Voxelwise Inference using Convolution Random Fields

Samuel Davenport, Fabian Telschow

University of Oxford

May 20, 2020

Voxelwise Inference using RFT

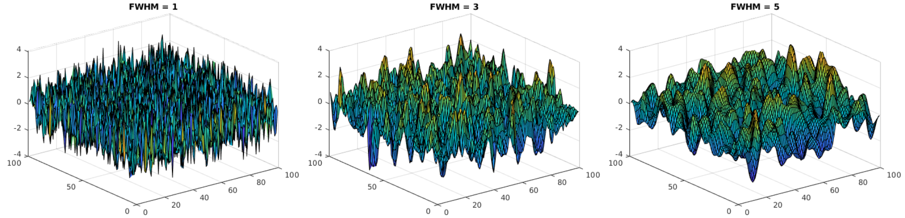
Random Fields and Test Statistics

Definition

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

$$Z : S \longrightarrow \mathbb{R}$$

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



Smoothing

Suppose we have N subjects and that for each subject we observe a 3D image data X_n on a lattice V. In fMRI smoothing is done in order to increase the signal to noise ratio. I.e. for each subject n, X_n is smoothed with a kernel K to give

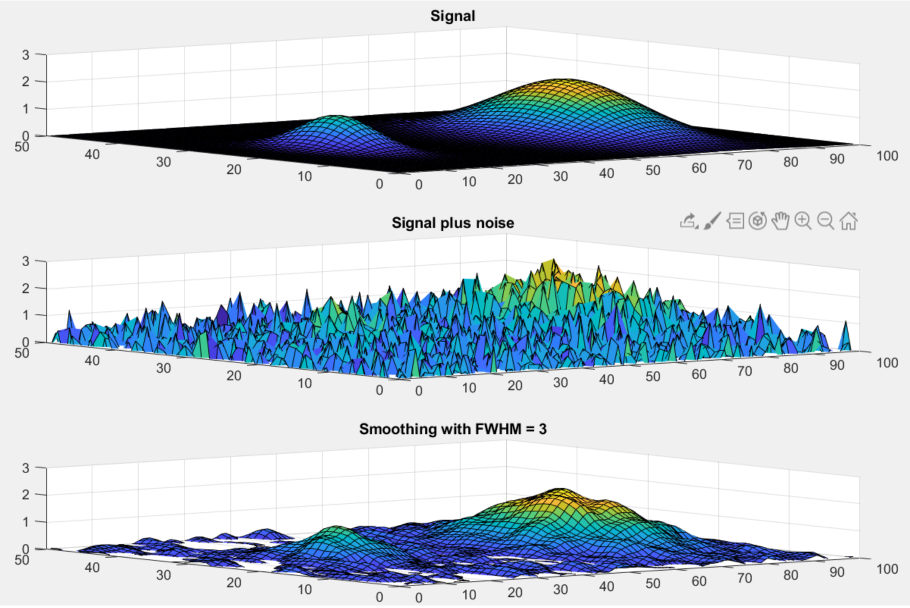
$$Y_n(v) = \sum_{l \in V} K(v-l) X_n(l)$$

at every voxel $v \in V$. An example of a typically used smoothing kernel is

$$K(x) = \frac{1}{(2\pi\sigma)^{n/2}} e^{-\frac{1}{2}x^T \Sigma^{-1} x}.$$

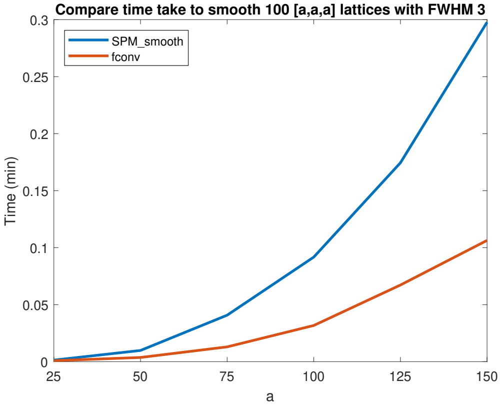
In fMRI it is typical to take $\Sigma = \sigma^2 I$ and to say that you're smoothing with $\mathrm{FWHM} = 2\sqrt{2\log(2)}\sigma$.

Why you should smooth



Smoothing in SPM

3D smoothing in SPM is typically done with spm_smooth.



fconv (implemented in the RFTtoolbox) performs smoothing significantly faster than SPM taking advantage of matlab's convn function and the fact that the Gaussian kernel is separable.

Brain Imaging

Having smoothed, in brain imaging we have image data from each subject and at each voxel v we fit a linear model:

$$Y(v) = A\beta(v) + \epsilon$$

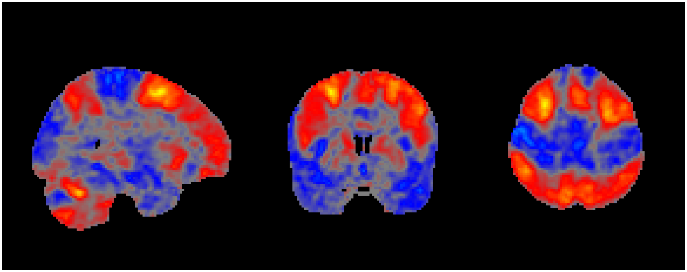
where $Y(v) = [Y_1(v), \dots, Y_N(v)]^T$ and A is some matrix. In this context we typically want to test whether $c^T\beta = 0$ for some contrast vector c. The simplest example of this is a one-sample t test. I.e. taking $A = [1, \dots, 1]^T$, $\beta \in \mathbb{R}$ we get

$$\hat{\beta} = \frac{1}{N}\sum_{n=1}^{N} Y_n(v)$$

where N is the number of subjects. We test the hypothesis that $\beta = 0$ using

$$T_L(v) = \frac{\hat{\beta}\sqrt{N}}{\hat{\sigma}} = \frac{\frac{1}{\sqrt{N}}\sum_{n=1}^{N} Y_n(v)}{\left(\frac{1}{N-1}\sum_{n=1}^{N}\left(Y_n(v) - \frac{1}{N}\sum_{n=1}^{N} Y_n(v)\right)^2\right)^{1/2}}.$$

t fields



Definition

Given $S \subset \mathbb{R}^D$, and $N \in \mathbb{N}$ and Gaussian random fields $Y_1, \dots, Y_N$, define the t-field to be $T : S \to \mathbb{R}$,

$$T(s) = \frac{\frac{1}{\sqrt{N}} \sum_{n=1}^{N} Y_n(s)}{\left(\frac{1}{N-1} \sum_{n=1}^{N} \left(Y_n(s) - \frac{1}{N} \sum_{n=1}^{N} Y_n(s) \right)^2 \right)^{1/2}}.$$

Voxelwise Inference: Controlling the FWER

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.

$$\mathbb{P}\left(\max_{v \in V_0} T_L(v) > u\right)$$

and we seek to control this at a level α.

Historically voxelwise RFT has assumed that there is a smooth random field T such that $T \approx T_L$ and

$$\max_{l \in V} T_L(l) \approx \sup_{s \in S} T(s)$$

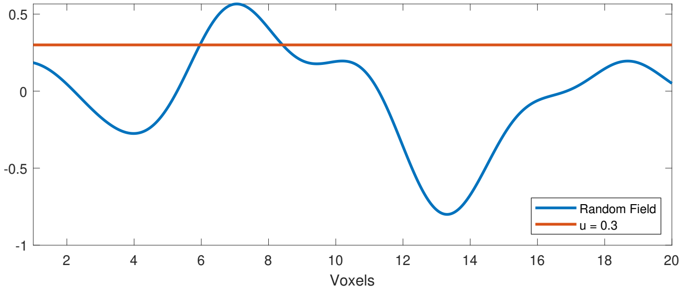
an assumption known as the **good lattice assumption**.

Voxelwise RFT

Let $M_u(T)$ be the number of local maxima of T above a threshold u then assuming that T is twice differentiable,

$$\mathbb{P}\left(\sup_{s\in S} T(s) > u\right) = \mathbb{P}(M_u(T) \geq 1) \leq \mathbb{E}[M_u(T)]$$

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



Expected Euler Characteristic and LKCs

The Euler Characteristic approximation

$\mathbb{E}[M_u(T)]$ is difficult to estimate and requires us to be clever. To do so, given $u \in \mathbb{R}$, define the **excursion set** to be

$$\mathcal{A}_u(T) = \{s \in S : T(s) \geq u\}$$

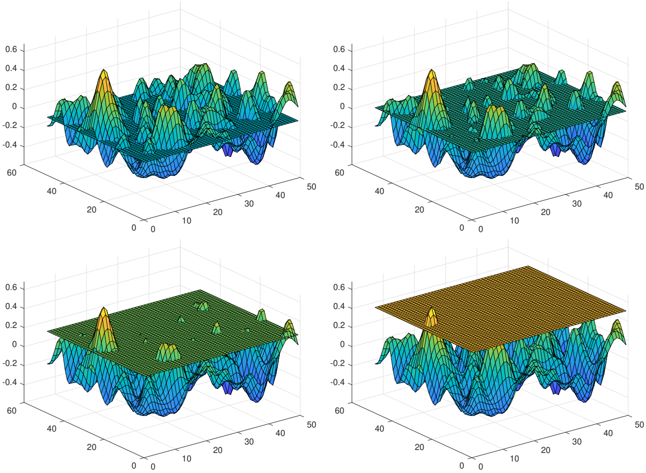
and let

$$\chi_u(T) = \chi(\mathcal{A}_u(T))$$

be the Euler characteristic of the excursion set. In 1D the Euler characteristic is the number of connected components. In 2D it's the number of connected components minus the number of holes.

The Euler Characteristic approximation

When there are no holes the Euler Char is the number of connected components i.e. clusters. At high thresholds it equals the number of local maxima.



Gaussian Kinematic Formula

Theorem

Let Y be a constant variance D-dimensional Gaussian random field with variance σ^2 then, under certain regularity conditions, for all $u \in \mathbb{R}$,

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

where $\mathcal{L}_0, \dots, \mathcal{L}_D$ are constants that depend on S and the covariance function of Y and $\rho_d : \mathbb{R} \to \mathbb{R}$ are functions, called the EC densities, such that for $u \in \mathbb{R}$,

$$\rho_d(u) = \frac{1}{(2\pi)^{(d+1)/2}} H_{d-1}(u) e^{-u^2/2}.$$

where H_{d-1} is the $(d-1)$th Hermite polynomial.

Gaussian Kinematic Formula - general fields

(?, ?) showed that this can be generalized to any function of Gaussian random fields.

Theorem

Let $Y_1, \ldots, Y_N$ be i.i.d D-dimensional unit variance Gaussian random fields and let $F : \mathbb{R}^N \to \mathbb{R}$. Let T be a random field such that

$$T(s) = F(Y_1(s), \ldots, Y_n(s))$$

for all $s \in S$. Then, under certain regularity conditions, for all $u \in \mathbb{R}$,

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

where $\mathcal{L}_0, \ldots, \mathcal{L}_D$ are constants and $\rho_d^F : \mathbb{R} \to \mathbb{R}$ are functions that depends on F and are easy to compute.

GKF - tFields

In particular, given Gaussian random fields $Y_1, \dots, Y_N$ (some $N \in \mathbb{N}$), we can write the one-sample t-field as:

$$\begin{aligned} T(s) &= \frac{\frac{1}{\sqrt{N}} \sum_{n=1}^N Y_n(s)}{\left(\frac{1}{N-1} \sum_{n=1}^N \left(Y_n(s) - \frac{1}{N} \sum_{n=1}^N Y_n(s) \right)^2 \right)^{1/2}} \\ &= F(Y_1(s), \dots, Y_N(s)) \end{aligned}$$

where $F : \mathbb{R}^N \to \mathbb{R}$ sends $y = (y_1, \dots, y_N)$ to

$$F(y) = \frac{\frac{1}{\sqrt{N}} \sum_{n=1}^N y_n}{\left(\frac{1}{N-1} \sum_{n=1}^N \left(y_n - \frac{1}{N} \sum_{n=1}^N y_n \right)^2 \right)^{-1/2}}.$$

So for T-fields, we have

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

where $\rho_d^F(u)$ has a closed form. Note that for all $s \in S$

$$\begin{aligned} T(s) &= \frac{\frac{1}{\sqrt{N}} \sum_{n=1}^N Y_n(s)}{\left(\frac{1}{N-1} \sum_{n=1}^N \left(Y_n(s) - \frac{1}{N} \sum_{n=1}^N Y_n(s)\right)^2\right)^{1/2}} \\ &= \frac{\frac{1}{\sqrt{N}} \sum_{n=1}^N \frac{Y_n(s)}{\sigma(s)}}{\left(\frac{1}{N-1} \sum_{n=1}^N \left(\frac{Y_n(s)}{\sigma(s)} - \frac{1}{N} \sum_{n=1}^N \frac{Y_n(s)}{\sigma(s)}\right)^2\right)^{1/2}} \end{aligned}$$

where $\sigma^2(s) = \mathrm{var}(Y_1(s))$, so in this case the assumption that the fields are constant variance doesn't matter.

LKCs under stationarity

The LKCs in general do not have easy closed forms. However if we assume stationarity, things are a lot easier.

Theorem

Let Y be a variance 1 stationary random field on a D-dimensional box $S = [0, T_1] \times \dots \times [0, T_D]$. Then for $d = 1, \dots, D$,

$$\mathcal{L}_d = \sum_{F \in \mathcal{F}_d} |F| \det(\Lambda_F)^{1/2}$$

where $\mathcal{F}_d$ is the set of faces of S of dimension d and Λ_F is a $d \times d$ submatrix of $\Lambda = cov(\nabla Y)$ corresponding to F and $\mathcal{L}_0 = 1$.

Stationary LKCs in 1D, 2D

We have,

$$\mathcal{L}_d = \sum_{F \in \mathcal{F}_d} |F| \det(\Lambda_F)^{1/2}.$$

Thus, for $D = 1, S = [0, T_1]$ and so has 1 face of dimension 1 namely S itself. As such (since $\Lambda \in \mathbb{R}$),

$$\mathcal{L}_1 = T_1 \Lambda^{1/2}.$$

For $D = 2$, $S = [0, T_1] \times [0, T_2]$ is a rectangle so the $2D$ face is the rectangle itself and the $1D$ faces are its sides. As such

$$\mathcal{L}_2 = |S| \det(\Lambda)^{1/2} = T_1 T_2 (\Lambda_{11}\Lambda_{22} - \Lambda_{12}^2)^{1/2}$$

and

$$\mathcal{L}_1 = T_1 \Lambda_{11}^{1/2} + T_2 \Lambda_{22}^{1/2}.$$

Assumptions of Voxelwise RFT

Cluster Failure - Clusterwise Inference

Clusterwise inference is a method for controlling false positives by deriving parametric distributions for the number of clusters above a threshold. There are two (main) types of RFT used in fMRI: voxelwise and clusterwise. In 2016 (?, ?) showed that clusterwise inference had inflated false positive rates.

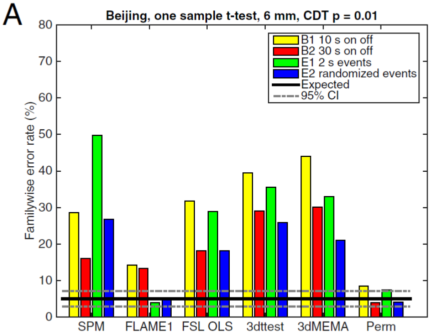


Figure 1: Clusterwise inference has inflated false positive rates

Cluster Failure - Voxelwise Inference

However they showed the opposite held true for voxelwise inference.

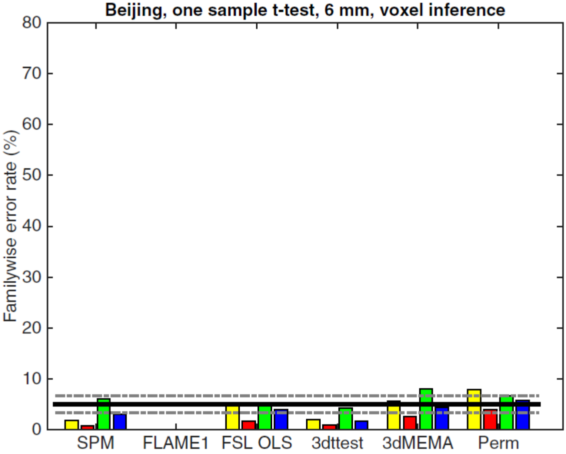
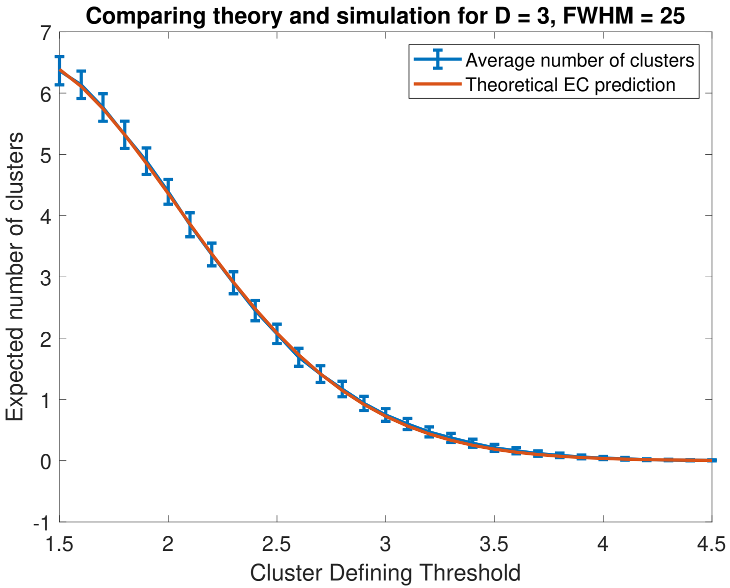


Figure 2: RFT voxelwise inference has conservative false positive rates.

The Euler characteristic approximation



Good Lattice Assumption

Historically voxelwise RFT has assumed that there is a smooth random field T such that $T \approx T_L$ and

$$\max_{l \in V} T_L(l) \approx \sup_{s \in S} T(s)$$

an assumption known as the **good lattice assumption**. With high enough smoothing this is not a problem. However at smoothing levels typically used in fMRI this fails. In particular given T_L, suppose that a random field T exists such that for all v, $T(v) = T_L(v)$ then

$$\max_{l \in V} T_L(l) = \max_{l \in V} T(l) < \sup_{s \in S} T(s).$$

Thus for any threshold u

$$\mathbb{P}(\max_{l \in V} T_L(l) > u) < \mathbb{P}(\max_{l \in V} T(l) > u)$$

so choosing thresholds for T_L based on T leads to conservativeness.

Assumptions of voxelwise RFT in SPM

- Good Lattice Assumption (i.e smoothness)
- Stationarity
- Gaussianity
- Accuracy of the Euler characteristic approximation

We will show that the good lattice assumption and stationarity can be completely dropped.

Gaussianity can also be ignored so long as you have enough subjects for the central limit theorem to work however it seems reasonable for fMRI.

For thresholds used for 0.05 error rate control the Euler chacteristic is almost always the same as the number of maxima. So this remaining assumption is barely a constraint.

For large number of subjects (when doing one-sample testing) this makes LESS stringent assumptions than permutation testing!

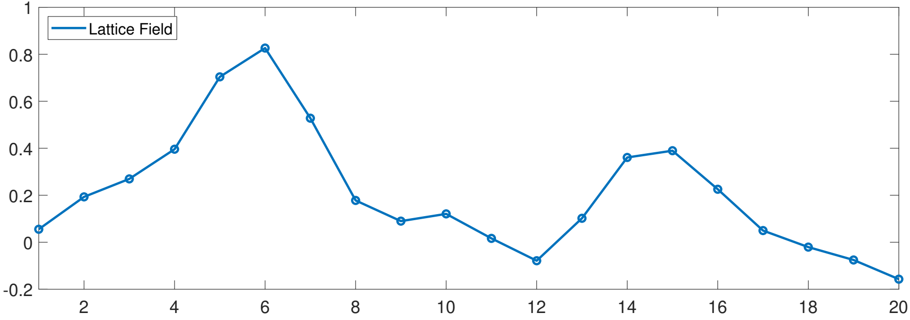
Convolution Fields and non-stationary LKCs

Lattice smoothing

In fMRI smoothing is done in order to increase the signal to noise ratio. To understand how this works, let $X(l)$ be random at every point l of a lattice L. Then smoothing X with a kernel K gives

$$Y(v) = \sum_{l \in L} K(v-l)X(l)$$

at every voxel $v \in L$.

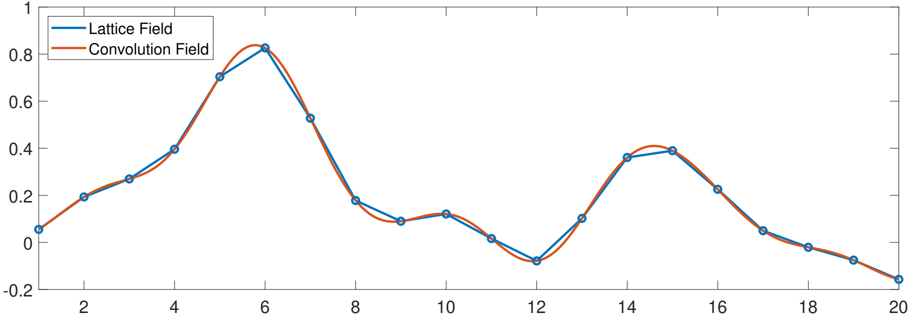


Convolution Random Fields

Definition

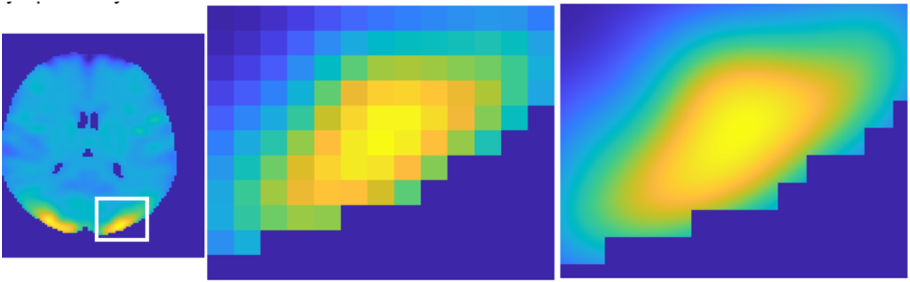
Given random data X on a lattice $L \subset \mathbb{R}^D$ for $s \in \mathbb{R}^D$ and some kernel K, define the convolution field to be $Y : \mathbb{R}^D \to \mathbb{R}$, s.t. for all $s \in S$,

$$Y(s) := (K \star X)(s) = \sum_{l \in L} K(s-l)X(l).$$



Convolution Fields in Brain Imaging

Applying convolution fields in 3D and taking a slice you get the following images!



Convolution t-fields

Given convolution random fields $Y_1, \ldots, Y_N$ for each subject the convolution t-field is just what you'd expect i.e.

$$
\begin{aligned}
T(s) &= \frac{\frac{1}{\sqrt{N}} \sum_{n=1}^{N} Y_n(s)}{\left(\frac{1}{N-1} \sum_{n=1}^{N} \left(Y_n(s) - \frac{1}{N} \sum_{n=1}^{N} Y_n(s) \right)^2 \right)^{1/2}} \\
&= F(Y_1(s), \ldots, Y_N(s))
\end{aligned}
$$

We can similarly define convolution F-fields and more complicated fields.

LKCs under non-stationarity

Because fMRI data is non-stationary we need to be able to estimate the LKCs uner non-stationarity. There has been some work on this though not much progress until recently. One thing that you can take advantage of is closed forms of the higher LKCs, in particular in D dimensions:

$$\mathcal{L}_D = \int_S \sqrt{\mathrm{var}(\nabla Y)}\, ds$$

(where Y) is the variance 1 random field from the GKF. Note that if we assume stationarity,

$$\mathcal{L}_D = \Lambda^{1/2}|S|$$

we recover our stationary formula! $\mathcal{L}_{D-1}$ also has a nice closed form in any dimension and

$$\mathcal{L}_0 = \chi(S)$$

i.e. the Euler characteristic of the domain. However these nice closed forms do not exist for the other LKCs! (A problem in 3D for instance.)

Hermite Projector Estimator

In 2020, (?, ?) proposed the Hermite Projector method for LKC estimation. This allows for LKC estimation is any dimension. In particular,

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

as ρ_0 is the Gaussian cdf. Then we can write

$$\mathcal{L}_d = \frac{(2\pi)^{d/2}}{(d-1)!} \int_{-\infty}^{\infty} H_{d-1}(u)(\mathbb{E}[\chi(\mathcal{A}_u(Y))] - \Phi(t)\chi(S))\, du$$

since $\rho_d(u) = \frac{1}{(2\pi)^{(d+1)/2}} H_{d-1}(u) e^{-u^2/2}$ and the hermite polynomials are orthogonal. Plugging in estimates for the expectation in the integral gives Hermite Projector estimates $\hat{\mathcal{L}}_d$ for the LKCs.

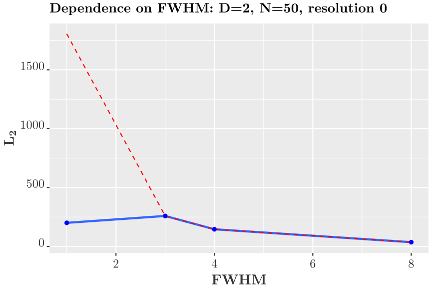
Results

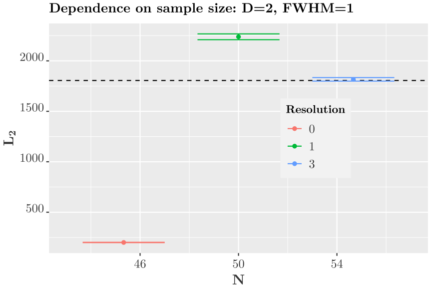
Estimating L2 in 2D - LKCconv

Recall in $2D$

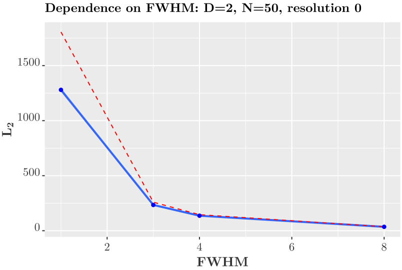
$$\mathcal{L}_2 = \int_S \sqrt{\mathrm{var}(\nabla Y)}\, ds \approx \sum_s \sqrt{\mathrm{var}(\nabla Y(s))}\Delta s$$

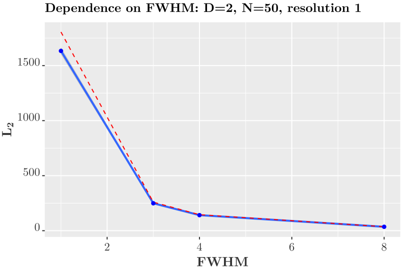
Resolution $= \frac{1}{\Delta x} - 1$.

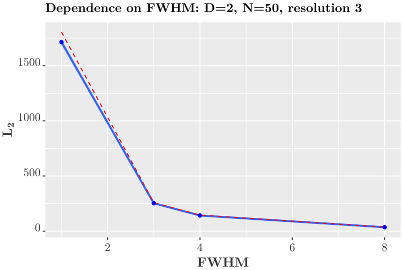




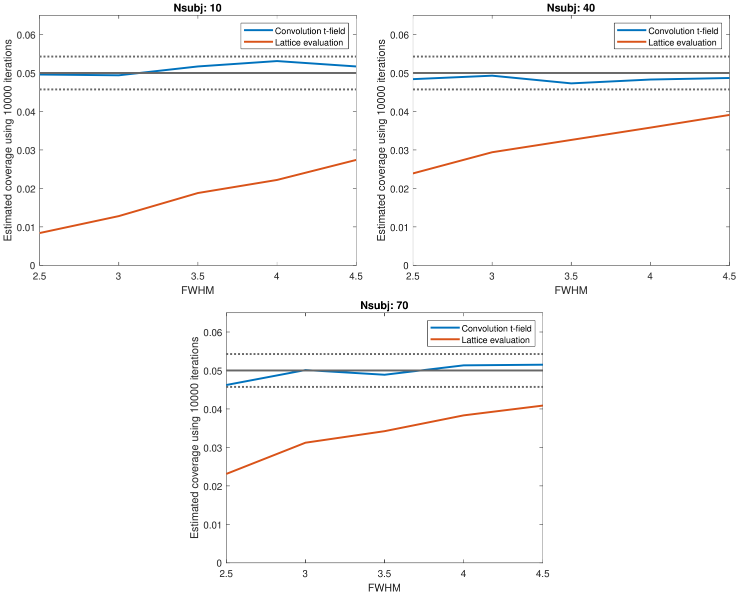
Estimating L2 in 2D - HPE







2D voxelwise inference



Conclusion

- Existing software (SPM, FSL, AFNI etc) only has LKC implementations under stationarity but the framework is more general.
- Using convolution fields exactly controls the FWER at the right level and allows you to drop the good lattice assumption.
- For large samples permutation testing makes stronger assumptions than RFT inference i.e. symmetry of the individual subjects (something that has been called into question recently (?, ?)).
- It's also much faster than existing non-parametric methods that are currently widely used but are slow and inefficient.
- Currently working (with Tom) on validating this on resting state data from the UK biobank.

Software and Slides

- Software available to run LKC estimation under non-stationary and to generate convolution fields is available at `sjdavenport.github.io/software`.
- Slides available at `sjdavenport.github.io/talks`.
- Practical on convolution fields available at `sjdavenport.github.io/tutorials`.

Bibliography