An unbiased estimator for the roughness of a multivariate Gaussian random field

K.J. Worsley

July 28, 2000

Abstract

Images from positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) are often modelled as stationary Gaussian random fields, and a general linear model is used to test for the effect of explanatory varaibles on a set of such images (Friston *et al.*, 1994; Worsley and Friston, 1995). Thompson *et al.* (1996) have modelled displacements of brain surfaces as a multivariate Gaussian random field. In order to test for significant local maxima in such fields using the theory of Adler (1981) and its recent refinements (Worsley, 1994, 1995a, 1995b; Siegmund and Worsley, 1995), we need to estimate the roughness of such fields. This is defined as the variance matrix of the derivative of the random field in each dimension. Some methods have been given by Worlsey *et al.* (1992) for the special case of stationary variance of the random field, and where the random field is sampled on a uniform lattice. In this note we generalise to the case of multivariate Gaussian data with unknown non-stationary variance matrix, and non-lattice sampling. This latter is particularly important for studying the displacement of brain surfaces, which will be the subject of a future publication.

Addresses of corresponding author:
Department of Mathematics and Statistics
McGill University
805 Sherbrooke St. West
Montreal
Québec
Canada H3A 2K6.
e-mail: worsley@math.mcgill.ca
web: http://www.math.mcgill.ca/∼keith
ph: 1-514-398-3842
fax: 1-514-398-3899.

1 Model

We shall study the general case of m correlated observations at each point and a multivariate linear model for the mean. The model for the $m \times 1$ vector of correlated observations for the ith image $Y_i(\mathbf{x})$ at position $\mathbf{x} = (x_1, \ldots, x_d)'$ in d dimensional space is

$$\mathbf{Y}_i(\mathbf{x})' = \mathbf{g}_i'\beta(\mathbf{x}) + \epsilon_i(\mathbf{x})'\mathbf{\Sigma}(\mathbf{x})^{1/2} \qquad (i = 1, \ldots, n),$$

where

- $\mathbf{g}_i$ is a $p \times 1$ vector of explanatory variables for the ith image,
- $\beta(\mathbf{x})$ is an $p \times m$ matrix of parameters at position $\mathbf{x}$,
- $\mathbf{\Sigma}(\mathbf{x})$ is the $m \times m$ variance matrix at position $\mathbf{x}$, which depends on position but not on image, and $\mathbf{\Sigma}(\mathbf{x})^{1/2}$ is the upper triangular Cholesky factor of $\mathbf{\Sigma}(\mathbf{x})$, defined by $\mathbf{\Sigma}(\mathbf{x})^{1/2\prime}\mathbf{\Sigma}(\mathbf{x})^{1/2} = \mathbf{\Sigma}(\mathbf{x})$,
- $\epsilon_i(\mathbf{x})$ is an $m \times 1$ vector error term whose components are independent stationary Gaussian random fields with zero mean, and unit standard deviation. We make the usual assumption that $\epsilon_1(\mathbf{x}), \ldots, \epsilon_n(\mathbf{x})$ are independent for fixed $\mathbf{x}$.

Note that this model is the only one for which the usual t and F statistics for testing components of $\beta(\mathbf{x})$ are stationary random fields whose null distributions do not depend on the unknown $\mathbf{\Sigma}(\mathbf{x})$.

2 Estimator of the roughness

To simplify the notation, we shall drop the argument $\mathbf{x}$ from now on. We are interested in estimating the roughness of the error term, measured by the $d \times d$ matrix

$$\mathbf{\Lambda} = \mathrm{Var}(\partial\epsilon/\partial\mathbf{x}),$$

where ϵ is any component of ϵ_i. A natural estimator is to consider the sample variance of the derivative of a predictor of ϵ_i, pooled over components. Let $\mathbf{Y} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_n)'$ be the $n \times m$ matrix of image measurements, and let $\mathbf{G} = (\mathbf{g}_1, \ldots, \mathbf{g}_n)'$ be the $n \times p$ matrix of explanatory variables. Then the usual least squares estimator of β is

$$\hat{\beta} = (\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'\mathbf{Y},$$

and the residual of the ith image is

$$\mathbf{r}_i' = \mathbf{Y}_i' - \mathbf{g}_i'\hat{\beta} \qquad (i = 1, \ldots, n).$$

Then define the normalised residual

$$\mathbf{u}_i' = \mathbf{r}_i' \left(\sum_{i=1}^{n} \mathbf{r}_i \mathbf{r}_i' \right)^{-1/2} \qquad (i = 1, \ldots, n),$$

where $\mathbf{M}^{-1/2} = (\mathbf{M}^{1/2})^{-1}$. The proposed estimator is based on the sum of squares and cross-products matrix of the derivatives of the normalised residuals. Let λ_{jk} be the jk element of $\mathbf{\Lambda}$, $j, k = 1, \ldots, d$. Then an unbiased estimator of λ_{jk} is

$$\hat{\lambda}_{jk} = \frac{\nu - m - 1}{m(\nu - m)N} \sum_{\mathbf{x}} \sum_{i=1}^{n} \frac{\partial \mathbf{u}_i'}{\partial x_j} \frac{\partial \mathbf{u}_i}{\partial x_k}, \tag{1}$$

where $\nu = n - p$ is the degrees of freedom of the model and N is the number of points $\mathbf{x}$ in the sum.

3 Proof of unbiasedness

We now show that $\hat{\lambda}_{jk}$ is unbiased for λ_{jk}. Let $\mathbf{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_n)'$. We shall use the dot notation combined with a subscript j to indicate differentiation with respect to x_j. Then the inner sumation of (1) can be written as $\mathrm{tr}(\dot{\mathbf{U}}_j'\dot{\mathbf{U}}_k)$. Let $\mathbf{Z} = (\mathbf{r}_1, \ldots, \mathbf{r}_n)'\mathbf{\Sigma}^{-1/2}$ so that we can write

$$\mathbf{U} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1/2}.$$

Then differentiating with respect to x_j

$$\dot{\mathbf{U}}_j = \{\mathbf{I}_n - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\}\dot{\mathbf{Z}}_j(\mathbf{Z}'\mathbf{Z})^{-1/2},$$

where $\mathbf{I}_n$ is the $n \times n$ identity matrix. Then

$$\dot{\mathbf{U}}_j'\dot{\mathbf{U}}_k = (\mathbf{Z}'\mathbf{Z})^{-1/2'}\dot{\mathbf{Z}}_j'\{\mathbf{I}_n - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\}\dot{\mathbf{Z}}_k(\mathbf{Z}'\mathbf{Z})^{-1/2}.$$

Let $\mathbf{R} = \mathbf{I}_n - \mathbf{G}(\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'$. Then $(\dot{\mathbf{Z}}_j, \dot{\mathbf{Z}}_k)$ is multivariate normal with mean zero and

$$\mathrm{Cov}(\dot{\mathbf{Z}}_j, \dot{\mathbf{Z}}_k) = \lambda_{jk}\mathbf{R} \otimes \mathbf{I}_m,$$

independent of $\mathbf{Z}$. Hence conditional on $\mathbf{Z}$

$$\begin{aligned} \mathrm{E}\{\mathrm{tr}(\dot{\mathbf{U}}_j'\dot{\mathbf{U}}_k) \mid \mathbf{Z}\} &= \lambda_{jk}\mathrm{tr}[\mathbf{R}\{\mathbf{I}_n - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\}]\mathrm{tr}\{(\mathbf{Z}'\mathbf{Z})^{-1}\} \\ &= \lambda_{jk}(\nu - m)\mathrm{tr}\{(\mathbf{Z}'\mathbf{Z})^{-1}\}. \end{aligned}$$

The last step is obtained by noting that $\mathrm{tr}(\mathbf{R}) = \nu$ and $\mathbf{RZ} = \mathbf{Z}$. Now $\mathbf{Z}'\mathbf{Z} \sim \mathrm{Wishart}_m(\mathbf{I}_m, \nu)$ and it can be shown that the diagonal elements of $(\mathbf{Z}'\mathbf{Z})^{-1}$ have the same distribution as the inverse of a $\chi^2_{\nu-m+1}$ random variable (see for example Anderson, 1984, page 130). Hence

$$\mathrm{E}\{\mathrm{tr}(\dot{\mathbf{U}}_j'\dot{\mathbf{U}}_k)\} = \lambda_{jk}(\nu - m)\frac{m}{\nu - m - 1}.$$

Dividing both sides by $(\nu - m)m/(\nu - m - 1)$ and averaging over $\mathbf{x}$ proves the result.

4 Correlated images from fMRI

The result can be extended to the case of correlations between images, as in fMRI data. Suppose the correlation matrix of the images, determined by the hemodynamic response function, is $\mathbf{V}$. Then following the above arguments,

$$\mathrm{Cov}(\dot{\mathbf{Z}}_j, \dot{\mathbf{Z}}_k) = \lambda_{jk}\mathbf{RVR} \otimes \mathbf{I}_m.$$

Hence conditional on $\mathbf{Z}$

$$\begin{aligned} \mathrm{E}\{\mathrm{tr}(\dot{\mathbf{U}}_j'\dot{\mathbf{U}}_k) \mid \mathbf{Z}\} &= \lambda_{jk}\mathrm{tr}[\mathbf{RVR}\{\mathbf{I}_n - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\}]\mathrm{tr}\{(\mathbf{Z}'\mathbf{Z})^{-1}\} \\ &= \lambda_{jk}[\mathrm{tr}\{\mathbf{RV}\} - \mathrm{tr}\{\mathbf{Z}'\mathbf{VZ}(\mathbf{Z}'\mathbf{Z})^{-1}\}]\mathrm{tr}\{(\mathbf{Z}'\mathbf{Z})^{-1}\}. \end{aligned}$$

Now $\mathrm{E}(\mathbf{Z}'\mathbf{Z}) = \mathrm{tr}(\mathbf{RV})\mathbf{I}_m$ and $\mathrm{E}(\mathbf{Z}'\mathbf{VZ}) = \mathrm{tr}(\mathbf{RVRV})\mathbf{I}_m$. Substituting these into the above gives

$$\mathrm{E}\{\mathrm{tr}(\dot{\mathbf{U}}_j'\dot{\mathbf{U}}_k)\} \approx \lambda_{jk}(\nu_{\mathrm{eff}} - m)\frac{m}{\nu_{\mathrm{eff}} - m - 1}.$$

where

$$\nu_{\mathrm{eff}} = \mathrm{tr}(\mathbf{RV})^2/\mathrm{tr}(\mathbf{RVRV})$$

is the effective degrees of freedom as defined by Worsley and Friston (1995). Thus replacing ν by ν_{eff} in (1) gives approximate unbiased estimators.

5 Applications

For lattice data, derivatives can be approximated numerically by differences of $\mathbf{u}_i$ between adjacent lattice points, divided by the lattice step size (see Worsley *et al.*, 1992, for the details). For non-lattice data, this provides an estimate of the variance of the derivative only in the direction of the vector joining the two points. We now show how to obtain an unbiased estimator of all the variances and covariances of the derivatives.

Restoring the dependence on $\mathbf{x}$, denote the coordinates of a pair of adjacent points by $\mathbf{x}$ and $\mathbf{x} + \delta\mathbf{h}$, where $\mathbf{h}$ is the unit vector joining the points and δ is the distance between them. Then following the above arguments it is straightforward to show from (1) that

$$\Delta = \frac{\nu - m - 1}{m(\nu - m)} \sum_{i=1}^{n} ||\mathbf{u}_i(\mathbf{x} + \delta\mathbf{h}) - \mathbf{u}_i(\mathbf{x})||^2/\delta^2$$

is an unbiased estimator of $\mathbf{h}'\mathbf{\Lambda}\mathbf{h}$, in the limit as $\delta \to 0$. In effect, then, each pair of points provides us with an unbiased estimator of a linear combination of the elements of $\mathbf{\Lambda}$.

This information can be combined to provide an unbiased estimator of all the elements of $\mathbf{\Lambda}$ by using least squares. Let $\boldsymbol{\Delta}$ be the $M \times 1$ vector of the values of Δ for all M pairs of adjacent points. Let vech be the operator that arranges the distinct elements of a $d \times d$ symmetric matrix into a $d(d-1)/2 \times 1$ vector. Let $\mathbf{H}$ be an $M \times d(d-1)/2$ matrix whose rows are $\mathrm{vech}(\mathbf{hh}')$ for all M corresponding pairs of adjacent points. Then the least squares estimator of $\mathbf{\Lambda}$ is given by

$$\mathrm{vech}(\hat{\mathbf{\Lambda}}) = (\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'\boldsymbol{\Delta}. \quad (2)$$

It is interesting to compare the estimator (2) to the more direct estimator of Worsley *et al.* (1992) for lattice data, which can be generalised to this setting as follows. The diagonal elements λ_{jj} are estimated by averaging Δ over all pairs of points separated by one lattice step in coordinate j. The off-diagonal elements λ_{jk} are estimated by averaging the product of differences in coordinates j and k averaged on the sides of a square of four adjacent lattice points; this avoids a possible bias in estimating the off-diagonal elements. It can be shown that this estimator is identical to (2) provided the list of adjacent points contains all points separated by one lattice step, and all points separated by a 'diagonal' lattice step, that is, one step in coordinate j and one step in coordinate k. Thus the estimator of Worsley *et al.* (1992) can be seen as a special case of the non-lattice estimator (2) presented here.

References

Adler, R.J. (1981). *The Geometry of Random Fields.* Wiley, New York.

Anderson, T.W. (1984). *An Introduction to Multivariate Statistical Analysis, 2nd edition.* Wiley, New York.

Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.B., Frith, C.D. and Frackowiak, R.S.J. (1995). Statistical parametric maps in functional imaging: A general approach. *Human Brain Mapping*, **2**, 189-210.

Siegmund, D.O & Worsley, K.J. (1995). Testing for a signal with unknown location and scale in a stationary Gaussian random field. *Annals of Statistics*, **23**, 608-639.

Thompson, P.M., Schwartz, C. and Toga, A.W. (1996). High-resolution random mesh algorithms for creating a probabilistic 3D surface atlas of the human brain. *NeuroImage*, **3**, 19-34.

Worsley, K.J. (1994). Local maxima and the expected Euler characteristic of excursion sets of χ^2, F and t fields. *Advances in Applied Probability*, **26**, 13-42.

Worsley, K.J. (1995a). Estimating the number of peaks in a random field using the Hadwiger characteristic of excursion sets, with applications to medical images. *Annals of Statistics*, **23**, 640-669.

Worsley, K.J. (1995b). Boundary corrections for the expected Euler characteristic of excursion sets of random fields, with an application to astrophysics. *Advances in Applied Probability*, **27**, 943-959.

Worsley, K.J., Evans, A.C., Marrett, S. and Neelin, P. (1992). A three dimensional statistical analysis for CBF activation studies in human brain. *Journal of Cerebral Blood Flow and Metabolism*, **12**, 900-918.

Worsley, K.J. and Friston, K.J. (1995). Analysis of fMRI time-series revisited - again. *NeuroImage*, **2**, 173-181.