Variance Heterogeneity in Genetic Mapping

Robert Wallace Corty

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Computer Science.

Chapel Hill
2018

Approved by:

Fernando Pardo Manuel de Villena, Ph.D., chair

James Evans, M.D., Ph.D.

Yun Li, Ph.D.

Lisa Tarantino, Ph.D.

William Valdar, Ph.D.

ABSTRACT

Robert Wallace Corty: Variance Heterogeneity in Genetic Mapping
(Under the direction of William Valdar)

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

Dedication…

ACKNOWLEDGEMENTS

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue duis dolore te feugait nulla facilisi. Lorem ipsum dolor sit amet,

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

QTL	quantitative trait locus
mQTL	mean-controlling quantitative trait locucs
vQTL	variance-controlling quantitative trait locucs
mvQTL	mean or variance controlling quantitative trait locucs
LMM	Linear Mixed Model
GLS	Generalized Least Squares
SLM	Standard Linear Model
DGLM	Double Generalized Linear Model
AYO	Add Your Own in alphabetic order...

CHAPTER 1

Introduction

	accommodate variance heterogeneity	detect variance heterogeneity	handle differential relatedness
LM	no	no	no
DGLM	yes	yes	no
LMM	no	no	yes
wLMM	yes	no	yes

CHAPTER 2

Genetic Mapping in Experimental Crosses with the Double Generalized Linear Model

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

CHAPTER 3

Genetic Mapping in Human Cohorts with the Double Generalized Linear Model

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

CHAPTER 4

The Heteroscedastic Linear Mixed Model

Whereas the previous chapters have dealt with differential relatedness obliquely, either assuming its absence or using a heuristic to correct for it, this chapter addresses it directly. The linear mixed model (LMM) [posulates] an observed phenotype, $\boldsymbol{y}$, as the sum of four unobserved, but estimated quantities: the population mean, $\boldsymbol{\mu}$, the fixed effect deviation, $\mathbf{X}\boldsymbol{\beta}$, the genomic value, $\mathbf{a}$, and the residual deviation, $\mathbf{e}$.

$$\boldsymbol{y} = \boldsymbol{\mu} + \mathbf{X}\boldsymbol{\beta} + \mathbf{a} + \mathbf{e} \tag{4.1}$$

with

$$\mathbf{a} \sim \mathrm{N}(0, \mathbf{K}\tau^2), \tag{4.2}$$

and

$$\mathbf{e} \sim \mathrm{N}(0, \mathbf{D}\sigma^2) \tag{4.3}$$

where $\mathbf{X}$ is a matrix of covariates, $\mathbf{K}$ is a known positive semi-definite genomic similarity matrix, and $\mathbf{D}$ is a known diagonal residual variance matrix. Of the estimated parameters, $\boldsymbol{\beta}$ and $\boldsymbol{\mu}$ are unconstrained, τ^2 and σ^2 are constrained to be non-negative, and $\mathbf{a}$ and $\mathbf{e}$ are constrained by their hierarchical specification (Equations 4.2 and 4.3).

This model can be useful in genetic mapping because In the context of genetic mapping, the goal is to fit this model for many different values of $\mathbf{X}$... and compare to one with no genetic in $\mathbf{X}$ to do LRT. The default way to do so would be to use Henderson's method each time, but this is very slow. A recent development described a matrix math trick to make this process fast when $\mathbf{D} = \mathbf{I}$. I first summarize that trick and then go on to describe a further trick that allows for fast fitting for any diagonal $\mathbf{D}$.

This model is equivalent to:

$$\boldsymbol{y} \sim \mathrm{N}(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}) \tag{4.4}$$

where

$$\boldsymbol{\Sigma} = \mathbf{K}\tau^2 + \mathbf{D}\sigma^2 \tag{4.5}$$

This model can be reparametrized by $h^2 = \frac{\tau^2}{\tau^2+\sigma^2}$ and $\lambda = \tau^2 + \sigma^2$, such that, after defining $\mathbf{V} = \mathbf{K}h^2 + \mathbf{D}(1 - h^2)$, it reduces to a generalized least squares (GLS) model.

$$\begin{aligned}
\boldsymbol{\Sigma} &= \left(\mathbf{K}\frac{\tau^2}{\tau^2+\sigma^2} + \mathbf{D}\frac{\sigma^2}{\tau^2+\sigma^2}\right)(\tau^2+\sigma^2) && (4.6)\\
&= \mathbf{K}h^2 + \mathbf{D}(1-h^2)\lambda && (4.7)\\
&= \mathbf{V}\lambda && (4.8)
\end{aligned}$$

and thus

$$\boldsymbol{y} \sim \mathrm{N}(\boldsymbol{\mu} + \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\lambda) \tag{4.9}$$

There is a known, closed-form, maximum likelihood estimator of $\boldsymbol{\beta}$ for the GLS model.

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}\boldsymbol{y} \tag{4.10}$$

We want to do LRT, so we need to estimate maximum likelihood values for beta, sigma, and tau. Many SNPs, so need to fit it fast.

Imagine we knew of a multiplier matrix, $\mathbf{M}$ such that

$$\mathbf{M}^{\mathrm{T}}\mathbf{M} = \mathbf{V}^{-1} \tag{4.11}$$

By now considering a new phenotype vector, $\boldsymbol{y}_m = \mathbf{M}\boldsymbol{y}$ and a new covariate matrix, $\mathbf{X}_m = \mathbf{M}\mathbf{X}$, we can make progress toward fast fitting of the real phenotype vector and covariate matrix.

The expectation and variance of $\boldsymbol{y}_m$ are:

$$\mathrm{E}(\boldsymbol{y}_m) = \mathrm{E}(\mathbf{M}\boldsymbol{y}) \tag{4.12}$$

$$= \mathbf{M}\mathrm{E}(\boldsymbol{y}) \tag{4.13}$$

$$\mathrm{Var}(\boldsymbol{y}_m) = \mathrm{Var}(\mathbf{M}\boldsymbol{y}) \tag{4.14}$$

$$= \mathbf{M}\mathrm{Var}(\boldsymbol{y})\mathbf{M}^{\mathrm{T}} \tag{4.15}$$

Recalling that We are now considering a model of the form:

Note the similarity of their log likelihoods:

$$\ell(\boldsymbol{\beta}, \sigma^2; \mathbf{X}, \boldsymbol{y}, \mathbf{V}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\mathbf{V}\sigma^2| - \frac{1}{2\sigma^2}(\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta})^{\mathrm{T}}\mathbf{V}^{-1}(\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta}) \tag{4.16}$$

$$= -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\mathbf{V}| - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta})^{\mathrm{T}}\mathbf{V}^{-1}(\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta}) \tag{4.17}$$

$$\ell(\boldsymbol{\beta}_m, \sigma_m^2; \mathbf{X}_m, \boldsymbol{y}_m) = \tag{4.18}$$

4.1 Maximum Likelihood Estimation for the Homoscedastic Model

[D = I] Kang proposed the following [cite EMMA].

4.2 Maximum Likelihood Estimation for the Heteroscedastic Model

Want to have a diagonal D, but not necessarily I.

$$\mathbf{M} = (\boldsymbol{\Lambda}_{\mathbf{L}} + \delta\mathbf{I})^{-\frac{1}{2}}\mathbf{U_L}^{\mathrm{T}}\mathbf{D}^{-\frac{1}{2}} \tag{4.19}$$

where

$$\mathbf{L} = \mathbf{D}^{-\frac{1}{2}}\mathbf{K}\mathbf{D}^{-\frac{1}{2}} \tag{4.20}$$

and

$$\mathbf{L} = \mathbf{U_L}\mathbf{\Lambda_L}\mathbf{U_L}^T \tag{4.21}$$

is its eigen decomposition

4.2.1 Validity

To be a valid multiplier matrix, $\mathbf{M}$ must have the property:

$$\mathbf{M}^{\mathrm{T}}\mathbf{M} = \mathbf{V}^{-1}$$

Proof. First, derive a useful form of $\mathbf{V}$.

$$\begin{aligned}
\mathbf{V} &= \mathbf{K} + \delta\mathbf{D} && \text{definition} \\
&= \mathbf{D}^{\frac{1}{2}}\mathbf{D}^{-\frac{1}{2}}(\mathbf{K} + \delta\mathbf{D}) && \text{pre-multiply by } \mathbf{D}^{\frac{1}{2}}\mathbf{D}^{-\frac{1}{2}} = \mathbf{I} \\
&= \mathbf{D}^{\frac{1}{2}}\mathbf{D}^{-\frac{1}{2}}(\mathbf{K} + \delta\mathbf{D})\mathbf{D}^{-\frac{1}{2}}\mathbf{D}^{\frac{1}{2}} && \text{post-multiply by } \mathbf{D}^{-\frac{1}{2}}\mathbf{D}^{\frac{1}{2}} = \mathbf{I} \\
&= \mathbf{D}^{\frac{1}{2}}(\mathbf{D}^{-\frac{1}{2}}\mathbf{K}\mathbf{D}^{-\frac{1}{2}} + \delta\mathbf{D}^{-\frac{1}{2}}\mathbf{D}\mathbf{D}^{-\frac{1}{2}})\mathbf{D}^{\frac{1}{2}} && \text{distribute } \mathbf{D}^{-\frac{1}{2}} \text{ in} \\
&= \mathbf{D}^{\frac{1}{2}}(\mathbf{D}^{-\frac{1}{2}}\mathbf{K}\mathbf{D}^{-\frac{1}{2}} + \delta\mathbf{I})\mathbf{D}^{\frac{1}{2}} && \text{definition of root inverse} \\
&= \mathbf{D}^{\frac{1}{2}}(\mathbf{L} + \delta\mathbf{I})\mathbf{D}^{\frac{1}{2}} && \text{define: } \mathbf{L} = \mathbf{D}^{-\frac{1}{2}}\mathbf{K}\mathbf{D}^{-\frac{1}{2}} \\
&= \mathbf{D}^{\frac{1}{2}}(\mathbf{U_L}\mathbf{\Lambda_L}\mathbf{U_L}^{\mathrm{T}} + \delta\mathbf{I})\mathbf{D}^{\frac{1}{2}} && \text{eigen decomposition of } \mathbf{L} \\
&= \mathbf{D}^{\frac{1}{2}}(\mathbf{U_L}\mathbf{\Lambda_L}\mathbf{U_L}^{\mathrm{T}} + \delta\mathbf{U_L}\mathbf{U_L}^{\mathrm{T}})\mathbf{D}^{\frac{1}{2}} && \text{property of eigen vectors} \\
&= \mathbf{D}^{\frac{1}{2}}\mathbf{U_L}(\mathbf{\Lambda_L} + \delta\mathbf{I})\mathbf{U_L}^{\mathrm{T}}\mathbf{D}^{\frac{1}{2}} && \text{distributive property}
\end{aligned}$$

and invert it

$$\begin{aligned}
\mathbf{V}^{-1} &= \left(\mathbf{D}^{\frac{1}{2}}\mathbf{U_L}(\mathbf{\Lambda_L} + \delta\mathbf{I})\mathbf{U_L}^{\mathrm{T}}\mathbf{D}^{\frac{1}{2}}\right)^{-1} && \text{definition} \\
&= \left(\mathbf{D}^{\frac{1}{2}}\right)^{-1}\left(\mathbf{U_L}(\mathbf{\Lambda_L} + \delta\mathbf{I})\mathbf{U_L}^{\mathrm{T}}\right)^{-1}\left(\mathbf{D}^{\frac{1}{2}}\right)^{-1} && \text{inverse of product} \\
&= \mathbf{D}^{-\frac{1}{2}}\left(\mathbf{U_L}(\mathbf{\Lambda_L} + \delta\mathbf{I})\mathbf{U_L}^{\mathrm{T}}\right)^{-1}\mathbf{D}^{-\frac{1}{2}} && \text{inverse of diagonal matrix} \\
&= \mathbf{D}^{-\frac{1}{2}}\mathbf{U_L}(\mathbf{\Lambda_L} + \delta\mathbf{I})^{-1}\mathbf{U_L}^{\mathrm{T}}\mathbf{D}^{-\frac{1}{2}} && \text{inverse of eigen decomposition}
\end{aligned}$$

and compare to $\mathbf{M}^{\mathrm{T}}\mathbf{M}$

$$\begin{aligned}
\mathbf{M}^{\mathrm{T}}\mathbf{M} &= \left((\mathbf{\Lambda_L} + \delta\mathbf{I})^{-\frac{1}{2}}\mathbf{U_L}^{\mathrm{T}}\mathbf{D}^{-\frac{1}{2}}\right)^{\mathrm{T}} \left((\mathbf{\Lambda_L} + \delta\mathbf{I})^{-\frac{1}{2}}\mathbf{U_L}^{\mathrm{T}}\mathbf{D}^{-\frac{1}{2}}\right) && \text{definition} \\
&= \left(\mathbf{D}^{-\frac{1}{2}}\mathbf{U_L}(\mathbf{\Lambda_L} + \delta\mathbf{I})^{-\frac{1}{2}}\right) \left((\mathbf{\Lambda_L} + \delta\mathbf{I})^{-\frac{1}{2}}\mathbf{U_L}^{\mathrm{T}}\mathbf{D}^{-\frac{1}{2}}\right) && \text{transpose of product} \\
&= \mathbf{D}^{-\frac{1}{2}}\mathbf{U_L}\left((\mathbf{\Lambda_L} + \delta\mathbf{I})^{-\frac{1}{2}}(\mathbf{\Lambda_L} + \delta\mathbf{I})^{-\frac{1}{2}}\right)\mathbf{U_L}^{\mathrm{T}}\mathbf{D}^{-\frac{1}{2}} && \text{associative property} \\
&= \mathbf{D}^{-\frac{1}{2}}\mathbf{U_L}(\mathbf{\Lambda_L} + \delta\mathbf{I})^{-1}\mathbf{U_L}^{\mathrm{T}}\mathbf{D}^{-\frac{1}{2}} && \text{definition of root inverse}
\end{aligned}$$

$\square$

4.3 Simulation Results

4.4 Software

CHAPTER 5

Conclusion and Future Directions

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

BIBLIOGRAPHY