Likelihood Theory and Methods for Generalized Linear Mixed Models

by Aishwarya Bhaskaran

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy in Mathematics

under the supervision of **Prof. Matt P. Wand** and **Dr. Joanna Wang**

University of Technology Sydney
Faculty of Science

Submitted in September 2022

Certificate of original authorship

I, **Aishwarya Bhaskaran**, declare that this thesis, is submitted in fulfilment of the requirements for the award of Doctor of Philosophy in Mathematics, in the School of Mathematical and Physical Sciences at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by an Australian Government Research Training Program.

Signature:

Production Note:
Signature removed prior to publication.

Date: September 23, 2022

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisor Matt Wand. This journey was definitely a challenging one and I would like to thank you for all your patience, guidance and constant encouragement throughout the research process. I have indeed learnt a lot from your valuable insights and advice. Thank you so much for always looking out for my well-being as well. This milestone would not have been possible without your help and I am truly grateful.

Next, I would like to thank all my dear friends whom I have met throughout my life. Thank you all for constantly cheering me up, being such great listening ears and for always encouraging me to go the extra mile. I am really glad to have met each and every one of you.

Last but definitely not the least, I would like to give special thanks to my dear family. Thank you for always supporting me in my decisions and for offering me advice when I needed it. Thanks for also being so understanding especially these past three years and for always making sure that I am alright. I really appreciate everything that you all have done and will continue to do for me.

List of papers/publications

The following list of paper and publications awards relate to the work presented in this thesis:

- Jiang, J., Wand, M.P. and Bhaskaran, A. (2022), Usable and precise asymptotics for generalized linear mixed model analysis and design. Journal of the Royal Statistical Society, Series B, 84: 55-82. DOI: 10.1111/rssb.12473.
- Bhaskaran, A. and Wand, M.P.(2023), Dispersion parameter extension of precise generalized linear mixed model asymptotics. Statistics and Probability Letters, 193, Article 109691.

Notation

In this chapter, we introduce acronyms that are frequently used throughout this thesis.

Acronyms

Table 1: Table with acronyms used in the thesis with their meanings.

Acronym	**Meaning**
GLMM	Generalized linear mixed model
GVA	Gaussian variational approximation
MLE	Maximum likelihood estimation
TAP	Thouless-Anderson-Palmer

Contents

Abstract

Generalized linear mixed models are an essential group of models for analysing many present-day complex data sets, especially those that contain non-normal and correlated response data. Despite the large volume of research concerning this group of models, there is very little theory concerning the statistical properties of maximum likelihood estimators for generalized linear mixed models. Existing theoretical results available for the asymptotic variance-covariance matrix for such estimators contain limits and expectations over the response distribution, hence such results are not in ready-to-use forms when carrying out tasks such as constructing studentized confidence intervals or optimal design determination. In this thesis, we derive precise asymptotic results for likelihood-based generalized linear mixed model analysis. The novel asymptotic normality results are derived for both cases involving either a canonical or noncanonical link function. In our approach, we derive the exact leading term behaviour of the Fisher information matrix when both the number of groups and number of observations within each group diverge. This leads to asymptotic normality results with explicit and simple studentizable forms. The implications of these results in optimal design theory is also explored, leading to simpler and more direct determination of approximate locally D-optimal designs. Towards the end of this thesis, a Thouless-Anderson-Palmer approach is introduced for modern statistical inference for generalized linear mixed models. Such methods have proven to provide accurate approximations to problems arising in machine learning contexts. However, statistical applications such as generalized linear mixed model analysis have not been investigated. Thus, we derive results for implementing the Thouless-Anderson-Palmer frequentist variational approach to generalized linear mixed models and analyse the accuracy of its variational estimates.