Fast Approximate Inference for Longitudinal and Multilevel Data Analysis

Cathy Yuen Yi Lee

GStat MBiostat BMathAdv(Hons)

A PhD thesis submitted in fulfilment of the requirements for
Doctor of Philosophy in Mathematics
in the School of Mathematical and Physical Sciences,
University of Technology Sydney

2016

© Cathy Yuen Yi Lee, 2016. All rights reserved.

Permission is herewith granted to University of Technology Sydney to circulate and to have copied for non-commercial purposes, at its discretion, the above title upon request of individuals and institutions.

Certificate of Original Authorship

I, Cathy Yuen Yi Lee, 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

Fear protects us in many ways, but what has served me is diving into my own obliviousness. Being more confident than I should be trying things that you never would have tried. Your inexperience is an asset in that it will make you think in original, unconventional ways. Accept your lack of knowledge and use it as your asset. Natalie Portman

I never thought I would go to university, let alone complete a PhD degree. I am always and forever grateful to my lovely parents for instilling the value of education in me and for making countless sacrifice to always provide for their children. I wouldn't be who I am today if it wasn't for the undying love and support my parents have unconditionally given to me.

Words are simply not enough to express my sincere gratitude to my academic parents, Matt and Louise, for believing in me and seeing my good qualities way before I ever could. Through watching them, I learned dedication and perseverance to succeed. I am no longer afraid of following my dream I had set since I was 16. Thank you for guiding me to reach a stage where I can be proud of who I am and what I have achieved for the past three years.

I must express my gratitude to Jason, my companion, for his enthusiasm and continuous support and encouragement. I was amazed by his willingness and determination to support me in every possible way throughout this journey. He continually challenged me to dig deeper and learn more, and I greatly appreciate everything he has done to make this possible.

"Friends are the bold essence of life". Completing this work would have been more difficult were it not for the support and friendship provided by my colleagues at the School of Mathematical and Physical Sciences. A special thanks to Chris, who has proofread countless pages of my work. Also a special thanks to Marianne, who has always been there for me and experienced all of the laugher and tears from my special journey.

I finish here with Hon's favourite quote "Stay Hungry, Stay Foolish". I believe this quote will continue to motivate me in life for many years to come.

Publications

Lee, C. Y . Y. and Wand, M. P. (2016). Streamlined mean field variational Bayes for longitudinal and multilevel data analysis. *Biometrical Journal*. DOI: 10.1002/bimj.201500007

Bentley, J. and Lee, C. Y. Y. (2016). Bayesian Inference for Gaussian Semiparametric Multilevel Models. *SAS Global Forum*, April 18-21, Las Vegas, United States. `http://support.sas.com/resources/papers/proceedings16/7200-2016.pdf`

Lee, C. Y. Y. and Wand, M. P. (2015). Variational methods for fitting complex Bayesian mixed effects models to health data. *Statistics in Medicine*. DOI: 10.1002/sim.6737.

Lee, C. Y. Y., Homer, C., Bisits, A. and Ryan, L. (2015). Increasing variation in hospital caesarean section rates among low-risk nulliparous women in Australia, from 1994 to 2010. *Birth in revision*.

Chen, J., Lee, C. Y. Y., Coull, B., Valeri, L., Gruskin, S., Beckfield, J., Singh, N. and Krieger, N. (2015). 50-year trends and state variation in socioeconomic and racial/ethnic inequalities in US infant death rates, 1960 to 2010. *Under preparation*.

The last listed publication related to my 3-week academic visit at the Harvard School of Public Health during March-April 2014. Throughout my visit, I collaborated with a research team of world-renowned social epidemiologists, biostatisticians and research scientists on a study under an overarching project entitled *The Public Health Disparities Geocoding Project Monograph*. For more information, please refer to https://www.hsph.harvard.edu/thegeocodingproject/.

Presentations

Lee, C. Y. Y and Bentley, J. P. Fast variational Bayesian inference for Gaussian semi-parametric multilevel models in SAS/IML®, Las Vegas, United States, 2016. (Accepted for a 20-min workshop presentation)

Lee, C. Y.Y. Fast approximate inference for longitudinal and multilevel data analysis. *The 16th J.B. Douglas Awards, Statistical Society of Australia*, Sydney, Australia, 2015. (Awarded Equal First Price for oral presentation)

Lee, C. Y. Y., Wand, M. P. and Ryan, L. Fast approximate inference for longitudinal and multilevel data analysis. *The 16th J.B. Douglas Awards, Statistical Society of Australia*, Sydney, Australia, 2015. (Awarded Equal First Price for oral presentation)

Lee, C. Y. Y., Chen, J., Coull, B., Valeri, L., Gruskin, S., Beckfield, J., Singh, N. and Krieger, N. 50-year trends and state variation in socioeconomic and racial/ethnic inequalities in US infant death rates, 1960 to 2010. *Joint Statistical Meeting*, Seattle, United States, 2015.

Lee, C. Y. Y. and Wand, M. P. Fast approximate variational inference for Bayesian multilevel models. *Young Statisticians Conference*, Adelaide, Australia, 2015. (Awarded 3rd Price for oral presentation)

Ryan, L. and Lee, C. Y. Y. Spatio-temporal variation in cesarean rates in New South Wales, Australia, from 1994 to 2010. *Joint Statistical Meeting*, Boston, United States, 2014.

Lee, C. Y. Y. and Wand, M. P. Fast approximate variational inference for Bayesian multilevel models. *Australian Statistical Conference in conjunction with the Institute of Mathematical Statistics Annual Meeting*, Sydney, Australia, 2014.

The fourth listed presentation related to my 3-week academic visit at the Harvard School of Public Health during March-April 2014. Throughout my visit, I collaborated with a research team of world-renowned social epidemiologists, biostatisticians and research scientists on a study under an overarching project entitled *The Public Health Disparities Geocoding Project Monograph.* For more information, please refer to https://www.hsph.harvard.edu/thegeocodingproject/.

Lee, C. Y. Y. and Wand, M. P. Fast approximate variational inference for Bayesian longitudinal and multilevel semiparametric regression models. *University of Queensland*, Brisbane, Australia, 2014.

Contents

1 Introduction **1**

1.1 Aim and motivation 1
1.2 Structure of the thesis 3
1.3 Semiparametric regression 4
1.3.1 Mixed model representation 6
1.4 Graphical models 6
1.5 Bayesian inference 9
1.5.1 A Bayesian viewpoint of probability and statistics 9
1.5.2 Markov chain Monte Carlo 10
1.5.3 A brief introduction to variational approximations 11
1.6 Terminological and notational conventions 17
1.7 Common distributions 18
1.8 Distributional results 18
1.9 Vector differential calculus 20
1.10 Special functions 20
1.11 Matrix results 21
1.12 Statistical software 22

2 Mean Field Variational Bayes Approximations for Longitudinal and Multilevel Data Analysis **23**

2.1 Introduction 23
2.2 A brief introduction to longitudinal and multilevel models 25
2.2.1 Graph theoretical viewpoint of longitudinal and multilevel models . 27
2.2.2 Random intercepts models 27
2.2.3 Random intercepts and slopes models 28
2.2.4 Extension to semiparametric regression 29
2.3 Mixed model representation 29
2.3.1 Sample size and subscript notation 31
2.4 Prior distributions for fixed and random effects 33

2.5 Gaussian semiparametric mixed models 34
2.5.1 Approximate Bayesian inference via mean field variational Bayes . . 34
2.5.2 Approximating density functions of entries of $\boldsymbol{\Sigma}^{\mathrm{R}}$ 39
2.6 Student-$\boldsymbol{t}$ semiparametric mixed models 41
2.6.1 Approximate Bayesian inference via mean field variational Bayes . . 41
2.7 Bernoulli semiparametric mixed models 45
2.7.1 Approximate Bayesian inference via mean field variational Bayes . . 46
2.8 Poisson semiparametric mixed models 49
2.8.1 Approximate Bayesian inference via non-conjugate variational message passing 49
2.9 Displaying approximate posterior means of regression function fits 55
2.10 Numerical evaluation 56
2.10.1 Fitting a Bayesian hierarchical model in Stan 57
2.10.2 Assessment of accuracy 59
2.10.3 Assessment of coverage 60
2.10.4 Assessment of speed 60
2.11 Concluding remarks 64
2.A Optimal q-densities derivation for Gaussian semiparametric mixed models . 65
2.A.1 Derivation of the marginal log-likelihood lower bound 71
2.B Optimal q-densities derivation for Student-t semiparametric mixed models . 75
2.B.1 Derivation of the marginal log-likelihood lower bound 76
2.C Optimal q-densities derivation for Bernoulli semiparametric mixed models . 80
2.C.1 Derivation of the marginal log-likelihood lower bound 81
2.D Optimal q-densities derivation for Poisson semiparametric mixed models . . 82
2.D.1 Derivation of the marginal log-likelihood lower bound 84

3 Streamlining Mean Field Variational Bayes Algorithms And Three-Level Model Extensions 85
3.1 Introduction 85
3.2 Computational challenges in naïve MFVB algorithms 86
3.3 Predictor structure and matrix notation 88
3.4 Streamlining mean field variational Bayes algorithms 89
3.4.1 Streamlining update expressions involving $\boldsymbol{\Sigma}_{q(\boldsymbol{\beta},\boldsymbol{u})}$ 91
3.4.2 Streamlining lower bound expression involving $\boldsymbol{\Sigma}_{q(\boldsymbol{\beta},\boldsymbol{u})}$ 93
3.5 Computational speed: naïve versus streamlined 94
3.6 Real data applications 101
3.6.1 Application to smoking data 101
3.6.2 Application to student assessment data 103
3.6.3 Application to German health care data 104

3.7 Extension to three-level semiparametric mixed models . . . 105
3.7.1 Mixed model representation . . . 110
3.7.2 Naïve mean field variational Bayes algorithms . . . 113
3.7.3 Streamlined mean field variational Bayes algorithms . . . 114
3.8 Concluding remarks . . . 114

4 Measurement Error, Missing Data and Real-Time Extensions 118
4.1 Introduction . . . 118
4.2 Gaussian semiparametric mixed models with measurement error problems . 119
4.2.1 Notation . . . 121
4.2.2 Approximate Bayesian inference via mean field variational Bayes . . 124
4.3 Gaussian semiparametric mixed models with missing data problems . . . 126
4.3.1 Notation . . . 129
4.3.2 Approximate Bayesian inference via mean field variational Bayes . . 130
4.4 Numerical evaluation . . . 134
4.4.1 Assessment of accuracy . . . 135
4.4.2 Assessment of coverage . . . 136
4.4.3 Assessment of speed . . . 136
4.5 Real-time mean field variational Bayes algorithms . . . 136
4.6 Concluding remarks . . . 142
4.A Optimal q-densities derivation for measurement error problems . . . 147
4.A.1 Derivation of the marginal log-likelihood lower bound . . . 154
4.B Optimal q-densities derivation for missing data problems . . . 157

5 Extension to Group-Specific Curve Models With Contrasting 161
5.1 Introduction . . . 161
5.2 Mean field variational Bayes approximations to caesarean section data . . . 164
5.2.1 Random intercept and slope model . . . 166
5.2.2 Group-specific curve model . . . 170
5.2.3 Factor-by-curve interactions . . . 171
5.3 Numerical evaluation . . . 176
5.3.1 Assessment of accuracy . . . 176
5.3.2 Assessment of speed . . . 177
5.4 Application to caesarean section data . . . 178
5.5 Concluding remarks . . . 182
5.A Derivation of the mean field variational Bayes algorithm for model (5.11) . 185
5.A.1 Mixed model representation . . . 186
5.A.2 Streamlined mean field variational Bayes algorithm . . . 188
5.A.3 Fitting the group-specific curve model (5.11) in `Rstan` . . . 188

6 Alternative Approach Based on Variational Message Passing 193
6.1 Introduction 193
6.2 Natural canonical forms for exponential family densities 194
6.3 Primitive integrals and function definitions 195
6.4 Factor graphs 197
6.5 Variational message passing 198
6.6 Illustrative example of simple mixed effects regression 200
6.7 Arbitrarily large models viewpoint 205
6.8 Concluding remarks 209

7 Conclusions and Future Directions 211

List of Figures

1.1 Directed acyclic graph involving six random variables: $x_1, \ldots, x_6$. The open circles represent the random ("hidden") nodes. The shaded circle represents the observed ("evidence") node. The grey dashed line polygon represents the Markov blanket of the node x_4. 8

1.2 Left: The smallest ancestral subgraph corresponding to Figure 1.1. Right: The moral graph corresponding to Figure 1.1. 8

1.3 Factor graph corresponding to the DAG in Figure 1.1. 15

2.1 Panel of scatterplots of the 11-year old mathematics test scores versus the eight-year old test scores. Each panel corresponds to a school. Details of the data are described in Goldstein (2010). 26

2.2 Directed acyclic graph corresponding to model (2.1). The open circles represent the random nodes. The shaded circle represents the observed node. The vector $\boldsymbol{y} = (y_{11}, \ldots, y_{mn_m})$ is represented by a single node, and β_0 and β_x are represented by the $\boldsymbol{\beta}$ node. Similarly, $\boldsymbol{u}^{\mathrm{R}} = (u_1^{\mathrm{R}} \ldots, u_m^{\mathrm{R}})$. 28

2.3 Directed acyclic graph for the two-level Bayesian semiparametric mixed models with the Gaussian and Student-t responses. The shaded node corresponds to the observed data vector and the open nodes correspond to the random or auxiliary variables. The pink open nodes $\boldsymbol{b}$ and ν are additional parameters for the Student-t response model. The numbered grey dashed polygons indicate the corresponding update expressions in Algorithm 3. . . 35

2.4 Directed acyclic graph for the two-level Bayesian semiparametric mixed models with the Bernoulli and Poisson responses. The shaded node corresponds to the observed data vector and the open node corresponds to the random or auxiliary variables. The numbered grey dashed polygons indicate the corresponding update expressions in Algorithms 4 and 6. . . . 35

2.5 Life cycle of nodes for mean field variational Bayes: (a) directed acyclic graph for model (2.11), (b) moral graph for model (2.11), (c) modification of Figure 2.5b with eight edges removed to impose q-density product restriction. In each graph, shading is used to signify the observed data vector. 37
2.6 Variational representation of the function $-\log(1+e^x)$ as the maximum of a family of parabolas, corresponding to (2.20). 47
2.7 Summary of MCMC samples for fitting the two-level Bayesian semiparametric mixed model with Gaussian response to the simulated data. The columns are: parameter name, trace plot of the MCMC sample, plot of sample against its lag-one sample, sample autocorrelation function, kernel density estimate of posterior density function and numerical summaries of posterior density function. 61
2.8 Approximate posterior density functions for the covariance parameters obtained via MFVB approximation (orange curves) and MCMC (blue curves) for a single replication from each of the simulation studies described in the text. The green vertical lines represent the true parameter values. The accuracy score on the top right of each plot represents the accuracy of MFVB approximation compared against a MCMC benchmark. 62
2.9 Approximate posterior means of the penalised regression functions and corresponding pointwise 95% credible sets obtained via MFVB approximation (orange curves) and MCMC (blue curves) for a single replication of the simulation study described in the text. The red curves represent the true mean functions and the sky blue circles represent the simulated data. 63
2.10 Side-by-side boxplots of accuracy scores for MFVB approximation compared against MCMC over 30 runs for the two-level Bayesian semiparametric mixed models. 63

3.1 Approximate posterior means of the penalised regression functions and corresponding pointwise 95% credible sets obtained via MFVB approximation (orange curves) and MCMC (blue curves) for the two-level Bayesian semiparametric mixed models to the real datasets. The sky blue circles represent the real data. 106
3.2 Approximate posterior density functions for the model parameters obtained via MFVB approximation (orange curves) and MCMC (blue curves) for the Gaussian response model to the smoking data. The accuracy score on the top right of each plot represents the accuracy of MFVB approximation compared against the MCMC benchmark. 107

3.3 Approximate posterior density functions for the model parameters obtained via MFVB approximation (orange curves) and MCMC (blue curves) for the Bernoulli response model to the Program of International Student Assessment data. The accuracy score on the top right of each plot represents the accuracy of MFVB approximation compared against the MCMC benchmark. 108
3.4 Approximate posterior density functions for the model parameters obtained via MFVB approximation (orange curves) and MCMC (blue curves) for the Poisson response model to the German health care data. The accuracy score on the top right of each plot represents the accuracy of MFVB approximation compared against the MCMC benchmark. 109
3.5 Directed acyclic graphs for model (3.17). The shaded node corresponds to the observed data vector. The colour keys at the top of the figure indicate the components of the graph corresponding to each response outcome. . . . 112
3.6 Different $\boldsymbol{Z}$ structures for the two-level and three-level models with $m = 2$. The $\boldsymbol{Z}$ matrix starts off as having a simple block-diagonal structure and, as the level of hierarchy increases, it grows into a "nested" block-diagonal structure. The definitions of matrices are described in the text. 113

4.1 Directed acyclic graph corresponds to model (4.1). The shaded node corresponds to the observed data vector and the open nodes correspond to the random or auxiliary variables. The fragments shown in grey are presented in the Gaussian model DAG as shown in Figure 2.3. 122
4.2 Directed acyclic graph corresponds to model (4.7). The shaded node corresponds to the observed data vector and the open nodes correspond to the random or auxiliary variables. The fragments shown in grey are presented in the Gaussian model DAG as shown in Figure 2.3. 122
4.3 Side-by-side boxplots of accuracy scores for the MFVB approximation compared against MCMC over 30 runs for the two-level Bayesian semiparametric mixed model with Gaussian response subject to classical measurement errors. 137
4.4 Side-by-side boxplots of accuracy scores for the MFVB approximation compared against MCMC over 30 runs for the two-level Bayesian semiparametric mixed model with Gaussian response subject to missingness. 138
4.5 Approximate posterior means of the penalised regression functions and corresponding pointwise 95% credible sets obtained via MFVB approximation (orange curves) and MCMC (blue curves) for a single replication of the simulation study described in the text. The red curves represent the true mean functions. The grey circles represent the observed data and the sky blue circles represent the unobserved or missing data. 139

4.6 Diagrammatic description of the real-time/online variational analysis. The first phase is batch phase, where we start by running a small subset of the initial data in the batch algorithm to initialise model parameters and obtain starting values for the data sufficient statistics (these are all we keep). The second phase is the real-time/online phase where the MFVB algorithm updates as each new data point arrives. Real-time updates use only the new data and summary statistics from previous iterations rather than the full set of available data. 141
4.7 Real-time MFVB fitting of the two-level Bayesian semiparametric mixed models with Gaussian response for two different sample sizes. The sky blue circles represent the batch data and the pink circles represent the new data. 145

5.1 Trends in caesarean section rates for low-risk nulliparous women in the largest state of Australia, New South Wales, between 1994 and 2010. The dotted line is the overall mean trend and the solid lines are the selected hospital-specific trends. 162
5.2 Trends in caesarean section rates for low-risk nulliparous women aged less than 25 years and those aged greater or equal to 25 years in New South Wales, Australia, from 1994 to 2010. The dotted line is the overall mean trend and the solid lines are the selected hospital-specific trends. 172
5.3 Various structures of the $\boldsymbol{Z}$ matrix across logistic mixed models (5.4), (5.9) and (5.11), with the number of groups $m = 2$. The $\boldsymbol{Z}$ matrix starts off as having a simple block-diagonal structure and, as the model increases in complexity, it grows into a "nested" block-diagonal structure. The definitions of matrices are described in Subsection 5.2.2. 175
5.4 The approximate posterior density functions obtained from MFVB and MCMC for a single replication of the simulation study described in the text. Each pair of density function corresponds to a model parameter $f(Q_k)$, $1 \leq k \leq 3$ and $g_i(Q_k)$, $k = 2$, where Q_k is the kth sample quintile of the xs. The vertical lines represent the true parameter values. The accuracy scores on the top right of on each plot show the accuracy of MFVB approximation compared against a MCMC benchmark. 178
5.5 Side-by-side boxplots of accuracy scores for the MFVB approximation compared against MCMC over 30 runs. Each boxplot corresponds to a model parameter $f(Q_k)$, $1 \leq k \leq 3$ and $g_i(Q_k)$, $k = 2$, where Q_k is the kth sample quintile of the xs. 179

5.6 The MFVB fitted hospital-specific probability functions of caesarean section for low-risk nulliparous women of aged less than 25 years and those of aged greater or equal to 25 years, as a function of time for each hospital. The dashed curves represent pointwise 95% credible sets. Each panel corresponds to a different hospital. 180
5.7 The MFVB estimated overall and selected hospital-specific contrast curves defined by (5.12), corresponding to the odds of caesarean section for low-risk nulliparous women aged greater or equal to 25 years, as a function of time, compared with those aged less than 25 years. The shaded regions correspond to pointwise 95% credible sets. 181
5.8 The MFVB estimated hospital-specific odds ratios of caesarean section for low-risk nulliparous women of older age compared with those of younger age in the year 2010. 182
5.9 Directed Acyclic Graph for the Bernoulli group-specific curve model with contrasting. The shaded node corresponds to the observed data vector and the open nodes correspond to the random or auxiliary variables. The fragments shown in grey are present in the Bernoulli model DAG graph as shown in Figure 2.4. 186

6.1 A factor graph corresponding to the function $g(x_1, x_2, x_3, x_4, x_5)$ defined by (6.3). 197
6.2 A factor graph corresponding to a Bayesian model with stochastic nodes $\boldsymbol{\theta}_1, \ldots \boldsymbol{\theta}_9$ and factors $f_1, \ldots f_{10}$. 199
6.3 Each of the messages between neighbouring nodes on the factor graph for the random intercepts and slopes model illustrative example. The pink arrows depict the direction from the stochastic node to factor, while the grey arrows depict the direction from the factor to stochastic node. 201
6.4 Diagrammatic depiction of the extension from simple mixed effects regression to three-level penalised spline mixed regression. The fragments shown in darker blue are the ones that are additional to the fragments in the simple mixed model. The fragments shown in grey are presented in Figure 6.3. . . 208

List of Tables

1.1 Distributions used in this thesis and their corresponding probability density or mass functions. 19

2.1 Average (standard error) elapsed of the computing times in seconds for MCMC and MFVB fitting of the two-level Bayesian semiparametric mixed models to the simulated data. 60
2.2 Percentage coverage of true parameter values by approximate 95% credible sets based on the MFVB approximate posterior density functions. The percentages are based on 1000 replications. 64

3.1 Average (standard error) elapsed of the computing times in seconds for the simulation described in the text, using the naïve Algorithm 3, streamlined Algorithm 7 and `gam`() function in R `mgcv` package. The ratio of naïve over streamlined and the ratio of `gamm()` over streamlined are also presented. . . 101
3.2 Description of the United States National Center for Health Statistics perinatal health data as presented in Abrevaya (2006). 102
3.3 Description of the 2000 Program for International Student Assessment conducted by the Organisation for Economic Cooperation and Development. . 103
3.4 Description of the German health care data as presented in Winkelmann (2004). 104

4.1 Percentage coverage of true parameter values by approximate 95% credible sets based on MFVB approximate posterior density functions. A value of $RR = 0.9$ corresponds to a small amount of measurement error and $RR = 0.7$ corresponds to a substantial corruption of the predictor. Low missingness for the MNAR model corresponds to $(\phi_0, \phi_1) = (2.95, -2.95)$ and high missingness to $(\phi_0, \phi_1) = (0.85, -1.05)$. The percentages are based on 1000 replications. 140

4.2 Average (standard error) elapsed of the computing times in seconds for the MFVB and MCMC fitting of the two-level Bayesian semiparametric mixed models with Gaussian response subject to classical measurement error or missing data problem. 140

5.1 Average (standard error) elapsed of the computing times in seconds for the streamlined MFVB Algorithm15 in the simulation setting described in (5.14).177

5.2 The MFVB estimated odds ratios of caesarean section (averaged across hospitals) for low-risk nulliparous women aged greater than or equal to 25 years compared with those aged less than 25 years for selected years of birth.182

6.1 Expressions for natural statistics, natural parameter vectors and natural canonical forms of common exponential family densities. 195

Abstract

Generalised linear mixed models are the cornerstone of longitudinal and multilevel data analysis. However, exact inference for Bayesian mixed models with semiparametric extensions is typically intractable, requiring approximate inference methods for use in practice. Markov chain Monte Carlo or MCMC is one of the most commonly used approximate inference methods in this setting, but can be computationally intensive and often suffers from poor convergence in complex models. A faster, deterministic alternative to MCMC is variational approximations, a class of deterministic algorithms that is based on reformulating the problem of computing the posterior distribution as an optimisation problem, simplifying that problem and finding solutions to the perturbed problem. In this thesis, we work with a particular class of variational approximations, known as the mean field variational Bayes (MFVB). In essence, MFVB approximations are based upon optimising the Kullback-Leibler divergence with respect to the so-called approximating distribution. We derive MFVB algorithms for a wide variety of Bayesian semiparametric mixed models with Gaussian, Student-t, Bernoulli and Poisson responses. In order to overcome the computational cost of the direct naïve approach to the underlying MFVB calculations for models, we introduce a novel, streamlined approach that involves matrix permutation and block decomposition. Through a series of numerical studies, we demonstrate that the MFVB algorithms achieve a good level of accuracy compared to a MCMC benchmark (our gold standard). Furthermore, our developed streamlined algorithms are shown to have a complexity that is linear in the number of groups at each level, representing a two orders of magnitude improvement over the naïve approach. More importantly, the modularity of MFVB allows relatively simple extensions to more complicated scenarios, including higher-level random effects, measurement error and/or missing data problems, models with group-specific curves and real-time or online data processing. Illustrations from various real data examples are provided.