Notes on Random Effects Analysis

W.D. Penny

Wellcome Department of Cognitive Neurology,
University College London.

June 29, 2001

Abstract

For balanced designs, a random effects analysis of neuroimaging data can be implemented using Statistical Parametric Mapping (SPM) by employing a simple two-level inference procedure.

1 Introduction

Random effects models [1] are effected within SPM using a multi-stage approach. For a balanced design (i.e. the same design matrix for each subject), each subjects data is surmised with an appropriate single image. This "single scan per subject" data is then entered into a second model, i.e. a model at the between subject level. Usually this will be a simple one-sample or two-sample t-test. The variance of these single images from subject to subject consists of contributions from both the between and within subject components of variance, in the correct proportions, and it can be shown (see below) that the resulting analysis is mathematically identical to the appropriate random effects analysis of these data. The procedure was proposed in [2] and makes use of the 'mean summary statistic approach' [1].

2 Maximum-Likelihood approach

Given a data set of contrasts from N subjects with n scans per subject, the population contrast is modelled by a two level process

$$\begin{aligned} w_{ij} &= w_i + e_{ij} \\ w_i &= w_{pop} + z_i \end{aligned} \tag{1}$$

where $\mathsf{Var}[e_{ij}] = \sigma_w^2$ is the within-subject variance and $\mathsf{Var}[z_i] = \sigma_b^2$ is the between-subject variance. We also have $\mathsf{E}[e_{ij}] = \mathsf{E}[z_i] = 0$. Collapsing into a single-level gives

$$w_{ij} = w_{pop} + z_i + e_{ij} \tag{2}$$

[1]Strictly speaking, we should call this a mixed effects analysis as some effects are fixed (ie. deterministic variables such as experimental condition) and some effects are random (ie. random variables such as the mean activation of a particular subject) - but we will continue to abuse the terminology, as it has now gained a foothold in neuroimaging.

We therefore have $p(w_{ij}) = \mathsf{N}(w_{pop}, \sigma^2_{pop})$ where the population variance is $\sigma^2_{pop} = \sigma^2_b + \sigma^2_w$. The maximum-likelihood estimate of the population mean is

$$\hat{w}_{pop} = \frac{1}{Nn} \sum_{i=1}^{N} \sum_{j=1}^{n} w_{ij} \tag{3}$$

This estimate has a mean $\mathsf{E}[\hat{w}_{pop}] = w_{pop}$ and a variance given by

$$\begin{aligned} \mathsf{Var}[\hat{w}_{pop}] &= \mathsf{Var}\left[\sum_{i=1}^{N} \frac{1}{N} \sum_{j=1}^{n} \frac{1}{n}(w_{pop} + z_i + e_{ij})\right] \\ &= \mathsf{Var}\left[\sum_{i=1}^{N} \frac{1}{N} z_i\right] + \mathsf{Var}\left[\sum_{i=1}^{N} \frac{1}{N} \sum_{j=1}^{n} \frac{1}{n} e_{ij}\right] \\ &= \frac{\sigma^2_b}{N} + \frac{\sigma^2_w}{Nn} \end{aligned} \tag{4}$$

3 Mean summary statistic approach

In SPM-99 we use the two-level model

$$\begin{aligned} \bar{w}_i &= w_i + n_i \\ w_i &= w_{pop} + z_i \end{aligned} \tag{5}$$

where in the first level we consider the variation of the sample mean for each subject around the true mean for each subject. The corresponding variance is $\mathsf{Var}[n_i] = \sigma^2_w/n$. At the second level we consider the variation of the true subject means about the population mean where $\mathsf{Var}[z_i] = \sigma^2_b$. We also have $\mathsf{E}[n_i] = \mathsf{E}[z_i] = 0$. Consequently

$$\bar{w}_i = w_{pop} + z_i + n_i \tag{6}$$

The population mean is then estimated as

$$\hat{w}_{pop} = \frac{1}{N} \sum_{i=1}^{N} \bar{w}_i \tag{7}$$

This estimate has a mean $\mathsf{E}[\hat{w}_{pop}] = w_{pop}$ and a variance given by

$$\begin{aligned} \mathsf{Var}[\hat{w}_{pop}] &= \mathsf{Var}\left[\sum_{i=1}^{N} \frac{1}{N} \bar{w}_i\right] \\ &= \mathsf{Var}\left[\sum_{i=1}^{N} \frac{1}{N} z_i\right] + \mathsf{Var}\left[\sum_{i=1}^{N} \frac{1}{N} n_i\right] \\ &= \frac{\sigma^2_b}{N} + \frac{\sigma^2_w}{Nn} \end{aligned} \tag{8}$$

which is the same as for the two-level model in the previous section. We can see this more easily by simply substituting

$$\bar{w}_i = \frac{1}{n} \sum_{j=1}^{n} w_{ij} \tag{9}$$

into equation 7 which then becomes identical to equation 3. The above variance decomposition can be written as

$$\mathsf{Var}[\hat{w}_{pop}] = \frac{\mathsf{Var}[\bar{w}_i]}{N} \tag{10}$$

where $\mathsf{Var}[\bar{w}_i] = \sigma_b^2 + \sigma_w^2/n$ (see [2]).

3.1 Inference

If we estimate $\mathsf{Var}[\bar{w}_i]$ using

$$s_i^2 = \frac{1}{N-1}\sum_{i=1}^{N}(\bar{w}_i - \hat{w}_{pop})^2 \tag{11}$$

and let

$$t = \frac{\hat{w}_{pop}}{\sqrt{s_i^2/(N-1)}} \tag{12}$$

then t has a T-distribution with $N-1$ degrees of freedom. The corresponding t-test reveals whether or not the observed population mean is significantly non-zero.

4 Summary

Therefore, in SPM99, the basic procedure for a random effects analysis is as follows:

1. First, fit the model for each subject - you can model all the subjects together, provided you use a subject separable model (a multiple-subject fixed-effects analysis with subject separable models) - strictly speaking, the individual subject models should be identical (i.e. this is a balanced design)

2. Define the effect of interest for each subject with a t-contrast. - Each contrast will write a 'con' image containing the contrast of the parameter estimates

3. Armed with one contrast image per subject, proceed to the second (between-subject) level, feeding the contrast images into one of the "Basic stats" models as appropriate (ie. one-sample t-test, two-sample t-test).

If there are, say, two populations of interest and one is interested in making inferences about differences in the populations then a two-sample t-test is used at the second level. It is not necessary that the numbers of subjects in each population be the same (but it is necessary to have the same design matrices for subjects in the same population - ie. balanced designs at the first-level).

[2]Note that a similar decomposition does not hold for equation 3 because the error z_i is common to all the w_{ij} whereas the errors e_{ij} are specific to each w_{ij}. This means we can't write $\mathsf{Var}[\hat{w}_{pop}] = \mathsf{Var}[w_{ij}]/Nn$.

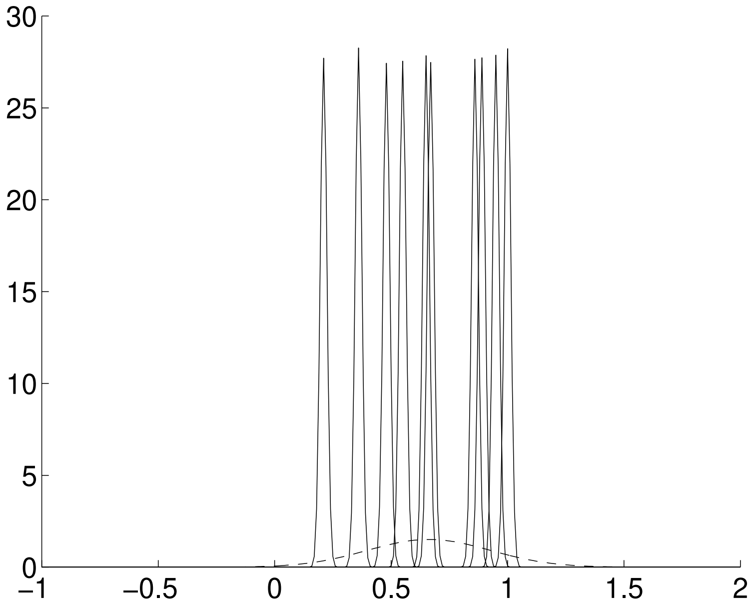

Figure 1: *Gaussian distributions of contrasts at a voxel in the left fusiform gyrus from an fMRI study for each of 10 subjects. The solid lines are the within-subject distributions and the dashed line is the between-subject distribution.*

5 Acknowledgements

Most of this material is taken from emails and help files written by Andrew Holmes.

A Transforming Variances

Under a linear transform $y = ax + b$, the variance of x changes according to

$$\mathsf{Var}[ax + b] = a^2 \mathsf{Var}[x] \tag{13}$$

If $\mathsf{Var}[x_i] = \mathsf{Var}[x]$ for all i then

$$\mathsf{Var}\left[\frac{1}{N}\sum_{i=1}^{N} x_i\right] = \frac{1}{N}\mathsf{Var}[x] \tag{14}$$

References

[1] L. Frison and S.J. Pocock. Repeated measures in clinical trials: An analysis using mean summary statistics and its implications for design. *Statistics in medicine*, 11:1685–1704, 1992.

[2] A.P. Holmes and K.J. Friston. Generalisability, random effects and population inference. In *NeuroImage*, volume 7, page S754, 1998.