Generalisability, Random Effects & Population Inference

A.P. Holmes, K.J. Friston

Wellcome Department of Cognitive Neurology, Institute of Neurology, London, UK

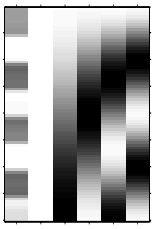
Introduction: This paper concerns the generalisability of inferences drawn from multi-subject functional neuroimaging experiments. Loosely speaking, a classical statistical hypothesis test proceeds by comparing the difference between the observed and hypothesised effect against the "yardstick" of variance. Most current methods of assessing functional neuroimaging data utilise only inter-scan variance, either explicitly via parametric statistics or implicitly via non-parametric methods. When subjects respond differentially, as they almost certainly do, this intra-subject variance is inappropriate for inter-subject inference. For inference to the populations from which subjects were drawn, a *random effects* analysis is required, accounting for both the inter-scan error variance *and* the inter-subject *component* of variance. This is particularly an issue in fMRI, where the inter-scan variability within an imaging session is very small in comparison to the variability of responses from subject to subject.

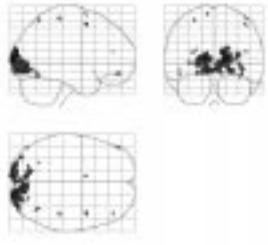
Theory: Consider a simple two condition fMRI activation experiment on n subjects: A basic *General Linear Model* [1] for the timecourses Y_{ij} at a single voxel is: $Y_{ij} = \gamma_i + \alpha_i f(j) + \cdots + \epsilon_{ij}$ for scan j on subject i. Here $f(\bullet)$ is a reference function, such as a box-car. ϵ_{ij} are the residual errors, usually assumed $\epsilon_{ij} \sim \mathcal{N}(0, \sigma_\epsilon^2)$. The parameters γ_i $(i = 1, \ldots, n)$ are additive subject (block) effects; α_i the magnitude of the activation for subject i. Current analyses [2] assume the parameters are *fixed effects*, and test for significant mean activation $\overline{\hat{\alpha}}_\bullet$ *for these subjects*. An experiment assessed thus can only be regarded as a case study. Typically group comparisons are similarly analysed, comparing average activation effects for subgroupings of the subjects, and inference cannot be extended beyond the subjects studied.

To extend inference to the population one must treat the subject activations α_i as *random effects*. Subjects are randomly sampled from the population, so the α_i are also randomly chosen. A simple model is $\alpha_i \sim \mathcal{N}(\alpha, \sigma_\alpha^2)$, where α is the *population* mean activation, about which we wish to infer. Thus we have a *hierarchical* model – an inter-subject level model on the parameters of the standard intra-subject level model presented above. In general such *mixed effects* models are difficult to assess, since we must estimate and account for both the inter-subject component of variance σ_α^2 (due to sampling subjects), and the intra-subject component σ_ϵ^2 (due to estimating the α_i's from the fMRI time series), in appropriate amounts [3].

Fortunately, most models are separable by subject (with corresponding design matrix subpartitions mutually orthogonal), such that the parameter estimates for each subject are independent. Then the two model levels can be separated, permitting implementation within existing functional neuroimaging software via a two-stage procedure. First, the individual subject level models are fitted, and the estimated activations $\hat{\alpha}_i$ written out as images. These are then assessed using a one sample t-test, implying a second level model: $\hat{\alpha}_i = \alpha + \epsilon_i$ with $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$, and it can be shown that σ^2 contains the appropriate variance components. This is intuitively appropriate, since the estimated activations are being assessed against their variance across subjects. More sophisticated hierarchical analyses are possible within this framework, using general inter-subject level models on contrast images derived from intra-subject models, permitting new types of experiment.

Example: Consider a six subject fMRI visual activation experiment (BA; 10 scans/epoch, B=fixation point, A="starfield simulation" with fixation; 8 cycles; TR=3.2s). For each subject i a simple model with smoothed box-car reference function is fitted (left panel), orthogonalised with respect to a "high-pass" filter of discrete cosine basis functions [4]. Images of the estimated activation parameter α_i are written out as an images (centre panel). The inter-subject level of the analysis proceeds using these subject activation images as "raw" data for a one sample t-test. The right panel depicts a maximum intensity projection of the resulting t-statistic image (5 d.f.), thresholded at p=0.001 (uncorrected).





Conclusion: Appropriate random effects analyses for multi-subject functional neuroimaging data are essential for valid population inference. They are standard in other fields, and are beginning to be demanded by discerning journals, especially for group comparisons. Such analyses can be effected using separable models in a multi-level approach as outlined above. The power of these analyses increases with sample size, pointing to experimental designs with more subjects and possibly fewer scans per subject.

1. Friston, K.J., Holmes, A.P. *et al.* (1995) *NeuroImage* **2**:45–53
2. Frackowiak, R.S.J. *et al.* (1997) *Human Brain Function* Academic Press
3. Searle, S.R. *et al.* (1992) *Variance Components* Wiley
4. Holmes, A.P. *et al.* (1997) *NeuroImage* **5**:S480