THE UNIVERSITY OF ADELAIDE

School of Agriculture and Wine
BiometricsSA

Scale Parameter Modelling of the t-distribution

Doctor of Philosophy
2005

Julian Taylor
Supervisor: Arunas P. Verbyla

Contents

Acknowledgements

Firstly, I would like to acknowledge and thank my supervisor, Ari Verbyla, for the interesting topic and the generous help throughout my candidature.

Thanks to Adelchi Azzalini for supplying and assisting with an example used in this thesis and a subsequent paper.

I would also like to thank the BiometricsSA staff and postgraduates for their helpful suggestions and comments.

I am also grateful to my fellow Adelaide University postgraduate students, David McInerney and Stuart Johnson for their support in the first stages of this degree.

Many thanks to my ex-girlfriend Megan Haynes for her love and support throughout the first part of my PhD. Similarly, I'm very grateful to my ex-girlfriend Allison Flaherty for her love, support and friendship over the last two years.

Many thanks to all my family and friends who have kept me motivated through the extreme bad times and have helped propel this thesis to its conclusion.

Peace out!

Declaration

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made.

I give consent to this copy of my thesis, when deposited in the University Library, being available for loan and photocopying.

SIGNED: .... **DATE:** ....

Abstract

This thesis considers location and scale parameter modelling of the heteroscedastic t-distribution. This new distribution is an extension of the heteroscedastic Gaussian and provides robust analysis in the presence of outliers as well accommodates possible heteroscedasticity by flexibly modelling the scale parameter using covariates existing in the data.

To motivate components of work in this thesis the Gaussian linear mixed model is reviewed. The mixed model equations are derived for the location fixed and random effects and this model is then used to introduce Restricted Maximum Likelihood (REML). From this an algorithmic scheme to estimate the scale parameters is developed.

A review of location and scale parameter modelling of the heteroscedastic Gaussian distribution is presented. In this thesis, the scale parameters are a restricted to be a function of covariates existing in the data. Maximum Likelihood (ML) and REML estimation of the location and scale parameters is derived as well as an efficient computational algorithm and software are presented.

The Gaussian model is then extended by considering the heteroscedastic t distribution. Initially, the heteroscedastic t is restricted to known degrees of freedom. Scoring equations for the location and scale parameters are derived and their intimate connection to the prediction of the random scale effects is discussed. Tools for detecting and testing heteroscedasticity are also derived and a computational algorithm is presented. A mini software package "hett" using this algorithm is also discussed.

To derive a REML equivalent for the heteroscedastic t asymptotic likelihood theory is discussed. In this thesis an integral approximation, the Laplace approximation, is presented and two examples, with the inclusion of ML for the heteroscedastic t, are discussed. A new approximate integral technique called Partial Laplace is also discussed and is exemplified with linear mixed models. Approximate marginal likelihood techniques using Modified Profile Likelihood (MPL), Conditional Profile Likelihood (CPL) and Stably Adjusted Profile Likelihood (SAPL) are also presented and offer an alternative to the approximate integration techniques.

The asymptotic techniques are then applied to the heteroscedastic t when the degrees of freedom is known to form two distinct REMLs for the scale parameters. The first approximation uses the Partial Laplace approximation to form a REML for the scale parameters, whereas, the second uses the approximate marginal likelihood technique MPL.

For each, the estimation of the location and scale parameters is discussed and computational algorithms are presented. For comparison, the heteroscedastic t for known degrees of freedom using ML and the two new REML equivalents are illustrated with an example and a comparative simulation study.

The model is then extended to incorporate the estimation of the degrees of freedom parameter. The estimating equations for the location and scale parameters under ML are preserved and the estimation of the degrees of freedom parameter is integrated into the algorithm. The approximate REML techniques are also extended. For the Partial Laplace approximation the estimation of the degrees of freedom parameter is simultaneously estimated with the scale parameters and therefore the algorithm differs only slightly. The second approximation uses SAPL to estimate the parameters and produces approximate marginal likelihoods for the location, scale and degrees of freedom parameters. Computational algorithms for each of the techniques are also presented. Several extensive examples, as well as a comparative simulation study, are used to illustrate ML and the two REML equivalents for the heteroscedastic t with unknown degrees of freedom.

The thesis is concluded with a discussion of the new techniques derived for the heteroscedastic t distribution along with their advantages and disadvantages. Topics of further research are also discussed.