Quadratic tests for local changes in random fields with applications to medical images

K. J. WORSLEY and A. C. VANDAL

McGill University[1]

Key words and phrases: Random fields, image analysis, change points, repeated measures, shrinkage estimators.
AMS 1980 subject classifications: Primary 62J10; secondary 62M99.

ABSTRACT

Positron emission tomography (PET) produces images of blood flow that can be used to find the regions of the brain that are activated by a linguistic task or a sensory stimulus. Several images are obtained on each subject in a randomised block design, and the first step is to test if any change in mean or 'activation' has occurred. We first treat the images as repeated measures in space and propose an ANOVA-like quadratic test statistic for testing whether any activation has taken place. We show that it has some optimality properties for detecting a single peak of increased mean activation at an unknown location in the image. We also investigate a statistic that is used in the medical literature which is based on the proportion of the image that exceeds a fixed threshold value. For both test statistics we provide a simple approximate null distribution in which the effective degrees of freedom, or effective number of independent observations in the image, depends on the volume of the brain multiplied by a known measure of the resolution of the PET camera. Once activation is detected, we propose a Stein-type shrinkage estimator for the mean change that has lower mean squared error than the usual sample average. The methods are illustrated on a set of cerebral blood flow images from an experiment in pain perception.

[1]Department of Mathematics and Statistics, McGill University, 805 rue Sherbrooke ouest, Montréal, Québec, Canada, H3A 2K6. The authors would like to thank Alan Evans, Sean Marrett, Peter Neelin, Cathy Bushnell and Gary Duncan of the Montreal Neurological Institute for permission to use data collected in their experiments. This research was supported by the Natural Sciences and Engineering Research Council of Canada, la Formation de Chercheurs et l'Aide à la Recherche du Québec and the Medical Research Council of Canada.

1 Introduction

The PET technique uses positron-emitting isotope labelled carriers, created in an on-site cyclotron, to produce an image of brain activity such as glucose utilization, oxygen consumption or blood flow. The PET technique has been studied in the statistics literature by Vardi, Shepp and Kaufmann (1985). Worsley, Evans, Strother and Tyler (1991) have modelled the spatial correlation of PET images of glucose utilization. In the last three years, researchers have succeeded in carrying out a new type of experiment in which a group of subjects in a randomised block design are given a set of stimuli such as a painfull heat stimulus or word recognition. The number of subjects and stimuli is small, typically 10 and 6 respectively. The unusual statistical feature is that each observation is a three-dimensional image of blood flow in the brain. By careful alignment it is possible to subtract the blood flow image under one stimulus condition from that under another to look for changes in blood flow, or activation between the two stimuli. Such an experiment is fully described in Worsley, Evans, Marrett and Neelin (1992) and the relevant details of the analysis will now be given.

Suppose that there are n independent subjects and that for the ith subject, $Z_i(\mathbf{t})$ is the value of this subtracted image at a point $\mathbf{t} \in C$, where $C \subset \mathrm{I\!R}^3$ is the brain, $i = 1, \cdots, n$. We propose the model

$$Z_i(\mathbf{t}) = \mu_Z(\mathbf{t}) + \sigma(\mathbf{t})\epsilon_i(\mathbf{t})$$

where $\mu_Z(\mathbf{t})$ is the mean change in blood flow between the two stimuli, $\sigma(\mathbf{t})$ is the standard deviation, and $\epsilon_i(\mathbf{t})$ is a zero mean, unit variance, stationary Gaussian random field, $i = 1, \cdots, n$. All three functions $\mu_Z(\mathbf{t})$, $\sigma(\mathbf{t})$ and $\epsilon_i(\mathbf{t})$ are assumed to be smooth functions of $\mathbf{t}$, $i = 1, \cdots, n$. We shall refer to $\mu_Z(\mathbf{t})$ as the signal, and $\sigma(\mathbf{t})\epsilon_i(\mathbf{t})$ as the noise in the image. We are interested in detecting a non-zero signal $\mu_Z(\mathbf{t})$ that is thought to be high in a small number of isolated 'regions of activation' and zero elswhere in the brain. The location and indeed the number of these regions of activation is unknown.

In section 2 we shall take an ANOVA approach, treating the images as repeated measures in space on the same subject, and we shall work with simple mean squares test statistics. Similar ANOVA methods have been used by Raz and Fein (1992) for EEG data taken over time, and by Smoot et al. (1992) to test for fluctuations in the cosmic microwave background. In section 3 we shall study a test used by Friston et al. (1990) based on the proportion of the image that exceeds a fixed threshold value. Worsley et al. (1993) have proposed likelihood ratio methods which we compare with the quadratic test and the exceedence proportion test in section 4. Once activation has been detected, we propose in section 5 a Stein-type shrinkage estimator of $\mu_Z(\mathbf{t})$ that has lower mean squared error than the sample average. Our results depend on the correlation function of the images, and in section 6 we shall suggest methods for determining it using the known resolution of the PET camera. The specificity and sensitivity of the tests will be compared using simulations in section 7, followed by an application to a PET experiment in pain perception.

2 ANOVA approach

2.1 Known σ

Let $\bar{Z}(\mathbf{t})$ be the sample mean of the differenced images, that is

$$\bar{Z}(\mathbf{t}) = \sum_{i=1}^{n} Z_i(\mathbf{t})/n.$$

We shall assume that the standard deviation is stationary, that is $\sigma(\mathbf{t}) = \sigma$, say, and that σ is known. The sample mean image is then standardised to give

$$X(\mathbf{t}) = n^{\frac{1}{2}} \bar{Z}(\mathbf{t})/\sigma = \mu(\mathbf{t}) + \epsilon(\mathbf{t}), \tag{2.1}$$

where

$$\mu(\mathbf{t}) = n^{\frac{1}{2}} \mu_Z(\mathbf{t})/\sigma \quad \text{and} \quad \epsilon(\mathbf{t}) = \sum_{i=1}^{n} \epsilon_i(\mathbf{t})/n^{\frac{1}{2}}.$$

Note that $X(\mathbf{t})$ is sufficient for $\mu(\mathbf{t})$, and that the noise component $\epsilon(\mathbf{t})$ is also a stationary Gaussian field with zero mean and unit variance.

The images can be regarded as repeated measures on a single individual. This suggests an ANOVA approach using a test statistic based on a quadratic function of $X(\mathbf{t})$. The simplest choice is the mean sum of squares

$$U = \int_C X(\mathbf{t})^2 d\mathbf{t}/|C|, \tag{2.2}$$

where $|C|$ is the volume of C. However we would like the test statistic to have good power at detecting a signal $\mu(\mathbf{t})$ that is high in a few isolated regions and zero elsewhere. We now show that the test statistic (2.2) is asymptotically locally most powerful amongst a large class of 'stationary' quadratic test statistics, where $\mu(\mathbf{t})$ is a single peak of unknown height and location, but with the same shape as the correlation function. Let

$$R(\mathbf{h}) = \text{Cov}\{X(\mathbf{t}), X(\mathbf{t} + \mathbf{h})\}$$

be the correlation function of $X(\mathbf{t})$.

Lemma 1 *Suppose that $\epsilon(\mathbf{t})$ is stationary and periodic inside an interval $I = [\mathbf{0}, \mathbf{c}) \subset \mathbb{R}^N$, that is $R(\mathbf{h} + \mathbf{c}_i) = R(\mathbf{h})$ where $\mathbf{c}_i$ is an N-vector with ith component equal to the ith component of $\mathbf{c}$ and zero elsewhere, $i = 1, \ldots, N$, $\mathbf{h} \in \mathbb{R}^N$. Let $D(\mathbf{h})$ be any symmetric periodic function, that is $D(\mathbf{h}) = D(-\mathbf{h})$ and $D(\mathbf{h} + \mathbf{c}_i) = D(\mathbf{h})$, $i = 1, \ldots, N$, $\mathbf{h} \in \mathbb{R}^N$, and let*

$$Q = \int_I \int_I X(\mathbf{t}_1) D(\mathbf{t}_1 - \mathbf{t}_2) X(\mathbf{t}_2) d\mathbf{t}_1 d\mathbf{t}_2.$$

We shall require that $\text{Var}(Q)/|I|$ is finite and bounded away from zero in the limit as I becomes large in each dimension. Then the statistic U evaluated on $C = I$ is asymptotically locally uniformly most powerful amongst all test statistics of the form Q for testing

$$H_0 : \mu(\mathbf{t}) = 0 \quad \text{against} \quad H_1 : \mu(\mathbf{t}) = \xi R(\mathbf{t} - \tau),$$

where $\xi > 0$ is an unknown peak height, and $\tau \in C$ is an unknown location.

Proof. Let $\star$ denote convolution over the interval I and $\langle f, g\rangle$ denote the inner product $\int_I f(\mathbf{t})g(\mathbf{t})d\mathbf{t}$. Then we can write $Q = \langle X, D \star X\rangle$. Then by expanding and taking expectations it can be shown using the periodicity of D and R that

$$\mathrm{E}(Q) = |I|\langle D, R\rangle + \langle \mu, D \star \mu\rangle,$$

$$\mathrm{Var}(Q) = 2|I|\langle D \star R, D \star R\rangle + 4\langle \mu, D \star R \star D \star \mu\rangle,$$

(see for example Seber (1977), Theorem 1.8, page 14 for the analogous result for quadratic forms of vectors). In the limit as I becomes large it can be shown that Q approaches a normal distribution, as follows. The statistic Q is simply the sum of squares of the periodogram of X weighted by the periodogram of D, that is a weighted sum of squares of independent normal random variables. It is easily shown that this converges locally to normality provided the sums of sqaures of the products of the periodogram of D and the periodogram of R are finite and greater than zero. This condition is equivalent to $\mathrm{Var}(Q)/|I|$ being finite and bounded away from zero. Hence the local power of Q is an increasing function of

$$\Delta = \{\mathrm{E}(Q|H_1) - \mathrm{E}(Q|H_0)\}/\mathrm{Var}(Q|H_0)^{1/2} = \langle \mu, D \star \mu\rangle/(2|I|\langle D \star R, D \star R\rangle)^{1/2}.$$

Now since μ is proportional to R shifted by an amount τ, and because R is periodic, then we can write

$$\Delta = \xi^2\langle R, D \star R\rangle/(2|I|\langle D \star R, D \star R\rangle)^{1/2} \le \xi^2\langle R, R\rangle/(2|I|)^{1/2}$$

by the Cauchy-Shwartz inequality, with equality if $D \star R$ is proportional to R, that is if $Q = \langle X, X\rangle/|I| = U$. □

If $\epsilon(\mathbf{t})$ is not periodic, and C is not an interval but C is large and τ is far from the boundary of C, then we may expect this result to hold approximately. The above result is also useful for deciding how to proceed in the case when the signal is not proportional to the correlation function. Suppose that we can observe a hypothetical high-resolution 'white noise' image of independent Gaussian variates at each voxel, to which is added an unknown signal $\xi\kappa(\mathbf{t} - \tau)$ say. Then if this image is smoothed, or convoluted, with a kernel proportional to $\kappa(\mathbf{h})$ then it is straightforward to check that the resulting image has a mean which is proportional to the correlation function centred at τ. Thus in this case the above lemma applies and U is optimal. Thus a powerful test should be obtained by using the mean sum of squares of the white noise image smoothed with the signal itself. This implies that broad regions of activation are best detected by smoothing the image, and the broader the region, the more the image should be smoothed. In practice this is actually done; brain images are arbitrarily smoothed to a resolution that is approximately three times lower than that actually obtainable.

Working directly from the definition of U, it can be shown that under H_0

$$\mathrm{E}(U) = 1, \quad \text{and} \quad \mathrm{Var}(U) = 2\int_C\int_C R(\mathbf{t}_2 - \mathbf{t}_1)^2 d\mathbf{t}_1 d\mathbf{t}_2 \Big/ |C|^2. \tag{2.3}$$

For large volumes we can approximate the null distribution of U by a normal random variable with unit mean and variance (2.3), provided the variance is finite. Since U is a mean sum of squares it is natural to try a χ^2 approximation with ν degrees of freedom, given by

$$\nu U \sim \chi^2_\nu, \text{ and } \nu = |C|^2 \Big/ \int_C\int_C R(\mathbf{t}_2 - \mathbf{t}_1)^2 d\mathbf{t}_1 d\mathbf{t}_2. \tag{2.4}$$

A similar approximation for variance components dates back at least to Satterthwaite (1946). It has the advantage of being accurate either if C is very large or if C is very small, since in the latter case $X(\mathbf{t})$ is approximately constant and $R(\mathbf{t}_2 - \mathbf{t}_1) \approx 1$ inside C, and so U has an approximate χ^2 distribution with $\nu \approx 1$ degree of freedom. Note that the effective degrees of freedom ν increases as the image becomes less correlated, or as the resolution increases, and in applications it is typically greater than 100.

For large C we can find a good lower bound for ν. Let

$$\tilde{\nu} = |C| \Big/ \int_{\mathbb{R}^N} R(\mathbf{h})^2 d\mathbf{h}.$$

By substituting $\mathbf{t}_2 = \mathbf{t}_1 + \mathbf{h}$ into (2.4) and integrating out $\mathbf{t}_1$ we have

$$\nu \geq |C|^2 \Big/ \int_{\mathbb{R}^N} \int_C R(\mathbf{h})^2 d\mathbf{t}_1 d\mathbf{h} = \tilde{\nu}.$$

with equality in the limit as C becomes large.

2.2 Unknown σ

In practice σ is unknown so we must replace it with an estimator. Let $m = n - 1$ and let $S^2(\mathbf{t})$ be the sample variance over subjects, defined by

$$S^2(\mathbf{t}) = \sum_{i=1}^{n} \{Z_i(\mathbf{t}) - \bar{Z}(\mathbf{t})\}^2 / m.$$

The variance σ^2 is estimated independently and unbiasedly by pooling the sample variance over all $\mathbf{t} \in C$ to give

$$\hat{\sigma}^2 = \int_{\mathbf{t} \in C} S^2(\mathbf{t}) d\mathbf{t} / |C|.$$

We can now replace σ by $\hat{\sigma}$ to give an analogue of the usual ANOVA F statistic:

$$F = U / \hat{\sigma}^2.$$

The above arguments suggest that the null distribution of F can be well approximated by an F-distribution with ν and $m\nu$ degrees of freedom; Box (1954) made the analogous approximation for F statistics from ANOVA with correlated observations.

It is interesting to note that the test based on F shows some robustness against non-stationary $\sigma(\mathbf{t})$. The reason is that the numerator U and the denominator $\hat{\sigma}^2$ both have the same null expectation, which then cancels in the F ratio. However the degrees of freedom ν should be decreased as follows. Let $\star$ denote convolution over C and $\langle f, g \rangle$ denote the inner product $\int_C f(\mathbf{t}) g(\mathbf{t}) d\mathbf{t}$, so we can write $U = \langle X, X \rangle / |C|$. Let σ now denote the function $\sigma(\mathbf{t})$. Then by expanding and taking expectations we have

$$\mathrm{E}(U) = \mathrm{E}(\hat{\sigma}^2) = \langle \sigma, \sigma \rangle / |C|$$

$$\mathrm{Var}(U) = m \mathrm{Var}(\hat{\sigma}^2) = 2 \langle \sigma^2, R^2 \star \sigma^2 \rangle / |C|^2.$$

The same reasoning as above suggests approximating F by an F distribution with $\nu_{\rm ns}$ and $m\nu_{\rm ns}$ degrees of freedom, where

$$\nu_{\rm ns} = \beta/\gamma, \quad \beta = \langle \sigma, \sigma \rangle^2 \quad \text{and} \quad \gamma = \langle \sigma^2, R^2 \star \sigma^2 \rangle.$$

Following Huynh and Feldt (1976) we can find unbiased estimators for β and γ by the method of moments. Let

$$\hat{\beta} = \langle S^2, 1/\{1 + 2R^2/m\} \star S^2 \rangle \text{ and } \hat{\gamma} = \langle S^2, R^2/\{1 + 2R^2/m\} \star S^2 \rangle. \tag{2.5}$$

Then $\hat{\beta}$ and $\hat{\gamma}$ are unbiased for β and γ respectively, since

$$\mathrm{E}\{S(\mathbf{t}_1)^2 S(\mathbf{t}_2)^2\} = \sigma(\mathbf{t}_1)^2 \sigma(\mathbf{t}_2)^2 \{1 + 2R(\mathbf{t}_2 - \mathbf{t}_1)^2/m\}. \tag{2.6}$$

This suggests replacing $\nu_{\rm ns}$ by the estimator

$$\hat{\nu}_{\rm ns} = \hat{\beta}/\hat{\gamma}.$$

In practice the calcalations for $\hat{\nu}_{\rm ns}$ are quite laborious. Instead we can use the simpler spatial analogues of the Huynh-Feldt estimator of $\nu_{\rm ns}$, which can be calculated in the following way. Let $\mathbf{S}$ be the $n \times n$ sample variance matrix with (i,j) element $S_{ij} = \langle Z_i - \bar{Z}, Z_j - \bar{Z} \rangle / (|C|m)$, $i, j = 1, \ldots, n$, so that $\hat{\sigma}^2 = \mathrm{tr}(\mathbf{S}) = \sum_i S_{ii}/m$. Then

$$\hat{\nu}_S = \mathrm{tr}(\mathbf{S})^2/\mathrm{tr}(\mathbf{S}^2) = \hat{\sigma}^4/(\sum_{i,j} S_{ij}^2) \tag{2.7}$$

is the sample analogue of $\nu_{\rm ns}$. The Huynh-Feldt estimator of $\nu_{\rm ns}$ is then

$$\hat{\nu}_{\rm HF} = \{(m+1)\hat{\nu}_S - 2\}/(m - \hat{\nu}_S). \tag{2.8}$$

Although $\hat{\nu}_{\rm HF}$ is derived under no assumptions about the correlation structure, it is very sensitive to heteroscedasticity of the subjects, as we shall see in the next section.

2.3 Heteroscedasticity

Suppose now the subject standard deviations are unequal, that is

$$Z_i(\mathbf{t}) = \mu_Z(\mathbf{t}) + \sigma_i \epsilon_i(\mathbf{t}),$$

but $\sigma(\mathbf{t})$ is stationary within the same subject. Then if we replace σ by $\bar{\sigma} = (\sum_{i=1}^n \sigma_i^2)^{1/2}$ in the definition of $X(\mathbf{t})$ (2.1) then the null distribution of U remains the same as before. We would, of course, expect a slight loss of power because $X(\mathbf{t})$ is no longer the best linear unbiased estimator of $\mu(\mathbf{t})$. When the subject standard deviations are unknown it is straightforward to show that $\bar{\sigma}^2$ is unbiasedly estimated by $\hat{\sigma}^2$. Thus the numerator and denominator of the F statistic have equal expectations under the null hypothesis, and we can try approximating the distribution of F by an F-distribution. Clearly the effective numerator degrees of freedom are the same as before, ν, but we now show that the effective denominator degrees of freeedom are less than $m\nu$.

Let $\mathbf{V} = \text{diag}(\sigma_1^2, \ldots, \sigma_n^2)$, and let $\mathbf{M}$ be the $n \times n$ centering matrix divided by m, with diagonal elements $1/n$ and off-diagonal elements $-1/(nm)$. Let $\mathbf{Z}(\mathbf{t}) = (Z_1(\mathbf{t}), \ldots, Z_n(\mathbf{t}))'$ be the n-vector of all observations at point $\mathbf{t}$. Then we can write

$$\hat{\sigma}^2 = \int_{\mathbf{t} \in C} \mathbf{Z}(\mathbf{t})'\mathbf{M}\mathbf{Z}(\mathbf{t}) d\mathbf{t}/|C|,$$

and so

$$\text{E}(\hat{\sigma}^2) = \int_{\mathbf{t} \in C} \text{tr}(\mathbf{MV}) d\mathbf{t}/|C| = \text{tr}(\mathbf{MV}),$$

and

$$\text{Var}(\hat{\sigma}^2) = \int_C \int_C 2\text{tr}(\mathbf{MVMV}) R(\mathbf{t}_2 - \mathbf{t}_1)^2 d\mathbf{t}_1 d\mathbf{t}_2/|C|^2 = 2\text{tr}(\mathbf{MVMV})/\nu.$$

The effective degrees of freedom of $\hat{\sigma}^2$ is thus

$$2\text{E}(\hat{\sigma}^2)/\text{Var}(\hat{\sigma}^2) = m_{\text{het}}\nu,$$

say, where

$$m_{\text{het}} = \text{tr}(\mathbf{MV})^2/\text{tr}(\mathbf{MVMV}) = m/\{1 + \text{CV}^2(m-1)/(m+1)\}, \qquad (2.9)$$

and CV is the coefficient of variation of $\sigma_1^2, \ldots, \sigma_n^2$; clearly $m_{\text{het}} \leq m$.

Following the procedure in the previous section we use the method of moments to find an estimator of m_{het}. An obvious candidate is the sample equivalent found by replacing $\mathbf{MVM}$ in (2.9) by its unbiased estimator $\mathbf{S}$, suggesting that m_{het} could be estimated by $\hat{\nu}_S$ from (2.7). However the numerator and denominator of ν_S are slightly biased, and better unbiased estimators can be found by the method of moments, as for the Huynh-Feldt estimator. In this case we get

$$\begin{aligned} \text{E}\{\text{tr}(\mathbf{S})^2\} &= 2\text{tr}(\mathbf{MVMV})/\nu + \text{tr}(\mathbf{MV})^2, \\ \text{E}\{\text{tr}(\mathbf{S}^2)\} &= \{\text{tr}(\mathbf{MV})^2 + \text{tr}(\mathbf{MVMV})\}/\nu + \text{tr}(\mathbf{MVMV}). \end{aligned}$$

Removing expectations and solving for m_{het} gives

$$\hat{m}_{\text{het}} = \{(\nu+1)\hat{\nu}_S - 2\}/(\nu - \hat{\nu}_S), \qquad (2.10)$$

which approaches $\hat{\nu}_S$ for large ν. In practice we would of course use $\tilde{\nu}$ as an approximation to ν. Note the close resemblance between $\hat{m}_{\text{het}}$ (2.10) and the Huynh-Feldt estimator $\hat{\nu}_{\text{HF}}$ (2.8); ν in (2.10) replaces m in (2.8).

2.4 Robustness

If $\sigma(\mathbf{t})$ is non-stationary, then we saw in section 2.2 that both numerator and denominator degrees of freedom should be reduced. Some idea of the reduction can be gained by supposing that $\sigma(\mathbf{t})^2 = \sigma^2 \chi^2_\eta(\mathbf{t})/\eta$, where $\chi^2_\eta(\mathbf{t})$ is a realisation of a χ^2 random field with η degrees of freedom formed by taking the sum of squares of η zero mean, unit variance stationary Gaussian random fields with the same correlation structure as $Z_i(\mathbf{t})$. We may then gain some idea of the size of $\nu_{\rm ns}$ by finding the average of a and b over all 'random' realisations of $\sigma(\mathbf{t})^2$. Suppose the root mean squared deviation of $\sigma(\mathbf{t})^2$ about its average value of one is δ, so that $\eta = 2/\delta^2$, then in view of (2.6), we have

$$\nu_{\rm ns} \approx \frac{\nu + \delta^2}{1 + \nu\delta^2 \int\int R(\mathbf{t}_1 - \mathbf{t}_2)^4 d\mathbf{t}_1 d\mathbf{t}_2} = \frac{\nu + \delta^2}{1 + \delta^2/2^{N/2}}$$

for the Gaussian correlation function introduced in section 6. In most applications ν is large, and so even with $\delta = 0.5$ in $N = 3$ dimensions the value of $\nu_{\rm ns}$ is approximately 92% of that of ν; this translates into an increase in false positive rate from 0.05 to 0.057 for $\nu = 300$, a value typical for the medical applications discussed in section 7. This suggests that F is quite robust against unequal voxel variances.

F is also quite robust against heteroscedasticity between the subjects since only the denominator degrees of freedom should be reduced. However the Huynh-Feldt estimator of the degrees of freedom is not recommended since it is highly sensitive to heteroscedasticity; from (2.8) and (2.9) we can see that $\hat{\nu}_{\rm HF} \approx n/{\rm CV}^2$. However the corrected degrees of freedom for unequal voxel standard deviations, $\hat{\nu}_{\rm ns}$, is quite robust against heteroscedasticity between the subjects, as the following argument shows. If we allow

$$Z_i(\mathbf{t}) = \mu_Z(\mathbf{t}) + \sigma_i \sigma(\mathbf{t}) \epsilon_i(\mathbf{t}) \qquad (2.11)$$

then it can be shown that

$$\mathrm{E}\{S(\mathbf{t}_1)^2 S(\mathbf{t}_2)^2\} = \bar{\sigma}^4 \sigma(\mathbf{t}_1)^2 \sigma(\mathbf{t}_2)^2 \{1 + 2R(\mathbf{t}_2 - \mathbf{t}_1)^2/m_{\rm het}\},$$

so that $\hat{\nu}_{\rm ns}$ should only be altered by replacing m with $m_{\rm het}$ in (2.5). A treatment of the distribution of F under the model (2.11) for unequal voxel and subject variances, along the same lines as in this section, seems very difficult.

Finally it is worth mentionning that F is quite robust against non-normality of $Z_i(\mathbf{t})$ provided the number of subjects is large, since the Central Limit Theorem will then guarantee approximate normality of $\bar{Z}(\mathbf{t})$.

3 Exceedence proportions

If the mean $\mu(\mathbf{t})$ is high in some 'region of activation' and low elsewhere then a reasonable test statistic can be based on the proportion of C where $X(\mathbf{t})$ exceeds a fixed threshold value x. Using the notation advocated by Knuth (1992), where a logical expression in parentheses takes the value one if the expression is true and zero otherwise, we can write this as

$$P = \int_C \{X(\mathbf{t}) \geq x\} d\mathbf{t}/|C|.$$

This statistic has been used by Friston et al. (199o) although the null distribution given in that paper is incorrect. We shall now derive the correct asymptotic distribution under the assumption of a stationary standard deviation $\sigma(\mathbf{t}) = \sigma$ say. It is straightforward to show that

$$\mathrm{E}(P) = \mathrm{P}\{X(\mathbf{t}) \geq x\} = \int_x^\infty (2\pi)^{-1/2} \exp(-z^2/2) dz = \Phi(-x),$$

say, and

$$\mathrm{Var}(P) = \int_C \int_C [\mathrm{P}\{X(\mathbf{t}_1) \geq x, X(\mathbf{t}_2) \geq x\} - \Phi(-x)^2] d\mathbf{t}_1 d\mathbf{t}_2 / |C|^2.$$

Let (X_1, X_2) be a bivariate normal random variable with zero mean, unit variance, and correlation ρ and let

$$f(x; \rho) = \mathrm{P}(X_1 \geq x, X_2 \geq x) - \Phi(-x)^2,$$

so that

$$\mathrm{Var}(P) = \int_C \int_C f\{x; R(\mathbf{t}_2 - \mathbf{t}_1)\} d\mathbf{t}_1 d\mathbf{t}_2 / |C|^2.$$

Then provided that $R(\mathbf{h})$ is non-negative we can write, substituting $\mathbf{t}_2 = \mathbf{t}_1 + \mathbf{h}$,

$$\mathrm{Var}(P) \leq \int_{\mathbb{R}^N} \int_C f\{x; R(\mathbf{h})\} d\mathbf{t}_1 d\mathbf{h} / |C|^2 = g(x)/|C|, \tag{3.1}$$

where

$$g(x) = \int_{\mathbb{R}^N} f\{x; R(\mathbf{h})\} d\mathbf{h}.$$

If the above integral is finite and bounded away from zero in the limit as C becomes large then (3.1) becomes an equality, and if $\epsilon(\mathbf{t})$ is weakly dependent then the null distribution of P tends to normality with mean $\Phi(-x)$ and variance (3.1).

A convenient expression for $f(x; \rho)$ can be found as follows. From the Appendix of Worsley (1983) we have

$$\mathrm{P}(X_1 \geq x, X_2 \geq x) = \int_{(\cos^{-1}\rho)/2}^{\pi/2} \pi^{-1} \exp(-x^2 \sec^2\theta/2) d\theta.$$

Transforming to $r = \cos 2\theta$ we obtain

$$\mathrm{P}(X_1 \geq x, X_2 \geq x) = \int_{-1}^{\rho} (2\pi)^{-1} (1 - r^2)^{-1/2} \exp\{-x^2/(1+r)\} dr.$$

Since $\Phi(-x)^2$ equals the above integral with $\rho = 0$ then we obtain

$$f(x; \rho) = \int_0^{\rho} (2\pi)^{-1} (1 - r^2)^{-1/2} \exp\{-x^2/(1+r)\} dr. \tag{3.2}$$

For the case of an isotropic field we can simplify $g(x)$ as follows. Suppose $R(\mathbf{h}) = R_{\mathrm{I}}(y)$, where $y = (\mathbf{h}'\mathbf{h})^{1/2}$. Then we can write

$$g(x) = \int_0^{\infty} 2\pi^{N/2} / \Gamma(N/2) y^{N-1} f\{x; R_{\mathrm{I}}(y)\} dy. \tag{3.3}$$

Using (3.2) and integrating by parts over y we get

$$g(x) = \int_0^{\infty} \frac{\pi^{N/2-1} y^N \exp\{-x^2/(1+r)\}}{N\Gamma(N/2)(1 - r^2)^{1/2}} \left(-\frac{dr}{dy} \right) dy, \tag{3.4}$$

where $r = R_{\mathrm{I}}(y)$.

It is straightforward to show that the local power of the test based on P depends on

$$\begin{aligned} \mathrm{E}(P|H_1) - \mathrm{E}(P|H_0) &= \int_C [\Phi\{\mu(\mathbf{t}) - x\} - \Phi\{-x\}] d\mathbf{t}/|C| \\ &= (2\pi)^{-1/2} \exp(-x^2/2) \int_C \mu(\mathbf{t}) d\mathbf{t}/|C| + O\left(\int_C \mu(\mathbf{t})^2 d\mathbf{t}/|C|\right) \end{aligned} \quad (3.5)$$

which depends locally on $\mu(\mathbf{t})$ only through $\bar{\mu} = \int_C \mu(\mathbf{t}) d\mathbf{t}/|C|$. Now the uniformly most powerful test of $H_1 : \bar{\mu} > 0$ is based on $\bar{X} = \int_C X(\mathbf{t}) d\mathbf{t}/|C|$, hence P is locally less powerful than $\bar{X}$. From (3.5) the local asymptotic relative efficiency of P against $\bar{X}$ is

$$\mathrm{ARE} = \int_{\mathbb{R}^N} R(\mathbf{h}) d\mathbf{h} (2\pi)^{-1} \exp(-x^2)/g(x). \quad (3.6)$$

For isotropic fields we can show this is less than one directly as follows. Since $-1 < r < 1$ then

$$\exp\{-x^2/(1+r)\}(1-r^2)^{-1/2} \geq \exp(-x^2)$$

and so substituting in (3.4) and integrating by parts,

$$g(x) \geq \int_0^\infty \frac{\pi^{N/2-1} y^N}{N\Gamma(N/2)} \left(-\frac{dr}{dy}\right) dy \exp(-x^2) = \int_{\mathbb{R}^N} R(\mathbf{h}) d\mathbf{h} (2\pi)^{-1} \exp(-x^2),$$

confirming the fact that ARE is less than one. However in applications we are interested in power comparisons when the power is close to one, and because of the higher order terms in the average of $\mu(\mathbf{t})^2$ in (3.5) it is not then clear which statistic is more powerful.

There appears to be no easy way of correcting P for non-stationary standard deviations, even though it seems desirable to do so, since it should be prone to the largest standard deviation, as for $X_{\max}$, particularly when the threshold is high. It will also be sensitive to non-normality of the observations, as for $X_{\max}$ and $T_{\max}$, since it depends on the extreme tails of the distribution of the observations.

4 Maximum of the random field

Since activation is expected to produce a few isolated peaks in $\mu(\mathbf{t})$, then Worsley et al. (1992) have the proposed the global maximum

$$X_{\max} = \max_{\mathbf{t} \in C} X(\mathbf{t})$$

as a test statistic. Worsley et al. (1993) show that for the particular alternative where $\mu(\mathbf{t})$ is proportional to the correlation function of $X(\mathbf{t})$ centred at an unknown location then $X^+_{\max}$ is the square root of the -2log(likelihood ratio) test statistic for H_0: $\mu(\mathbf{t}) = 0$, where $x^+ = x$ if $x > 0$ and zero otherwise. If the standard deviation is stationary the P-value of $X_{\max}$ is approximately

$$\mathrm{P}(X_{\max} \geq x) \approx [c/(2\pi)^{(N+1)/2}] \mathrm{He}_{N-1}(x) e^{-x^2/2}, \quad (4.1)$$

for large x, where $\mathrm{He}_{N-1}(x)$ is the Hermite polnomial of degree $N-1$ in x and

$$c = |C| \det \left\{ \mathrm{Var}\left(\frac{\partial X(\mathbf{t})}{\partial \mathbf{t}}\right) \right\}^{1/2}.$$

See Adler (1981) Theorem 6.9.1, page 160, for a more careful statement of this result and the conditions under which it holds.

If σ is unknown we can now replace it by $\hat{\sigma}$ in (2.1), and since typically the size of C is large relative to the spread of the covariance function then $\hat{\sigma}$ is approximately constant and the null distribution (4.1) will be almost unchanged. Unfortunately this test statistic is quite sensitive to non-stationary standard deviations, since the null distribution of the maximum of a Gaussian random field depends assymptotically only on the maximum standard deviation (see Adler, 1981, Theorem 6.9.2, page 161). A solution to this is to replace σ by $S(\mathbf{t})$ in (2.1). Provided that $n > N$ then this produces a t-field with m degrees of freedom at every point, and Friston et al. (1991) have suggested using the maximum

$$T_{\max} = \max_{\mathbf{t} \in C} n^{1/2} \bar{Z}(\mathbf{t}) / S(\mathbf{t})$$

as a test statistic. Worsley (1994) shows that a good approximate P-value for large t is

$$\mathrm{P}(T_{\max} \geq t) \approx \frac{c}{(2\pi)^{3/2}} \left(1 + \frac{t^2}{m}\right)^{-(m-1)/2} \frac{\Gamma\left(\frac{m+1}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\left(\frac{m}{2}\right)^{1/2}} t, \qquad (4.2)$$

if $N = 2$ and $m \geq 2$, and if $N = 3$ and $m \geq 3$ then

$$\mathrm{P}(T_{\max} \geq t) \approx \frac{c}{(2\pi)^2} \left(1 + \frac{t^2}{m}\right)^{-(m-1)/2} \left(\frac{m-1}{m} t^2 - 1\right). \qquad (4.3)$$

If $m = N$ then $T_{\max}$ is infinite with probability greater than zero.

Some idea of the sensitivity of $X_{\max}$ to non-stationary standard deviations can be obtained by supposing that $\sigma(\mathbf{t})^2 = \sigma^2(\eta - 2)/\chi^2_\eta(\mathbf{t})$, where $\chi^2_\eta(\mathbf{t})$ is a realisation of a χ^2 random field with η degrees of freedom as defined in section 2.4. Then allowing $\sigma(\mathbf{t})$ to be random implies that when $|C|$ is large, $\hat{\sigma} \approx \sigma$ and the root mean squared deviation of $\sigma(\mathbf{t})^2$ about σ^2 is $\delta \approx [2/(\eta - 4)]^{1/2}$. Moreover $X(\mathbf{t})[\eta/(\eta - 2)]^{1/2}$ is then a t-field with η degrees of freedom and so (4.2,4.3) can be used to find the exceedence probability of $X_{\max}$. For example, if $N = 3$ and $c = 1500$, a typical value for the medical applications discussed in section 7, then under stationary standard deviations $\mathrm{P}(X_{\max} \geq 4.38) \approx 0.05$. However for moderate departures from non-stationarity such as $\delta = 0.2$ then $\mathrm{P}(X_{\max} \geq 4.38) \approx 0.27$, a substantial increase; even quite mild departures such as $\delta = 0.1$ give $\mathrm{P}(X_{\max} \geq 4.38) \approx 0.08$.

Finally we expect both $X_{\max}$ and $T_{\max}$ to be quite sensitive to non-normality, even when the number of subjects is large, since they depend very strongly on the extreme tails of the distribution of the observations.

5 Stein-type shrinkage estimators

Once activation has been detected we could use the image $X(\mathbf{t})$ itself as an estimator of $\mu(\mathbf{t})$. In this section we shall show that a biased Stein-type shrinkage estimator has lower mean squared error. Let a be fixed and consider the following shrinkage estimator of $\mu(\mathbf{t})$ which shrinks $X(\mathbf{t})$ towards zero by an amount that increases as the test statistic U becomes smaller,

$$\hat{\mu}(\mathbf{t}, a) = (1 - a/U) X(\mathbf{t}),$$

and a mean squared error loss function

$$L(\hat{\mu}, \mu) = \int_C \{\hat{\mu}(\mathbf{t}, a) - \mu(\mathbf{t})\}^2 d\mathbf{t}/|C|. \tag{5.1}$$

This choice of shrinkage estimator and loss function seems the most approporiate for practical problems. To find a value for a that guarantees that $\hat{\mu}(\mathbf{t}, a)$ has smaller expected loss than $X(\mathbf{t})$ we shall need the following lemma. Similar results can be found in Berger (1976).

Lemma 2 *Let $\mathbf{X} \in \mathbb{R}^d$ have a multivariate normal distribution with mean θ and variance matrix $\mathbf{\Sigma}$. Let $\lambda_{\max}$ be the largest eigen value of $\mathbf{\Sigma}$. Then provided $a = 1 - 2\lambda_{\max}/\mathrm{tr}(\mathbf{\Sigma}) > 0$,*

$$\hat{\theta}(a) = \left(1 - \frac{a\mathrm{tr}(\mathbf{\Sigma})}{\mathbf{X}'\mathbf{X}}\right)\mathbf{X}$$

has smaller expected squared error loss than $\mathbf{X}$ for estimating θ.

Proof. Let $\mathbf{A}$ be an orthogonal matrix of eigenvectors of $\mathbf{\Sigma}$ so that $\mathbf{A}'\mathbf{\Sigma}\mathbf{A} = \mathrm{diag}(\lambda_1, \ldots, \lambda_d)$ is a diagonal matrix of eigenvalues of $\mathbf{\Sigma}$. Then $\mathbf{Y} = \mathbf{A}'\mathbf{X}$ is multivariate normal with mean $\mathbf{m} = \mathbf{A}'\theta$ and variance matrix $\mathrm{diag}(\lambda_1, \ldots, \lambda_d)$. The expected squared error loss of $\hat{\theta}(a)$ can be written in terms of $\mathbf{Y}$ as

$$\mathrm{E}\{[\hat{\theta}(a) - \theta]'[\hat{\theta}(a) - \theta]\} = \mathrm{tr}(\mathbf{\Sigma})\left[1 - 2a\mathrm{E}\left\{\frac{\mathbf{Y}'(\mathbf{Y} - \mathbf{m})}{\mathbf{Y}'\mathbf{Y}}\right\} + a^2\mathrm{E}\left\{\frac{\mathrm{tr}(\mathbf{\Sigma})}{\mathbf{Y}'\mathbf{Y}}\right\}\right]. \tag{5.2}$$

The components of $\mathbf{Y} = (Y_1, \ldots, Y_d)'$ are independent normal random variables with $\mathrm{E}(Y_i) = m_i$, say, and $\mathrm{Var}(Y_i) = \lambda_i$, $i = 1, \ldots, d$. Using integration by parts the second term of (5.2) can be written as

$$\begin{aligned}
\mathrm{E}\left(\frac{\mathbf{Y}'(\mathbf{Y} - \mathbf{m})}{\mathbf{Y}'\mathbf{Y}}\right) &= \sum_{i=1}^d \mathrm{E}\left(\frac{Y_i}{\sum_j Y_j^2}(Y_i - m_i)\right) = \sum_{i=1}^d \mathrm{E}\left(\frac{\partial}{\partial Y_i}\frac{Y_i}{\sum_j Y_j^2}\lambda_i\right) \\
&= \sum_{i=1}^d \mathrm{E}\left(\frac{\sum_j Y_j^2 - 2Y_i^2}{(\sum_j Y_j^2)^2}\lambda_i\right) \\
&\geq \mathrm{E}\left(\frac{\sum_j \lambda_j - 2\lambda_{\max}}{\sum_j Y_j^2}\right) = a\mathrm{E}\left(\frac{\mathrm{tr}(\mathbf{\Sigma})}{\mathbf{Y}'\mathbf{Y}}\right).
\end{aligned}$$

Substituting in (5.2) we have

$$\mathrm{E}\{[\hat{\theta}(a) - \theta]'[\hat{\theta}(a) - \theta]\} \leq \mathrm{tr}(\mathbf{\Sigma})[1 - a^2\mathrm{E}\{\mathrm{tr}(\mathbf{\Sigma})/\mathbf{Y}'\mathbf{Y}\}], \tag{5.3}$$

which is less than the mean squared error of $\mathbf{X}$. □

We can apply Lemma 2 to images as follows. Let $w(\mathbf{t}), \mathbf{t} \in C$ be any weight function such that

$$\int_C w(\mathbf{t})^2 d\mathbf{t} = 1 \tag{5.4}$$

and let

$$V_{\max} = \max_{w(\mathbf{t})} \mathrm{Var}\left(\int_C w(\mathbf{t})X(\mathbf{t})d\mathbf{t}\right) \tag{5.5}$$

be the maximum variance of any such weighted integral of $X(\mathbf{t})$. Suppose $\mathbf{X} \in \mathbb{R}^d$ is a vector of values of the image $X(\mathbf{t})$ sampled at a finite number d of points in C, and all integrals in (2.2), (5.1), (5.4) and (5.5) are replaced by means over these points multipled by $|C|$. Since $R(0) = 1$ then

$$\lambda_{\max}/\mathrm{tr}(\mathbf{\Sigma}) = V_{\max}/|C|.$$

If we let

$$a = 1 - 2V_{\max}/|C| \tag{5.6}$$

then provided $a > 0$ Lemma 2 shows that $(1-a/U)X(\mathbf{t})$ has smaller mean squared error loss than $X(\mathbf{t})$, sampled at these points. If C is an interval I say and $X(\mathbf{t})$ is periodic on I then the eigenvalues are proportional to the spectral density function $\phi(\omega)$, say, of $X(\mathbf{t})$, provided it exists. If $a = 1 - 2\max_\omega \phi(\omega) > 0$ then $(1 - a/U)X(\mathbf{t})$ will have a smaller expected loss. If the d sampled values of $X(\mathbf{t})$ are actually independent then $V_{\max}/|C| = 1/d$, so that $d_{\mathrm{eff}} = |C|/V_{\max}$ can be regarded as the 'effective' dimension of $X(\mathbf{t})$.

When H_0 is true the reduction in mean squared error can be considerable. If we apply the χ^2 approximation of (2.4) to $\mathbf{Y}'\mathbf{Y}$ then $\mathrm{E}\{\mathrm{tr}(\mathbf{\Sigma})/\mathbf{Y}'\mathbf{Y}\} \approx \nu/(\nu - 2) \geq 1$, and so from (5.3,5.6) the mean squared error of the shrunken estimator relative to $X(\mathbf{t})$ is approximately $4V_{\max}/|C| = 4/d_{\mathrm{eff}}$. However when H_0 is false then the shrunken estimator is not very useful. It is guaranteed to have lower errors averaged over the whole of C, particularly outside the regions of activation where the mean is close to zero. This is achieved at the expense of larger errors in the regions of activation where the mean is much larger than zero. Thus the estimator shrinks $X(\mathbf{t})$ towards zero in just the regions where it should not be shrunk. The fault lies in the fact that, from an empirical Bayes point of view, the estimator is based on a prior belief that the mean is close to zero. However we suspect that the mean is a single peak or set of peaks at unknown locations, and so a better estimator might be found by shrinking $X(\mathbf{t})$ towards this instead.

6 Gaussian correlation function

Worsley et al. (1992) made the assumption that the image under no activation can be generated by convolving a white noise gaussian random field with a kernel $\kappa(\mathbf{h})$. In PET imaging the kernel or 'point response function' is determined by placing a point source of isotope in the PET camera and measuring the response. It can be reasonably approximated by a Gaussian density of the form

$$\kappa(\mathbf{h}) \propto \exp\{-\mathbf{h}'\mathbf{K}^{-1}\mathbf{h}\}. \tag{6.1}$$

The correlation function is then

$$R(\mathbf{h}) = \kappa \star \kappa(\mathbf{h})/\kappa \star \kappa(\mathbf{0}) = \exp\{-\mathbf{h}'\mathbf{K}^{-1}\mathbf{h}/2\}. \tag{6.2}$$

The $N \times N$ positive definite shape matrix $\mathbf{K}$ is usually approximated by a diagonal matrix with diagonal elements measured in terms of the 'full width at half maximum', or the width of the kernel at half its maximum value. It is straightforward to show that if F_j is the full width at half maximum in dimension j then the jth diagonal element of $\mathbf{K}$ is $F_j^2/(4\log_e 2)$,

$j = 1, \ldots, N$. Worsley et al. (1992) introduced the unitless quantity 'RESELS' or 'resolution elements', defined as the volume $|C|$ divided by the product of the full width at half maxima, RESELS$= |C|/\prod_j F_j$. Typically in $N = 3$ dimensions, $100 <$ RESELS < 500. Then for the Gaussian correlation function (6.2) the approximate degrees of freedom of U can be expressed only in terms of the RESELS:

$$\tilde{\nu} = |C|\det(\mathbf{K})^{-1/2}\pi^{-N/2} = \text{RESELS}(4\log_e 2/\pi)^{N/2}. \tag{6.3}$$

For the maximum image statistics, Worsley et al. (1992) show that

$$c = \tilde{\nu}\pi^{N/2}.$$

We can evaluate the variance of the exceedence proportion as follows. Let $\mathbf{K}^{1/2}$ be any $N \times N$ square root of $\mathbf{K}$ chosen so that $\mathbf{K}^{1/2\prime}\mathbf{K}^{1/2} = \mathbf{K}$. Transform to $\mathbf{z} = \mathbf{K}^{-1/2\prime}\mathbf{h}$, so that $d\mathbf{h} = \det(\mathbf{K})^{1/2}d\mathbf{z}$ and

$$\int_{\mathbb{R}^N} f\{x; R(\mathbf{h})\}d\mathbf{h} = \det(\mathbf{K})^{1/2}\int_{\mathbb{R}^N} f\{x; r([\mathbf{z}'\mathbf{z}]^{1/2})\}d\mathbf{z} = \det(\mathbf{K})^{1/2}g_{\mathrm{I}}(x),$$

where $r(y) = \exp(-y^2/2)$ and

$$g_{\mathrm{I}}(x) = \int_0^\infty \frac{\pi^{N/2-1}y^{N+1}}{N\Gamma(N/2)(1-\exp(-y^2))^{1/2}}\exp[-x^2/\{1+\exp(-y^2/2)\} - y^2/2]dy.$$

The limiting value of $\mathrm{Var}(P)$ is then

$$\mathrm{Var}(P) \approx g_{\mathrm{I}}(x)/\{|C|\det(\mathbf{K})^{-1/2}\} = g_{\mathrm{I}}(x)/(\tilde{\nu}\pi^{N/2}). \tag{6.4}$$

It is useful to express this in terms of the 'effective' number of independent observations that would produce the same variance of P. From (3.1) this is

$$\mathrm{E}(P)\{1-\mathrm{E}(P)\}/\mathrm{Var}(P) \approx \tilde{\nu}\pi^{N/2}\Phi(-x)\Phi(x)/g_{\mathrm{I}}(x).$$

For $N = 3$ and $x = 1.64$, 2.33 and 2.58, typical of those used by Friston et al. (1990), the corresponding values of $\pi^{N/2}\Phi(-x)\Phi(x)/g_{\mathrm{I}}(x)$ are 0.97, 1.71 and 2.16, respectively. The local asymptotic relative efficiency from (3.6) is ARE $= (2\pi)^{N/2-1}\exp(-x^2)/g_{\mathrm{I}}(x)$, and values corresponding to x above are ARE=0.62, 0.35 and 0.26, respectively.

For a periodic image $X(\mathbf{t})$ with a periodic Gaussian correlation function the spectral density is also periodic Gaussian with peak at zero. Thus the weight $w(\mathbf{t})$ corresponding to $V_{\max}$ is stationary and so

$$V_{\max} = \int_C\int_C R(\mathbf{t}_1 - \mathbf{t}_2)d\mathbf{t}_1 d\mathbf{t}_2/|C|^2.$$

Thus for large C the effective dimension of $X(\mathbf{t})$ is

$$d_{\mathrm{eff}} \approx |C|\Big/\int_{\mathbb{R}^N} R(\mathbf{t})d\mathbf{t} = |C|\det(\mathbf{K})^{-1/2}(2\pi)^{-N/2} = \tilde{\nu}/2^{N/2}. \tag{6.5}$$

7 Specificity and power

7.1 Specificity

Talbot et al. (1991) carried out an experiment in which PET cerebral blood flow images were obtained for $n = 8$ subjects while a thermistor was applied to the forearm at both warm ($42^o C$) and hot ($48^o C$) temperaratures, each condition being studied twice on each subject. The purpose of the experiment was to find regions of the brain that were activated by the hot stimulus, compared to the warm stimulus. Individual images were aligned and sampled on a $128 \times 128 \times 80$ lattice of voxels, separated at approximately $\delta_1 = 1.4$mm, $\delta_2 = 1.7$mm and $\delta_3 = 1.5$mm on the front-back, left-right and vertical axes, respectively. The region of the brain C of interest occupied a volume of $|C| = 1091\text{cm}^3$. The image was reconstructed to a resolution of $F_1 = 20$mm, $F_2 = 20$mm and $F_3 = 7.6$mm, giving RESELS = 359.

For simulation work, we created 200 Gaussian random fields with zero mean and unit variance. This was sampled on a $128 \times 128 \times 64$ lattice of voxels, seperated at approximately $\delta_1 = \delta_2 = \delta_3 = 1.5$mm in every direction, inside a hemisphere C of radius 75mm and volume $|C| = 884\text{cm}^3$, which roughly approximated the brain region used in the study of Talbot et al. (1991). The Gaussian random fields were created by convolving Gaussian white noise with a Gaussian smoothing kernel of resolution $F_1 = 18$mm, $F_2 = 18$mm and $F_3 = 7.5$mm, chosen to approximate the resolution of the PET camera, giving RESELS = 364. Convolution was achieved via the Fast Fourier Transform.

For the U, $X_{\max}$ and P tests which do not use the sample standard deviation, the sample size was taken as $n = 1$ so that no averaging was necessary. For $X_{\max}$ and P the negative of the random field was used as an extra simulation, giving a total of 400 simulated values in these cases. For the F statistic, a value of $n = 10$ was chosen for the number of subjects, and 10 images were chosen by sampling at random without replacement from the 200 simulated images available, and this was repeated 400 times. Thus all simulated statistics except U are not truly independent. Nominal critical points from sections 2-4 at levels $\alpha = 0.1$, 0.05 and 0.01 for all the tests with non-random degrees of freedom are shown in Table 1. The effective degrees of freedom of U was $\nu = 361$ and its approximation was $\tilde{\nu} = 301$; only the latter was used in the simulations. The average $\hat{\nu}_{\text{HF}}$ was 366, very close to ν, and the average $\hat{m}_{\text{het}}$ was 9.04, very close to $m = 9$. The proportion of times that our simulated statistics exceeded the critical values are shown in Table 2. The degrees of freedom used for U was $\tilde{\nu}$, and the degrees of freedom used for F was $\tilde{\nu}$ in the numerator and $m\tilde{\nu}$ in the denominator. We have omitted tests with random degrees of freedom since they gave very similar results. It can be seen that the specificity of all tests is in good agreement with the nominal levels, except for $T_{\max}$, due perhaps to the very sharp peaks of the t-field falling between the sampled points.

Table 1. *Nominal critical points of tests*

Test	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
U	1.106	1.138	1.199
F	1.113	1.147	1.213
P, $x = 1.64$	0.0669	0.0715	0.0802
P, $x = 2.33$	0.0155	0.0171	0.0200
P, $x = 2.58$	0.00842	0.00942	0.0113
$X_{\max}$	4.21	4.39	4.78
$T_{\max}$	11.08	12.57	16.71

Table 2. *Specificity of tests, equal standard deviations*

Test	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
U	0.0975	0.0425	0.0050
F	0.0875	0.0400	0.0075
P, $x = 1.64$	0.1075	0.0425	0.0050
P, $x = 2.33$	0.0975	0.0525	0.0125
P, $x = 2.58$	0.0975	0.0575	0.0175
$X_{\max}$	0.1000	0.0450	0.0025
$T_{\max}$	0.0450	0.0200	0.0050

To study the effect of non-stationary $\sigma(\mathbf{t})$ on the specificity of the test statistics, we created a standard deviation function $\sigma(\mathbf{t})$ by sampling once from a Gaussian random field with mean 1, standard deviation 0.1, and the same correlation structure as the above 200 simulated images. This corresponds roughly to $\delta = 0.2$ in the discussion of robustness in section 2.4 and section 4. All images were then multiplied by the same $\sigma(\mathbf{t})$ and the above tests were repeated. The F test using $\hat{\nu}_{\text{ns}}$ was not simulated, since $\hat{\nu}_{\text{ns}}$ is computationally too intensive. We found $\nu_{\text{ns}} = 368$, very close to ν. The average $\hat{\nu}_{\text{HF}}$ was 359 and the average $\hat{m}_{\text{het}}$ was 9.03, both almost unchanged by the non-stationary standard deviations. Obviously the exceedence proportions of $T_{\max}$ are completely unaffected by non-stationary $\sigma(\mathbf{t})$, but for the other tests the results are shown in Table 3, again omitting tests with random degrees of freedom since they gave very similar results. We can see that U and F are quite robust against non-stationary $\sigma(\mathbf{t})$, but the tests based on $X_{\max}$ and P for high threshold are highly sensitive to non-stationary $\sigma(\mathbf{t})$.

Table 3. *Specificity of tests, non-stationary* $\sigma(\mathbf{t})$

Test	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
U	0.0925	0.0500	0.0075
F	0.0925	0.0425	0.0075
P, $x = 1.64$	0.1150	0.0475	0.0125
P, $x = 2.33$	0.1375	0.0775	0.0150
P, $x = 2.58$	0.1550	0.1025	0.0300
$X_{\max}$	0.5150	0.3625	0.1300

To study the effect of heteroscedasticity we let $\sigma_1, \ldots, \sigma_{10}$ take values from a sample of 10 independent univariate Gaussian random variables with mean 1 and standard deviation

0.2, which gave CV= 0.32. As a result the average $\hat{\nu}_{\mathrm{HF}}$ was greatly reduced to 92, the critical values of F were increased, and the false positive rates fell to 0.0075, 0.0050 and 0.0000 for nominal level 0.10, 0.05 and 0.01 tests respectively. The average corrected degrees of freedom was $\hat{m}_{\mathrm{het}} = 8.37$, still close to $m = 9$, indicating little effect of heteroscedasticity; using this gave false positive rates 0.0950, 0.0450 and 0.0075 respectively, very close to the nominal levels.

7.2 Power

An 'optimal' (in the sense of Lemma 1) signal $\mu_1(\mathbf{t})$ was created which was proportional to the Gaussian correlation function $R(\mathbf{t} - \tau)$ (6.2):

$$\mu_1(\mathbf{t}) = \xi \exp\{-(\mathbf{t} - \tau)'\mathbf{K}^{-1}(\mathbf{t} - \tau)/2\}.$$

The shape matrix $\mathbf{K}$ was chosen to be a diagonal matrix with jth diagonal element equal to $F_j^2/(4 \log_e 2)$ and the location τ was chosen to lie in the anterior cingulate close to where activation was in fact detected in Talbot et al. (1991). The peak height was chosen to be ξ=4. Another signal $\mu_3(\mathbf{t})$ was created with three peaks identical in shape and height to $\mu_1(\mathbf{t})$ but centred in the anterior cingulate, the primary and secondary somatosensory regions of the brain, close to where activation was detected in Talbot et al. (1991). Finally a 'broad' signal $\mu_B(\mathbf{t})$ was created by convolving a $3 \times 4.5 \times 3$cm region of uniform height with the Gaussian point spread function (6.1) to create a broader region of activation as opposed to a sharp peak. The maximum height of this region was chosen to be the same as the peak heights above. The region was located in the right hemisphere in roughly the same place where activation was detected by Talbot et al. (1991).

The three signals were added to each simulated image and the tests were repeated; the standard deviations of the images were not perturbed by multiplying by $\sigma(\mathbf{t})$ created as in the previous section, so the tests with random degrees of freedom which are designed to account for this are not shown. The power of the tests was then estimated by the proportions of simulated values exceeding the nominal critical values in Table 1 and the results are shown in Table 4. The degrees of freedom used for U was $\tilde{\nu}$, and the degrees of freedom used for F was $\tilde{\nu}$ in the numerator and $m\tilde{\nu}$ in the denominator. We note that the $X_{\max}$ test is by far the most sensitive at finding a single peak, closely followed by the exceedence proportion P at high threshold; all tests are better at finding three peaks or a broad peak. For F there does not seem to be much sacrifice for estimating σ from the data, but for $T_{\max}$ the effect is severe and the test has very poor power.

Table 4(a). *Power of tests against a single peak, $\mu_1(\mathbf{t})$*

Test	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
U	0.250	0.160	0.045
F	0.230	0.130	0.040
P, $x = 1.64$	0.310	0.200	0.040
P, $x = 2.33$	0.430	0.315	0.150
P, $x = 2.58$	0.495	0.380	0.225
$X_{\max}$	0.615	0.560	0.360
$T_{\max}$	0.075	0.040	0.005

Table 4(b). *Power of tests against three peaks,* $\mu_3(\mathbf{t})$

Test	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
U	0.725	0.555	0.300
F	0.635	0.535	0.250
P, $x = 1.64$	0.730	0.605	0.400
P, $x = 2.33$	0.910	0.850	0.715
P, $x = 2.58$	0.935	0.905	0.815
$X_{\max}$	0.920	0.895	0.720
$T_{\max}$	0.175	0.090	0.010

Table 4(c). *Power of tests against a broad peak,* $\mu_B(\mathbf{t})$

Test	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
U	1.000	0.995	0.955
F	0.990	0.985	0.945
P, $x = 1.64$	0.995	0.985	0.940
P, $x = 2.33$	1.000	1.000	1.000
P, $x = 2.58$	1.000	1.000	1.000
$X_{\max}$	0.995	0.975	0.885
$T_{\max}$	0.355	0.230	0.045

There is a straightforward way of assessing the power of all tests against a signal which is a realisation of a Gaussian random field with the same correlation structure as the noise. The power, averaged over all such random realisations, is equivalent to just the P-value for the same statistic but with σ replaced by $\sigma(1+b^2)^{1/2}$, where b is the root mean square amplitude of the signal relative to that of the noise. Powers calculated in this way are shown in Table 5 for $b = 0.5$. Now the quadratic tests are most powerful, followed by the exceedence proportion for highest threshold; $X_{\max}$ and particularly $T_{\max}$ have low power. Note that F is almost as powerful as U, showing that there is only a small price to pay for not knowing σ. For comparison with Tables 4(a-c), the root mean squared amplitudes of their signals are b =0.231, 0.385 and 0.644 respectively.

Table 5. *Power of tests against a realisation of a Gaussian random field*

Test	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
U	0.926	0.867	0.684
F	0.906	0.834	0.629
P, $x = 1.64$	0.607	0.492	0.286
P, $x = 2.33$	0.680	0.591	0.414
P, $x = 2.58$	0.676	0.593	0.429
$X_{\max}$	0.467	0.275	0.079
$T_{\max}$	0.182	0.092	0.019

Finally we averaged the loss of the shrinkage estimator $\hat{\mu}(\mathbf{t}, a)$ for no signal and the three signals above. The effective number of independent observations was $d_{\text{eff}} = 106$ from

(6.5). The mean squared errors, relative to no shrinkage ($a = 0$) were 0.006, 0.056, 0.134 and 0.299 respectively. Obviously the shrunken estimator performs extremely well when no signal is present, and better on the smaller signals which are close to zero. However the shinkage factors are considerable: 0.016, 0.063, 0.140 and 0.302 on average, respectively, thus over-shrinking the esimator where the signal is high.

7.3 Application

To check the specificity of our tests, the difference images of the two warm conditions were used as a dataset which should have an expectation of zero throughout. To detect activation, we analysed the difference between the average of the two hot conditions and the average of the two warm conditions to search for activation due to the painful heat stimulus. One subject was scanned only once in the hot condition so this subject was dropped leaving 7 subjects for this dataset. All tests except U were applied, and the results are shown in Table 6. For the $X_{\max}$ and P tests, the known σ was replaced by the pooled estimate $\hat{\sigma}$.

Table 6. *Application*

d.f.	Null		Activation	
m	7		6	
ν	348.1		348.1	
$\tilde{\nu}$	297.5		297.5	
$\hat{\nu}_{\rm ns}$	327.2		336.8	
$\hat{\nu}_S$	4.96		5.10	
$\hat{\nu}_{\rm HF}$	18.4		37.4	
$\hat{m}_{\rm het}$	5.05		5.20	
Test	Statistic	P-value	Statistic	P-value
F, $(\nu, m\nu)$ d.f.	1.0863	0.146	1.871	8.3×10^{-17}
F, $(\tilde{\nu}, m\tilde{\nu})$ d.f.	1.0863	0.164	1.871	1.2×10^{-14}
F, $(\hat{\nu}_{\rm ns}, m\hat{\nu}_{\rm ns})$ d.f.	1.0863	0.154	1.871	2.5×10^{-16}
F, $(\hat{\nu}_{\rm HF}, m\hat{\nu}_{\rm HF})$ d.f.	1.0863	0.373	1.871	0.0030
F, $(\tilde{\nu}, \hat{m}_{\rm het}\tilde{\nu})$ d.f.	1.0863	0.210	1.871	3.0×10^{-14}
P, $x = 1.64$	0.0655	0.114	0.1087	2.3×10^{-6}
P, $x = 2.33$	0.0200	0.011	0.0373	2.8×10^{-10}
P, $x = 2.58$	0.0117	0.008	0.0248	5.7×10^{-13}
$X_{\max}$	4.20	0.103	5.06	0.0028
$T_{\max}$	7.4	2.65	15.6	0.72

The conclusions are as follows. The F statistic does not detect any activation in the null dataset, but detects it in the activated dataset. The numerator degrees of freedom ν, $\tilde{\nu}$ and $\hat{\nu}_{\rm ns}$ are about the same whether corrected or uncorrected for non-stationary standard deviations, but the uncorrected sample estimator $\hat{\nu}_S$ and the Huynh-Feldt estimator $\hat{\nu}_{\rm HF}$ are much lower probably due to heteroscedasticity among the subjects. The denominator degrees of freedom corrected for heteroscedasticity $\hat{m}_{\rm het}$ are lower, as expected, but the effect on P-values is small. The exceedence proportions P detect activation in both null and activation datasets, particularly for higher thresholds, perhaps due to non-stationary standard deviations or non-normal obervations. The $X_{\max}$ test almost detects activation in the null dataset, perhaps due to non-stationary standard deviations or non-normality, and clearly detects it in the activation dataset, but $T_{\max}$ detects nothing, perhaps because of its poor power; note that the nominal P-value (4.3) exceeded one for the null dataset since t was small.

8 Conclusions

Of the tests presented here, the maximum of the random field appears to be the most powerful at detecting a single peak, followed by the exceedence proportions for high threshold. When several peaks are present, or the signal is a broad peak, the quadratic tests appear to be almost as powerful. All tests are quite robust against unequal subject variances, provided the Huynh-Feldt estimator of the degrees of freedom is avoided, but of course there is a small loss in power. Unfortunately the most powerful tests are also the most sensitive to non-stationary standard deviations and non-nomality, and so for straight testing purposes we reccomend the F test which appears to be quite robust. It is of some theoretical interest to note that we can find a Stein-type shrinkage estimator of the signal with guaranteed lower mean squared error, though this seems to be of little use in practice.

REFERENCES

Berger, J.O. (1976). Admissable minimax estimation of a multivariate normal mean with arbitrary quadratic loss. *Annals of Statistics*, **4**, 223-226.

Box, G.E.P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems: II. Effects of inequality of variance and of correlation between errors in the two-way classification. *Annals of Mathematical Statistics*, **25**, 484-498.

Friston, K.J., Frith, C.D., Liddle, P.F. and Frackowiak, R.S.J. (1991). Comparing functional (PET) images: the assessment of significant change, *Journal of Cerebral Blood Flow and Metabolism*, **11**, 690-699.

Friston, K.J., Frith, C.D., Liddle, P.F., Lammertsma A.A., Dolan, R.D. and Frackowiak, R.S.J. (1990), The relationship between local and global changes in PET scans *Journal of Cerebral Blood Flow and Metabolism*, **10**, 458-466.

Huynh, H and Feldt, L.S. (1976) Estimation of the Box correction for degrees of freedom from sample data in the randomized block and split-plot designs. *Journal of Educational Statistics*, **1**, 69-82.

Knuth, D.E. (1992). Two notes on notation. *The American Mathematical Monthly*, **99**, 403-422.

Raz, J. and Fein, G. (1992). Testing for evoked potential signals using an approximation to an exact permutation test. *Biometrics*, **48**, 1069-1080.

Satterthwaite, F.E. (1946). An approximate distribution of estimates of variance components. *Biometrics*, **2**, 110-114.

Seber, G.A.F. (1977). *Linear Regression Analysis*. Wiley, New York.

Smoot, G.F., Bennett, C.L., Kogut, A., Wright, E.L., Aymon, J., Boggess, N.W., Cheng, E.S., De Amici, G., Gulkis, S., Hauser, M.G., Hinshaw, G., Jackson, P.D., Janssen, M., Kaita, E., Kelsall, T., Keegstra, P., Lineweaver, C., Lowenstein, K., Lubin, P., Mather, J., Meyer, S.S., Moseley, S.H., Murdock, T., Rokke, L., Silverberg, R.F., Tenorio, L., Weiss, R. and Wilkinson, D.T. (1992). Structure in the *COBE* differential microwave radiometer first-year maps. *Astrophysical Journal*, **396**, L1-L5.

Talbot, J.D., Marrett, S., Evans, A.C., Meyer, E., Bushnell, M.C. and Duncan, G.H. (1991). Multiple representations of pain in human cerebral cortex. *Science*, **251**, 1355-1358.

Vardi, Y., Shepp, L.A. and Kaufmann, L. (1985). A statistical model for positron emission tomography. *Journal of the American Statistical Association*, **80**, 8-20.

Worsley, K.J. (1983). Tests for a two-phase multiple regression. *Technometrics*, **25**, 35-42.

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. (1986). Confidence regions and tests for a change-point in a sequence of exponential family random variables. *Biometrika*, **73**, 91-104.

Worsley, K.J., Evans, A.C., Strother, S.C. and Tyler, J.L (1991). A linear spatial correlation model, with applications to positron emission tomography. *Journal of the American Statistical Association*, **86**, 55-67.

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., Evans, A.C., Marrett, S. and Neelin, P. (1993). Detecting changes in random fields and applications to medical images. *Journal of the American Statistical Association*, submitted for publication.