NeuroImage xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Technical Note

Bayesian model reduction and empirical Bayes for group (DCM) studies

Karl J. Friston [a], Vladimir Litvak [a], Ashwini Oswal [a], Adeel Razi [a,b], Klaas E. Stephan [a,c], Bernadette C.M. van Wijk [a], Gabriel Ziegler [a], Peter Zeidman [a,*]

[a] The Wellcome Trust Centre for Neuroimaging, UCL, 12 Queen Square, London, UK
[b] Department of Electronic Engineering, NED University of Engineering & Technology, Karachi, Pakistan
[c] Translational Neuromodeling Unit (TNU), Institute for Biomedical Engineering, University of Zurich and ETH Zurich, Switzerland

ARTICLE INFO

Article history:
Received 12 March 2015
Accepted 6 November 2015
Available online xxxx

Keywords:
Empirical Bayes
Random effects
Fixed effects
Dynamic causal modelling
Classification
Bayesian model reduction
Hierarchical modelling

ABSTRACT

This technical note describes some Bayesian procedures for the analysis of group studies that use nonlinear models at the first (within-subject) level – e.g., dynamic causal models – and linear models at subsequent (between-subject) levels. Its focus is on using Bayesian model reduction to finesse the inversion of multiple models of a single dataset or a single (hierarchical or empirical Bayes) model of multiple datasets. These applications of Bayesian model reduction allow one to consider parametric random effects and make inferences about group effects very efficiently (in a few seconds). We provide the relatively straightforward theoretical background to these procedures and illustrate their application using a worked example. This example uses a simulated mismatch negativity study of schizophrenia. We illustrate the robustness of Bayesian model reduction to violations of the (commonly used) Laplace assumption in dynamic causal modelling and show how its recursive application can facilitate both classical and Bayesian inference about group differences. Finally, we consider the application of these empirical Bayesian procedures to classification and prediction.

Introduction

This paper introduces some potentially useful procedures for the analysis of data from group studies using nonlinear models; for example, dynamic causal models of neurophysiological timeseries. Its key contribution is to finesse the problems that attend the inversion or fitting of hierarchical models in a nonlinear setting. This is achieved by using *Bayesian model reduction* that allows one to compute posterior densities over model parameters, under new prior densities, without explicitly inverting the model again. For example, one can invert a nonlinear (dynamic causal) model for each subject in a group and then evaluate the posterior density over group effects, using the posterior densities over parameters from the single-subject inversions. This application can be regarded as a generalisation of the standard summary statistic approach; however, instead of just using point estimators as summaries of first (within-subject) level effects, one can take the full posterior density to the second (between-subject) level. Furthermore, this *empirical Bayes* procedure can be applied to any model inversion scheme that furnishes posterior densities, which can be summarised with a multivariate Gaussian distribution.

Bayesian model reduction refers to the Bayesian inversion and comparison of models that are reduced (or restricted) forms of a full (or parent) model. It can be applied whenever models can be specified in terms of (reduced) prior densities. A common example would be switching off a parameter in a full model by setting its prior mean and variance to zero. The important aspect of Bayesian model reduction is that models differ only in their priors, which means that the posterior of a reduced model can be derived from the posterior of the full model. In this paper, we will use Bayesian model reduction to evaluate empirical priors to provide an *empirical Bayesian model reduction* scheme.

Empirical Bayes refers to the Bayesian inversion or fitting of hierarchical models. In hierarchical models, constraints on the posterior density over model parameters at any given level are provided by the level above. These constraints are called *empirical priors* because they are informed by empirical data. In this paper, we will consider an empirical Bayesian approach to any hierarchical model that can be expressed in terms of an arbitrary (nonlinear) model at the first level and a standard (parametric) empirical Bayesian (PEB) model at higher levels (Efron and Morris, 1973; Kass and Steffey, 1989). In other words, if the parameters of a nonlinear model of subject-specific data are generated by adding random (Gaussian) effects to group means, then the procedures of this paper can be applied. Crucially, these procedures are very efficient because each hierarchical level of the model requires only the posterior density over the parameters of the level below. This means, one can invert deep hierarchical models without having to revisit lower levels. This aspect of the scheme rests on Bayesian model reduction, a procedure that we have

* Corresponding author at: The Wellcome Trust Centre for Neuroimaging, Institute of Neurology, 12 Queen Square, London, UK, WC1N 3BG.
E-mail address: peter.zeidman@ucl.ac.uk (P. Zeidman).

http://dx.doi.org/10.1016/j.neuroimage.2015.11.015
1053-8119/

previously described in the context of *post hoc* model optimisation and discovery (Friston and Penny, 2011; Friston et al., 2011; Rosa et al., 2012). Here, it is put to work in the context of empirical Bayes and, as we will see later, evaluating predictive posterior densities for classification.

We envisage empirical Bayesian model reduction will be applied primarily to group Dynamic Causal Modelling (DCM) studies, where subjects are assigned to groups according to factors such as behaviour, diagnosis or genetics (e.g. Bernal-Casas et al., 2012). However, the ideas presented here are not limited to DCM. They can be applied to any nonlinear model and, interestingly, any inversion scheme at the first (within-subject) level. This may be particularly important for harnessing the computational investment of schemes that use stochastic methods to evaluate first level posteriors (Sengupta et al., in press). Bayesian model reduction resolves (or at least frames) a number of issues in the inversion and interpretation of group DCM studies. Consequently, we will take the opportunity to illustrate some of these issues using a worked example. These include the problem of local maxima when evaluating different models for Bayesian model comparison—and the fundamental distinction between random (between-subject) effects at the level of models and their parameters. In contrast to our previous treatment of random model effects at the between-subject level (Stephan et al., 2009), this paper considers random parameter effects in the setting of parametric empirical Bayes. We will also look at the fundamental difference between classical and Bayesian inference about group effects. Finally, we will briefly consider Bayesian classification of single subjects and touch on (leave-one-out) cross validation.

This paper comprises four sections. The first reviews Bayesian model reduction and introduces its application in a hierarchical or empirical Bayesian setting. This section reviews the basic theory, which generalises conventional approaches to random effects modelling. The second section applies the theory of the first to group studies, providing specific expressions for the procedures used in subsequent sections. The third section considers Bayesian model reduction using a worked example based on a (simulated) DCM study of mismatch negativity. The focus of this section is the utility of Bayesian model reduction in finessing (e.g., local maxima) problems that are often encountered when inverting nonlinear models. We will see that Bayesian model reduction provides more robust estimates of posterior probabilities than fitting models to the data separately, because it is less susceptible to violations of (e.g., Laplace) assumptions. This application of Bayesian model reduction provides Bayesian model averages that could be used for classical inference with the standard summary statistic approach, which we illustrate using canonical covariates analysis. However, one can go further in terms of model comparison and classification using empirical Bayesian model reduction. The last section revisits the worked example to illustrate model comparison and averaging at the second (between-subject) level. Our focus here is on inference about group effects and classification using the posterior predictive density afforded by empirical priors. The worked example was chosen to be representative of real DCM studies—so that the procedures could be illustrated in a pragmatic way. We will therefore refer to specific (Matlab) routines that implement the procedures. These routines are part of the academic SPM software available from http://www.fil.ion.ucl.ac.uk/spm.

Methods and theory

Bayesian model reduction

Bayesian model reduction refers to the Bayesian inversion of reduced models using only the posterior densities of a full model. Bayesian model reduction provides an efficient way to invert large numbers of (reduced) models, following the (usually computationally expensive) inversion of a full model. Consider a generative model that is specified in terms of its likelihood and priors. For example, models with additive Gaussian noise have the following form:

$$\begin{aligned}\ln p(y,\theta|m) &= \ln p(y|\theta,m) + \ln p(\theta|m)\\ p(y|\theta,m) &= \mathcal{N}(\Gamma(\theta),\Sigma(\theta))\\ p(\theta|m) &= \mathcal{N}(\eta,\Sigma)\end{aligned} \tag{1}$$

Here, $\Gamma(\theta)$ is a possibly nonlinear mapping from the parameters of a model to the predicted response y. Gaussian assumptions about observation noise, with a parameterised covariance $\Sigma(\theta)$, define the likelihood model that, when equipped with (Gaussian) priors, specifies the generative model. The generative model provides a probabilistic mapping from model parameters to observed data. Inference corresponds to the inversion of this mapping; from data to parameters. Usually, this inversion uses some form of approximate Bayesian inference.

Approximate Bayesian inference can always be cast as maximising the (negative) variational free energy with respect to the sufficient statistics $\tilde{q}$ of an approximate posterior $q(\theta|\tilde{q})$: see (Roweis and Ghahramani, 1999; Friston, 2008) for a fuller discussion. In this paper, a tilde (~) denotes the set of sufficient statistics of the prior $\tilde{p}$ and posterior $\tilde{q}$. Under the Laplace assumption (used throughout this work), the sufficient statistics correspond to the mean and covariance of each density. Using $\tilde{p}=(\eta,\Sigma)$ for the sufficient statistics of the prior, approximate Bayesian inference therefore corresponds to the optimisation problem:

$$\begin{aligned}\tilde{q}^* &= \arg\max_q F(\tilde{p},\tilde{q})\\ F(\tilde{p},\tilde{q}) &= \underbrace{E_q[\ln p(y|\theta)]}_{\text{accuracy}} - \underbrace{D_{KL}[q(\theta|\tilde{q})||p(\theta|\tilde{p})]}_{\text{complexity}}\end{aligned} \tag{2}$$

Here, we have expressed the free energy in terms of accuracy (first term) and complexity (second term), which is the Kullback–Leibler divergence between the (approximate) posterior and prior. Usually, this optimisation would proceed using a Fisher scoring scheme or related gradient ascent: see (Friston et al., 2007) and the appendix for details. After the negative free energy has been maximised the following approximate equalities provide an estimate of the posterior density over unknown model parameters and the log evidence or (marginal) likelihood of the model itself:

$$\begin{aligned}q(\theta|\tilde{q}^*) &\approx p(\theta|y,\tilde{p})\\ F(\tilde{p},\tilde{q}^*) &\approx \ln p(y|\tilde{p})\end{aligned} \tag{3}$$

By expressing the free energy as a function of the sufficient statistics of the prior and approximate posterior, it can be seen that the free energy depends on the prior, which in turn, specifies our beliefs about a model.

Now, say we wanted to estimate the posterior under a new model after eliminating some parameters to produce a reduced model. This is commonplace in classical statistics and corresponds to evaluating the treatment and residual sum of squares for a new contrast of parameters. Exactly the same idea can be applied to Bayesian inference. This rests upon the definition of a reduced model as a likelihood model with reduced priors. Consider Bayes rule replicated for reduced and full models (m_R, m_F):

$$\begin{aligned}p(\theta|y,m_R) &= \frac{p(y|\theta,m_R)p(\theta|m_R)}{p(y|m_R)}\\ p(\theta|y,m_F) &= \frac{p(y|\theta,m_F)p(\theta|m_F)}{p(y|m_F)}\end{aligned} \Leftrightarrow \begin{aligned}p(\theta|y,m_R)p(y|m_R) &= p(y|\theta,m_R)p(\theta|m_R)\\ p(\theta|y,m_F)p(y|m_F) &= p(y|\theta,m_F)p(\theta|m_F)\end{aligned} \tag{4}$$