 and Q_2 to update $\boldsymbol{\beta}^{(k+1)}$ and $\boldsymbol{\Sigma}^{(k+1)}$.
The **E** and **M** steps are alternated until convergence. Unfortunately, the expectations in (30) cannot be computed in closed form for GLMMs. However, they may be approximated by Markov chain Monte Carlo (MCMC). In light of this, McCulloch (1997) developed a Monte Carlo EM (MCEM) algorithm. The Metropolis-Hastings algorithm is used for drawing samples from difficult-to-calculate density functions.

To illustrate Metropolis algorithm, a proposal distribution $g(\boldsymbol{u})$ is selected, from which an initial value of $\boldsymbol{u}$ is drawn. The new candidate value $\boldsymbol{u}' = (u_1, u_2, \ldots, u_{k-1}, u_k', u_{k+1}, \ldots, u_Q)$, which has all elements the same as previous values expect the kth, is accepted (as opposed to keeping the previous value) with probability

$$A_k(\boldsymbol{u}', \boldsymbol{u}) = \min\left\{1, \frac{f(\boldsymbol{u}'|\boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\Sigma})g(\boldsymbol{u})}{f(\boldsymbol{u}|\boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\Sigma})g(\boldsymbol{u}')}\right\} \tag{31}$$

If we choose $g(\boldsymbol{u}) = \phi(\boldsymbol{u}|\boldsymbol{\Sigma})$, then the ratio term in (31) can be simplified to

$$\begin{aligned} &\frac{f(\boldsymbol{u}'|\boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\Sigma})g(\boldsymbol{u})}{f(\boldsymbol{u}|\boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\Sigma})g(\boldsymbol{u}')} \\ &= \left[\frac{f(\boldsymbol{u}', \boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\Sigma})}{f(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\Sigma})}\phi(\boldsymbol{u}|\boldsymbol{\Sigma})\right] / \left[\frac{f(\boldsymbol{u}, \boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\Sigma})}{f(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\Sigma})}\phi(\boldsymbol{u}'|\boldsymbol{\Sigma})\right] \\ &= \frac{f(\boldsymbol{y}|\boldsymbol{u}', \boldsymbol{\beta}, \boldsymbol{\Sigma})\phi(\boldsymbol{u}'|\boldsymbol{\Sigma})\phi(\boldsymbol{u}|\boldsymbol{\Sigma})}{f(\boldsymbol{y}|\boldsymbol{u}, \boldsymbol{\beta}, \boldsymbol{\Sigma})\phi(\boldsymbol{u}|\boldsymbol{\Sigma})\phi(\boldsymbol{u}'|\boldsymbol{\Sigma})} \\ &= \frac{f(\boldsymbol{y}|\boldsymbol{u}', \boldsymbol{\beta}, \boldsymbol{\Sigma})}{f(\boldsymbol{y}|\boldsymbol{u}, \boldsymbol{\beta}, \boldsymbol{\Sigma})} \end{aligned} \tag{32}$$

The MCEM procedure combines the EM steps and Metropolis algorithm in estimating the fixed parameters and variance components as follows:

step 1 Choose the starting value of $\boldsymbol{\beta}^{(0)}, \boldsymbol{\Sigma}^{(0)}$. Set $b = 0$

step 2 Generate the sequence $\boldsymbol{u}^{(1)}, \boldsymbol{u}^{(2)}, \ldots, \boldsymbol{u}^{(B)}$ from the conditional distribution of $\boldsymbol{u}$ given y with Metropolis algorithm.

step 3 Maximize $\sum_{b=1}^{B} \log f(\boldsymbol{y}|\boldsymbol{u}^{(b)}, \boldsymbol{\beta})/B$ and $\sum_{b=1}^{B} \log \phi(\boldsymbol{u}^{(b)}|\boldsymbol{\Sigma})/B$ to obtain $\boldsymbol{\beta}^{(m+1)}$ and $\boldsymbol{\Sigma}^{(m+1)}$

step 4 Iterate between step 2 and step 3 until convergence is reached.

This method can be easily extended to allow for multiple random effects. But the advantage comes at a price. A major drawback of MCEM is the computational intensity. First, the convergence of *EM* algorithm is usually very slow, especially at the neighborhood of maximum of marginal likelihood. Second, the chain in Metropolis algorithm has to run long enough for reliable estimation.

In the Bayes framework, there are some alternatives for estimation. Interested readers are referred to *Monte Carlo Newton-Raphson* (MCNR, McCulloch (1997)), *MCMC* (Hadfield et al., 2010).

1.5 Multiple Hypothesis Testing

Multiple hypothesis testing procedures deal with type I error rates in a family of tests. The problems arise when we consider a set of statistical inference simultaneously. For each of the individual tests or confidence intervals, there is a type I error which can be controlled by the experimenter. If the family of tests contains one or more true null hypotheses, the probability of rejecting one or more of these true null increases.

While traditional multiple testing procedures focus on modest number of tests, a different set of techniques are needed for large-scale inference, in which tens or even hundreds of thousands of tests are performed simultaneously. For example, in genomics study, expression levels of 50,000 genes for each of 100 individuals can be measured using modern technologies such as microarray or RNA-Sequencing. In testing differential expression (DE), 50,000 tests need to be conducted against the null that there is no DE between treatment/control. This has brought new challenge to the field of multiple hypothesis testing. Benjamini and Hochberg (1995) points out that the control of familywise error rate (FWER), i.e. the probability of making one or more false discovery in a set of tests, tends to have substantially less power.

False discovery rate (FDR), introduced by Benjamini and Hochberg (1995), is the expected proportion of false positives among all significant calls (null rejected). FDR has been studied extensively (Benjamini and Yekutieli (2001), Storey and Tibshirani (2003), Efron (2004), Efron (2010) and more) over the past two decades. FDR is

equivalent to FWER (Benjamini and Hochberg, 1995) when all hypotheses are true but smaller if there are some true discoveries to be made. We will focus our attention on FDR in this part.

Let m, m_0 and m_1 be the number of tests, true nulls and true alternatives respectively. Let also F and T be the number of true nulls and true alternatives among S tests that are declared as significant. Table (1.5) shows the relation among them. The FDR is

	Called significance	Called not significant	Total
Null True	F	$m_0 - F$	m_0
Alternative true	T	$m_1 - T$	m_1
total	S	$m - S$	m

1.6 Disertation Objective

2 Identification of stably expressed genes

Section text.

3 gene set enrichment analysis

4 Conclusion

Conclusion text.

References

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 289–300.

Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of statistics*, pages 1165–1188.

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., and White, J.-S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in ecology & evolution*, 24(3):127–135.

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88(421):9–25.

Breslow, N. E. and Lin, X. (1995). Bias correction in generalised linear mixed models with a single component of dispersion. *Biometrika*, 82(1):81–91.

Efron, B. (2004). Large-scale simultaneous hypothesis testing. *Journal of the American Statistical Association*, 99(465).

Efron, B. (2010). *Large-scale inference: empirical Bayes methods for estimation, testing, and prediction*, volume 1. Cambridge University Press.

Hadfield, J. D. et al. (2010). Mcmc methods for multi-response generalized linear mixed models: the mcmcglmm r package. *Journal of Statistical Software*, 33(2):1–22.

Littell, R. C., Stroup, W. W., Milliken, G. A., Wolfinger, R. D., and Schabenberger, O. (2006). *SAS for mixed models*. SAS institute.

Liu, Q. and Pierce, D. A. (1994). A note on gauss—hermite quadrature. *Biometrika*, 81(3):624–629.

McCullagh, P. and Nelder, J. A. (1989). Generalized linear models.

McCulloch, C. E. (1997). Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American statistical Association*, 92(437):162–170.

Myers, R. H., Montgomery, D. C., Vining, G. G., and Robinson, T. J. (2012). *Generalized linear models: with applications in engineering and the sciences*, volume 791. John Wiley & Sons.

Pinheiro, J. C. and Bates, D. M. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of computational and Graphical Statistics*, 4(1):12–35.

Pinheiro, J. C. and Chao, E. C. (2006). Efficient laplacian and adaptive gaussian quadrature algorithms for multilevel generalized linear mixed models. *Journal of Computational and Graphical Statistics*, 15(1).

Rencher, A. C. and Schaalje, G. B. (2008). *Linear models in statistics*. John Wiley & Sons.

Storey, J. D. and Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences*, 100(16):9440–9445.