Variance Components Analysis of a Multi-Site fMRI Study

Technical Report No. 04-14
Information and Computer Science Department,
University of California, Irvine

Seyoung Kim, Hal Stern, Padhraic Smyth, First BIRN

Information and Computer Science
University of California, Irvine
CA 92697-3425
sykim@ics.uci.edu, sternh@uci.edu, smyth@ics.uci.edu

1 Introduction

Data from a pilot study of human control subjects collected by the FIRST BIRN (Functional Imaging Research on Schizophrenia Test-bed Biomedical Informatics Research Network), also known as Function BIRN or fBIRN, provide a unique opportunity to assess the sources of variability in fMRI data. MR scans of five control subjects were recorded at 11 different sites over two visits to each site. Each visit comprised ten different runs/tasks (four runs of a sensorimotor task and two runs each of a cognitive task, breath holding task, and rest task). In this report we analyze the data from the sensorimotor task. Variability is observed among different runs in a single visit, between visits at a single site, among sites and among subjects. The goal of our analysis is to assess the contributions of these different sources of variability, providing information that can be used in the design and analysis of future multi-site studies.

The raw functional images were preprocessed using SPM. Preprocessing steps included correcting each task for head motion, registration to the subject's high-resolution anatomical scan, normalization to a common brain shape, and spatial smoothing. Each run was analyzed using a general linear model that yields an estimated effect β for each predictor included in the linear model. For example, the sensorimotor task involves a sequence of 30 second blocks in which the subject alternates performing the task and resting. The regression coefficient corresponding to the task indicator (on/off) is one measure of functional response for the task. Alternative functional response measures, e.g., percent signal response or t-score, are also possible.

Our focus is on analyzing the β-maps of the sensorimotor task to decompoe the total variance into components attributable to the different sources described above. In Section 2, we describe a variance component model that we use in this work. In Section 3, we present results for the sensorimotor task, and conclude with some discussion in Section 4.

2 Variance Component Model for a Multi-Site fMRI Study

A traditional variance components analysis can be used to decompose the observed variation in signal into portions attributable to subjects, sites, their interaction, day-to-day variability, and run-to-run variability within a visit. One such model that we choose to use in our analysis is

$$Y_{ijkl} = \mu + a_i + b_j + ab_{ij} + v_{ijk} + r_{ijkl},$$

where Y_{ijkl} is the response measure, μ is an overall mean, a_i is a subject effect, b_j is a site effect, ab_{ij} allows for subject-site interactions, v_{ijk} is a visit effect and r_{ijkl} is a run effect (essentially all sources of variation not included in previous terms). We use the notation Y_{ijkl} for the response which varies in the analyses that follow depending on the size and location of the brain region being considered. Each of the effects (subject, site, interaction, visit, run) is further modeled as a Gaussian random variable with mean zero and variance parameter that characterizes the contribution to overall response variability of the particular source. It is the variance parameters, or the variance components, that are of interest.

Parameters are estimated as medians of the posterior distribution of the variance components from a Bayesian analysis with weak prior distributions on the variance components. The prior distributions for the analyses reported here are $N(0, 10^{10})$ for μ (a common vague prior for the mean of a Gaussian distribution) and gamma$(0.01, 0.01)$ on the precision parameters (reciprocals of the variances). The latter is a gamma distribution with mean 1 and variance 100. We are currently studying sensitivity to this choice of the prior distribution for the variance parameters. Samples of the variance parameters from their posterior distribution given the data are obtained using WinBugs, a freely available software implementation of the Gibbs sampler. The Gibbs sampler was run for 500,000 iterations by which time the draws were determined to be representative of the posterior distribution. The medians of the last 100,000 draws are reported as estimates for the variance components.

3 Results

We next provide results from variance components analyses of the sensorimotor tasks using the linear model in Section 2. Data from eight sites were available when this study was begun and thus we restrict attention to those eight sites throughout. Six regions of interest (ROIs) that are known to be activated by sensorimotor tasks were identified by our collaboraters and voxels in these regions extracted from β-maps of the entire brain. These ROIs include left and right precentral gyrus (thought to be activated during motor activity), left and right superior temporal gyrus (thought to be activated by auditory response), and left and right occipital lobe (thought to be activated for visual response). The response Y_{ijkl} varies in the analyses that follow. We consider analyses of data from a single voxel in each ROI, a $5 \times 5 \times 5$ cube surrounding that single voxel, or the entire brain region. For the whole region we perform separate analyses for the overall mean value and for the average of the highest 10% (most active) β values.

3.1 Variance Components Results

Table 1 gives the estimated variance components and the proportion of the total variance (the sum of the estimated components) attributable to each component for the average β over two regions of interest (left precentral gyrus, left superior temporal gyrus). The proportions of the total variance attributable to each source are easier to interpret than the actual variance components so we concentrate on these. The table (and similar tables for other tasks and regions) indicates that 15-30% of the observed variation is visit-to-visit and run-to-run variability for the same person at the same site. The largest source of variance is inter-individual variability but there is also substantial variance by site. Table 1 also gives estimated results for the active regions, defined here as the top 10% of the regression coefficients. Note the variance components are larger when we focus on the active voxels due to the fact that the top 10% does not include the large numbers of non-activated voxels with near zero effect. Interestingly, the variance proportions are similar for the overall means and the active voxels. We also note that in both the average of the entire region and the top 10%, the variance proportions are similar across regions.

Table 1: Variance Components for entire brain regions

	Left PCG (motor)		Left STG (auditory)	
	variance	proportion	variance	proportion
Region avg.	mean β value is .49		mean β value is .72	
subj	.0124	.46	.0125	.42
site	.0056	.21	.0056	.19
subj.site	.0043	.16	.0030	.10
visit	.0017	.06	.0033	.11
run	.0027	.10	.0054	.18
Top 10% avg.	mean β value is.553		mean β value is. 835	
subj	.0112	.12	.0332	.15
site	.0309	.32	.0562	.26
subj.site	.0297	.31	.0934	.43
visit	.0197	.20	.0173	.08
run	.0056	.06	.0172	.08

Table 2: Varying the size of regions (left precentral gyrus)

	Variance estimate			Proportion of total variance		
Region (with mean β)	One voxel (.266)	$5 \times 5 \times 5$ cube (.248)	Whole region (.049)	One voxel	$5 \times 5 \times 5$ cube	Whole region
subj	.0264	.0243	.0124	.29	.34	.46
site	.0289	.0217	.0056	.32	.30	.21
subj.site	.0196	.0149	.0043	.22	.21	.16
visit	.0041	.0036	.0017	.05	.05	.06
run	.0105	.0075	.0027	.12	.10	.10

Table 3: Varying the number of runs used for estimation (The response is the mean of the left precentral gyrus.)

	Variance estimate			Proportion of total variance		
Runs (with mean β)	First 2 runs (.049)	Last 2 runs (.051)	All 4 runs (.049)	First 2 runs	Last 2 runs	All 4 runs
subj	.0119	.0129	.0124	.47	.45	.46
site	.0054	.0061	.0056	.22	.21	.21
subj.site	.0041	.0046	.0043	.16	.16	.16
visit	.0017	.0025	.0017	.07	.09	.06
run	.0022	.0029	.0027	.09	.10	.10

Table 2 explores the effect of the size of the region analyzed on the variance components. Variance components are computed using the functional response measure (regression coefficient) for a single voxel, the average of the voxel-wise regression coefficient over a $5 \times 5 \times 5$ cube surrounding the single voxel, and the average of the voxel-wise regression coefficients over the entire region. The Table gives results for the left precentral gyrus; again, similar results are obtained for other regions. Variance proportions are consistent between the single voxel and the $5 \times 5 \times 5$ cube, but the proportions for the whole region are different. The run-to-run variability and visit-to-visit variability, which speak most to the reproducibility of the fMRI measures, are quite consistent across the regions of different size whereas the contributions due to subject, site and their interaction vary more.

Table 3 reproduces the analysis of Table 1, computing variance components for the average effect over the entire region, but using the first two runs of the sensorimotor task and the last two runs separately. The results are remarkably consistent among the first two runs, the last two runs, and all of the four runs, and suggest that only two runs of the sensorimotor task during a visit are needed to provide representative data. The same result was found in other regions of the brain. This resulted supported the design of the protocol for future data collection which we include two sensorimotor runs to be used in calibrating results from different sites.

3.2 A Predictive Experiment

One goal of the fBIRN collaboration is to develop methods that facilitate pooling results across sites. To assess whether the variance components model may be a useful tool in this regard we carried out a small predictive experiment. The predictive experiment also provides some information about the quality of fit of the variance components model which assumes an additive relationship and no particular trends among runs/visits (i.e., no learning or boredom effects). Our predictive approach uses cross-validation. We remove data from one subject at one site, fit a variance components model on the remaining data, and use the fitted model to predict what would be seen for the removed combination of subject and site.

Table 4: Predictive ability (One voxel, left precentral gyrus)

		Observed mean and s.d.	Posterior predicted mean response	Naive predicted response
Subject 5	Site 2	.245 (.083)	.282	.304
	Site 3	.391 (.180)	.491	.381
	Site 4	.396 (.067)	.316	.312
	Site 5	.167 (.057)	.284	.308
	Site 6	.317 (.187)	.289	.305
	Site 7	.858 (.145)	.476	.357
	Site 8	.565 (.142)	.573	.408
	Mean Abs Error		.107	.138
Site 8	Subject 1	.486 (.062)	.472	.344
	Subject 2	.615 (.087)	.407	.310
	Subject 3	.791 (.153)	.462	.348
	Subject 4	.040 (.070)	.349	.244
	Subject 5	.565 (.142)	.573	.408
	Mean Abs Error		.174	.250

Table 5: Predictive ability ($5 \times 5 \times 5$ mean, left precentral gyrus)

		Observed mean and s.d.	Posterior predicted mean response	Naive predicted response
Subject 5	Site 2	.256 (.062)	.254	.277
	Site 3	.330 (.158)	.469	.358
	Site 4	.380 (.061)	.301	.290
	Site 5	.169 (.048)	.267	.285
	Site 6	.307 (.160)	.261	.277
	Site 7	.728 (.136)	.441	.333
	Site 8	.489 (.113)	.500	.365
	Mean Abs Err		.095	.115
Site 8	Subject 1	.343 (.053)	.390	.295
	Subject 2	.562 (.070)	.383	.298
	Subject 3	.704 (.120)	.420	.324
	Subject 4	.055 (.056)	.274	.212
	Subject 5	.489 (.113)	.500	.365
	Mean Abs Err		.148	.194

Table 4 gives the cross validation results on β values for a single voxel in the left precentral gyrus. Each row of the table represents a separate cross-validation run. The combination of subject and site shown in the first two columns is removed and the posterior predictive mean for the response of the omitted subject-site combination is reported along with the mean and standard deviation of the eight runs (four runs on each of two days) that were omitted. The averages over all other measurements for the removed subject and site are provided in the last column, and can be thought of as a naive prediction method that avoids modeling assumptions. In addition we give the mean absolute value of the prediction errors across the different sites for a single subject and across the different subjects at a single site. Table 5 show the same analysis using mean β values over the $5 \times 5 \times 5$ cubes in the left precentral gyrus.

In both Table 4 and Table 5, we find that the variance components model has more predictive ability than the naive method, suggesting that the model captures some important aspects of the data. However, the distance between the predicted response and the observed value is still quite large suggesting room for improvement.

4 Conclusions

We analyzed the amount of variability due to site, subject, visit, and run effects in fMRI data from the fBIRN study of human control subjects using a variance components model. In the sensorimotor task, the results were consistent over different regions of brain. The size of the region used to define the functional response affects the absolute size of the estimated variance components but not the proportions of the variance components relative to the total variance. Also, in all of our analysis, we found that 15-30% of variance in response (as measured by regression coefficient values) is run-to-run or visit-to-visit variation.

The results from our predictive experiment show that the additive model for variance components is useful despite its simplicity. To improve the predictive power, we will explore alternative variance component models and other predictive approaches.