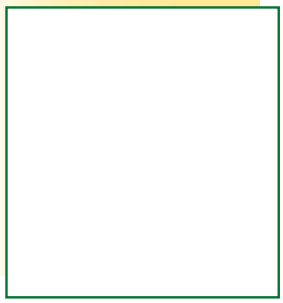

© PHOTO CREDIT

Modeling and Inference of Multisubject fMRI Data

Using Mixed-Effects Models for Joint Analysis

JEANETTE MUMFORD AND THOMAS NICHOLS

Functional magnetic resonance imaging (fMRI) is a rapidly growing technique for studying the brain in action. Since its creation [1], [2], cognitive scientists have been using fMRI to understand how we remember, manipulate, and act on information in our environment. Working with magnetic resonance physicists, statisticians, and engineers, these scientists are pushing the frontiers of knowledge of how the human brain works.

The design and analysis of single-subject fMRI studies has been well described. For example, [3], chapters 10 and 11 of [4], and chapters 11 and 14 of [5] all give accessible overviews of fMRI methods for one subject. However, while the appropriate manner to analyze a group of subjects has been the topic of several recent papers, we do not feel it has been covered well in introductory texts and review papers. Therefore, in this article, we bring together old and new work on so-called group modeling of fMRI data using a consistent notation to make the methods more accessible and comparable.

The analysis of single-subject fMRI data has drawn heavily on signal processing techniques. As discussed in the following, linear time invariant systems are the standard way to specify the model for the experimentally related signal in fMRI. When more than one subject is considered, the model must account for differing response magnitudes in each subject. While it is easy to specify a multisubject model that fits different responses for each subject, standard inference procedures do not account for the random subject-to-subject variation in response magnitude. When this random variation is neglected, the inferences are specific to the cohort of subjects studied. As most experimenters want to make inference on the population average magnitude, inference methods must account for heterogeneity in the population, and specifically, a significant result must be based on statistical confidence that the population from which these subjects were drawn shows a given effect on average. Population inference is the goal of group modeling, and it is a statistical challenge not met by direct application of methods found in a first-year statistics course. Basic statistics and regression usually only cover ordinary least squares (OLS), linear regression, and other fixed-effects models that do not yield population inferences.

In the next section, we distinguish fixed-effects models from mixed-effects models and will motivate the importance of a mixed-effects model for group fMRI analysis. The sections following that describe single-subject modeling and show a general method for estimating the group model.

Fixed-Effects Versus Mixed-Effects

To motivate the need of a mixed-effects analysis, we use a simple nonimaging example. Instead of measuring brain activation, perhaps we wish to compare hair length between genders. We wish to determine if there is evidence that American men and women have different length hair. It isn't feasible to measure every American, so we will randomly select men and women from the whole population. Based on just these two samples, we will try to make a statement, or inference, about all Americans. In order to make this comparison, we need the distributions of hair length for both men and women, and once these are obtained, a statistical comparison can determine whether or not they differ.

The experiment is conducted by randomly choosing four men and four women and for each randomly selecting a single hair from their heads and measuring it (in the following, we consider measuring multiple hairs). For each group, note that there are two sources of variation: within individual and between individual. The between-individual variation stems from each person having a different hair cut and hence different hair length, while the within-individual variation is present since, on any one person, the length of each hair varies over the head. Let σ_W^2 be the within-subject variance and σ_B^2 the between-subject variance. The top eight distributions in Figure 1 show the hair length distributions for the four men and four women. These distributions describe the relative frequency of hair length of a randomly selected hair from a single individual. Here we have assumed that the variation of a given individual's hair length is 1 in ($\sigma_W^2 = 1$).

If our population of interest is precisely these eight men and women, then between-subject variation can be neglected and a fixed-effects analysis can be used. The question to be answered is: How does the hair length of these particular four men compare to that of these particular four women? The resulting fixed-effects distributions are shown in Figure 1, below the individuals' distributions. Each gender's fixed-effect variance is $\sigma_{FFX}^2 = (1/4)\sigma_W^2 = 0.25$.

If we are not just interested in these eight men and women but the comparison of hair length between all men and

0739-5175/06/$20.00©2006IEEE

women, the next step is to construct population hair length distributions. A mixed model treats the four men and four women as randomly selected, not as the entire population of interest, and it takes into consideration the between-individual variances as well as the within-individual variances. The bottom of Figure 1 shows hair length distributions for men and women when the between-individual variance is $\sigma_B^2 = 49$ in. The variance of each gender's group distributions has two contributions:

$$\sigma_{MFX}^2 = \frac{\sigma_W^2}{4} + \frac{\sigma_B^2}{4} = \frac{1}{4} + \frac{49}{4} = 12.5.$$

Note that if the fixed-effects distributions were wrongly used to make a conclusion about all men and women, they would show that males have shorter hair than females. In fact, the mixed-effects distributions show considerable overlap, and we would not, based on this small sample, be able to conclude that men and women have different hair length.

One simplification here is that we only measured one hair per person. It would be better to randomly select multiple hairs, measure each, and take the average. If we instead had measured 25 hairs per person, then the distribution of each subject's average would have variance $\sigma_W^2/25$; for the fixed effect distribution

$$\sigma_{FFX}^2 = \frac{1}{4} \times \frac{\sigma_W^2}{25} = 0.01,$$

and for the mixed-effects distribution

$$\sigma_{MFX}^2 = \frac{1}{4} \times \frac{\sigma_W^2}{25} + \frac{1}{4}\sigma_B^2 = 12.26.$$

Observe that since σ_B^2 is so much larger than σ_W^2, increasing intrasubject precision has little impact on the mixed-effects variance.

Returning to fMRI, the basic issues are essentially the same. Instead of measuring multiple hairs, we are measuring the brain activation at a particular brain location multiple times. In multiple subject fMRI studies, most often, the interest is in making conclusions about populations and not specific subjects, and hence a mixed-effects method is necessary to get valid inferences in group fMRI.

Single-Subject fMRI Analysis

The basis of fMRI is the blood-oxygen-level-dependent (BOLD) effect. Due to differential magnetic susceptibility of oxygenated (oxygen-rich) hemoglobin and deoxygenated hemoglobin, the BOLD effect results in greater MRI intensity when brain activity increases (see, e.g., [6] for details). Since the BOLD effect is related to blood flow and volume, which do not change instantaneously, the BOLD response is temporally blurred and delayed relative to the experimental stimuli presented to the subject. Any intrasubject model must account for these effects (see Figure 2).

Consider a specific experiment, which we will revisit throughout the article [7]. In 12 healthy subjects, the investigator wanted to study the activation of higher-level motor areas during visually cued right-hand finger movement where the task was either tapping the index finger only, sequentially tapping the fingers, or randomly tapping the fingers. The hypothesis is that motor-related brain activity would increase with the complexity of the task, where index tapping is the most simple and random tapping is the most complex task. The design consisted of 30-s pseudorandomly ordered blocks of rest and the three visually cued finger tapping tasks [7]. Figure 2 displays the experimental design for this study.

All modeling discussed here is applied in a voxelwise fashion (a voxel is a single volume-element), i.e., each model is fitted to the data associated with each voxel separately. Intrasubject fMRI modeling is generally based on a linear time-invariant systems approach to the BOLD response. The experimental stimuli are represented by $x(t)$, consisting of just zeros and ones, indicating when a stimulus is present. For example, a block design consists of stimuli that are on for a duration of 2–30 s and would be represented in $x(t)$ by a box-car, and an event-related design consists of transient stimuli, which are represented by delta functions. In practice, there may be multiple experimental conditions, each with indicator $x_j(t)$, $j = 1, \ldots, J$. Using an assumed hemodynamic response function (HRF) $h(t)$, the noiseless predicted response is then the convolution of $x_j(t)$ and $h(t)$—$(h \otimes x_j)(t)$. The predicted response is discretized into $\{x_{jt}\}_{t=1}^T$ and used to create a predictor for the observed data $\{y_t\}_{t=1}^T$. The fitted model is then $y_t = \beta_0 + \sum x_{jt}\beta_j + \epsilon_t$, where ϵ_t is mean zero random error. Figure 2 shows both the experimental stimuli and the experimental predictors that result from the convolution with an HRF for the finger tapping experiment described previously.

In matrix notation, for the kth of N subjects, we write $Y_k = X_k\beta_k + \epsilon_k$, where Y_k is a T_k vector of the observed data, X_k is the $T_k \times p$ ($p = J + 1$) predictor matrix, and ϵ_k is the T_k vector of random errors [Figure 3(a), top]. Note that column 1 of X_k consists of a 1 for the intercept followed by columns $\{x_{tj}\}_{t=1}^T$, $j = 1, \ldots, J$.

If the errors ϵ_k are independent and have homogeneous variance σ_k^2, then the Gauss-Markov theorem [8] gives the minimum variance, unbiased estimate of β_k as

$$\hat{\beta}_k^{OLS} = \left(X_k^T X_k\right)^{-1} X_k^T Y_k, \tag{1}$$

which has variance $\text{Cov}(\hat{\beta}_k^{OLS}) = (X_k^T X_k)^{-1}\sigma_k^2$. The OLS residuals are $R_k = A_k Y_k$, where $A_k = I - X_k(X_k^T X_k)^{-1} X_k^T$ is the residual forming matrix. The unbiased estimate of the variance of the errors is

$$\hat{\sigma}_k^{2OLS} = \frac{1}{\nu_k} A_k^T A_k, \tag{2}$$

where $\nu_k = T_k - p$ are the degrees of freedom. This method is known as OLS. As found by many authors, residual error in fMRI is not independent and exhibits excess variation at low frequencies (sometimes called $1/f$-type autocorrelation) [9]–[12]. When $\text{Cov}(\epsilon_k) = V_k\sigma_k^2 \neq I\sigma_k^2$, where V_k is the correlation matrix, estimates obtained from (1) will still be unbiased ($\text{E}(\hat{\beta}_k) = \beta_k$, where $\text{E}(\cdot)$ denotes expectation) but will not have optimal precision (minimum variance), and the estimate of residual variance (2) will be biased.

The optimal approach with dependent errors is whitening, or decorrelation of the data and model. Instead of working directly

with (1), we premultiply by a matrix that renders the errors independent: $V_k^{(-1/2)} Y_k = V_k^{(-1/2)} X_k \beta_k + V_k^{(-1/2)} \epsilon_k$, where $V_k^{(-1/2)}$ is a matrix such that $V_k^{(-1/2)} V_k V_k^{-(1/2)T} = I$. We rewrite this as

$$Y_k^* = X_k^* \beta_k + \epsilon_k^*, \tag{3}$$

where Y_k^*, X_k^*, ϵ_k^* are the whitened data, model, and errors, respectively. The Gauss-Markov estimate of β_k is now just the OLS estimate using Y_k^* and X_k^*:

$$\hat{\beta}_k^{\mathrm{GLS}} = \left(X_k^{*T} X_k^*\right)^{-1} X_k^{*T} Y_k^* \tag{4}$$

and is referred to as the generalized least squares (GLS) estimate. The estimate $\hat{\beta}_k^{\mathrm{GLS}}$ has optimal variance, given by

$$\mathrm{Cov}\left(\hat{\beta}_k^{\mathrm{GLS}}\right) = \sigma_k^2 \left(X_k^{*T} X_k^*\right)^{-1}. \tag{5}$$

Similarly, the unbiased estimate of whitened error variance is

$$\hat{\sigma}_k^{\mathrm{GLS}} = \frac{1}{\nu_k} \left(Y_k^* - X_k^* \hat{\beta}_k^{\mathrm{GLS}}\right)^T \left(Y_k^* - X_k^* \hat{\beta}_k^{\mathrm{GLS}}\right). \tag{6}$$

In short, with knowledge of the whitening matrix $V_k^{-\frac{1}{2}}$, optimal estimates for β_k can be found with GLS.

There are two important details to single-subject modeling. First, whitening assumes that the true error correlation V_k is known precisely. In practice, V_k must be estimated from the data, and an estimate may be biased and highly variable, potentially corrupting the whitening process and yielding estimates of β_k and σ_k^2 that are worse than OLS. In fMRI, it is generally acknowledged that some sort of spatial regularization of $\hat{V}_k$ is required [12]–[16]. This approach reduces the variability in $\hat{V}_k$ by pooling over space either locally [12]–[14] or globally [15], [16].

The other important detail is the use of contrasts to summarize evidence for a particular effect. Rarely does an investigator have interest in all p elements of β_k. Rather, interest typically focuses on one condition versus another or an average of conditions versus another. For example, in the finger tapping experiment, we may only be interested in whether activation from random finger tapping (Condition 3) is greater than sequential finger tapping (Condition 2), in which case we define contrast $c = [0 \;\; 0 \;\; -1 \;\; 1]$ and estimate the quantity $c\beta_k = \beta_{k_3} - \beta_{k_2}$ with $c\hat{\beta}_k$ [Figure 3(a), bottom]. The variance of the estimated contrast is

$$\mathrm{Cov}(c\hat{\beta}_k) = c(\mathrm{Cov}(\hat{\beta}_k))c^T. \tag{7}$$

Given user's interest in contrasts of β_k, in the remainder of this article we focus on inference of $c\beta_k$.

Inference on $c\beta_k$ is made with a ratio of the estimate $c\hat{\beta}_k$ to its standard error. In contrast to the true standard deviation ($\sqrt{\mathrm{Cov}(c\hat{\beta}_k)}$), the standard error of an estimator is its estimated standard deviation ($\sqrt{\widehat{\mathrm{Cov}}(c\hat{\beta}_k)}$). If the estimated BOLD response magnitude is large relative to its standard error, we conclude that the result was unlikely to have arisen by chance. When the random errors have a Gaussian distribution, the ratio follows a Student's T distribution, and it forms the basis of inference for linear models.

Model

The starting point for our statistical modeling of group fMRI data is voxel-aligned data. That is, at each voxel, we have data from each subject that have been motion-corrected, aligned to a standard atlas brain, and perhaps smoothed (see "Praparation for Multisubject Modeling"). Before we set out the various models used for group modeling, we define notation.

Notation

It is useful to specify the complete model in stages, a first or lower level, where a model is fit for each subject, and a second level, which combines the different subjects.

As shown previously, the general linear model for the kth subject of N subjects is

$$Y_k = X_k \beta_k + \epsilon_k, \tag{8}$$

where Y_k is the $T_k \times 1$ vector of fMRI response data, X_k is the $T_k \times p$ design matrix, β_k is a vector of p parameters, and the error vector of length T_k is Gaussian distributed with variance σ_k^2 and correlation V_k, $\epsilon_k \sim N(0, \sigma_k^2 V_k)$. Subjects are independent, and so $\mathrm{Cov}(\epsilon_k, \epsilon_k') = 0$ for $k \neq k'$. Note that while each subject can have a differing number of scans (T_k), all of the design matrices X_k must have the same number of columns, each column expressing the same effect in each subject's data. In general, V_k is not diagonal and will express the autocorrelation that is present in fMRI data; a typical assumption is **<au: Please spell out.>** AR(1) noise, such that $(V_k)_{ij} = \rho_k^{|i-j|}$, where ρ_k is the first-order autocorrelation.

These N first-level models can be concisely expressed as

$$Y = X\beta + \epsilon, \tag{9}$$

where $Y = [Y_1^T, \ldots, Y_N^T]^T$, $X = \mathrm{diag}(X_1, \ldots, X_N)$, $\beta = [\beta_1^T, \ldots, \beta_N^T]^T$, and $\epsilon = [\epsilon_1^T, \ldots, \epsilon_N^T]^T$ with covariance $V = \mathrm{Cov}(\epsilon) = \mathrm{diag}(\sigma_1^2 V_1, \ldots, \sigma_N^2 V_N)$ ($\mathrm{diag}(\cdot)$ defines a block-diagonal matrix); let $T = \sum T_k$ be the total number of scans in the entire dataset.

The second-stage analysis is used to relate subject-specific parameters β_k to population parameters β_g:

$$\beta = X_g \beta_g + \epsilon_g. \tag{10}$$

Assuming all first-level parameters are taken to the second level, X_g is a $Np \times p_g$ second-level design matrix, β_g is a vector of length p_g that contains the second-level parameters, and $\epsilon_g \sim N(0, \sigma_g^2 V_g)$, where V_g is a block-diagonal matrix with blocks V_{g_k}; note that we separate overall group variance σ_g^2 from the correlation matrix V_g. Typically, X_g has a very simple form, with columns of ones to test the mean response over subjects. The estimation of the parameters in the two-stage analysis is a challenge, since β occurs in both (9) and (10), and β is not observed. While there are standard methods for fitting this so-called hierarchical model [17], they are based on all T data points and involve iterative optimization. Since a typical group analysis can have $T = 20,000$ scans, with each scan having 100,000 voxels, direct application of these methods is generally not practical. A more computationally efficient approach is to build a group model-based summary statistics, described next.

Summary Statistics Approach

The summary statistics approach is a natural approach that involves first estimating β from (9) then estimating β_g using a modified version of (10). Because the first-level model (9) is separable by subject, $\hat{\beta} = [\hat{\beta}_1^T, \ldots, \hat{\beta}_N^T]^T$ can be found subject-by-subject with (4). The group model based on $\hat{\beta}$ is

$$\begin{aligned} \hat{\beta} &= X_g\beta_g + \epsilon_g + (\hat{\beta} - \beta) \\ &= X_g\beta_g + \epsilon_{\tilde{g}}. \end{aligned} \tag{11}$$

Note that (10) models the unobservable, true mean responses β for each subject, while (11) models the observed, estimated responses $\hat{\beta}$ for each subject. The summary statistic model's errors $\epsilon_{\tilde{g}}$ have mean 0 variance

$$V_{\tilde{g}} = (X^T V^{-1} X)^{-1} + \sigma_g^2 V_g, \tag{12}$$

where the first component can also be written $\mathrm{diag}(\{\sigma_k^2 (X_k^{*T} X_k^*)^{-1}\})$, and reflects the intrasubject variance-covariance of the $\hat{\beta}_k$s, while the second component indicates how variable the true effect is between subjects. If $V_{\tilde{g}}$ is known, then the GLS estimate of β_g is given by,

$$\hat{\beta}_g = \left(X_g^{*T} X_g^*\right)^{-1} X_g^{*T} \hat{\beta}^* \tag{13}$$

$$Cov\left(\hat{\beta}_g\right) = \sigma_{\tilde{g}}^2 \left(X_g^{*T} X_g^*\right)^{-1}, \tag{14}$$

where $X_g^* = V_{\tilde{g}}^{-\frac{1}{2}} X_g$ and $\hat{\beta}^* = V_{\tilde{g}}^{-\frac{1}{2}} \hat{\beta}$. Assuming Gaussian ϵ and ϵ_g, it can be shown that this summary-statistic-based estimate is identical to that found using all of the data [18].

A crucial observation is that this summary statistic approach requires both the subject-level parameter estimates $\hat{\beta}_k$ and their variances $\sigma_k^2 (X_k^{*T} X_k^*)^{-1}$. If OLS is used with (11), ignoring the covariances, often the estimates will be suboptimal and the standard errors incorrect. An important special case when second-level OLS and GLS estimates coincide involves contrasts.

As discussed previously, the goal is usually inference on a particular contrast of parameters $c\beta_k$. In this case, the whole $\hat{\beta}$ doesn't need to be brought to the second level, only the N contrasts [18]; $\hat{\beta}$ becomes $\hat{\beta}_{\mathrm{cont}} = [c\hat{\beta}_1, \ldots, c\hat{\beta}_N]^T$, V becomes a diagonal matrix with entries $\sigma_k^2 c (X_k^* X_k^*)^{-1} c^T$, and V_g will have a simple form, typically just I_N. If the intrasubject contrast variance is homogeneous, i.e.,

$$\sigma_k^2 c (X_k^{*T} X_k^*)^{-1} c^T = \sigma_{k'}^2 c (X_{k'}^{*T} X_{k'}^*)^{-1} c^T$$

for $k \neq k'$, then the OLS and GLS estimators β_g are equivalent [19].

Figure 3(b) shows an example of a second-level model consisting of a single contrast from each of 12 subjects. This model produces group-level estimates of the contrast for the group of the first six subjects (β_{g1}) and the last six subjects (β_{g2}). The group model is given by (11), except the dependent variable is $\hat{\beta}_{\mathrm{cont}}$.

The following sections introduce different summary statistics methods that have been developed. Due to the massive size of fMRI datasets, standard statistical software is not useful, and custom software is required. Because of this, the first three sections are organized around statistical methods implemented in three widely used software packages, FSL (http://www.fmrib.ox.ac.uk/fsl), fMRIstat (http://www.math.mcgill.ca/keith/fmristat), and SPM (http://www.fil.ion.ucl.ac.uk/spm). While all of the methods use the model described above, they differ in how they find estimates for the between-subject variance $V_{\tilde{g}}$.

FSL

The FMRIB software library (FSL) uses the summary statistics group model described previously [(9) and (11)] [18], with the restriction that only a single contrast per subject is taken to the second level. They use Bayesian methods to estimate β_g while accounting for uncertainty in the estimates of σ_g^2 (see "Bayesian Versus Classicial Inference"). First we review FSL's first-level modeling methods.

As indicated previously, the autocorrelation V_k is needed to find optimal intrasubject estimates $\hat{\beta}_k$ (4). FSL uses three steps to obtain $\hat{V}_k$ for each voxel. First, a high-pass filter is applied to data and the model to remove low-frequency noise and reduce nonstationarity. Second, OLS residuals $(Y_k - \hat{\beta}^{\mathrm{OLS}})$ are used to estimate an autocorrelation function (ACF) which is regularized with a Tukey taper. Finally, the voxelwise ACFs are further regularized with a spatial smoothing; since autocorrelation tends to vary more between tissue type and less within, a nonstationary spatial smoothing is used, which accounts for tissue type as determined by functional image intensity. The resulting autocorrelation estimate $\hat{V}_k$ is used in (4) and (5).

Inference on β_g is based on its posterior distribution conditional on the data Y, $p(\beta_g|Y)$. However, the posterior $p(\beta_g|Y)$ doesn't have a closed form, so a two-stage method is used to find a posterior mean estimate $\hat{\beta}_g^{\mathrm{FSL}}$; first, a fast approximation is used, followed by a slower Markov chain Monte Carlo (MCMC).

At each voxel, the posterior of β_g is approximated as a multivariate T with noncentrality parameter $\hat{\beta}_g$, variance parameters $\mathrm{Cov}(\hat{\beta}_g)$ [see (13) and (14)], and degrees of freedom ν_g. The noncentrality and variance parameters depend on the unknown mixed-effects covariance $V_{\tilde{g}}$. The fast method assumes large intrasubject degrees of freedom ν_k, so the intrasubject contribution to $V_{\tilde{g}}$ is assumed known without error, leaving only intersubject variance σ_g^2 to be estimated. A maximum a posteriori**<AU: correct?>** estimate $\hat{\sigma}_g^{2\mathrm{FSL}}$ is found using iterative optimization, and the degrees of freedom are estimated conservatively as $N - p_g$.

The point estimates of σ_g and β_g are used to find the posterior probability of a positive response**<AU: should P in the equation be in italics?>** $\mathrm{P}(\beta_g > 0|Y)$. By equating the posterior probability to a Z statistic via a P-value, voxelwise Z statistics are created that offer classical tests of the null hypothesis $\mathcal{H}_o : \beta_g = 0$; only voxels with Z statistics close to the desired significance threshold continue on to the next stage. The second stage employs a slower, more accurate MCMC method of estimation [18], which accounts for uncertainty in σ_g by estimating the effective degrees of freedom ν_g of the posterior. This stage produces more accurate test statistics for the voxels that were near the threshold in the first stage, and these are used to locate voxels where the group-level parameters are significant.

fMRIstat

Worsley et al. [14] developed a summary statistics approach

that is implemented in the fMRIstat package. As with FSL's method, GLS is used to estimate the parameters of the first level, and only a single contrast per subject is taken from the first to the second level. One important aspect of the method is that the random-effects variance $\sigma_{\tilde{g}}^2$ is estimated using restricted maximum likelihood (ReML), the standard classical variance estimation method (see "Maximum Likelihood and Restricted Maximum Likelihood"). Another unique aspect is the regularization $\hat{\sigma}_g^2$, which is used to increase the effective degrees of freedom of the variance estimate.

At the first level, an AR autocorrelation model is fit to a sample covariance matrix of the OLS residuals. The OLS residuals have covariance $R_k V_k R_k^T \neq V_k$, and so the AR coefficients are biased. After applying a bias correction, the AR coefficients are spatially smoothed and then used to create $\hat{V}_k^{(-1/2)}$ (see [20] and [21] for recent work on this smoothing step). GLS estimates for β_k and its variance are as before [(4) and (5)].

The second term of $\mathrm{Cov}(\epsilon_{\tilde{g}})$, the between-subject variance, is estimated with ReML, $\hat{\sigma}_2^{2\mathrm{ReML}}$. Since group size N is often small, this variance estimate is itself very variable; equivalently, it has very low degrees of freedom. In contrast, the pooled fixed-effects variance

$$\hat{\sigma}_F^2 = \sum_k \frac{\nu_k}{\sum_k \nu_k} \hat{\sigma}_k^2 c \left(X_k^{*T} X_k^*\right)^{-1} c^T \tag{15}$$

has very high degrees of freedom, $\sum_k \nu_k$. To borrow strength from this high-precision variance estimate, Worsley et al. [14] considered the following manipulation of the mixed-effect variance

$$\hat{\sigma}_F^2 + \hat{\sigma}_g^2 = \frac{\hat{\sigma}_F^2 + \hat{\sigma}_g^2}{\hat{\sigma}_F^2} \hat{\sigma}_F^2 \approx \mathrm{smooth}\left(\frac{\hat{\sigma}_F^2 + \hat{\sigma}_g^2}{\hat{\sigma}_F^2}\right) \hat{\sigma}_F^2. \tag{16}$$

That is, since the ratio of mixed- to fixed-effect variance appeared to have little structure, they smooth that ratio. By solving for random-effect variance, they obtain an estimate consisting of a smooth image times**<AU: multiplied by?>** a high-degree-of-freedom variance estimate:

$$\hat{\sigma}_g^{2\mathrm{fmristat}} = \mathrm{smooth}\left(\frac{\hat{\sigma}_g^2}{\hat{\sigma}_F^2}\right) \hat{\sigma}_F^2. \tag{17}$$

Whereas FSL used MCMC to find accurate degrees-of-freedom, fMRIstat selects the full width at half maximum (FWHM) of variance-ratio smoothing to obtain effective degrees of freedom of at least 100. Of course, this decrease in variability of the variance estimate comes at the cost of an increase in the bias of the variance estimate.

Finally, the T statistics are formed by the ratio of $\hat{\beta}_g$ and its standard error (using $\hat{\sigma}_g^{2\mathrm{fmristat}}$) and are used to make inferences on the activation within each voxel.

SPM2

The SPM2 package employs the theory developed by Friston et al. in [22] and [23]. SPM2 also uses GLS with an estimate of $\hat{V}_k$ to estimate the first level. It differs from the previous two methods by only requiring the estimates of the mean parameters β_k, not both the mean and covariance parameters, to be taken from the first level into the second level. Such a simplification, though, requires an additional assumption, that of homogeneous intrasubject variance (over subjects). The benefit is that this allows more than a single contrast from each subject to be estimated at the second level. The second-level model is then estimated using ReML.

First, we review SPM2's first-level modeling. The intrasubject autocorrelation V_k is modeled with a two-term Taylor series approximation to an AR(1) model, $\rho = 0.2$. The autocorrelation estimates are based on the sample covariance of the raw data Y_k; this avoids the bias due to the covariance of residuals but can introduce bias if a strong signal is present. While FSL and fMRIstat both estimate V_k separately for each voxel, SPM2 assumes the autocorrelation is the same for all voxels. To bias the global estimate towards the most important voxels, only those voxels surviving an overall F test at level 0.001 contribute to the sample covariance matrix. Note that while V_k is global, σ_k^2 is estimated separately at each voxel. The resulting $\hat{V}_k$ is used to find GLS estimates for β_k and its variance, as above.

At the second level, SPM2 is capable of obtaining group-level estimates of all parameters in β_k or subsets of parameters from the first level simultaneously, instead of only one contrast at a time as in FSL and fMRIstat. This allows for both group-level t-tests that test the significance of one contrast at a time and F tests that allow testing of multiple contrasts simultaneously. F tests are not possible in FSL or fMRIstat, since more than one contrast is simultaneously required.

To omit the first-level covariances at the second level, SPM2 must assume that the intrasubject variances are the same for every subject, $\sigma_k^2 (X_k^{*T} X_k^*)^{-1} = \sigma_{k'}^2 (X_{k'}^{*T} X_{k'}^*)^{-1}$ for $k \neq k'$. In this case, the summary statistics covariance $V_{\tilde{g}}$ takes the form of a block diagonal matrix with identical blocks, $V_{\tilde{g}k} = \sigma_k^2 (X_k^{*T} X_k^*)^{-1} + \sigma_g^2 V_{gk}$, where $V_{\tilde{g}k}$ and V_{gk} are the kth block of their respective matricies. SPM2 uses ReML to estimate the common covariance in each block, V_{gk}, without ever separately estimating within- and between-subject variance. As with the first-level, this ReML estimation only takes place on subset voxels, those with significant overall F statistics.

An important special case is when only one contrast is of interest, β_{cont}, as in the previous two sections. In that setting, $\sigma_g^2 V_{\tilde{g}}$ will be identity times**<AU: multiplied by?>** a scalar, and the second-level estimate (13) reduces to the OLS estimate.

Generalized Estimating Equations

Another summary statistics approach that has been studied involves using generalized estimating equations (GEE) to estimate the second level [23]. Similarly, to the second level of SPM2, this method only requires the mean parameter estimates from the first level, and all parameters or subsets of parameters from the first level may be analyzed at the second level. Previous GEE analysis used first-level results that were estimated with SPM2 [23]. A benefit of the GEE approach is that it does not assume the covariance of $\hat{\beta}_g$ is heterogeneous across space as SPM2 does but estimates covariance separately for each voxel.

Just as with SPM2, the first-level covariance estimates are not needed due to the assumption that intrasubject variance is the same across all subjects, and so $V_{\tilde{g}}$ is a block diagonal matrix with identical blocks. For our description, we assume all first-level parameters continue to the second level, and so

X_g is an $Np \times p_g$ matrix. To estimate the second level, the GEE method uses two variance estimates, the first being the working correlation V_W, which is an approximate estimate of $\sigma_g^2 V_{\tilde{g}}$ [24] and need not have the structure of the true correlation. If we use V_W to estimate β_g, we have

$$\hat{\beta}_g = \left(X_g^T V_W^{-1} X_g\right)^{-1} X_g^T V_W^{-1} \hat{\beta}.$$

Although V_W can be used to find an unbiased estimate of $\hat{\beta}_g$, an additional, more accurate estimate of $V_{\tilde{g}}$ is incorporated into the estimate of the variance of the parameter estimates, known as the sandwich estimator:

$$\widehat{\mathrm{Cov}}\left(\hat{\beta}_g\right) = \left(X_g^T V_W^{-1} X_g\right)^{-1} X_g^T V_W^{-1} \hat{V}_{\tilde{g}} V_W^{-1} X_g \left(X_g^T V_W^{-1} X_g\right)^{-1}.$$

The estimate, $\hat{V}_{\tilde{g}}$, has a block-diagonal structure given by

$$\hat{V}_{\tilde{g}} = \mathrm{diag}(\hat{V}_m, \ldots, \hat{V}_m). \tag{18}$$

The blocks on the diagonal are the $p \times p$ mixed-effects covariance matrix estimates

$$\hat{V}_m = \sum_{k=1}^{N} (\hat{\beta}_k - X_{gk}\hat{\beta}_g)(\hat{\beta}_k - X_{gk}\hat{\beta}_g)^T/(N-1),$$

where X_{gk} is the portion of X_g that corresponds to subject k (in this case, rows $p(k-1)+1$ through pk of the design matrix X_g). Since $\hat{V}_m$ is not fitted to a covariance structure such as an AR or ARMA model, this is referred to as an unstructured covariance. The benefit of using both V_W and $\hat{V}_{\tilde{g}}$ is that our estimate of $\mathrm{Cov}(\hat{\beta}_g)$ is robust and tends, asymptotically, to the true value of $\mathrm{Cov}(\hat{\beta}_g)$, even if the working correlation is misspecified. Also, the estimate of β_g is unbiased, regardless of the choice of the working correlation; therefore, V_W does not need to be very complicated, and even the identity matrix could be used. The benefit of this method is that it does not assume that the covariance is spatially homogeneous, and so the variance of the parameter estimates is calculated separately for each voxel, which reduces bias of the variance estimate.

Discussion

When making group-level inference on fMRI data, it is important to use a mixed models approach so that both the within-subject variation and the between-subject variation are accounted for. The summary statistics approach is a popular approach for group-level modeling of fMRI data. All four of the methods presented here are summary statistics methods, with one of the differences between the methods being how the variance of the group-level error $\epsilon_{\tilde{g}}$ is estimated.

FSL and fMRIstat take similar approaches to estimate $V_{\tilde{g}}$. Both methods use the estimate of the covariance from the first level to determine the within-subject variation, $(X^T V^{-1} X)^{-1}$. By allowing only contrasts of parameter estimates from the first level into the second level, the second component of the covariance is simplified from $\sigma_g^2 V_g$ to $\sigma_g^2 I$. From this point, the two methods differ in how they estimate σ_g^2, where FSL uses a two-stage estimating approach including MCMC, and fMRIstat uses the EM algorithm.

SPM2 and the GEE method differ from both FSL and fMRIstat in that they do not use the first-level covariance estimates at the second level due to the assumption that $V_{\tilde{g}}$ has identical blocks along the diagonal. GEE estimates the covariance of the group-level parameters by use of the sandwich estimator, which leads to a consistent estimate of the variance of $\hat{\beta}_g$, and SPM2 uses a spatially homogeneous covariance estimate that is pooled over a subset of voxels.

With both SPM2 and GEE, there is no constraint on the dimension of X_g, multiple parameters can be estimated at the group level, and, therefore, it is possible to carry out multiple t-tests and F tests. F tests allow multiple contrasts to be tested at once. For example, if your group-level model had three parameters, $\beta_g = [\beta_{g1}, \beta_{g2}, \beta_{g3}]^T$, an F test could be used to simultaneously test if any of these parameters **were** zero.

Conclusions

We have reviewed four commonly used approaches to group modeling in fMRI. The methods differ in their computational intensity (FSL with its two-level estimation including MCMC being the most intense) and assumptions (SPM2 with its assumption of spatially homogeneous covariance V_g being the most restrictive).

Acknowledgments

Thanks to Karl Friston and Andrew Holmes for the hair length example, to Richard Leahy and Dimitrios Pantazis for comments on the manuscript, and to Heidi Johansen-Berg for the data used in the example.

<au: Please provide short biographies and address for correspondence.>

References

[1] K.K. Kwong, J.W. Belliveau, D.A. Chesler, I.E. Goldberg, R.M. Weisskoff, B.P. Poncelet, D.N. Kennedy, B.E. Hoppel, M.S. Cohen, and R. Turner, "Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation," *Proc. Nat. Acad. Sci..*, vol. 89, pp. 5675–5679, Jun. 1992. **<au: Please provide issue number.>**

[2] S. Ogawa and T. Lee, "Functional brain mapping with physiologically sensitive image signals," *J. Magn. Reson. Imaging*, vol. 2(P)-WIP (Suppl), p. S22, 1992.**<au: Please provide issue number.>**

[3] S. Rabe-Hesketh, E.T. Bullmore, and M.J. Brammer, "The analysis of functional magnetic resonance images," *Statistical Methods Med. Res.*, vol. 6, pp. 215–237, 1997. **<au: Please provide issue number.>**

[4] R.S.J. Frackowiak, Ed., *Human Brain Function.*, 2nd ed. New York: Academic, 2003.

[5] P. Jezzard, P.M. Matthews, and S.M. Smith, Eds., *Functional MRI: An Introduction to Methods*. London, U.K.: Oxford Univ. Press, 2003.

[6] C. Moonen and P. Bandettini, Eds., *Functional MRI*. New York: Springer-Verlag, 2000.

[7] H. Johansen-Berg, M.F.S. Rushworth, M.D. Bogdanovic, U. Kischka, S. Wimalaratna, and P.M. Matthews, "The role of ipsilateral premotor cortex in hand movement after stroke," *Proc. Nat. Acad. Sci. U.S.A.*, vol. 99, pp. 14518–14523, Oct. 2002. **<au: Please provide issue number.>**

[8] F.A. Graybill, *Theory and Application of the Linear Model*. Duxbury Press, 1976. **<au: Please provide location of publisher.>**

[9] E. Zarahn, G.K. Aguirre, and M. D'Esposito, "Empirical analyses of BOLD fMRI statistics. I. Spatially unsmoothed data collected under null-hypothesis conditions," *NeuroImage*, vol. 5, pp. 179–197, Apr. 1997. **<au: Please provide issue number.>**

[10] P.L. Purdon and R.M. Weisskoff, "Effect of temporal autocorrelation due to physiological noise and stimulus paradigm on voxel-level false-positive rates in fMRI," *Hum. Brain Map.,* vol. 6, no. 4, pp. 239–249, 1998.

[11] J.L. Marchini and S.M. Smith, "On bias in the estimation of autocorrelations for fMRI voxel time-series analysis," *NeuroImage*, vol. 18, pp. 83–90, Jan. 2003. **<au: Please provide issue number.>**

[12] M.W. Woolrich, B.D. Ripley, M. Brady, and S.M. Smith, "Temporal autocorrelation in univariate linear modeling of fMRI data," *NeuroImage*, vol. 14, pp. 1370–1386, Dec. 2001.**<au: Please provide issue number.>**

[13] J.L. Marchini and B.D. Ripley, "A new statistical approach to detecting significant activation in functional MRI," *NeuroImage*, vol. 12, pp. 366–380, Oct. 2000.**<au: Please provide issue number.>**

[14] K.J. Worsley, C.H. Liao, J. Aston, V. Petre, G.H. Duncan, F. Morales, and A.C. Evans, "A general statistical analysis for fMRI data," *NeuroImage,* vol. 15, pp. 1–15, Jan. 2002.**<au: Please provide issue number.>**

[15] K.J. Friston, D.E. Glaser, R.N.A. Henson, S. Kiebel, C. Phillips, and J.

Ashburner, "Classical and bayesian inference in neuroimaging: Applications," *NeuroImage*, vol. 16, pp. 484–512, Jun. 2002.**<au: Please provide issue number.>**
[16] G.K. Aguirre, E. Zarahn, and M. D'Esposito, "Empirical analyses of BOLD fMRI statistics. II. Spatially smoothed data collected under null-hypothesis and experimental conditions," *NeuroImage*, vol. 5, pp. 199–212, Apr. 1997.**<au: Please provide issue number.>**
[17] V.G and M. G, *Linear Mixed Models for Longitudinal Data*. New York: Springer-Verlag, 2000. **<au: Please provide last name of M.G.>**
[18] C.F. Beckmann, M. Jenkinson, and S.M. Smith, "General multilevel linear modeling for group analysis in fmri," *NeuroImage*, vol. 20, pp. 1052–1063, 2003. **<au: Please provide issue number.>**
[19] A. Holmes and K. Friston, "Generalisability, random effects and population inference," in *NeuroImage*, vol. 7, p. S754, 1998. **<au: Please provide issue number.>**
[20] K.J. Worsley, "Spatial smoothing of autocorrelations to control the degrees of freedom in fMRI analysis," *NeuroImage*, vol. 26, pp. 635–641, Jun. 2005.**<au: Please provide issue number.>**
[21] T. Gautama and M.M. Van Hulle, "Optimal spatial regularisation of autocorrelation estimates in fMRI analysis," *NeuroImage*, vol. 23, pp. 1203–1216, Nov. 2004.**<au: Please provide issue number.>**
[22] K.J. Friston, W. Penny, C. Phillips, S. Kiebel, G. Hinton, and J. Ashburner, "Classical and bayesian inference in neuroimaging: Theory," *NeuroImage*, vol. 16, pp. 465–483, Jun. 2002. **<au: Please provide issue number.>**
[23] W.-L. Luo, M.J, W.T, and N. T, "Robust and Local Nonsphericity modeling for secnd level PET and fMRI analysis," *NeuroImage*, vol. 22, No. 1 Suppl., p. S47, 2005. **<au: Please provide full name of all authors.>**
[24] K.-Y. Liang and S.L. Zeger, "Longitudinal data analysis using generalized linear models," *Biometrika*, vol. 73, pp. 13–22, 1986.**<au: Please provide issue number.>**
[25] R.P. Woods, S.T. Grafton, C.J. Holmes, S.R. Cherry, and J.C. Mazziotta, "Automated image registration: I. general methods and intrasubject, intramodality validation," *J. Comput. Aided Tomography*, vol. 22, pp. 141–154, 1998. **<au: Please provide issue number.>**
[26] K.J. Friston, J. Ashburner, C.D. Frith, J.-B. Poline, J.D. Heather, and R.S.J. Frackowiak, "Spatial registration and normalization of images," *Hum. Brain Map.*, vol. 2, pp. 165–189, 1995.**<au: Please provide issue number.>**
[27] M. Jenkinson, P. Bannister, J.M. Brady, and S.M. Smith, "Improved optimisation for the robust and accurate linear registration and motion correction of brain images," *NeuroImage*, vol. 17, pp. 825–841, 2002. **<au: Please provide issue number.>**
[28] J. Ashburner and K.J. Friston, "Nonlinear spatial normalization using basis functions," *Hum. Brain Map.*, vol. 7, pp. 254–266, 1999. **<au: Please provide issue number.>**
[29] R.P. Woods, S.T. Grafton, J.D.G. Watson, N.L. Sicotte, and J.C. Mazziotta, "Automated image registration: Ii. intersubject validation of linear and nonlinear models," *J. Comput. Aided Tomography*, vol. 22, pp. 155–165, 1998. **<au: Please provide issue number.>**
[30] G. Christensen, R. Rabbit, and M. Miller, "Deformable templates using large deformation kinematics," *IEEE Trans. Image Processing*, vol. 5, pp. 1435–1447, 1996. **<au: Please provide issue number.>**
[31] P. Thompson and A. Toga, "A surface-based technique for warping 3-dimensional images of the brain," *IEEE Trans. Med. Imag.*, vol. 15, pp. 1–616, 1996.**<au: Please provide issue number.>**

Callouts

Linear time invariant systems are the standard way to specify the model for the experimentally related signal in fMRI.

Due to differential magnetic susceptibility of oxygenated hemoglobin and deoxygenated hemoglobin, the BOLD effect results in greater MRI intensity when brain activity increases.

Due to the massive size of fMRI datasets, standard statistical software is not useful, and custom software is required.

When making group-level inference on fMRI data, it is important to use a mixed models approach.

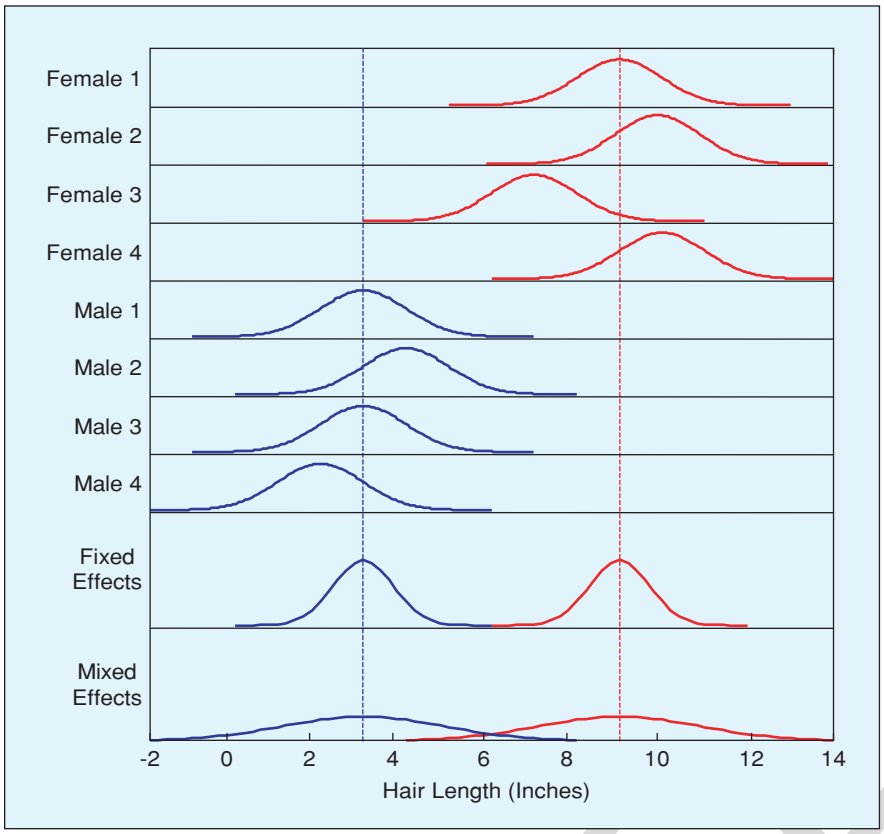

Fig. 1. These distributions illustrate the difference between fixed- and mixed-effects analysis, where blue and red distributions refer to males and females, respectively. The top eight distributions are subject-specific distributions, followed by the group distributions stemming from fixed-effects and mixed-effects analysis. The vertical lines indicate the sample means for the two groups.

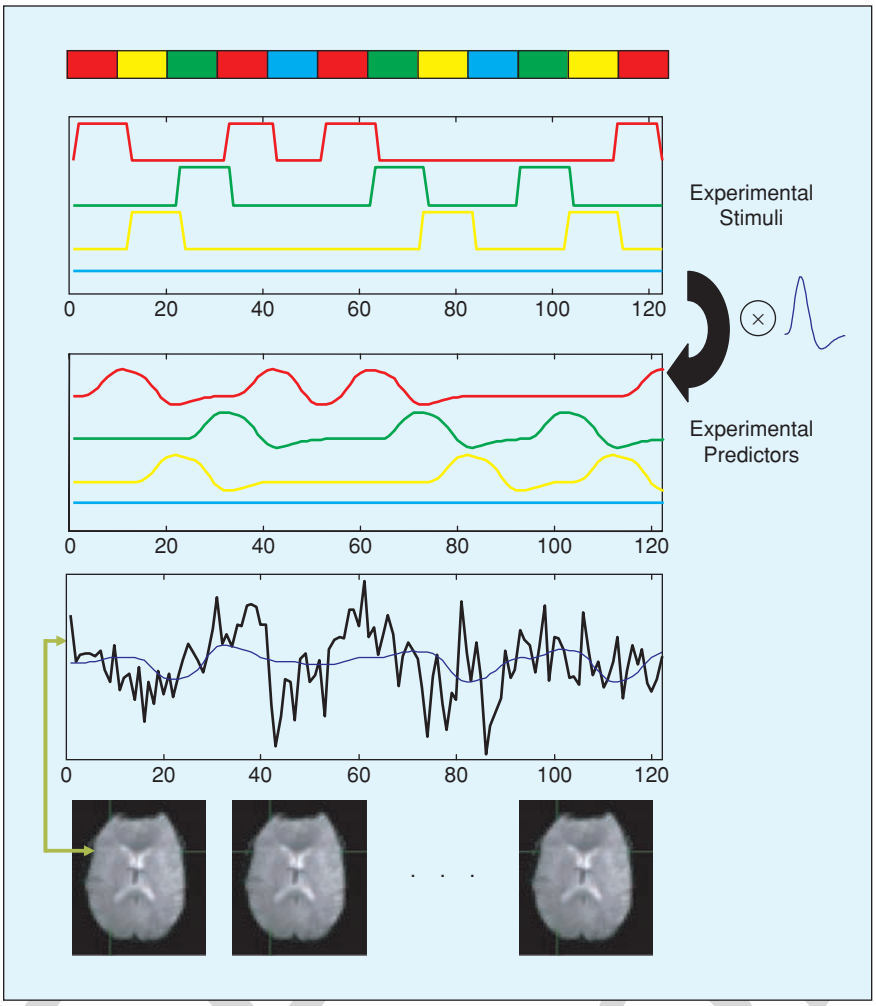

Fig. 2. The experimental stimuli and predictors associated with the BOLD response from a singe voxel (volume-element) over time. The top color bar indicates when the subject was cued to tap their fingers randomly (red), sequentially (green), only the index (yellow), or not at all (blue). The associated experimental stimuli are shown as well as the experimental predictors that are created by convolving the stimuli with an HRF. Finally, the original BOLD response (black) is shown with the predicted model fit (blue) based on the model formed with the experimental predictors.

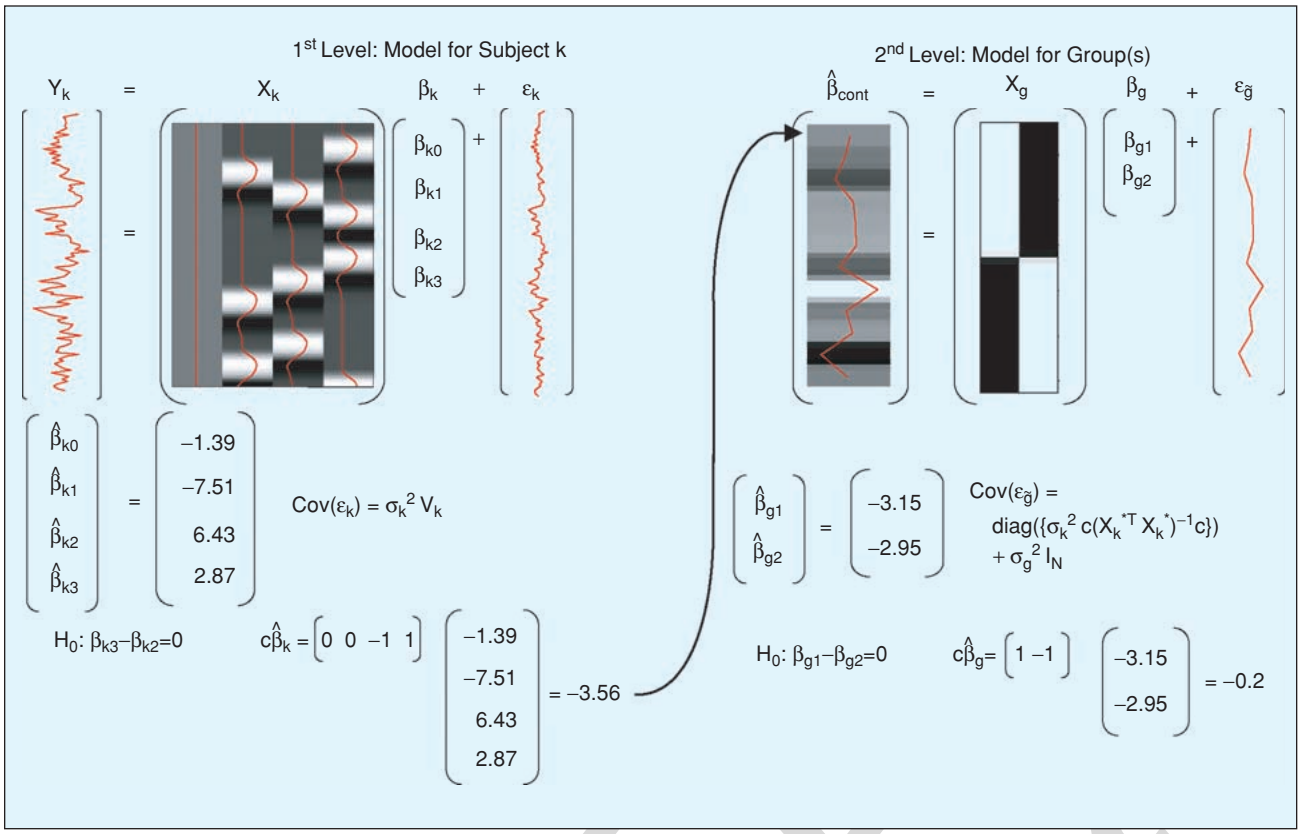

Fig. 3. Two-stage model in the case where a single contrast from each subject is taken from level 1 to level 2. (a) The model for one subject of the finger tapping experiment, including the contrast that is applied to the parameter estimates to test if the activation of sequential finger tapping is different from random finger tapping. (b) The second-level model incorporating the first-level contrasts from 12 subjects, where the model produces group-level estimates pertaining to the first six subjects and the last six subjects. The contrast at this level compares the two groups of subjects.

STATISTICS TERMS

Efficiency

The efficiency of an estimator is the inverse of variance. If you have two estimators $\hat{E}_1$ and $\hat{E}_2$, where $\hat{E}_1$ is more efficient than $\hat{E}_2$, this implies that $Var(\hat{E}_1) < Var(\hat{E}_2)$.

Bayesian versus Classical Inference

Classical statistical inference is the basis for most widely known statistical procedures. Also known as frequentist inference, the approach assumes that there is a fixed, unknown parameter that describes a feature of a population (say, the mean BOLD response in a given brain region in a given experiment). The data, which is a random process over repetitions of the experiment, is collected to learn about this parameter. Classical inference is couched in terms of unlimited repeated samples of the population (in our case, of fMRI subjects). For example, the interpretation of a confidence interval (an interval about an estimate that expresses its uncertainty) requires reference to an infinite number of hypothetical replications of the experiment: A level 95% confidence interval will contain the true (fixed) parameter 95% of the time with many repetitions of the experiment.

Bayesian statistical inference regards the parameters as random instead of as fixed. Before any data is collected, the parameters are assigned an a priori distribution, called the prior. After the experiment, the prior is updated into a posterior, based on what has been learned about the parameter; the posterior is the distribution of the parameter conditional on the observed data. Bayesian inference is based on the posterior distribution. For example, a Bayesian confidence interval is an interval that has a given probability of containing the (random) parameter after having seen the data. There is no reference to the frequency of an event over ad infinitum repetitions of the experiment.

While Bayesian methods offer intuitive probabilistic statements about unknown quantities of interest (the parameters), they can be controversial. Different investigators may have different beliefs and so use different priors, and then get different results based on the same data. To address this, many authors use so-called noninformative priors, which exert as little influence on the posterior as possible. The two approaches, fortunately, can be reconciled. For most problems, with more and more data, the prior becomes less and less important, and Bayesian and classical inferences will generally agree.

Maximum Likelihood and Restricted Maximum Likelihood

For the following illustration, we use the first-level model, $Y_k = X_k\beta_k + \epsilon_k$ where $\epsilon \sim N(0, \sigma_k^2 V_k)$ and Y_k has length T_k. One of the difficulties in estimation is that there are multiple parameters to estimate, the components of β and the components of $\sigma_k^2 V_k$. The maximum likelihood (ML) and ReML are two methods that are used to estimate these parameters. The starting point of both of these methods is the formation of a likelihood equation or the joint probability distribution function of the random variables. In the ML case, $Y_k \sim N(X_k\beta_k, \sigma_k^2 V_k)$ for $k = 1, \ldots, N$ are the random variables of interest, and the likelihood is a function of β_k and $\sigma_k^2 V_k$, given by the product of the N normal distribution functions:

$$L(\beta_k, \sigma_k^2 V_k) = \prod_{k=1}^{N} \left\{ (2\pi)^{-T_k/2} \left|\sigma_k^2 V_k\right|^{-1/2} \exp\left(-\frac{1}{2} (Y_k - X_k\beta_k)^T \left(\sigma_k^2 V_k\right)^{-1/2} (Y_k - X_k\beta_k) \right) \right\}. \tag{19}$$

Since this likelihood cannot be maximized for both β_k and $\sigma_k^2 V_k$ simultaneously, first an estimate of $\sigma_k^2 V_k$ is plugged into (19) and the likelihood is maximized to find $\hat{\beta}_k$, then this estimate of β_k is substituted into (19), and it is maximized to estimate $\sigma_k^2 V_k$. This process is repeated until the estimates converge to a solution. Usually a specific structure of V_k, for example AR(1), is assumed to simplify the estimation process by reducing the number of variance parameters that need to be estimated. One of the pitfalls of this method is that when estimating the variance, an estimate of β_k is used instead of the true value. This causes the variance estimate to be biased in the case of maximum likelihood. For example, the ML one sample variance estimate is $\frac{1}{N}\sum_i (Y_i - \overline{Y})^2$ and is biased by a factor of $\frac{N}{N-1}$.

ReML starts with a different random variable $R_k = A_k Y_k$, where $A_k = (I - X_k(X_k^T X_k)^{-1} X_k^T)$, which has $R_k \sim N(0, A_k \sigma_k^2 V_k A_k^T)$, and so the likelihood is only a function of $\sigma_k^2 V_k$. This likelihood is then maximized to get the estimate of $\sigma_k^2 V_k$, and since β_k was not involved, the result is an unbiased estimate of the variance. For example, it can be shown that the ReML one sample variance estimate is $\frac{1}{N-1}\sum_i (Y_i - \overline{Y})^2$ and is unbiased. The ReML method only supplies an estimate for the variance parameters that are substituted into (19), which is maximized to find the estimate for β_k. Further details on ReML and ML can be found in (17).

PREPARATIONS FOR MULTISUBJECT MODELING

There are various preprocessing steps that must be applied to fMRI data before group modeling can be performed. Of the three steps, intrasubject registration, intersubject registration, and spatial smoothing, the second is the most crucial to group modeling: without intersubject registration, different subjects' brains will not line up and group modeling will be impossible.

Intrasubject Registration—Movement Correction

Despite experimenters' and subjects' efforts, subjects invariably move their heads in the magnet. If uncorrected, movement can be a significant source of nuisance variability. Consider that we are interested in finding BOLD signal changes on the order of 0.1–5%, yet if a subject moves his head a distance of one-half voxel, a voxel at the edge of the brain will experience a 50% change in intensity. Hence, successful estimation and correction of movement is necessary to find the subtle effects of interest.

Motion correction methods are all generally rigid body, estimating three translation and three rotation parameters to match a given image to the reference image, typically the first image collected. This is a classic image processing problem (see, e.g., (25)–(27)). The principal differences between methods are on the cost function to measure image similarity (typically least squares or mutual information), the optimization method, and the interpolation method used (which may differ between estimation and the final application of movement parameters).

Intersubject Registration

Everyone, even identical twins, has a uniquely shaped brain. Before group modeling of fMRI data can be performed, all subjects must be spatially transformed into a common space. Some times known as *spatial normalization*, this process finds a transformation that best warps a subject into a common atlas brain space <AU: is there a word missing here? the sentence is unclear> best corresponds to location $T(x)$ in a subject's brain. Finding the best parameterization of the transformation T is an active area of research (see, e.g., (28)–(31)). For the purposes of this work, we simply assume that the functional data have been spatially transformed such that a given voxel in each subject corresponds to the same atlas location, as best as is possible.

Spatial Smoothing

Human anatomy is highly variable, and two brains cannot not necessarily be matched gyri-to-gyri even when the registration is done manually. To overcome these limitations of intersubject registration, spatial smoothing is applied to blur out residual anatomical differences. Commonly used are Gaussian kernels with full width at half maximum of 5–10 mm.