Random Field Theory

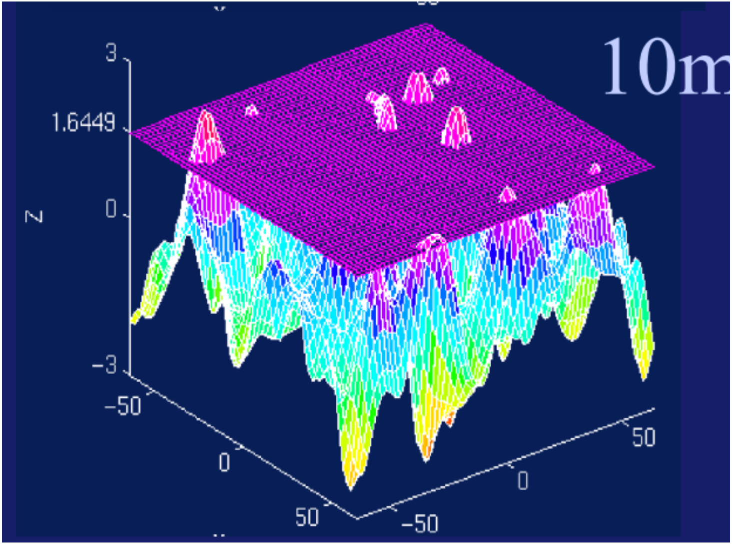
Samuel J. Davenport

University of Oxford

November 15, 2019

An Introduction to RFT

Illustration in 2D



Robert Adler - RFT

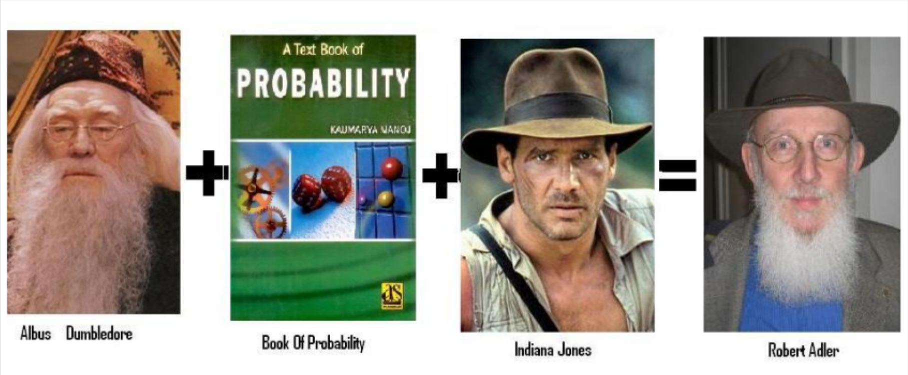


Figure 1: Robert Adler developed RFT in books such as the Geometry of Random Fields.

Neuroimagers - RFT

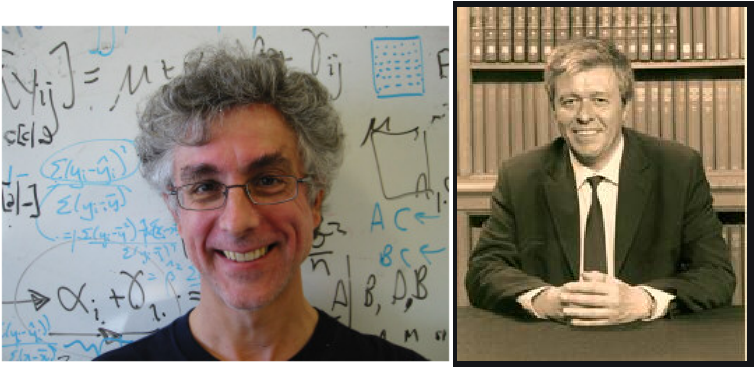


Figure 2: Keith Worsley (left) and Karl Friston (right) introduced RFT methods for determining the threshold.

Neuroimagers - Permutation

Figure 3: Thomas Nichols (left), Andrew Holmes (and others) developed non-parametric methods for estimating the threshold. Recently Anders Eklund (right) and Tom showed that RFT (as currently implemented) does not control the false positive rate while permutation does.

Random Fields

Definition

Given $N \in \mathbb{N}$ and $T \subset \mathbb{R}^N$, define an N-dimensional random field X to be a random function

$$X : T \longrightarrow \mathbb{R}$$

we say that X is a Gaussian random field if for all $k \in \mathbb{N}$, given $(t_1, \ldots, t_k) \in T$, $(X(t_1), \ldots, X(t_k))$ has a non-degenerate Gaussian distribution.

Controlling the FWER

- Let $V \subset T$ be the set of voxels
- Take $X(v)$ to be our test statistic at each $v \in V \subset T$ so we reject the null hypothesis that there is no activity at v if $\mathbb{P}(X(v) > u) < \alpha$.

There are typically large number (around 200000) of voxels and so multiple testing must be accounted for! In particular taking $\alpha = 0.05$ will mean around 10000 false discoveries!

Definition

Suppose that $V_0 \subset V$ is the set of voxels that are null. Then we define the FWER (family wise error rate) to be the probability of at least one false discovery. I.e.

$$\mathbb{P}\left(\max_{v \in V} X(v) > u\right)$$

and we seek to control this at a level α.

Note that for a fine enough lattice V,

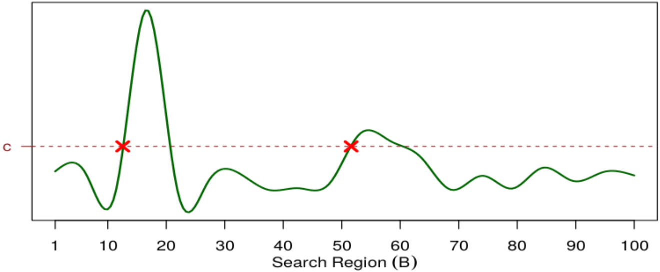
$$\mathbb{P}(\max_{v \in V} X(v) > u) \approx \mathbb{P}(\max_{t \in T} X(t) > u.$$

Voxelwise RFT

Let M be the number of local maxima of X above u then assuming that X is twice differentiable,

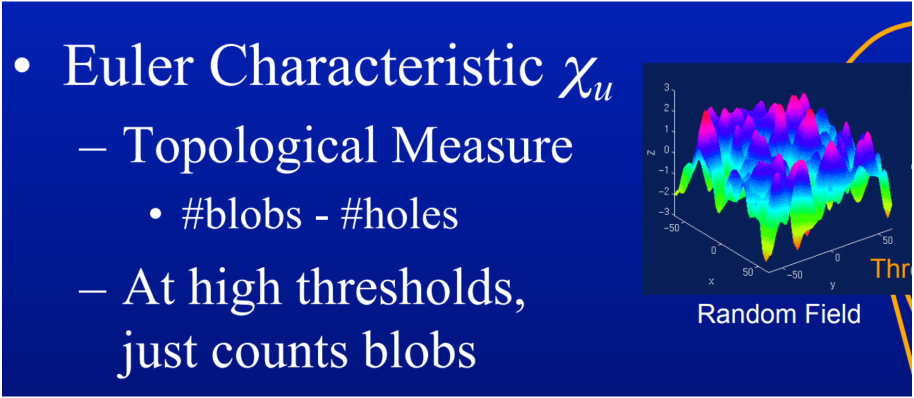
$$\mathbb{P}\left(\sup_{t \in T} X(t) > u\right) = \mathbb{P}(M \geq 1) \leq \mathbb{E}[M].$$

because X exceeds u if and only if there is at least one local maxima above u. This is best seen by looking at a picture

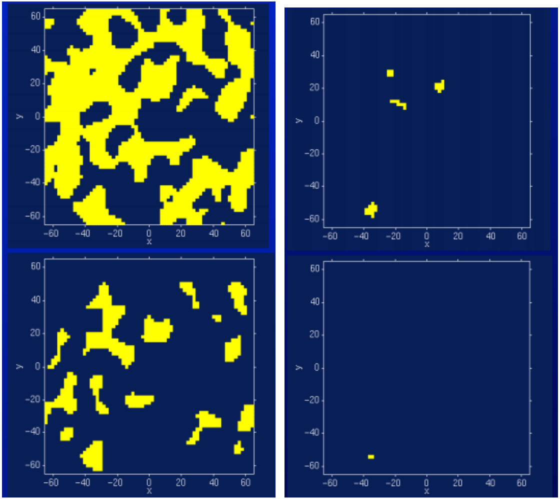


The Euler Characteristic

$\mathbb{E}[M]$ is difficult to estimate and requires us to be clever.
To do so we introduce a topological quantity called the Euler Characteristic $\chi_u$which looks at the excursion set and calculates the number of blobs minus the number of holes.



At High Thresholds, Euler Char is the number of Maxima



Using the Euler Characteristic

Definition

Given $u \in \mathbb{R}$, let $\mathcal{A}_u$ be the excursion set when the threshold is u ie

$$\mathcal{A}_u = \{t \in T : X(t) \geq u\}.$$

The Euler Characteristic $\chi(\mathcal{A}_u)$ is the number of blobs of $\mathcal{A}_u$ minus the number of its holes. Then for high thresholds, $M_u = \chi(\mathcal{A}_u)$. So

$$\mathbb{P}\left(\sup_{t \in T} X(t) > u\right) = \mathbb{P}(M \geq 1) \leq \mathbb{E}[M] \approx \mathbb{E}[\chi(\mathcal{A}_u)].$$

Note that for large enough u

$$\mathbb{P}(M \geq 1) \approx \mathbb{E}[M].$$

Calculating the expected Euler Characteristic

For stationary Gaussian fields the expected euler characteristic has a nice closed form (Adler, 1981):
So if $N = 1$,

$$\mathbb{E}[\chi(A)] = \mu(T)|\Lambda|^{1/2} \exp\left(-\frac{u^2}{2\sigma^2}\right)/2\pi\sigma$$

and if $N = 2$, then

$$\mathbb{E}[\chi(A)] = \mu(T)(2\pi)^{-3/2}\sigma^{-3}|\Lambda|^{1/2}u \exp\left(-\frac{u^2}{2\sigma^2}\right).$$

If $N = 3$, then

$$\mathbb{E}[\chi(A)] = \mu(T)(2\pi)^{-2}\sigma^{-5}|\Lambda|^{1/2}(u^2 - \sigma^2) \exp\left(-\frac{u^2}{2\sigma^2}\right).$$

where $\sigma = \mathrm{var}(X)^{1/2}$ and $\Lambda = \mathrm{cov}(\nabla X)$.

Notation

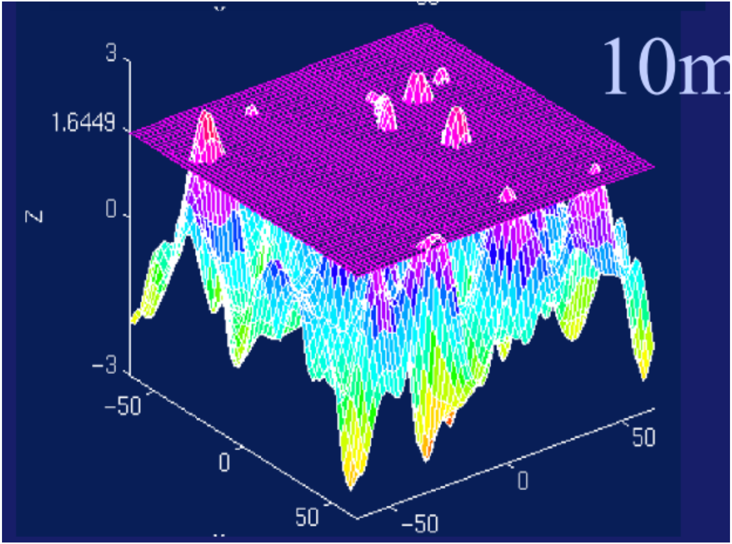
We typically take S to be compact (i.e the brain!) and given a threshold u

- let m be the number of clusters above u where the clusters are the connected components of the excursion set $\mathcal{A}_u$ and
- let $n_1, \dots, n_m$ be the number of voxels within each cluster.

note that $m, n_1, \dots, n_m$ are all random variables size the field X is random.

Note that the number of clusters is not in general equal to the number of local maxima but they are equal for large enough u. (I.e. $M \approx m$ for large enough u as you have one peak per cluster.)

Illustration in 2D



Assumptions

The use of RFT in brain imaging makes a number of assumptions. These are listed below

- The random field X is stationary.
- m the number of clusters above u has a Poisson distribution.
- The threshold u is reasonably high.
- The cluster sizes $n_1, \dots, n_m$ are independent with common distribution denoted by n and are independent of m. (This is certainly not true but may be reasonable for high thresholds.)
- The lattice approximation to the random field is good enough. (Good Lattice Assumption.)

we will discuss how RFT proceeds under these assumptions and then look at how some of the assumptions can now be dropped.

Distribution of the size of a cluster

We will derive the distribution n of the size of the largest cluster above a threshold. Let $n_{\max} = \max\{n_1, \dots, n_m\}$, then (K. J. Friston, Worsley, Frackowiak, Mazziotta, & Evans, 1994),

Theorem

Suppose that the assumptions above hold, then

$$\mathbb{P}(n_{\max} \geq k) = 1 - e^{-\mathbb{E}m\mathbb{P}(n \geq k)}$$

note that this is important as it allows us to perform FWER control on the cluster sizes. I.e if we take our test statistics to be $n_1, \dots, n_m$ then we can choose k such that

$$\mathbb{P}(n_{\max} \geq k) \leq \alpha$$

and reject cluster j if $n_j > k$.
Limitations...

Lemma

In order to prove this we require the following Lemma from (K. Friston, Holmes, Poline, Price, & Frith, 1996)

Lemma

Suppose that m is Poisson, the cluster sizes are i.i.d and independent of m. Let m_k be the number of clusters that are of size at least k, then

$$m_k \sim Po(\mathbb{P}(n \geq k)\mathbb{E}m)$$

where n is the distribution of the size of each cluster. Equivalently, for each $c \in \mathbb{N}$,

$$\mathbb{P}(m_k \geq c) = 1 - \sum_{i=0}^{c-1} \lambda(i, \mathbb{P}(n \geq k)\mathbb{E}m)$$

where

$$\lambda(i,t) = t^i e^{-t}/i!$$

Lemma proof

Proof.

Given $c \in \mathbb{N}$,

$$\begin{aligned}\mathbb{P}(m_k \geq c) = 1 - \mathbb{P}(m_k < c) &= 1 - \sum_{i=0}^{c-1} \mathbb{P}(m_k = i) \\ &= 1 - \sum_{i=0}^{c-1} \sum_{j=i}^{\infty} \mathbb{P}(m_k = i | m = j)\mathbb{P}(m = j) \\ &= 1 - \sum_{i=0}^{c-1} \sum_{j=i}^{\infty} \mathbb{P}(m = j) \binom{j}{i} \mathbb{P}(n \geq k)^i \mathbb{P}(n < k)^{j-i}\end{aligned}$$

where we have used the fact that $m_k \geq i$ implies that $m \geq i$.

Now, taking $p = \mathbb{P}(n \geq k)$ and $q = \mathbb{E}m$,

$$\begin{aligned}
\sum_{j=i}^{\infty} \mathbb{P}(m = j) \binom{j}{i} \mathbb{P}(n \geq k)^i \mathbb{P}(n < k)^{j-i} & \\
&= \sum_{v=0}^{\infty} \mathbb{P}(m = v + i) \frac{(v+i)!}{v!i!} p^i (1-p)^v \\
&= \sum_{v=0}^{\infty} \frac{q^{v+i} e^{-q}}{(v+i)!} \frac{(v+i)!}{v!i!} p^i (1-p)^v \\
&= \frac{(pq)^i e^{-q}}{i!} \sum_{v=0}^{\infty} (q(1-p))^v / v! \\
&= \frac{(pq)^i e^{-q}}{i!} e^{q(1-p)} = (pq)^i e^{-pq}/i! = \lambda(i, pq).
\end{aligned}$$

$\square$

Distribution of the maximum cluster size

Theorem

Suppose that the assumptions above hold, then

$$\mathbb{P}(n_{\max} \geq k) = 1 - e^{-\mathbb{E}m\mathbb{P}(n\geq k)}$$

Proof.

$$\begin{aligned}\mathbb{P}(n_{\max} \geq k) &= \sum_{i=0}^{\infty} \mathbb{P}(m = i, n_{\max} \geq k) = \sum_{i=0}^{\infty} \mathbb{P}(m = i)\mathbb{P}(n_{\max} \geq k | m = i)\\ &= \sum_{i=0}^{\infty} \mathbb{P}(m = i)(1 - \mathbb{P}(n < k | m = i))\\ &= \sum_{i=0}^{\infty} \mathbb{P}(m = i)(1 - \mathbb{P}(n < k)^i)\end{aligned}$$

Distribution of the maximum cluster size (continued)

Proof.

$$
\begin{aligned}
&= 1 - \sum_{i=0}^{\infty} \mathbb{P}(m = i)\mathbb{P}(n < k)^i \\
&= 1 - \sum_{i=0}^{\infty} \frac{1}{i!}\mathbb{E}(m)^i e^{-\mathbb{E}m}\mathbb{P}(n < k)^i = 1 - e^{-\mathbb{E}m}\sum_{i=0}^{\infty} \frac{1}{i!}\mathbb{E}(m)^i\mathbb{P}(n < k)^i \\
&= 1 - e^{-\mathbb{E}m}e^{\mathbb{E}m\mathbb{P}(n<k)} = 1 - e^{-\mathbb{E}m}e^{\mathbb{E}m(1-\mathbb{P}(n\geq k))} \\
&= 1 - e^{-\mathbb{E}m\mathbb{P}(n\geq k)}
\end{aligned}
$$

□

Cluster Failure

T-field

Definition

For $n > 1$, let $Y_1, \dots, Y_n$ be Gaussian random fields on $\mathbb{R}^D$. Then define the one-sample t-field to be

$$T(Y_1, \dots, Y_n) = \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i}{\left(\frac{1}{m-1} \sum_{i=1}^n (Y_i - \frac{1}{n} \sum Y_i)^2 \right)^{1/2}}$$

For a t-field it's possible to generate closed forms for $\mathbb{E}[\chi(\mathcal{A}_u)]$ and the distribution of n. So it is possible to use RFT to develop a threshold u which controls the FWER.
RFT relies on a number of assumptions, as such non-parameteric alternatives such as permutation are commonly used.

Algorithm 1 One Sample Permutation - Voxelwise

Input: Images $Y_1, \dots, Y_N$ on a set of voxels V, the number of permutations P and desired alpha level α
Let $T = T(Y_1, \dots, Y_n)$
for $p = 1, \dots, P$ **do**
    Generate a vector B of length N such that the entries are independent Bern(0.5) random variables.
    For $i = 1, \dots, n$, let $Y_i^* = (-1)^{B(i)} Y_i$
    Let $T_p = T(Y_1^*, \dots, Y_N^*)$, let $m_p = \max_{v \in V} T_p(v)$
end for
Let u be the upper 0.05 quantile of the empirical distribution given by $\{m_1, \dots, m_P\}$.
Reject v such that $T(v) > u$.

Clusterwise - Permutation Testing

Algorithm 2 One Sample Permutation - Clusterwise

1: **Input**: Images $Y_1, \ldots, Y_N$ on a set of voxels V, the number of permutations P and desired alpha level α and CDT u

2: Let $T = T(Y_1, \ldots, Y_n)$ and let $n_1, \ldots, n_m$ be the sizes of the clusters above the threshold u.

3: **for** $p = 1, \ldots, P$ **do**

4: Generate a vector B of length N such that the entries are independent Bern(0.5) random variables.

5: For $i = 1, \ldots, n$, let $Y_i^* = (-1)^{B(i)} Y_i$

6: Let $T_p = T(Y_1^*, \ldots, Y_N^*)$, let c_p be the size of the largest cluster of T_p above the threshold u.

7: **end for**

8: Let k be the upper 0.05 quantile of the empirical distribution given by $\{c_1, \ldots, c_P\}$.

9: Reject the clusters i such that $n_i > k$.

Testing using resting state data

(Eklund, Nichols, & Knutsson, 2016) validated Permutation and RFT inference using resting state data. They used subjects from 3 different resting state datasets: Cambridge Beijing and Oulu (with 198, 198 and 103 subjects respectively). For each dataset they took 1000 subsets of size 20, computed the one-sample t-statistic and compared this value to 0.05 level thresholds generated by RFT and permutation to control the false positive rates.

I.e for each dataset and for $I = 1, \dots, 1000$ they computed the percentage of datasets which recorded a false positive. Under the null we'd expect this to be 5%.

- p = 0.01 corresponds to a CDT of u =1 - $\Phi^{-1}(0.01) \approx 2.32$
- p = 0.001 corresponds to a CDT of u =1 - $\Phi^{-1}(0.001) \approx 3.09$

Cluster Failure - Clusterwise

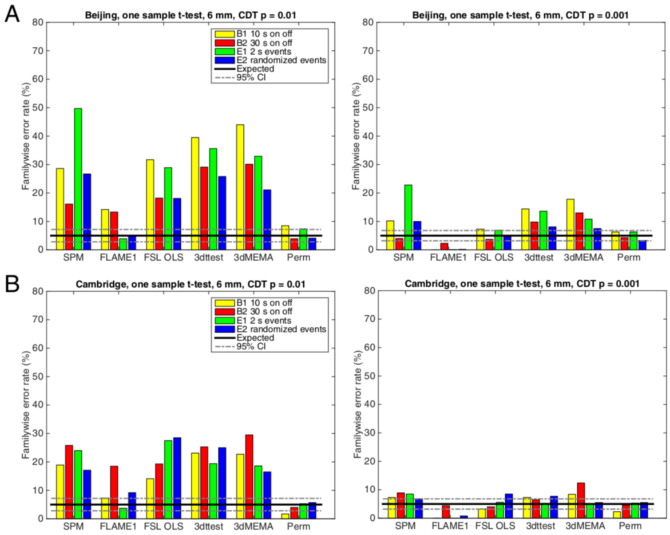


Figure 4: Results from (Eklund et al., 2016), preprocessing by software

Cluster Failure - Voxelwise

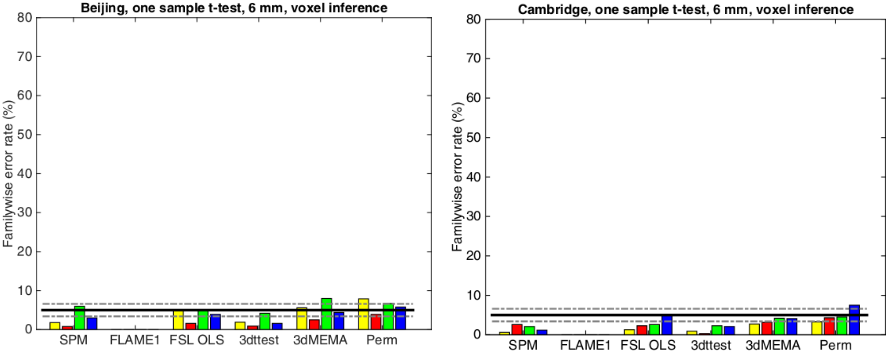


Figure 5: Results from (Eklund et al., 2016), preprocessing by software

Assumptions Breakdown - Smoothness

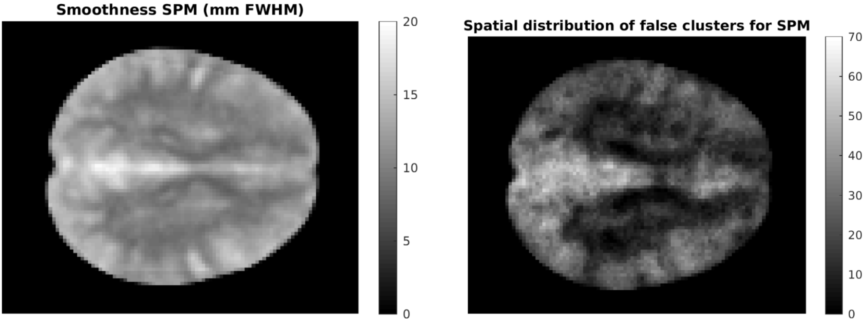


Figure 6: Smoothness is assumed to be the same everywhere (by stationarity) but in fact it's not. This leads to false clusters being detected in smooth regions since the smoother the region the greater the expected clustersize. Clearly a big problem.

Assumptions Breakdown - Cluster Extent Thresholds

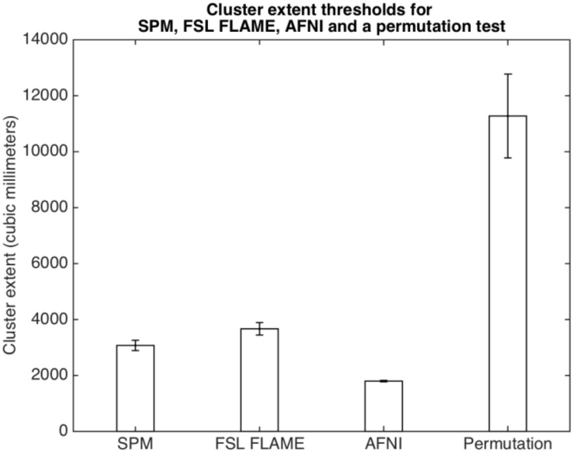


Figure 7: Cluster extent thresholds for $p = 0.01$

Assumptions Breakdown - Acf

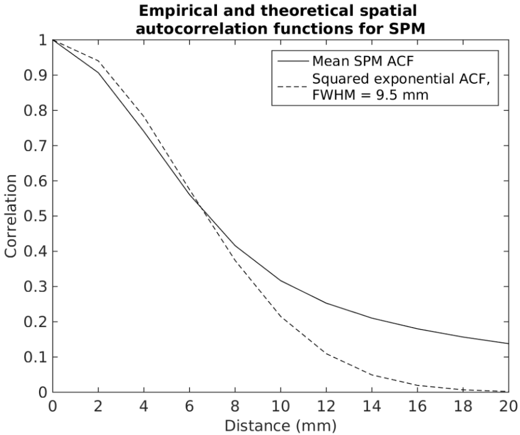


Figure 8: Acf is not Gaussian. They make a big deal out of this, but in fact this is only required top show the number of clusters is Poisson and even then is only needed near 0.

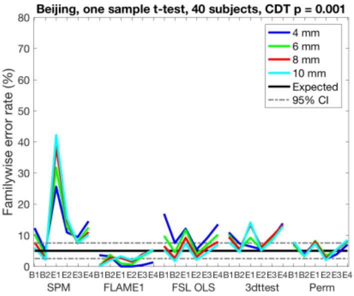
General Notes

- Essentially all these results show is that applying RFT when the assumptions don't hold won't work. Which shouldn't be a big surprise.
-

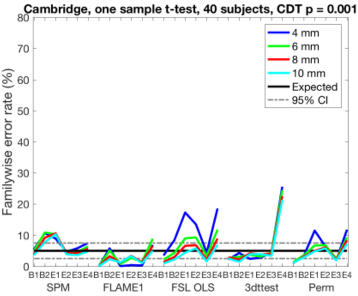
Nevertheless this is an extremely important paper.

Issues

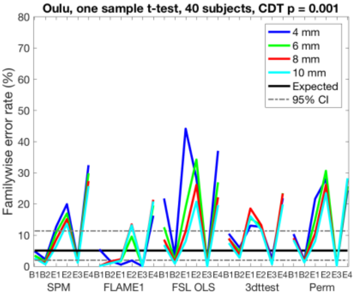
- Minor: The CDTs used are different for SPM and FSL. Let $\Phi_T(x) = \mathbb{P}(t_n \leq x)$ (where t_n is the t-distribution with n degrees of freedom) and $\Phi_T(x) = \mathbb{P}(N(0,1) \leq x)$. Then SPM uses a CDT of 1 - $\Phi^{-1}(0.01)$ and FSL uses the more correct: 1 - $\Phi_T^{-1}(0.01)$. (So the results are not comparable.)
- Major: While expressions for $\mathbb{E}[\chi(\mathcal{A}_u)]$ have been implemented for the t-statistic in SPM and FSL, expressions for $\mathbb{P}(n \geq k)$ have not been. See e.g. (?, ?).
- Major: Permutation is only calculated using preprocessed data from FSL.
- The draws are dependent which could cause an issue.


Beijing, one sample t-test, 40 subjects, CDT p = 0.001
Familywise error rate (%)
4 mm
6 mm
8 mm
10 mm
Expected
95% CI
B1B2E1E2E3E4B1B2E1E2E3E4B1B2E1E2E3E4B1B2E1E2E3E4B1B2E1E2E3E4
SPM
FLAME1
FSL OLS
3dttest
Perm

(a)


Cambridge, one sample t-test, 40 subjects, CDT p = 0.001
Familywise error rate (%)
4 mm
6 mm
8 mm
10 mm
Expected
95% CI
B1B2E1E2E3E4B1B2E1E2E3E4B1B2E1E2E3E4B1B2E1E2E3E4B1B2E1E2E3E4
SPM
FLAME1
FSL OLS
3dttest
Perm

(b)


Oulu, one sample t-test, 40 subjects, CDT p = 0.001
Familywise error rate (%)
4 mm
6 mm
8 mm
10 mm
Expected
95% CI
B1B2E1E2E3E4B1B2E1E2E3E4B1B2E1E2E3E4B1B2E1E2E3E4B1B2E1E2E3E4
SPM
FLAME1
FSL OLS
3dttest
Perm

Conclusions and Questions

- The permutations were preprocessed using FSL and look very similar to the FSL RFT values in shape. So potentially issues in the pre-processing. Permutations should really have been run with pre-processing in all settings.
- Does this mean the data is non-Gaussian?
- More likely the fact that the samples are dependent is the reason why there is a problem here. In fact Oulu has the lowest number of subjects and displays the highest false positive rates.
- Nevertheless RFT needs to be improved, the question is how?

Convolution Fields and Local Maxima

Expected Euler Characteristic - LKC calculation

Theorem

Under certain regularity conditions, for non-stationary random fields:

$$\mathbb{E}[\chi(\mathcal{A}_u)] = \sum_{d=0}^{D} \mathcal{L}_d \rho_d(u)$$

- $\rho_d : \mathbb{R} \longrightarrow \mathbb{R}$ are the euler-characteristic densities and are known
- $\mathcal{L}_d$ are the Lipshitz Killing Curvatures (LKCs) which depend on T and on the covariance structure and partial derivatives of X.
- The above formula holds for Gaussian, T and F fields so is very general.

Example LKCs in D = 1, 2

In 1D. Let We have:

$$\mathcal{L}_1 = \int_T \sqrt{\mathrm{var}\left(\frac{dX}{dt}(t)\right)}\, dt$$

In 2D in the case that $T \subset \mathbb{R}^2$ with piecewise C^2 -boundary ∂T parametrized by the piecewise C^2 -function $\gamma : [0, 1] \rightarrow \mathbb{R}^2$ the LKCs are given by

$$\mathcal{L}_1 = \frac{1}{2}\int_0^1 \sqrt{\frac{d\gamma^T}{dt}(t)\Lambda(\gamma(t))\frac{d\gamma}{dt}(t)}$$

and

$$\mathcal{L}_2 = \int_T \sqrt{\det(\Lambda(t))}\, dt$$

where $\Lambda(t) = \mathrm{cov}(\nabla X(t))$.
The details are not important the point is that they are computable!

FWHM

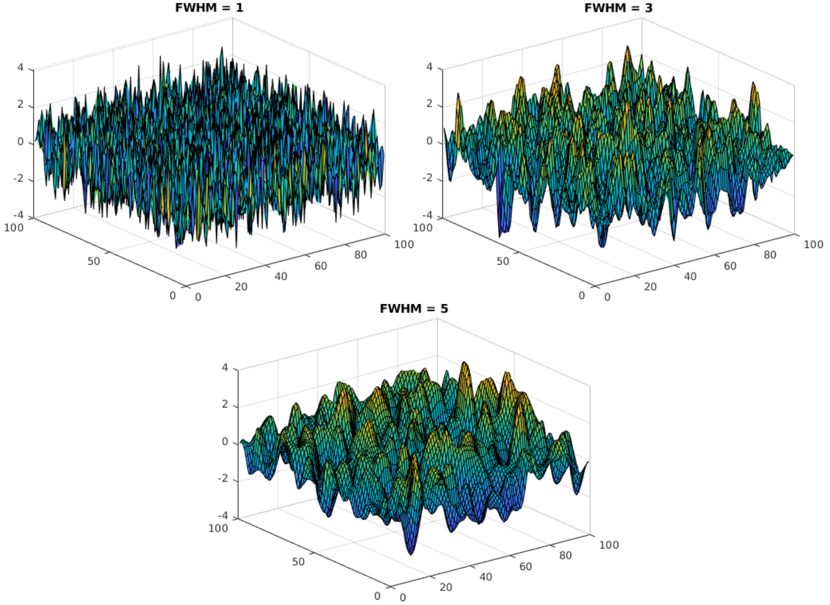
In Neuroimaging it is common to smooth your images with a Gaussian kernel. Given a kernel with parameter σ, the FWHM of the kernel is $\text{FWHM} = 2\sqrt{(}log(2))\sigma$.
If the field is iid white N(0,1) noise smoothed with a multivariate Gaussian kernel with parameter Σ then

$$\Lambda = \Sigma^{-1}/2$$

In 3D if Σ is diagonal then

$$\Lambda = \Sigma^{-1}/2 = \begin{pmatrix} 1/\text{FWHM}_x^2 & 0 & 0 \\ 0 & 1/\text{FWHM}_y^2 & 0 \\ 0 & 0 & 1/\text{FWHM}_z^2 \end{pmatrix} 4\log(2)$$

2D Fields with various FWHM



Smoothness estimation in SPM

When doing RFT one of the most important things to estimate is the smoothness. In 1D this is $\text{var}\left(\frac{dX}{dt}\right)$.
Even when the random fields are stationary current smoothness estimation in SPM/FSL is very bad. Given a random field on a lattice (with spacing h)the current estimates (in 1D) estimate this as

$$\frac{1}{|V|}\sum_{v\in V}\left(\frac{X(v+h)-X(v)}{h}\right)^2$$

more generally in ND where the lattice spacing is h_j in the jth direction, $\Lambda_{ij} = \text{cov}\left(\frac{\partial X}{\partial t_i}, \frac{\partial X}{\partial t_j}\right)$ is estimated as

$$\Lambda_{ij} = \frac{1}{|V|}\sum_{v\in V}\left(\frac{X(v+h_i)-X(v)}{h_i}\right)\left(\frac{X(v+h_j)-X(v)}{h_j}\right).$$

Kebel smoothness estimation

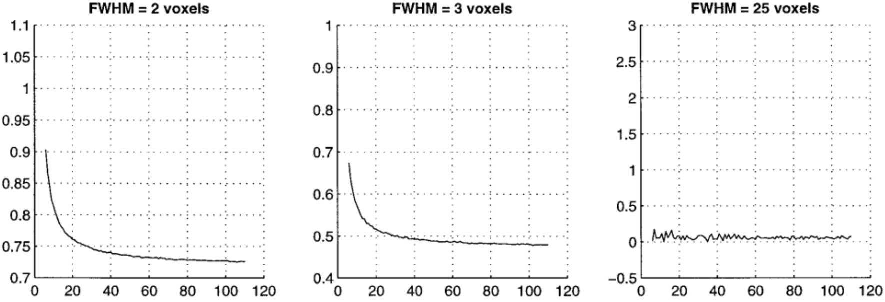


Figure 11: Estimate error versus FWHM as can be seen for low FWHM the estimate is terrible. (Typical FWHM in fMRI are between 3 and 6! SPM implementation.)

Solution: Convolution Fields

Definition

Given a set of voxels $V \subset T$, a random field X on V and a smooth kernel $K : \mathbb{R}^D \longrightarrow \mathbb{R}$ define the discrete convolution field to be $Y : S \longrightarrow \mathbb{R}$

$$Y(s) = (K * X)(s) := \sum_{v \in V} K(s - v)X(v).$$

This is a smooth random field defined on the whole of S. Note that typically people only evaluate Y on the lattice V however our proposal allows evaluation at every point.

$$\mathbb{E}[Y(s)] = \sum_{v \in V} K(s - v)\mathbb{E}[X(v)]$$

$$\mathrm{var}(Y(s)) = \sum_{v, v' \in V} K(s - v)K(s - v')\mathbb{E}\big[(X(v) - \mu(v))(X(v') - \mu(v'))\big].$$

FWHM and LKCs of Convolution Fields

Differentiating is fairly easy, in particular:

$$Y'(s) = (K * X)(s) := \sum_{v \in V} K'(s - v)X(v).$$

and so (assuming that the fields are zero mean),

$$\mathrm{var}(Y'(s)) = \sum_{v, v' \in V} K'(s - v)K'(s - v')\mathbb{E}\big[X(v)X(v')\big].$$

If the original field was just white noise on the lattice then this becomes:

$$\mathrm{var}(Y'(s)) = \sum_{v \in V} (K'(s - v))^2$$

Estimating the FWHM is then a matter of evaluating the variance of the derivatives. Note that

FWHM and LKCs of Convolution Fields - cont

Either you can use the pointwise estimate of $\text{var}(Y'(s))$ from above or you could average over some points in the field.
In 1D you get the LKC for free since

$$\mathcal{L}_1 = \int_T \sqrt{\text{var}\left(\frac{dX}{dt}(t)\right)}\, dt$$

similarly for 2D. In 3D it's much more difficult to compute the LKCs. Fabian Teleschow and I are currently working on this.

FWHM estimation comparison

False Positive Rates

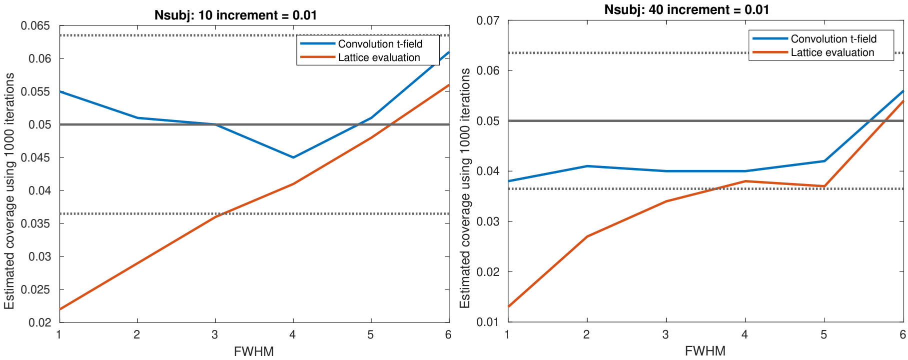


Figure 12: False Coverage in 1D

Kac-Rice formula

Theorem

Given random fields f and g, under certain regularity conditions, if

$$N_u := N_u(T) := N_u(f, g : T, B)$$

denotes the number of points in T for which $f(t) = u \in \mathbb{R}^N$ and $Z(t) \in B \subset \mathbb{R}^K$, and $p_t(x, \nabla y, v)$ denotes the joint density of $(f(t), \nabla f(t), g(t))$, we have, with $D = N(N+1)/2 + K$,

$$\mathbb{E}[N_u] = \int \mathbb{E}[|\det \nabla f(t)| 1_B(g(t)) | f(t) = u] p_t(u) dt,$$

where the p_t here is the density of $f(t)$.

Applying Kac-Rice

Theorem

Kac Rice:

$$\mathbb{E}[N_u] = \int \mathbb{E}[|\det \nabla f(t)| 1_B(g(t)) | f(t) = u] p_t(u) dt,$$

In particular taking $f = \nabla X$ and $g = \nabla^2 X$ and B to be the set of negative definite matrices, then $N_0 = M$ is the number of local maxima and by Kac-Rice:

$$\mathbb{E}[M] = \int \mathbb{E}\big[|\det \nabla^2 X(t)| 1_B(\nabla^2 X) | \nabla X(t) = 0\big] p_{\nabla X(t)}(0) dt.$$

Peak Height Distributia

In general we want to make inference on non-stationary random fields. One thing we can estimate here is the peak height distribution namely:

$$\mathbb{P}(f(t_0) > u | t_0 \text{ is a local maximum of } f(t))$$

the event has probability zero so we need to consider:

$$F_{t_0}(u) := \lim_{\epsilon \to 0} \mathbb{P}(f(t_0) > u | \exists \text{ a local maximum of } f(t) \text{ in } U_{t_0}(\epsilon))$$

Maxima Distribution Theorem

(Cheng & Schwartzman, 2015) has the following result:

Theorem

Let $\{f(t) : t \in T\}$ be a Gaussian random field. Then for each $t_0 \in T$ and $u \in \mathbb{R}$,

$$F_{t_0}(u) = \frac{\mathbb{E}\left[|\det \nabla^2 f(t_0)| 1(f(t_0) > u) 1(\nabla^2 f \in B) \middle| \nabla f(t_0) = 0\right]}{\mathbb{E}[|\det \nabla^2 f(t_0)| 1(\nabla^2 f \in B) | \nabla f(t_0) = 0]}$$

note both the numerator and denominator look like Kac-Rice and this is not a coincidence!

Cluster Extent

Now we're interested in cluster extent. As such we'd like to have an idea of the following distribution:

Proof of the Result

The proof in (Cheng & Schwartzman, 2015) is complicated. We will present a slightly simpler proof.

$$\mu(t_0, \epsilon) = \#\{t \in U_{t_0}(\epsilon) : \nabla f(t) = 0\}$$

and for $h > 0$, $u \in \mathbb{R}$

$$\mu_u(t_0, \epsilon, h) = \left\{t \in U_{t_0}(\epsilon) : \nabla f(t) = 0, f(t) > u, \nabla^2 f \in B\right\}$$

Extension to χ^2 fields

Conclusion

Conclusion

- Random Field Theory DOES work you just need to get around the assumptions. It can be developed in the non-stationary case as illustrated.
- We will very much encourage the use of convolution fields even when

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

Cheng, D., & Schwartzman, A. (2015). Distribution of the height of local maxima of Gaussian random fields. *Extremes*, *18*(2), 213–240. doi: 10.1007/s10687-014-0211-z

Eklund, A., Nichols, T. E., & Knutsson, H. (2016). Cluster failure: Why fmri inferences for spatial extent have inflated false-positive rates. *Proceedings of the national academy of sciences*, *113*(28), 7900–7905.

Friston, K., Holmes, A., Poline, J.-B., Price, C., & Frith, C. (1996). Detecting Activations in PET and fMRI: Levels of Inference and Power. *NeuroImage*, *4*(3), 223–235. Retrieved from `http://linkinghub.elsevier.com/retrieve/pii/S1053811996900749` doi: 10.1006/nimg.1996.0074

Friston, K. J., Worsley, K. J., Frackowiak, R. S. J., Mazziotta, J. C., & Evans, A. C. (1994). Assessing the significance of focal activations using their spatial extent. *Human Brain Mapping*, *1*, 214–220.