Variational Inference for Heteroscedastic and Longitudinal Regression Models

by

Marianne Menictas

Submitted to the School of Mathematical Sciences, Faculty of Science in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

at the

UNIVERSITY OF TECHNOLOGY SYDNEY

2015

CERTIFICATE OF ORIGINAL AUTHORSHIP

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text. I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Signature of Student:

Date:

Acknowledgements

My deepest thank you goes to my advisor and academic father, Matt. You opened my eyes to a new way of thinking. I could not have asked for a more inspirational, supportive and patient role model to guide me through this journey and will forever be grateful for what you have taught me. I extend my appreciation to my co-supervisor Peter, who spent many hours guiding me with helpful advice throughout my PhD journey. To my honours advisor Debbie, thank you for showing me the importance of rigorous research and for always having your door open to me. Cathy, my PhD buddy and great friend, discussing the intricacies of our work was the highlight of my day so thank you. To Mum and Nick, thank you for your constant unconditional love and support and for always being there for me. Finally, to Dad, my quest would not have taken place if it weren't for you encouraging me to be unreservedly curious and hungry for the pursuit of knowledge. You are the biggest inspiration in my life.

Contents

Abstract

The focus of this thesis is on the development and assessment of mean field variational Bayes (MFVB), which is a fast, deterministic tool for inference in a Bayesian hierarchical model setting. We assess the performance of MFVB via the use of comprehensive comparisons against a Markov chain Monte Carlo (MCMC) benchmark. Each of the models considered are special cases of semiparametric regression. In particular, we focus on the development and assessment of the performance of MFVB for heteroscedastic and longitudinal semiparametric regression models. Generally, the new MFVB methodology performs well in its assessment of accuracy against MCMC for the semiparametric and nonparametric regression models considered in this thesis. It is also much faster and is shown to be applicable to real-time analyses. Several real data illustrations are provided. Altogether, MFVB proves to be a credible inference tool and a good alternative to MCMC, especially when analysis is hindered by time constraints.