Adjusting the Parameter Estimation of the Parentage Analysis Software *MasterBayes* to the Presence of Siblings

A thesis presented in partial fulfillment of the requirements for the degree of
Master of Applied Statistics
at Massey University, Albany,
New Zealand

Florian Heller
2009

Abstract

Parentage analysis is concerned with the estimation of a sample's pedigree structure, which is often essential knowledge for estimating population parameters of animal species, such as reproductive success. While it is often easy to relate one parent to an offspring simply by observation, the second parent remains frequently unknown. Parentage analysis uses genotypic data to estimate the pedigree, which then allows inferring the desired parameters. There are several software applications available for parentage analysis, one of which is *MasterBayes*, an extension to the statistical software package R. *MasterBayes* makes use of behavioural, phenotypic, spatial and genetic data, providing a Bayesian approach to simultaneously estimate pedigree and population parameters of interest, allowing for a range of covariate models. *MasterBayes* however assumes the sample to be a randomly collected from the population of interest. Often however, collected data will come from nests or otherwise from groups that are likely to contain siblings. If siblings are present, the assumption of a random population sample is not met anymore and as a result, the parameter variance will be underestimated. This thesis presents four methods to adjust *MasterBayes'* parameter estimate to the presence of siblings, all of which are based on the pedigree structure, as estimated by *MasterBayes*. One approach, denoted as DEP, provides a Bayesian estimate, similar to *MasterBayes'* approach, but incorporating the presence of siblings. Three further approaches, denoted as W1, W2 and W3, apply importance sampling to re-weight parameter estimates obtained from *MasterBayes* and DEP. Though fully satisfying adjustment of the estimate's variance is only achieved at nearly perfect pedigree assignment, the presented methods do improve *MasterBayes*' parameter estimation in the presence of siblings considerably, when the pedigree is uncertain. DEP and W3 show to be the most successful adjustment methods, providing comparatively accurate,

though yet underestimated variances for small family sizes. W3 is the superior approach when the pedigree is highly uncertain, whereas DEP becomes superior when about half of all parental assignments are correct. Large family sizes introduce to all approaches a tendency to underestimate the parameter variance, the degree of underestimation depending on the certainty of pedigree. Additionally, the importance sampling schemes provide at large uncertainty of pedigree comparatively good estimates of the parameter's expected values, where the non importance sampling approaches severely fail.

Acknowledgements

I would like to acknowledge a number of people for the support provided during the course of this research project.

First and foremost, my special thanks go to my supervisor, Dr Beatrix Jones. Her deep knowledge, insightful criticism, patience and continued encouragement guided me through the process of this project.

I would further like to thank Dr Howard Edwards for providing me with the helpful course material of the Bayesian Statistics paper.

A big thanks goes also to Tim Napier, who went through the hassle of proof-reading the thesis.

Lastly, I would like to thank my family, my friends and especially my partner Chen Geng for their continued encouragement and support.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS