Multi-level Modelling and Spatial Inference for Large-Scale Neuroimaging Data

Supplementary Material B

Thomas J. Maullin-Sapey

Submitted in support of the thesis "Multi-level Modelling and Spatial Inference for Large-Scale Neuroimaging Data".

S1 Illustrative Example of BLMM Setup

To illustrate the construction described in Section 4.3.1.1 of Chapter 4 in practice, we provide a brief mock example consisting of a "small" longitudinal group-level analysis. The analysis design contains 3 subjects, each of which has had multiple image acquisitions taken irregularly across repeated visits and has recorded observations for "sex", "age", and "BMI". Only one grouping factor for the random effects is included in the model, the factor "subject", and acquisitions for the first, second and third subjects were taken across 2, 3 and 2 visits, respectively. (It is emphasized that the number of subjects and observations used in this example are far too small to draw meaningful inference from in practice and that this example has only been constructed for illustrative purposes). For this example, if the user wished to model a between-subject random intercept and a between-subject random slope for age using BLMM, the input for the matrices X, g_1 and z_1, and the full construction of Z, may appear as follows:

$$X = \begin{bmatrix} 1 & 1 & 19 & 18.7 \\ 1 & 1 & 21 & 19.2 \\ 1 & 1 & 18 & 24.8 \\ 1 & 1 & 20 & 25.1 \\ 1 & 1 & 24 & 24.6 \\ 1 & 0 & 21 & 20.9 \\ 1 & 0 & 22 & 20.1 \end{bmatrix}, \quad g_1 = \begin{bmatrix} 1 \\ 1 \\ 2 \\ 2 \\ 2 \\ 3 \\ 3 \end{bmatrix}, \quad z_1 = \begin{bmatrix} 1 & 19 \\ 1 & 21 \\ 1 & 18 \\ 1 & 20 \\ 1 & 24 \\ 1 & 21 \\ 1 & 22 \end{bmatrix}, \quad Z = \begin{bmatrix} 1 & 19 & 0 & 0 & 0 & 0 \\ 1 & 21 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 18 & 0 & 0 \\ 0 & 0 & 1 & 20 & 0 & 0 \\ 0 & 0 & 1 & 24 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 21 \\ 0 & 0 & 0 & 0 & 1 & 22 \end{bmatrix},$$

where the four columns of X correspond to an intercept, sex, age and BMI respectively, g_1 indicates which observations correspond to which subjects, and z_1 includes an intercept and age. In this example, the input for the response vector, Y, would be a text file containing a list of the filenames of seven NIfTI images corresponding to each visit across all subjects. During input specification, the user may also specify hypothesis tests to be conducted. For example, to perform an approximate T-test testing for the non-zero effect of BMI in the above example, the user may specify T as the statistic type and $[0,0,0,1]$ as the contrast vector.

S2 Random Mask Generation

In this section, we detail the process used to generate the random masks which were applied to the simulated data of Section 4.3.2. To create randomly deformed masks for the simulated analyses, the standard T1 2mm MNI152 brain mask, which is available in the FSL software package, was employed. Mask generation began by first eroding the standard T1 2mm mask by approximately three voxels. This was achieved by smoothing the binary mask with an isotropic Gaussian kernel of 4 Full Width Half Maximum (FWHM) and then re-thresholding the resultant smooth image at 0.7. The purpose of this erosion was to create a mask that was approximately the same shape and size as the standard T1 2mm MNI152 brain mask but with some voxels missing near cortical boundaries and the edge of the brain. Following this, random deformations in the size and shape of the eroded analysis mask were generated. This process was achieved by multiplying an image of $N(0,1)$ Gaussian noise by 8, adding the result to the eroded T1 2mm mask and smoothing the resultant sum with an isotropic Gaussian kernel of 10 FWHM. The final image was then re-thresholded at 0.6. This process was established empirically to ensure that the masks produced had approximately the same quantity of voxels as a standard fMRI analysis mask, with enough random deformation present to exhaustively test the missingness handling capacity of BLMM.

Using the masks generated by this process, the final analysis mask in each simulation instance (following the application of a 50% missingness threshold, c.f. Section 4.3.1.1) contained on average 217930.6 voxels, with a standard error of approximately 127.0 voxels across simulation instances. This meant that the final analysis mask occupied approximately 21.8% of each NIfTI volume. As noted earlier, this is, by design, similar to the 22.8% of the NIfTI volume that is occupied by the original FSL T1 2mm brain mask and may be expected to be occupied in a standard fMRI analysis. On average, in each simulation instance, approximately 87483.9 voxels had missing data (40.1% of the total number of voxels in the analysis mask), and 130446.7 voxels had full observations across all volumes generated (59.9%). We emphasize here that this extreme degree of subject-mask variability was deliberately simulated to stress test the missingness handling and time efficiency capabilities of BLMM.

S3 Mean Absolute Difference for Parameter Estimation (n=200)

Method	Missing-data voxels	Full-data voxels	All voxels
Simulation 1			
β	6.86×10^{-9}	4.75×10^{-9}	5.45×10^{-9}
	(1.32×10^{-11})	(9.14×10^{-12})	(1.02×10^{-11})
σ^2	1.43×10^{-9}	1.06×10^{-9}	1.18×10^{-9}
	(3.10×10^{-12})	(2.21×10^{-12})	(2.46×10^{-12})
D	8.26×10^{-6}	5.20×10^{-6}	6.20×10^{-6}
	(1.67×10^{-8})	(1.10×10^{-8})	(1.24×10^{-8})
Simulation 2			
β	4.39×10^{-5}	2.34×10^{-5}	3.05×10^{-5}
	(2.05×10^{-7})	(1.43×10^{-7})	(1.62×10^{-7})
σ^2	6.12×10^{-6}	3.33×10^{-6}	4.29×10^{-6}
	(2.44×10^{-8})	(1.70×10^{-8})	(1.92×10^{-8})
D	2.54×10^{-4}	1.40×10^{-4}	1.79×10^{-4}
	(1.24×10^{-5})	(8.59×10^{-6})	(9.78×10^{-6})
Simulation 3			
β	1.70×10^{-5}	8.01×10^{-6}	1.11×10^{-5}
	(1.13×10^{-7})	(7.53×10^{-8})	(8.73×10^{-8})
σ^2	1.27×10^{-6}	5.44×10^{-7}	7.95×10^{-7}
	(7.13×10^{-9})	(4.11×10^{-9})	(5.09×10^{-9})
D	6.02×10^{-3}	2.93×10^{-3}	3.99×10^{-3}
	(4.10×10^{-5})	(2.72×10^{-5})	(3.17×10^{-5})

Table S1: Mean absolute difference in the estimates produced by lmer and BLMM for β, σ^2 and D, averaged across voxels and simulation instances. Each reported average is an image-wide mean, taken across approximately $218,000$ voxels, further averaged across 1000 simulation instances. In each simulation instance, the model employed included 200 observations generated according to the methods outlined in Section 4.3.2. Empirical standard errors, taken across simulation instances, are also given in brackets underneath each entry in the table.

S4 Mean Absolute Difference for Parameter Estimation (n=500)

Method	Missing-data voxels	Full-data voxels	All voxels
Simulation 1			
β	1.54×10^{-9}	1.09×10^{-9}	1.26×10^{-9}
	(1.96×10^{-12})	(1.52×10^{-12})	(1.60×10^{-12})
σ^2	3.48×10^{-10}	2.54×10^{-10}	2.89×10^{-10}
	(2.51×10^{-13})	(1.69×10^{-13})	(1.78×10^{-13})
D	1.13×10^{-6}	6.64×10^{-7}	8.40×10^{-7}
	(1.26×10^{-9})	(8.30×10^{-10})	(9.06×10^{-10})
Simulation 2			
β	2.00×10^{-6}	6.63×10^{-7}	1.17×10^{-6}
	(1.13×10^{-8})	(5.62×10^{-9})	(7.60×10^{-9})
σ^2	2.61×10^{-7}	8.41×10^{-8}	1.51×10^{-7}
	(1.00×10^{-9})	(4.19×10^{-10})	(6.15×10^{-10})
D	1.35×10^{-3}	4.84×10^{-4}	8.13×10^{-4}
	(7.24×10^{-6})	(3.81×10^{-6})	(4.98×10^{-6})
Simulation 3			
β	8.10×10^{-7}	3.02×10^{-7}	4.94×10^{-7}
	(5.60×10^{-9})	(2.55×10^{-9})	(3.60×10^{-9})
σ^2	3.53×10^{-8}	1.00×10^{-8}	1.96×10^{-8}
	(1.45×10^{-10})	(3.60×10^{-11})	(7.30×10^{-11})
D	2.76×10^{-4}	9.67×10^{-5}	1.64×10^{-4}
	(2.06×10^{-6})	(9.58×10^{-7})	(1.35×10^{-6})

Table S2: Mean absolute difference in the estimates produced by lmer and BLMM for β, σ^2 and D, averaged across voxels and simulation instances. Each reported average is an image-wide mean, taken across approximately $218,000$ voxels, further averaged across 1000 simulation instances. In each simulation instance, the model employed included 500 observations generated according to the methods outlined in Section 4.3.2. Empirical standard errors, taken across simulation instances, are also given in brackets underneath each entry in the table.

S5 Mean Absolute Difference for Parameter Estimation (n=1000)

Method	**Missing-data voxels**	**Full-data voxels**	**All voxels**
Simulation 1			
β	6.33×10^{-10}	4.19×10^{-10}	5.05×10^{-10}
	(7.66×10^{-13})	(5.85×10^{-13})	(6.20×10^{-13})
σ^2	1.57×10^{-10}	1.10×10^{-10}	1.29×10^{-10}
	(8.10×10^{-14})	(5.84×10^{-14})	(5.15×10^{-14})
D	4.96×10^{-7}	2.49×10^{-7}	3.48×10^{-7}
	(5.66×10^{-10})	(3.38×10^{-10})	(3.91×10^{-10})
Simulation 2			
β	6.54×10^{-8}	1.20×10^{-8}	3.36×10^{-8}
	(2.95×10^{-10})	(4.35×10^{-11})	(1.33×10^{-10})
σ^2	6.71×10^{-9}	1.24×10^{-9}	3.46×10^{-9}
	(1.55×10^{-11})	(1.45×10^{-12})	(5.93×10^{-12})
D	6.25×10^{-5}	1.65×10^{-5}	3.51×10^{-5}
	(3.32×10^{-7})	(9.35×10^{-8})	(1.80×10^{-7})
Simulation 3			
β	5.96×10^{-8}	2.66×10^{-8}	3.99×10^{-8}
	(3.10×10^{-10})	(1.32×10^{-10})	(1.71×10^{-10})
σ^2	5.80×10^{-9}	3.78×10^{-9}	4.60×10^{-9}
	(1.29×10^{-11})	(6.80×10^{-12})	(7.17×10^{-12})
D	2.97×10^{-5}	2.07×10^{-5}	2.44×10^{-5}
	(1.02×10^{-7})	(3.29×10^{-8})	(5.74×10^{-8})

Table S3: Mean absolute difference in the estimates produced by lmer and BLMM for β, σ^2 and D, averaged across voxels and simulation instances. Each reported average is an image-wide mean, taken across approximately $218,000$ voxels, further averaged across 1000 simulation instances. In each simulation instance, the model employed included 1000 observations generated according to the methods outlined in Section 4.3.2. Empirical standard errors, taken across simulation instances, are also given in brackets underneath each entry in the table.

S6 Mean Absolute Difference for Maximized ReML Criteria

Method	Missing-data voxels	Full-data voxels	All voxels
Simulation 1			
$n = 200$	8.72×10^{-10}	6.16×10^{-10}	7.02×10^{-10}
	(2.98×10^{-12})	(2.23×10^{-12})	(2.21×10^{-12})
$n = 500$	6.24×10^{-10}	7.36×10^{-10}	6.95×10^{-10}
	(1.87×10^{-11})	(1.82×10^{-11})	(1.32×10^{-11})
$n = 1000$	1.11×10^{-9}	8.23×10^{-10}	9.36×10^{-10}
	(4.92×10^{-11})	(6.02×10^{-11})	(4.04×10^{-11})
Simulation 2			
$n = 200$	4.25×10^{-3}	2.42×10^{-3}	3.05×10^{-3}
	(4.14×10^{-5})	(3.04×10^{-5})	(3.39×10^{-5})
$n = 500$	2.03×10^{-4}	6.96×10^{-5}	1.20×10^{-4}
	(2.43×10^{-6})	(1.28×10^{-6})	(1.63×10^{-6})
$n = 1000$	1.02×10^{-5}	2.46×10^{-6}	5.58×10^{-6}
	(3.99×10^{-7})	(2.34×10^{-7})	(2.12×10^{-7})
Simulation 3			
$n = 200$	1.11×10^{-3}	6.72×10^{-4}	8.23×10^{-4}
	(1.29×10^{-5})	(9.51×10^{-6})	(1.05×10^{-5})
$n = 500$	2.55×10^{-4}	1.75×10^{-4}	2.05×10^{-4}
	(2.61×10^{-6})	(1.87×10^{-6})	(1.74×10^{-6})
$n = 1000$	5.78×10^{-5}	3.43×10^{-5}	4.38×10^{-5}
	(1.22×10^{-6})	(9.44×10^{-7})	(7.58×10^{-7})

Table S4: Mean absolute difference in the maximized ReML criteria produced by lmer and BLMM, averaged across voxels and simulation instances. Each reported average is an image-wide mean, taken across approximately $218,000$ voxels, further averaged across 1000 simulation instances. The number of observations present in the model used for each simulation setting is displayed on the left. All observations were generated according to the methods outlined in Section 4.3.2. Empirical standard errors, taken across simulation instances, are also given in brackets underneath each entry in the table.