Variational Bayesian Inference: Message Passing Schemes and Streamlined Multilevel Data Analysis

Tui Hiraka Nolan

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

Doctor of Philosophy (Mathematics)

at the

UNIVERSITY OF TECHNOLOGY SYDNEY

April, 2020

Certificate of Original Authorship

I, Tui Hiraka Nolan, declare that this thesis is submitted in fulfilment of the requirements for the award of Doctor of Philosophy (Mathematics), in the School of Mathematical and Physical Sciences, Faculty of Science 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 the Australian Government Research Training Program.

Signature: Production Note: Signature removed prior to publication.

Date: 24 April 2020

Acknowledgements

I wish to extend my deepest thanks to my supervisor, Professor Matt Wand, for his guidance and mentorship. His support has provided opportunities and career pathways that were otherwise inaccessible, and I am truly grateful. I am extremely fortunate to have him as a role model and a friend.

My doctoral studies were supported by a Research Training Program Scholarship and a Jumbunna Postgraduate Research Scholarship, which I very much appreciate.

I would like to extend my gratitude to UTS: Jumbunna Institute for Indigenous Education and Research and the Aurora Education Foundation. Both of these organisations have influenced my passion for promoting education within the Indigenous Australian community. My selection in the 2016 Aurora Indigenous Scholars International Study Tour opened my eyes to the international opportunities that are available through my doctoral studies.

Sincere thanks are due to my family, who have provided continual support and encouragement. A special thanks to my late father, John Nolan. Although he did not witness this academic journey, his legacy of hard work and dedication has motivated and guided me in the success of this research endeavour.

To my wife, Cassandra Nolan, thank you for your love and encouragement. Your companionship has been my most valued source of support, which has inspired me during the most difficult periods of this doctoral program.

Publications

Nolan, T., & Wand, M. (2019). *Solutions to sparse multilevel matrix problems*. (Submitted to Statistics and Probability Letters and available on arXiv at arXiv:1903.03089)

Nolan, T., Menictas, M., & Wand, M. (2019). *Streamlined computing for variational inference with higher level random effects*. (Submitted to the Journal of Machine Learning Research and available on arXiv at arXiv:1903.06616)

Menictas, M., Nolan, T., Simpson, D., & Wand, M. (2019). *Streamlined variational inference for higher level group-specific curve models*. (Submitted to Statistical Modelling and currently available on arXiv at arXiv:1903.04043)

Nolan, T., & Wand, M. (2017). Accurate logistic variational message passing: algebraic and numerical details. *Stat*, 6, 102–112.

Awards and Honours from my Doctoral Candidature

Fulbright Postdoctoral Scholarship (The Kinghorn Foundation and Western Sydney University, 2019)

Aboriginal and Torres Strait Islander Scientist Travelling Research Award (Australian Academy of Science and the Alan Turing Institute, 2019)

STEM Professional Early Career Award (CSIRO, 2019)

Research Training Program Scholarship (University of Technology Sydney and the Australian Government, 2016–2019)

Jumbunna Postgraduate Research Scholarship (UTS: Jumbunna Institute for Indigenous Education and Research, 2016–2019)

Aurora Indigenous Scholars International Study Tour (Aurora Education Foundation and the Aspiration Initiative, 2016)

Abstract

Mean-field variational Bayes (MFVB) is a deterministic technique for approximating intractable integrals arising in Bayesian inference. They are typically used for making approximate inference for parameters in complex statistical models. Most of its foundational literature and applications are in Machine Learning. However, in the age of "Big Data", and by extension large sample sizes, MFVB has become an important tool in Statistics.

The approximating schemes afforded by MFVB rely on heavy algebraic derivations across the model. The emergence of Big Data has resulted in more complex statistical models, making the process of formulating an MFVB algorithm cumbersome. Fortunately, the MFVB updating scheme can be simplified by representing the parameters of the statistical model in a probabilistic graph. The derivations are made more efficient by decomposing the required computations into calculations that are local to each node in the graph.

We focus on constructing variational Bayesian inference algorithms based on a modularised format known as variational message passing (VMP), which is founded upon the notion of messages passed between fragments on a factor graph. Primitive functions, which represent the localised messages over factor graph fragments, are derived and can be called upon for direct implementation into arbitrarily large statistical models. The MFVB and VMP approaches result in superficially different algorithms, but converge to identical posterior density function approximations because they are founded upon the same optimisation problem. For complex statistical models, VMP has the advantage that the iterative updates are adjusted into a modularised format by taking advantage of the localised computations afforded by variational Bayesian methods. The resulting algorithm is a sequence of fragment-based functions that represent a compartmentalisation of the required algebra and computer coding.

Despite the computational convenience of VMP algorithms over their MFVB counterparts, the speed of both classes is limited for multilevel data models, such as Gaussian response linear mixed models. Statistical inference on such models requires standard matrix operations, such as inversion and matrix-vector multiplication, on sparse matrices, which are difficult to achieve efficiently. Furthermore, computational storage issues restrict the size of such models. Streamlined matrix algebraic results are necessary for implementing fast frequentist and variational Bayesian inference, which is not inhibited by storage-greedy sparse matrix operations, on multilevel data models.

This thesis develops factor graph fragment functions that can be used to build complex statistical models and achieves streamlined matrix algebraic derivations for multilevel data analysis.

Contents

1 Introduction **2**

1.1 Notational Guide 3

1.1.1 Logic and Set Notation 3

1.1.2 Matrix Notation 3

1.1.3 Functional Notation 4

1.1.4 Algorithmic Notation 5

1.2 Vector Differential Calculus 7

1.2.1 Vector and Matrix Operations 7

1.2.2 A Primer on Vector Differential Calculus 8

1.3 Probabilistic Graph Theory 10

1.3.1 Directed Acyclic Graphs 11

1.3.2 Undirected Graphs 12

1.3.3 Conditional Independence 14

1.3.4 Applications of Graphical Models 15

1.4 Exponential Families 16

1.4.1 Basics of Exponential Families 16

1.4.2 Graphical Models in Exponential Form 17

1.5 Semiparametric Regression 18

1.5.1 Mixed Model Representation 21

1.6 Bayesian Inference 22

1.6.1 Bayesian Statistics 22

1.6.2 Markov Chain Monte Carlo 22

2 Variational Bayesian Inference **24**

2.1 Mean Field Variational Bayes 25

2.2 Variational Message Passing 26

2.3 Exponential Family Special Case 29

2.3.1 Normal Distribution Special Case 30

2.4 Fundamental Fragments for Arbitrarily Large Regression Models 31

2.4.1 Gaussian Prior Fragment 32

2.4.2 Inverse G-Wishart Prior Fragment 32

2.4.3 Iterated Inverse G-Wishart Fragment 33

2.4.4 Dirichlet Prior Fragment 34

2.4.5 Gaussian Penalisation Fragment 35
2.4.6 Gaussian Likelihood Fragment 37
2.4.7 A Note on Fragment Inputs 37
2.5 A Motivating Example 38
2.5.1 The MFVB Approach 39
2.5.2 The VMP Approach 41
2.5.3 Arbitrarily Large Model Viewpoint 42
2.5.4 Heuristic Justification of the Mean Field Approximation 44
2.6 Assessment of Accuracy for Variational Bayesian Approximations 46
2.7 Consequences of the Mean Field Approximation 47
2.A Exponential Family Results 47
2.A.1 Normal Distribution 48
2.A.2 Inverse G-Wishart Distribution 49
2.A.3 Beta and Dirichlet Distributions 51
2.A.4 Binomial and Multinomial Distributions 52

3 Normal Mixture Models **53**
3.1 Univariate Normal Mixture Model 54
3.2 Multivariate Normal Mixture Model 57
3.3 Extension to Regression Models with Missing Data 59
3.4 Simulations of the Univariate Normal Mixture Model 64
3.4.1 Assessment of Accuracy 64
3.4.2 Speed Comparisons 65
3.5 Discussion 65
3.A Derivation of Algorithm 20 66
3.B Derivation of Algorithm 21 68
3.C Derivation of Algorithm 22 69
3.D Derivation of Algorithm 23 71
3.E Derivation of Algorithm 24 73

4 Variational Approximations for Probit Regression Models **76**
4.1 Approximation Techniques for the Probit Likelihood Fragment 77
4.1.1 Albert-Chib 77
4.1.2 Ormerod-Wand 78
4.2 Simulations of Probit-Based Regression Models 80
4.2.1 Linear Bayesian Probit Regression Model 80
4.2.2 Semiparametric Bayesian Probit Regression Model 82
4.3 Discussion 85
4.A Details for the Ormerod-Wand Fragment 86
4.A.1 Adaptive Gauss Hermite Quadrature 86
4.A.2 Derivation of Algorithm 26 87

4.B Approximation of $\mathrm{Corr}(\beta_0, \beta_1 | \boldsymbol{y})$ . . . 88

5 Variational Approximations for Logistic Regression Models 89

5.1 Approximation Techniques for the Logistic Likelihood Fragment . . . 90
5.1.1 Jaakkola-Jordan . . . 90
5.1.2 Saul-Jordan . . . 92
5.1.3 Knowles-Minka-Wand . . . 93
5.2 Simulations of Logistic-Based Regression Models . . . 97
5.2.1 Linear Bayesian Logistic Regression Model . . . 97
5.2.2 Semiparametric Bayesian Logistic Regression Model . . . 100
5.3 Other Approximation Techniques . . . 101
5.4 Discussion . . . 102
5.A Details for the Saul-Jordan Fragment . . . 103
5.A.1 Proof of Lemma 5.1.2 . . . 103
5.A.2 Proof of Theorem 5.1.3 . . . 103
5.A.3 Derivation of Algorithm 28 . . . 104
5.B Details for the Knowles-Minka-Wand Fragment . . . 104
5.B.1 Proof of Corollary 5.1.4 . . . 104
5.B.2 Proof of Theorem 5.1.5 . . . 104
5.C Approximation of $\mathrm{Corr}(\beta_0, \beta_1 | \boldsymbol{y})$ . . . 106

6 Solutions to Sparse Multilevel Matrix Problems 107

6.1 Two Level Sparse Matrix Problems . . . 108
6.1.1 Least Squares Form and QR-decomposition Enhancement for Two Level Sparse Matrices . . . 111
6.2 Three Level Sparse Matrix Problems . . . 113
6.2.1 Least Squares Form and QR-decomposition Enhancement for Three Level Sparse Matrices . . . 116
6.3 Discussion . . . 119
6.A Proofs . . . 120
6.A.1 Proof of Theorem 6.1.1 . . . 120
6.A.2 Proof of Theorem 6.1.2 . . . 122
6.A.3 Proof of Theorem 6.2.1 . . . 124
6.A.4 Proof of Theorem 6.2.2 . . . 128
6.B Algorithms . . . 133
6.B.1 Functions for QR-Decompositions . . . 133

7 Two Level Gaussian Response Models 134

7.1 Two Level Best Linear Unbiased Prediction . . . 135
7.2 Two Level Mean Field Variational Bayes . . . 137
7.3 Two Level Variational Message Passing . . . 141
7.3.1 Streamlined Two Level Gaussian Likelihood Fragment Updates . . . 144

7.3.2 Streamlined Gaussian Penalisation Fragment Updates 145
7.3.3 Computations for the q-Density Functions 146
7.4 Simulations of the Two Level Model 147
7.4.1 Assessment of Accuracy 147
7.4.2 Assessment of Speed and Computational Storage 149
7.5 Discussion 150
7.A The Lower Bound on the Marginal Log-Likelihood for the Two Level Gaussian Response Model 151
7.B Proofs 151
7.B.1 Proof of Corollary 7.1.1 151
7.B.2 Proof of Corollary 7.2.1 152
7.B.3 Proof of Corollary 7.3.1 152
7.C Algorithms 153
7.C.1 TwoLevelIteratedInverseChiSquared 153
7.C.2 TwoLevelIteratedInverseGWishart 154
7.C.3 Derivation of Algorithm 37 155
7.C.4 Derivation of Algorithm 38 156
7.C.5 Derivation of Algorithm 39 158

8 Three Level Gaussian Response Models 161
8.1 Three Level Best Linear Unbiased Prediction 161
8.2 Three Level Mean Field Variational Bayes 164
8.3 Three Level Variational Message Passing 167
8.3.1 Streamlined Three Level Gaussian Likelihood Fragment 171
8.3.2 Streamlined Three Level Gaussian Penalisation Fragment 173
8.4 Simulations of the Three Level Model 173
8.4.1 Assessment of Accuracy 173
8.4.2 Assessment of Speed and Computational Storage 176
8.5 Discussion 176
8.A The Lower Bound on the Marginal Log-Likelihood for the Three Level Gaussian Response Model 177
8.B Proofs 179
8.B.1 Proof of Corollary 8.1.1 179
8.B.2 Proof of Corollary 8.2.1 179
8.B.3 Proof of Corollary 8.3.1 180
8.C Algorithms 180
8.C.1 ThreeLevelIteratedInverseChiSquared 180
8.C.2 ThreeLevelIteratedInverseGWishart 181
8.C.3 ThreeLevelNaturalToCommonParameters 182
8.C.4 Derivation of Algorithm 44 182
8.C.5 Derivation of Algorithm 45 183

8.C.6 Derivation of Algorithm 46 186

9 Closing Remarks and Future Work **190**

9.1 Variational Message Passing Schemes 190

9.2 Sparse Multilevel Data Analysis 191

References **193**