NeuroImage 84 (2014) 97–112

Contents lists available at ScienceDirect

NeuroImage

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

Incorporating spatial dependence into Bayesian multiple testing of statistical parametric maps in functional neuroimaging

D. Andrew Brown [a,*], Nicole A. Lazar [a], Gauri S. Datta [a], Woncheol Jang [b], Jennifer E. McDowell [c]

[a] Department of Statistics, University of Georgia, Athens, GA 30602, USA
[b] Department of Statistics, Seoul National University, Seoul, Republic of Korea
[c] Department of Psychology, University of Georgia, Athens, GA 30602, USA

ARTICLE INFO

Article history:
Accepted 13 August 2013
Available online 24 August 2013

Keywords:
Multiple testing problem
Bayesian statistics
Conditional autoregressive model
False discovery rate
fMRI
Saccades

ABSTRACT

The analysis of functional neuroimaging data often involves the simultaneous testing for activation at thousands of voxels, leading to a massive multiple testing problem. This is true whether the data analyzed are time courses observed at each voxel or a collection of summary statistics such as statistical parametric maps (SPMs). It is known that classical multiplicity corrections become strongly conservative in the presence of a massive number of tests. Some more popular approaches for thresholding imaging data, such as the Benjamini–Hochberg step-up procedure for false discovery rate control, tend to lose precision or power when the assumption of independence of the data does not hold. Bayesian approaches to large scale simultaneous inference also often rely on the assumption of independence. We introduce a spatial dependence structure into a Bayesian testing model for the analysis of SPMs. By using SPMs rather than the voxel time courses, much of the computational burden of Bayesian analysis is mitigated. Increased power is demonstrated by using the dependence model to draw inference on a real dataset collected in a fMRI study of cognitive control. The model also is shown to lead to improved identification of neural activation patterns known to be associated with eye movement tasks.

Published by Elsevier Inc.

Introduction

Functional neuroimaging provides a set of tools that record state-related brain signals that are subsequently used to generate maps of the neural circuitry activation associated with that state. The data collected in such imaging studies are of massive scale, exacerbating problems commonly encountered in statistical analyses. The structure of these data requires hypothesis tests on a large number of parameters simultaneously to infer the presence or absence of signal changes throughout the brain. Testing thousands of locations simultaneously greatly increases the chances of some data spuriously exhibiting a signal by random variation. Increased risk of false positives must be balanced against overcorrection of the problem, which can make it difficult to see positive signals at all. Complicating the situation is the fact that imaging data tend to be spatially correlated so that there is redundant information between the locations being tested. Failure to account for this dependence structure can have an adverse effect on the ability to detect true signals. The need to balance a high risk for false positives with sensitivity for signal detection then becomes a foremost concern, as does the development of appropriate techniques for dealing with spatial dependence.

A breakthrough in the statistical analysis of massive data sets came when Benjamini and Hochberg (1995) introduced the false discovery rate (FDR) along with a simple procedure for its control. Introduced to the neuroscience community by Genovese et al. (2002), the procedure controls the expected proportion of discoveries (hypothesis rejections) that are false, as opposed to controlling the overall probability of committing any Type I error, as is accomplished with the strong family-wise error rate. This makes it much easier to scale up to larger data sets without becoming overly conservative. It should be noted, though, that even this procedure will fail if the wrong null distribution is used to calculate *p*-values. This can mean the difference between identifying hundreds of interesting cases and none at all (Efron, 2007).

Methods have been proposed to remedy the null distribution problem, including computationally-intense permutation tests (Holmes et al., 1996; Nichols and Holmes, 2001) or empirical estimation of the null distribution of transformed *p*-values (Efron, 2004). Image analysis based on identifying clusters of activated voxels is introduced in Forman et al. (1995). Thresholding via the expected Euler characteristic of excursion sets is explored in Worsley (2003). Nichols and Hayasaka (2003) compare permutation tests, Bonferroni, and random field theory methods for controlling the family-wise error rate in functional neuroimaging. A conditional version of FDR, the positive false discovery rate, is introduced in Storey (2003), along with a Bayesian interpretation of it. A recent overview of techniques for large-scale testing is Efron (2010), including thresholding via local false discovery rates and combining results to draw inference on sets of observations with enrichment analysis.

* Corresponding author.
E-mail address: brownda@uga.edu (D.A. Brown).

1053-8119/$ – see front matter. Published by Elsevier Inc.
http://dx.doi.org/10.1016/j.neuroimage.2013.08.024

Bayesian statistics also provides means of dealing with high throughput data, affording researchers a way to richly model complex phenomena while maintaining interpretability. Early work in the Bayesian analysis of brain images appears in Genovese (2000), Gössl et al. (2001), and Friston and Penny (2003). In the context of analyzing statistical parametric maps (SPMs; Friston et al., 1995) in fMRI, Marchini and Presanis (2004) remark that both Bayesian and FDR procedures share the common characteristic of adapting to the features of the data, rather than prescribing a fixed threshold rule. They conclude, however, that the use of Bayesian modeling is the most powerful approach for identifying regions of activation when compared to thresholding via FDR control or random field theory. More recent work includes Smith and Fahrmeir (2007), who use a binary Markov random field to model spatial correlation among the voxel-specific indicators of activation, and Bowman et al. (2008), who take advantage of hierarchical structures to separate local dependence and inter-regional correlation of predetermined regions of interest. SPMs are treated as arising from cluster point processes with both population-level and individual-level centers in Xu et al. (2009), allowing for the estimation of the proportion of individuals exhibiting activation at certain locations. Point processes are also used in Kang et al. (2011) for Bayesian meta-analysis of locations of reported foci from imaging studies and the variability among participants. A review of Bayesian procedures in fMRI may be found in Woolrich (2012).

One of the biggest obstacles to Bayesian inference being more widely accepted in the neuroimaging community is the prohibitive computational burden it often imposes. This computing problem has become one of the primary concerns for researchers who work with massive data sets but still need reasonable computation time to get results. Early attempts to deal with this in fMRI analysis are Penny et al. (2003) and Penny et al. (2005), who use variational Bayes to obtain approximations to posterior distributions. Friston and Penny (2003) suggest using empirical Bayes methods as opposed to fully hierarchical Bayes to reduce computational loads. Some empirical Bayes approaches to the more general multiple testing problem may be found in, e.g., Bogdan et al. (2008), Muralidharan (2010), and Martin and Tokdar (2012). See Scott and Berger (2010) for a discussion of conditions under which a multiplicity adjustment can be induced with both empirical Bayes and hierarchical Bayes.

Many of the Bayesian models currently found in the imaging literature rely on modeling the entire time series collected at each voxel (e.g. Fahrmeir and Gössl, 2002; Genovese, 2000; Gössl et al., 2001; Harrison et al., 2008; Penny et al., 2003, 2005; Smith and Fahrmeir, 2007). The Bayesian model presented in this article works directly with reduced imaging data. In particular, each observation is a test statistic quantifying the change in blood-oxygenation-level-dependent (BOLD) signal over the course of an fMRI experiment, averaging over the temporal dimension and thus vastly reducing the size of the data sets to be analyzed. By applying Bayesian thresholding to SPMs, researchers gain the added flexibility of modeling complex spatial and hierarchical structures, while maintaining reasonable computation times.

While some work has been done in modeling the spatial structure of fMRI data (e.g. Bowman et al., 2008; Gössl et al., 2001; Hartvig and Jensen, 2000; Smith and Fahrmeir, 2007), much of the work thus far developed is based on the assumption of independence of the data. This assumption can yield quite different results from what would otherwise be obtained by accounting for the true dependence structure. Both FDR and Bayesian procedures are influenced by the assumed correlation (or lack thereof). While FDR control still works under positive regression dependence with the original Benjamini–Hochberg procedure and under arbitrary dependence structures using the algorithm introduced by Benjamini and Yekutieli (2001), it still tends to lose accuracy under dependence (Efron, 2007). Bayesian procedures may also suffer from falsely assuming independence, becoming overly conservative when modeling data that are truly dependent.

In this work we extend the Bayesian multiple testing model considered in Scott and Berger (2006) by modeling a continuous, underlying signal that manifests in areas of the brain that are activated during a stimulus. Spatial dependence is introduced via a Gaussian autoregressive model on the signal. This facilitates the sharing of information between voxels. We demonstrate this model's ability to improve upon detection of task-related activation. In particular, we show how the spatial correlation induces the identification of larger clusters of activated regions, which carries more physiological meaning than individually-selected voxels.

The remainder of this paper is organized as follows: the Motivation section introduces the data collected during an fMRI study of cognitive control. This data set is used as motivation for our model. We review in the Methods section the relevant background and necessary methods for constructing our spatial Bayesian testing model and propose the model itself. We perform a simulation study and present the results in the Simulation study section, analyzing the performance of both the Scott–Berger model and our own on simulated spatial data. In the Results section, we show the results from applying the model to the original motivating data set. We compare the results to the Bayesian model under an independence assumption as well as the results obtained from false discovery rate control under an arbitrary dependence structure. We conclude in the Discussion section with a discussion of these results and commentary regarding future research directions.

Motivation

It is a common practice to analyze BOLD fMRI data using a voxelwise general linear model (GLM) (Friston et al., 1995). This approach models the MR signal at voxel j using a model of the form

$$\mathbf{y}_j = X\beta_j + \mathbf{e}_j,$$

where $\mathbf{y}_j = (y_{j,1}, \ldots, y_{j,T})^T$ is the time course of MR signals at voxel j, X is the $T \times p$ design matrix, β_j is the $p \times 1$ vector of regression coefficients, and $\mathbf{e}_j$ is the (often serially correlated) error in the measurements. Inference then focuses on the coefficients corresponding to factors of interest, such as the stimulus time course, to determine which voxels are significantly associated with the experimental stimulus. Analyzing the full time course at each voxel in a three-dimensional image can involve processing tens of millions of observations, creating huge computational demands (Friston and Penny, 2003; Smith and Fahrmeir, 2007).

Alternatively, researchers may choose to analyze fMRI data through the use of statistical parametric maps. With this approach, each voxel is assigned a single summary statistic quantifying the effect of the factor of interest. Analysis then involves performing inference on the observed statistic at each voxel, which has a known distribution under a null hypothesis. One of the advantages of the SPM approach is that the data to be processed are collapsed over the temporal dimension. The analysis is thus computationally simpler than a full voxelwise GLM analysis.

The data we consider here are from a study by Camchong et al. (2008) investigating the differences in neural activation patterns associated with cognitive control tasks that require generation of volitional saccade. The task involved alternating between blocks of fixation (baseline) and the volitional saccade task known as an antisaccade. Antisaccades require that participants inhibit a glance towards a prepotent cue and generate one to the mirror image location (opposite side of the screen, same distance from center). During fixation blocks participants fixed gaze on a target point for a duration of 22.5 s. For the antisaccade blocks, a single central target was presented for 1.7 s followed by a cue presented 8° to the left or right for 1.25 s. The participants were asked to move their eyes to the mirror image of the cue as quickly and as accurately as possible. Each run consisted of nine fixation periods alternated with eight antisaccade trials. Two blocked runs were recorded for each participant.

Data were obtained using a GE Signa Horizon LX 1.5 T MRI scanner. Each functional run collected T_2^*-weighted images with in-plane resolution 3.75 × 3.75 mm, TE = 40 ms, TR = 1912 ms, 3.8 second