

Medical Image Analysis

Instructor: Moo K. Chung

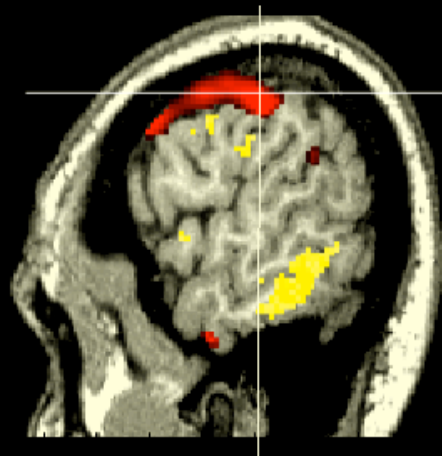
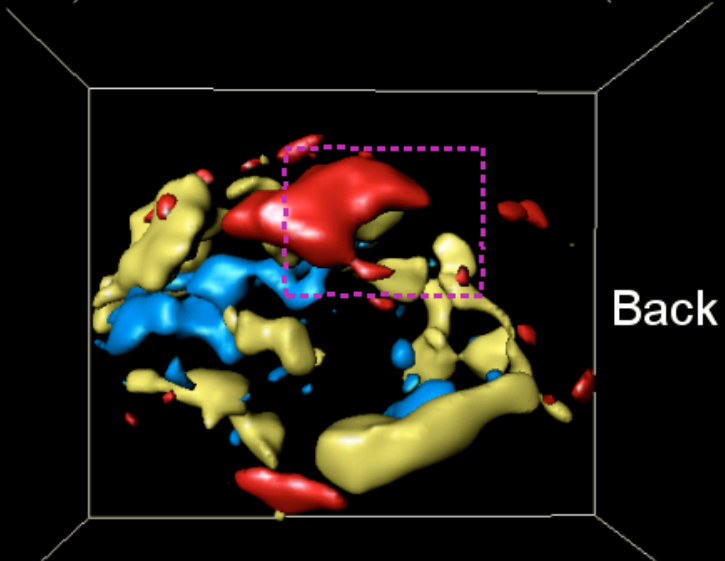
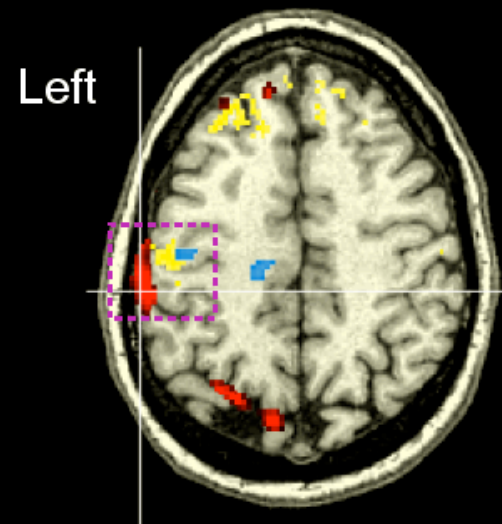
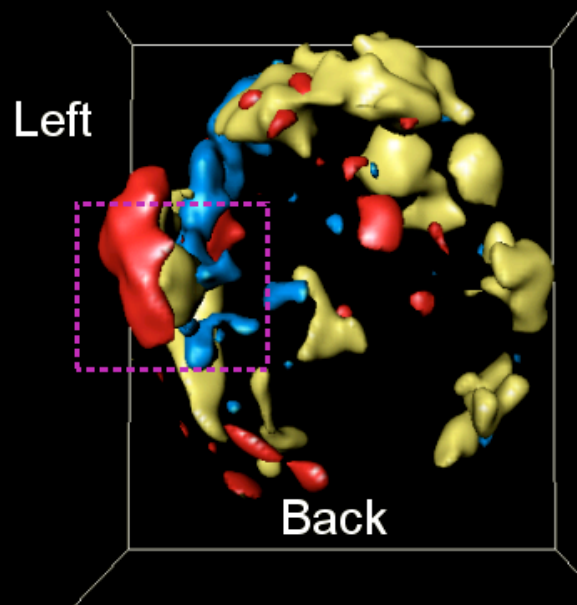
mchung@stat.wisc.edu

Lecture 10.

Multiple Comparisons

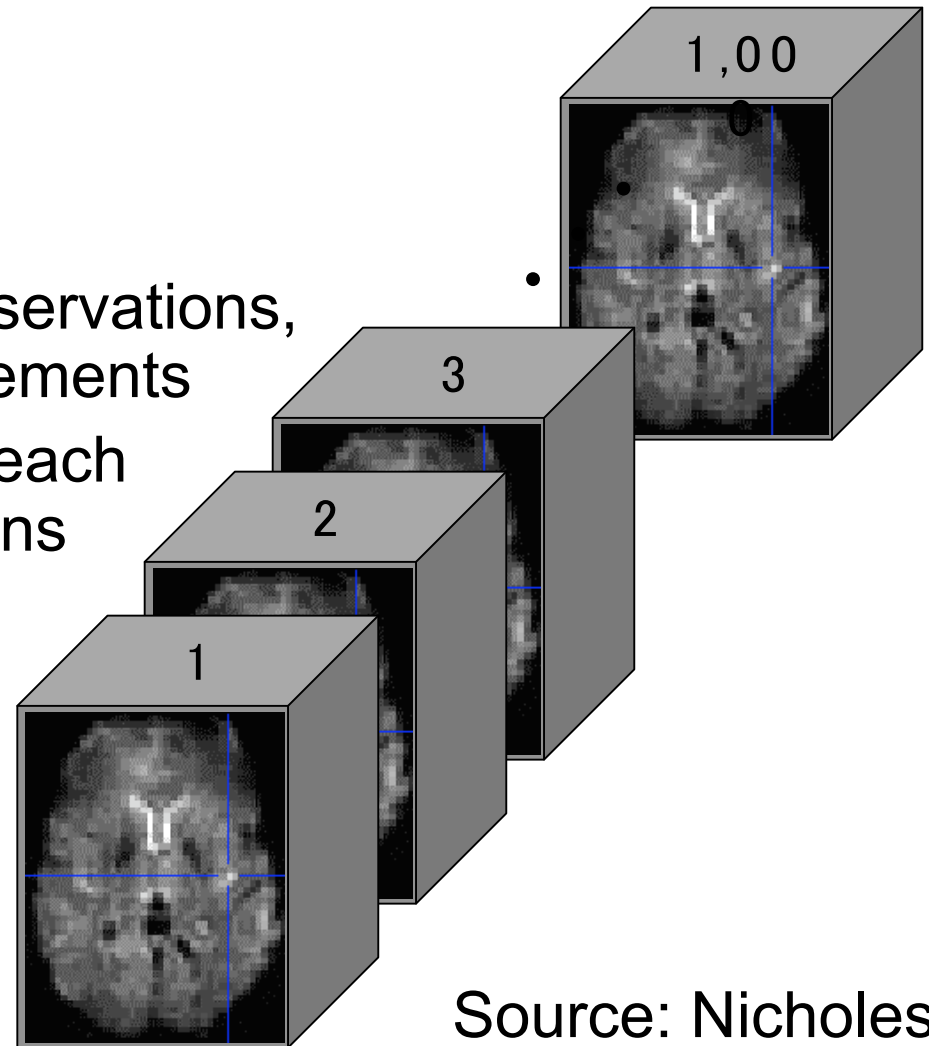
March 06, 2007

This lecture will show you how to construct P-value maps



fMRI Multiple Comparisons

- 4-Dimensional Data
 - 1,000 multivariate observations, each with 100,000 elements
 - 100,000 time series, each with 1,000 observations
- Massively Univariate Approach
 - 100,000 hypothesis tests



Source: Nicholes

Multiple Comparisons

$$X(t) = \mu(t) + e(t), t \in \Omega$$

Search region

$H_0 : \mu(t) = 0$ for all $t \in \Omega$ vs. $H_1 : \mu(t) > 0$ for some $t \in \Omega$.

If there is a point $\mu(t_0) > 0$, reject the null hypothesis

Note: For smooth image, clustered voxels will satisfy $\mu(t_0) > 0$.

Point-wise inference

$H_0 : \mu(t) = 0$ for all $t \in \Omega$ vs. $H_1 : \mu(t) > 0$ for some $t \in \Omega$

At each fixed t

$J_0(t) : \mu(t) = 0.$ vs. $J_1(t) : \mu(t) > 0.$

H_0 is true if $J_0(t)$ is true for all t : $H_0 = \bigcap_{t \in \Omega} J_0(t).$

H_1 is true if $J_1(t)$ is true for some t : $H_1^C = \bigcup_{t \in \Omega} J_0^C(t).$

Test statistic and rejection rule

- Hypothesis testing requires a test statistic and the corresponding rejection rule.
- For our example (one sample test), we can use Z-stat or T-stat. In many applications Z-stat is sufficient since T-stat is approximately Z-stat for large degrees of freedom.
- Then we construct a rejection rule:
Large T-stat value --> reject the null hypothesis
Small T-stat value --> accept the null hypothesis

Type-I error (alpha-level)

- The type-I error is the probability of rejecting the null hypothesis (there is no signal) when the alternate hypothesis (there is signal) is true.
- The type-I error is the probability of detecting **false positives**.
-
- The type-I error computation requires a statistic (Z-stat. t-stat, F-stat, Chi-square stat. etc).
- **Example:** Z-stat. If the Z-statistic value obtained from measurements is 1.65, $\alpha=0.05=P(Z>1.65)$.

Alpha-level for multiple comparisons

Family-wise error rate (FER)

$$\begin{aligned}\alpha &= P(\text{reject } H_0 | H_0 \text{ is true}) \\ &= P\left(\bigcup_{t \in \Omega} J_0^C(t) \mid H_0 \text{ is true}\right)\end{aligned}$$

$$\begin{aligned}\alpha &= P\left(\bigcup_{t \in \Omega} \{T(t) > h\}\right) \\ &= 1 - P\left(\bigcap_{t \in \Omega} \{T(t) \leq h\}\right) \\ &= 1 - P\left(\sup_{t \in \Omega} T(t) \leq h\right) \\ &= P\left(\sup_{t \in \Omega} T(t) > h\right).\end{aligned}$$

Corrected P-value

- P-value: the smallest alpha-level at which the null hypothesis is rejected.

Example: $P(Z > \text{z-stat. value})$.

- *Corrected P-value* is the P-value corrected for multiple comparisons.

$$\text{P - value} = P\left(\sup_{t \in \Omega} T(t) > t - \text{stat. value}\right)$$

Bonferroni correction

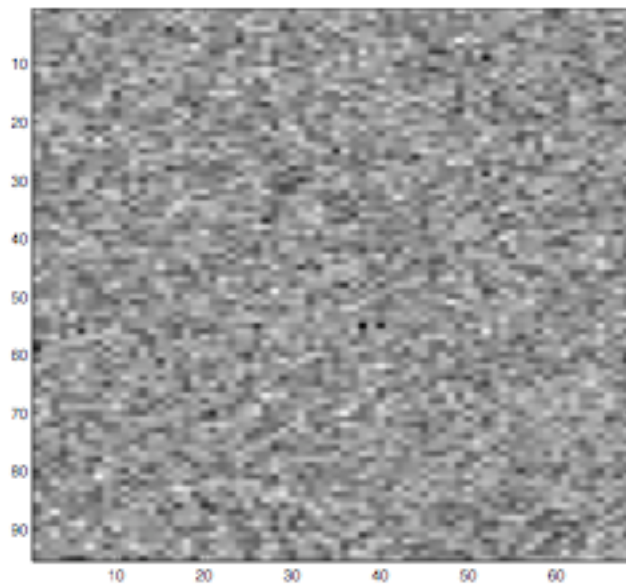
- Assume there are m -voxels in the search region: $t_1, \dots, t_m$

$$\begin{aligned}\alpha &= P\left(\bigcup_{j=1}^m \{T(t_j) > h\}\right) \\ &\leq \sum_{j=1}^m P(T(t_j) > h).\end{aligned}$$

This becomes exact if T statistics are not correlated. We control each T statistic separately.

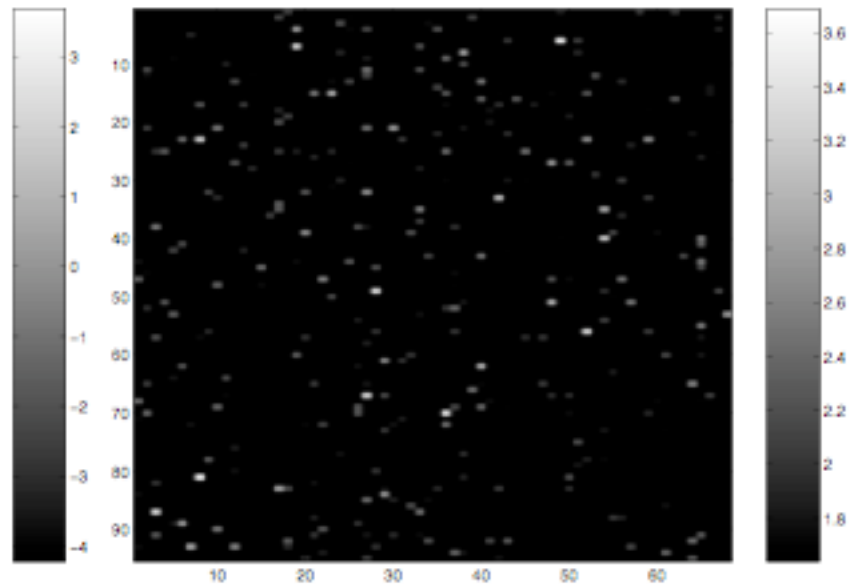
$$P(Y(t_j) > h) < \frac{\alpha}{m}$$

Bonferroni Correction Simulation Result



$Z \sim N(0,1)$ noise

$P(Z > 1.65) = 0.05$



**Bonferroni correction
thresholding at 1.65
5% false positives**

Motivation for FDR

- Traditional approach:

We do not want to make any Type-I errors by controlling the alpha-level (absolute number)

- New approach (FDR):

Maybe we are willing to make a couple of Type-I error as long as the rate of doing so is low.

So we control for the rate (relative number)

It is a bit of nonsense to compare these two numbers

False Discovery Rate (FDR)

- There are many variants of FDR.
- ***False discovery***: a null hypothesis is true (no signal) but we declare a significant effect anyway (reject null)
- **FDR**: the proportion of rejected voxels (significant tests) that have been falsely rejected (false discoveries).
- **FDR = # false discoveries / # rejected voxels**

FDR estimation

$$\text{FDR} = \frac{\# \text{ rejected null that is true}}{\# \text{ rejected null that is true} + \# \text{ rejected null that is false}}$$

Benjamini-Hochberg procedure for determining FDR threshold:

1. Set FDR threshold q between 0 and 1 (q-value)
2. Compute the point-wise P-value at each voxel.
3. Order P-values in increasing order $P_{(1)} < P_{(2)} < P_{(3)} \cdots < P_{(m)}$
4. Determine the largest index j that satisfies $P_{(j)} < \frac{jq}{m}$
5. Reject $P_{(1)}, P_{(2)}, P_{(3)}, \cdots, P_{(j)}$

FDR example

`pvals = [0.023, 0.001, 0.018, 0.0405, 0.006, 0.035, 0.044, 0.046, 0.021, 0.060]`

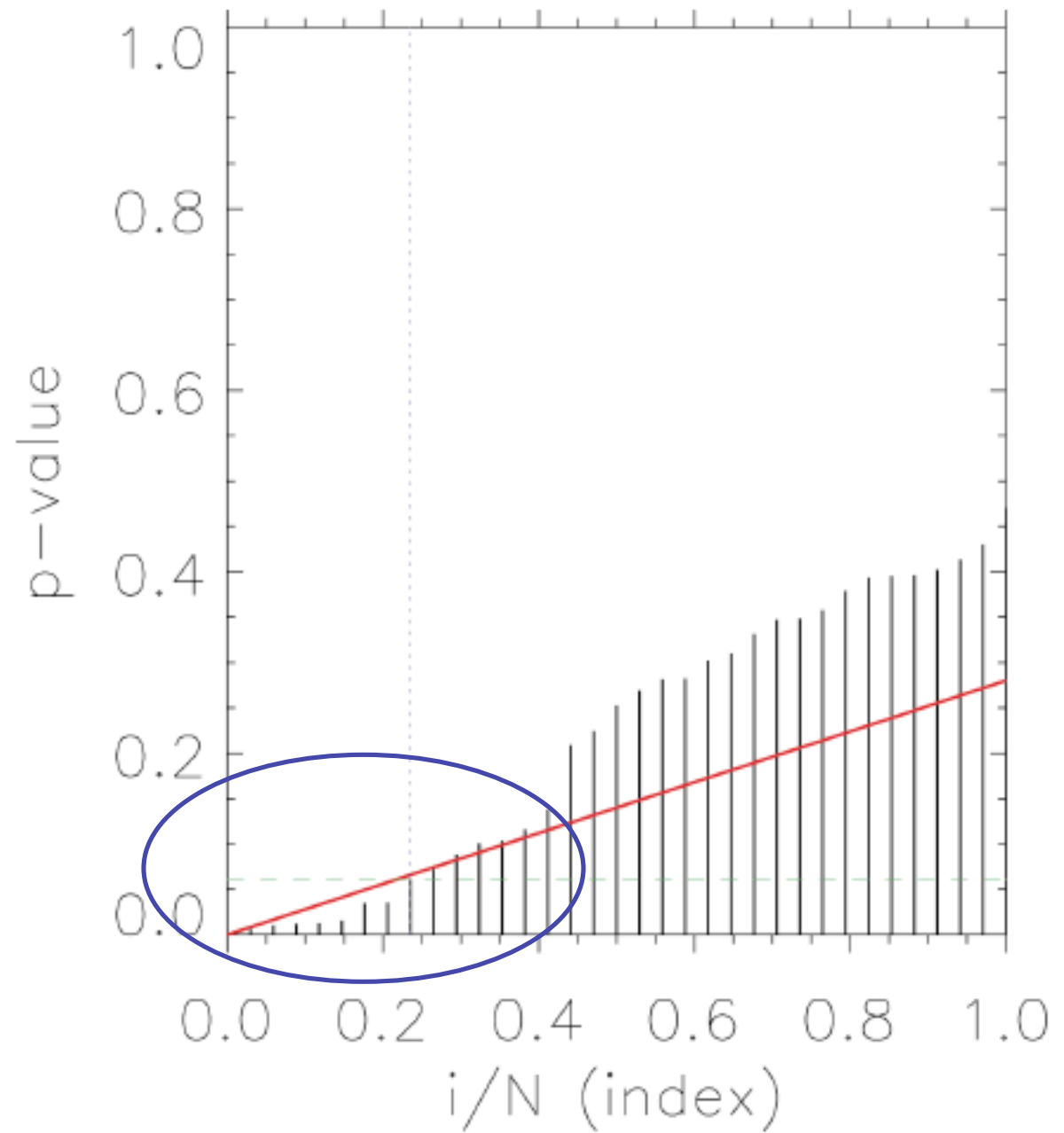
`sorted_pvals = [0.001, 0.006, 0.018, 0.021, 0.023, 0.035, 0.0405, 0.044, 0.046, 0.060]`

Compute $j\alpha/10$ for $j = 1, \dots, 10$ and $\alpha = 0.05$:

`j_alpha = [0.005, 0.010, 0.015, 0.020, 0.025, 0.030, 0.035, 0.040, 0.045, 0.050]`

`diff = [-0.004, -0.004, 0.003, 0.001, -0.002, 0.005, 0.0055, 0.004, 0.001, 0.010]`

**Rejection region
corrected by
FDR procedure**



FDR Software for SPM

Increases

SPM(T₂₃₈)

Design matrix

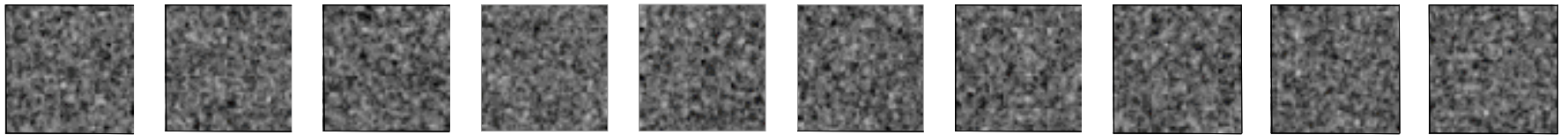
Design matrix

[illegible]

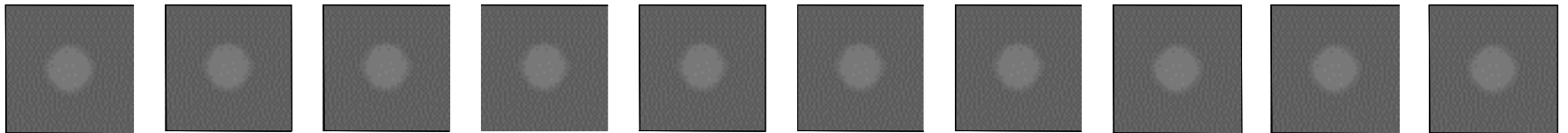
<http://www.sph.umich.edu/~nichols/FDR>

False Discovery Rate Illustration:

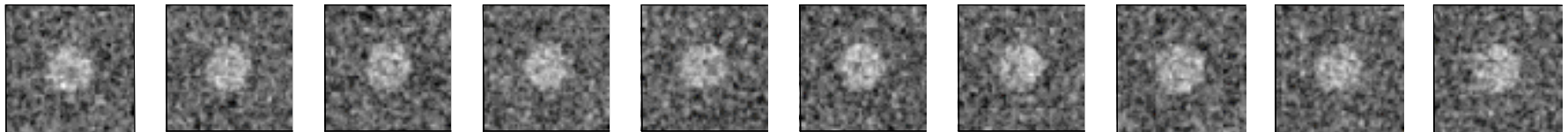
Noise



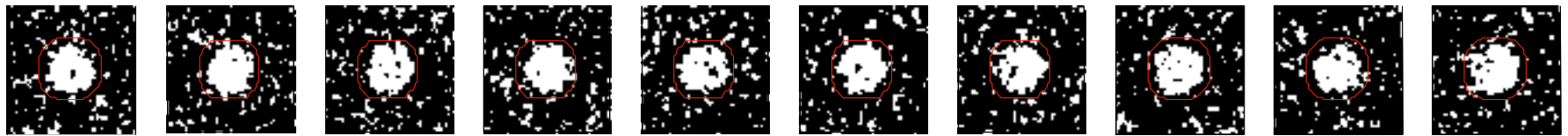
Signal



Signal+Noise



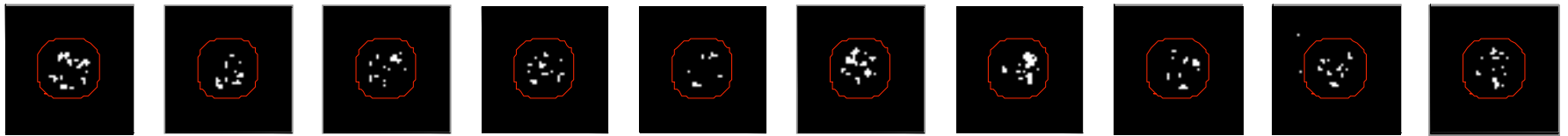
Control of Per Comparison Rate at 10%



11.3% 11.3% 12.5% 10.8% 11.5% 10.0% 10.7% 11.2% 10.2% 9.5%

Percentage of Null Pixels that are False Positives

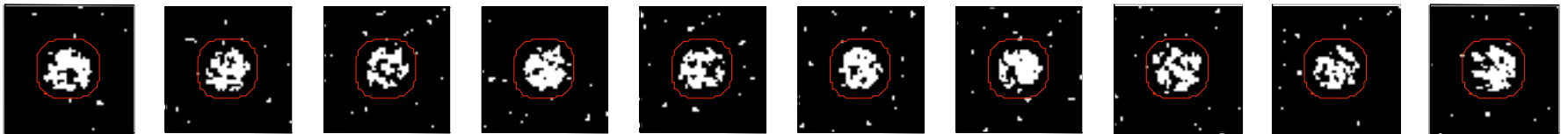
Control of Familywise Error Rate at 10%



FW E

Occurrence of Familywise Error

Control of False Discovery Rate at 10%



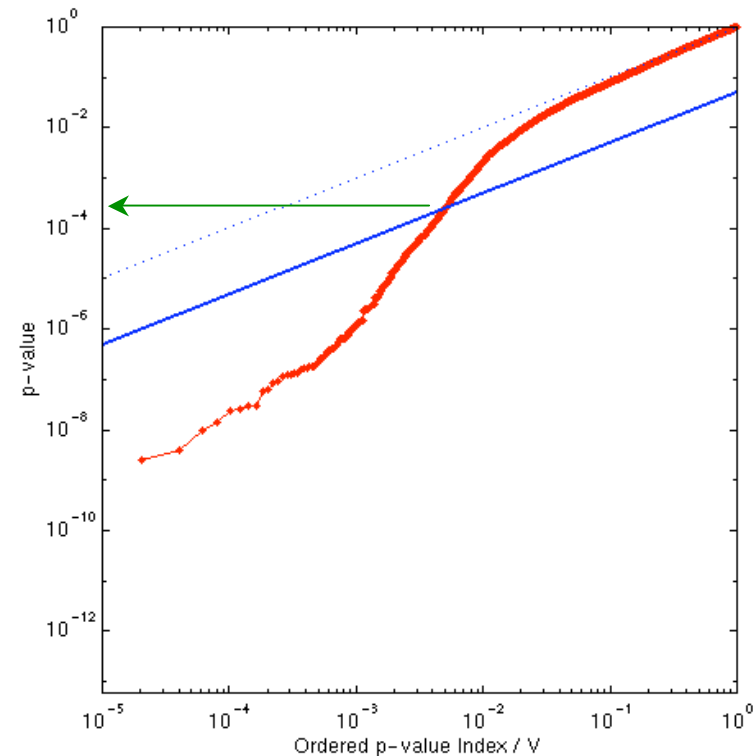
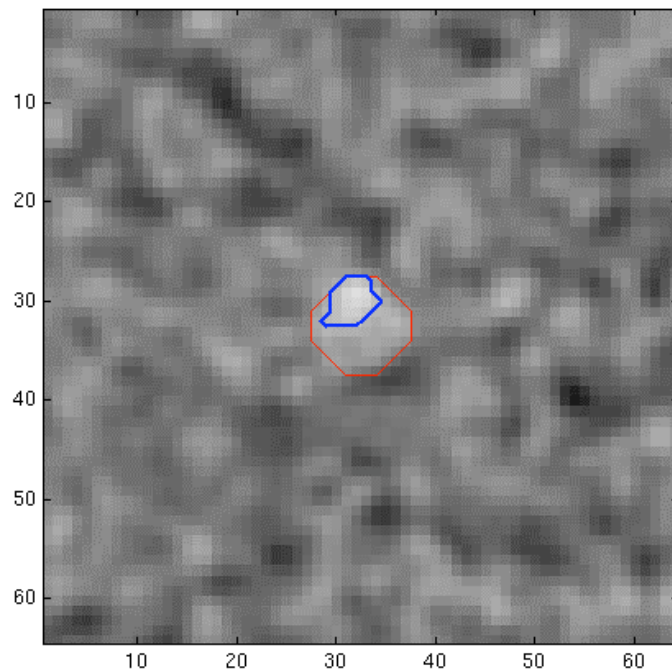
6.7% 10.4% 14.9% 9.3% 16.2% 13.8% 14.0% 10.5% 12.2% 8.7%

Percentage of Activated Pixels that are False Positives

Source: T. Nichols

Benjamini & Hochberg: Varying Signal Extent

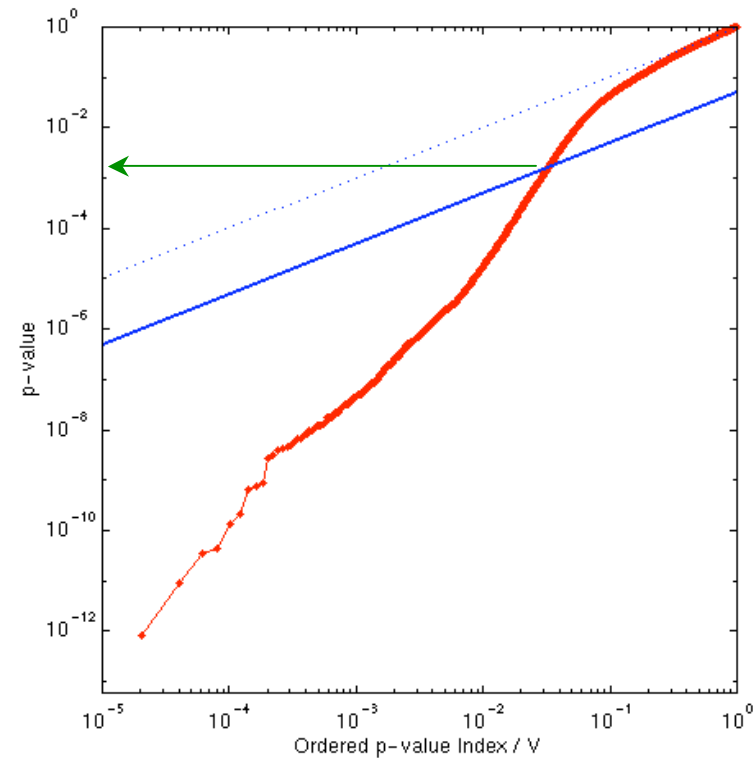
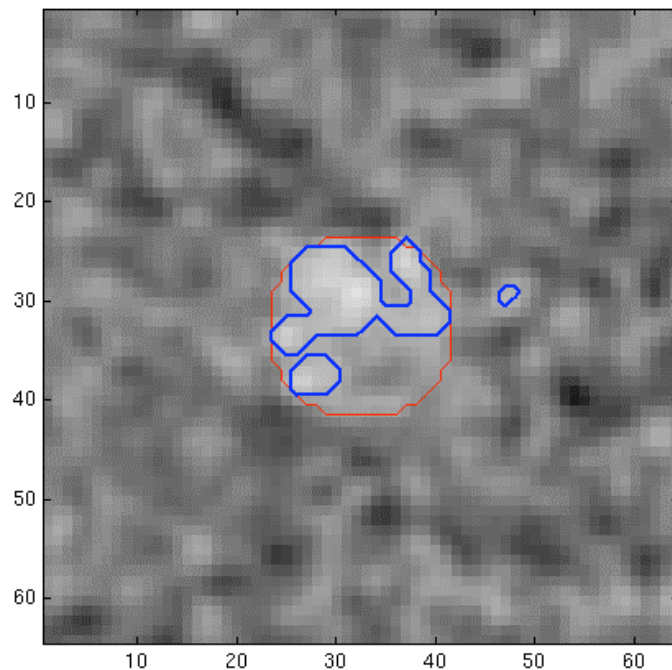
$$p = 0.000252 \quad z = 3.48$$



Signal Intensity 3.0 Signal Extent 5.0 Noise Smoothness 3.0

Benjamini & Hochberg: Varying Signal Extent

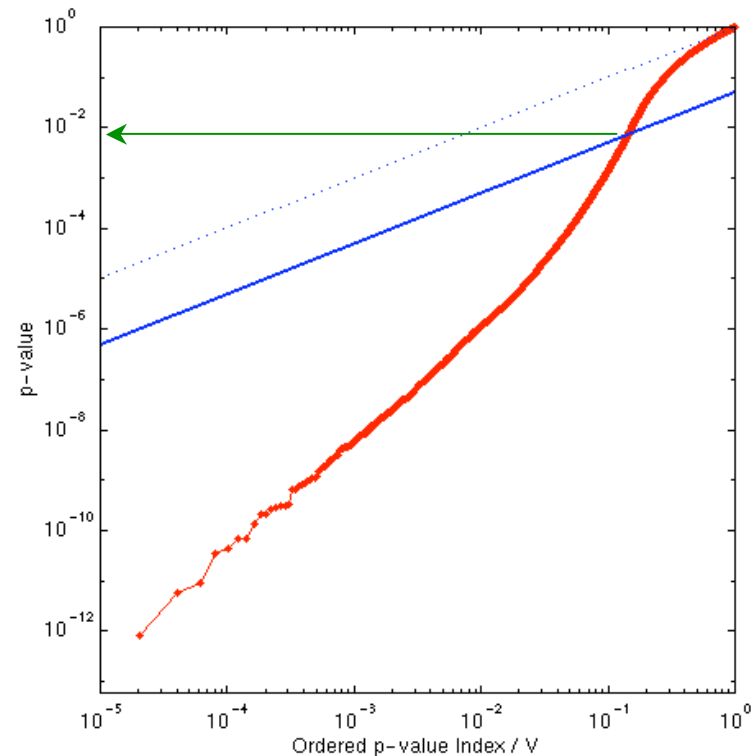
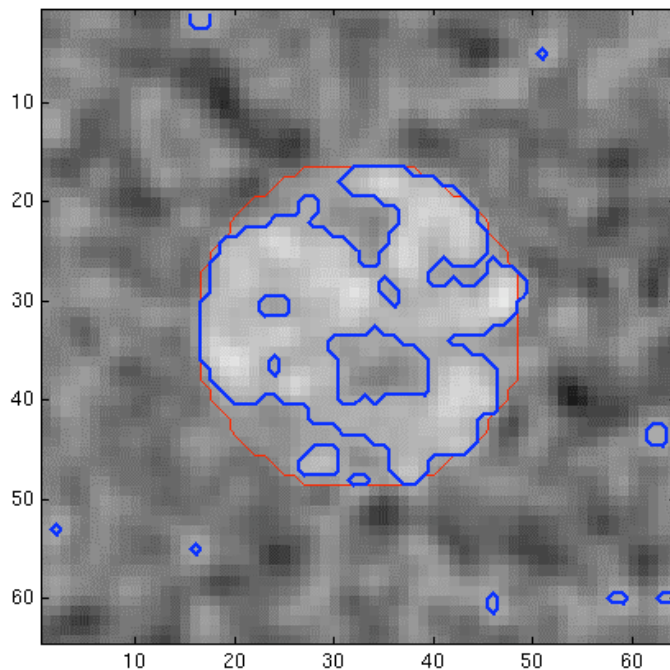
$$p = 0.001628 \quad z = 2.94$$



Signal Intensity 3.0 Signal Extent 9.5 Noise Smoothness 3.0

Benjamini & Hochberg: Varying Signal Extent

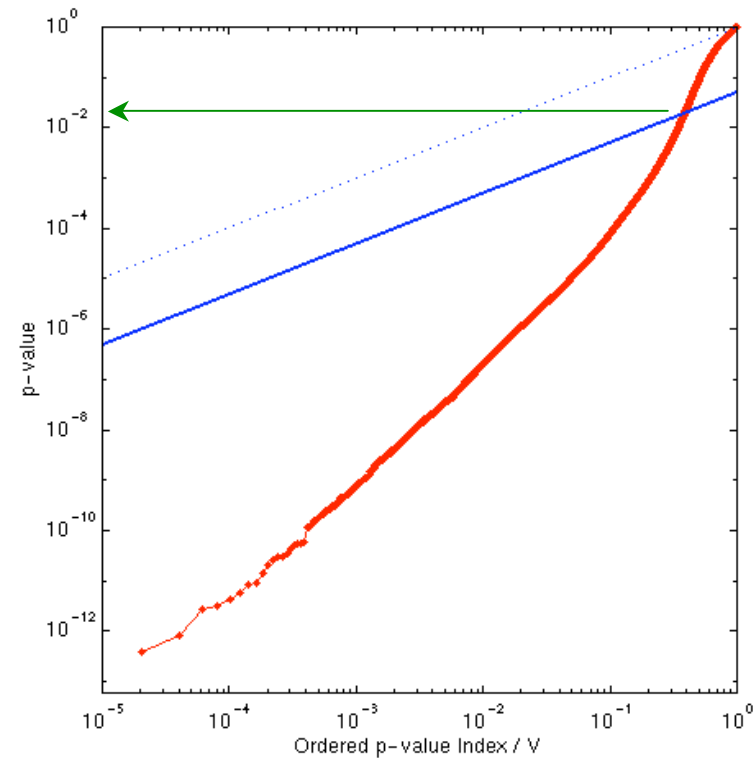
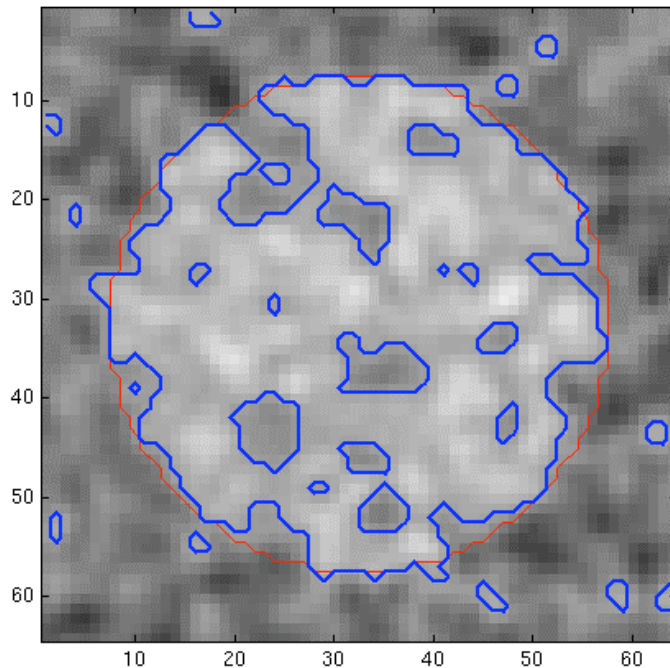
$$p = 0.007157 \quad z = 2.45$$



Signal Intensity 3.0 Signal Extent 16.5 Noise Smoothness 3.0

Benjamini & Hochberg: Varying Signal Extent

$$p = 0.019274 \quad z = 2.07$$



Signal Intensity 3.0 Signal Extent 25.0 Noise Smoothness 3.0

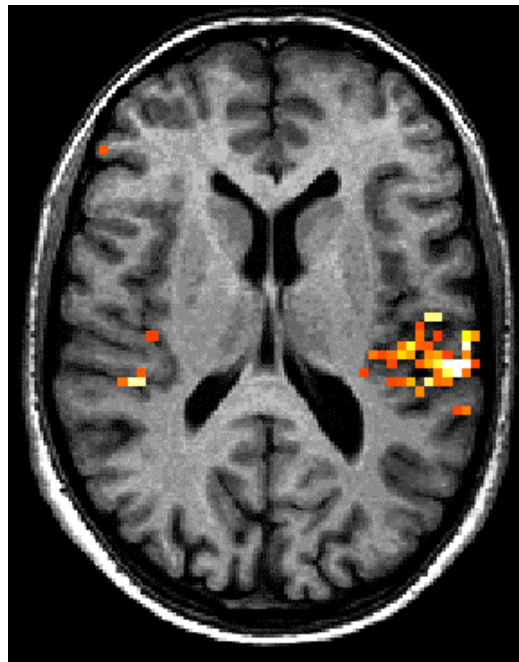
FDR properties

- FDR detects more signal while introducing more false positives.
- Larger the signal, the lower the threshold (easier to detect signal)
- Larger the signal, the more false positives

FDR: Example

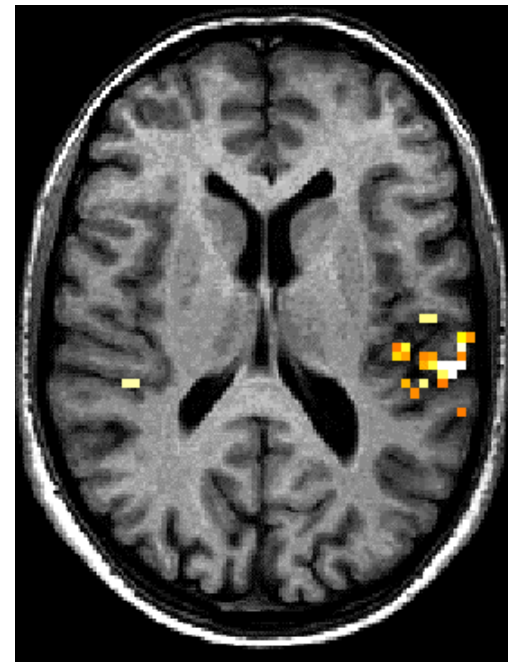
- Verbal fluency data
- 14 42-second blocks
 - ABABAB...
 - A: Two syllable words presented aurally
 - B: Silence
- Imaging parameters
 - 2Tesla scanner, TR = 7 sec
 - 84 64x64x64 images of 3 x 3 x 3 mm voxels

fMRI verbal fluency block design example



$\text{FDR} \leq 0.05$

$t_0 = 3.8119$



$\text{FWER} \leq 0.05$

Bonferroni

$t_0 = 5.485$

Random Field Theory Assumptions

- Images need to follow Gaussian
- Constructed statistics need to be sufficiently smooth. If underlying images are smooth, constructed statistics are smooth.
- The data need to be stationary (uniform FWHM within a search region). If not, we average FWHM across voxels.

Stationary Random Field

- The mean zero random field $T(t)$ is **stationary (homogeneous)** if the joint distribution of $T(t_1), \dots, T(t_n)$ are invariant under translations.

- A Stationary field has the following form of covariance function:

$$R(t,s) = E[T(t)T(s)] = E[T(t+a)T(s+a)] = R(t+a,s+a)$$



Stationarity

- Now let $a=-s$. $R(t,s) = R(t-s,0)$. This implies that the covariance function is a function of $t-s$ only.

Property of Stationary Random Field

- Let $R(t,s)=f(t-s)$.
- $\text{Var } T(t)=R(t,t)=f(t-t)=f(0)=\text{const.}$
- A stationary field has constant variance.

Isotropic Fields

- A special case of stationary fields is an **isotropic field**.

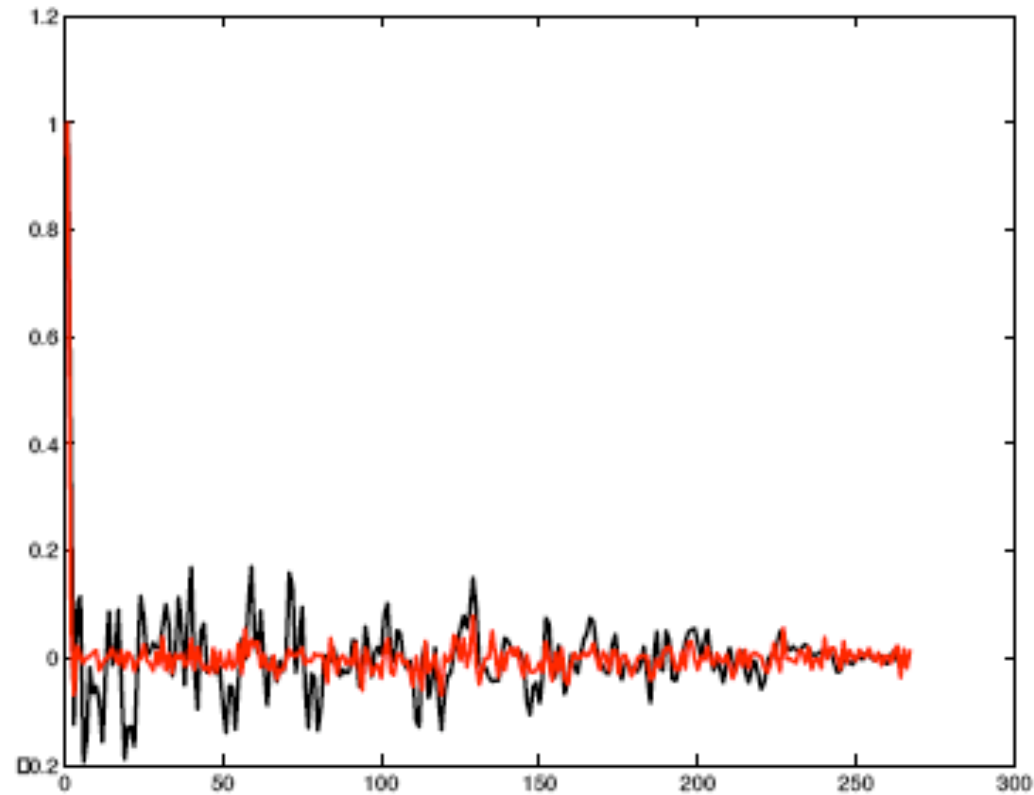
$$R(t,s)=f(||t-s||)$$

- The correlation of random variables $T(t)$ and $T(s)$ is given by

$$\rho(t, s) = \frac{\mathbb{E}[T(t)T(s)]}{[\mathbb{E}T^2(t)\mathbb{E}T^2(s)]^{1/2}} = \frac{f(||t - s||)}{f^2(0)}$$

It is natural to assume a decreasing correlation function.

Decreasing correlation function

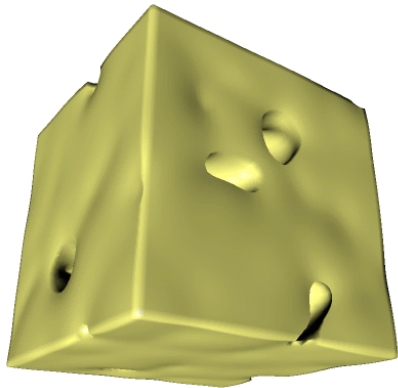


Black: fMRI autocorrelation in amygdala

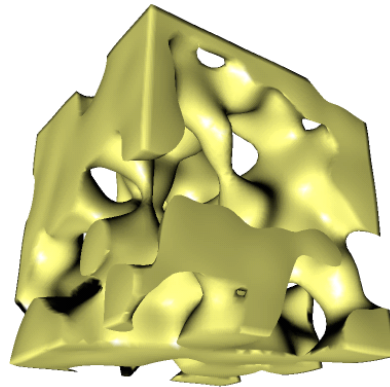
Red: autocorrelation after whitening process

Random Field Approach: Adler's formula

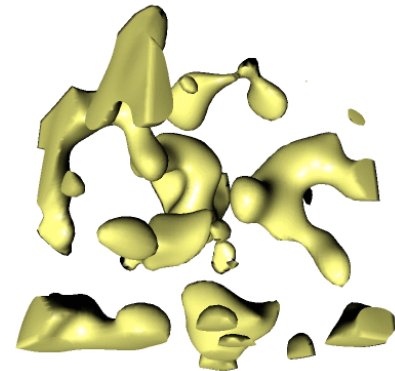
Excursion set $A_h = \{t : T(t) > h\}$



$h = -10$



$h = 0$



$h = 10$

$$P\left(\sup_{t \in \Omega} T(t) > h\right) \approx E(\chi(A_h))$$

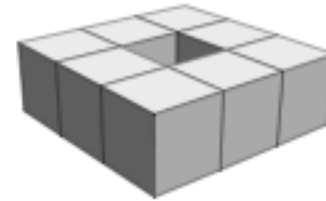
Computing Euler characteristic (topological measure)



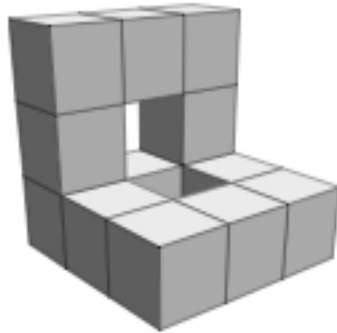
(a) $8-12+6-1=1$



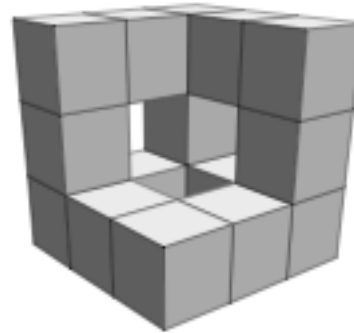
(b) $16-28+16-3=1$



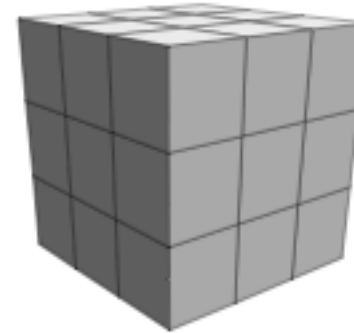
(c) $32-64+40-8=0$



(d) $48-100+64-13=-1$



(e) $56-120+78-16=-2$



(f) $64-144+108-26=2$

Partition search region into voxels.

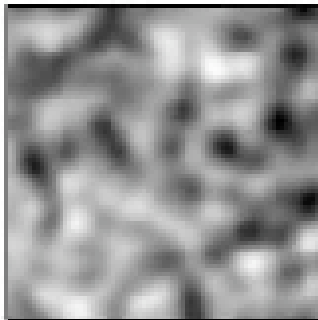
$$EC = \# \text{ volume} - \# \text{ faces} + \# \text{ edges} - \# \text{ vertices}$$

Heuristic Motivation for Adler's formula

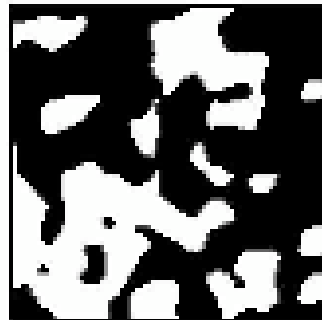
$$P\left(\sup_{t \in \Omega} T(t) > h\right) \approx P(\chi(A_h) > 0) \approx E(\chi(A_h))$$

For very high h value, the Euler characteristic counts the number of clusters, which is one.

Gaussian R.F. Z



Excursion set $Z > 0.52$



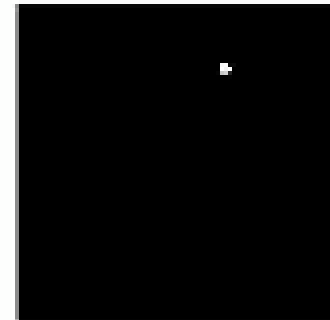
$$\chi_{\text{cl}} = 12 - 3 = 9$$

Excursion set $Z > 1.88$



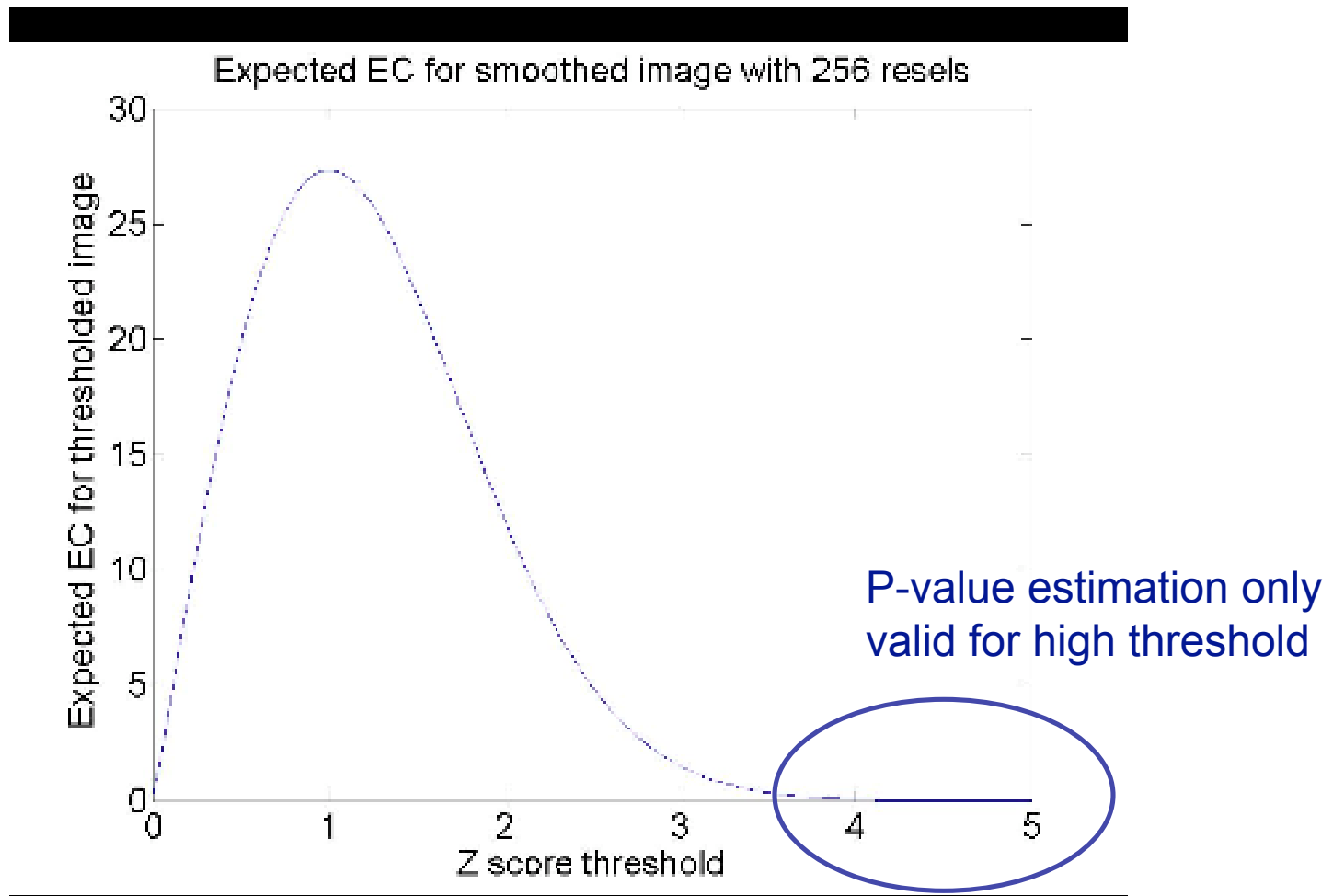
$$\chi_{\text{cl}} = 9 - 0 = 9$$

Excursion set $Z > 2.75$



$$\chi_{\text{cl}} = 1 - 0 = 1$$

Plot of expected EC over threshold



Expected Euler Characteristic

$$\mathbb{E}\chi(A_z) = P\left(Z(x) > z\right) + \sum_{i=1}^N \phi_i(\Omega) \rho_i(z)$$

ρ_i i -dimensional EC density

Smoothness of random field
Hidden in this formula

$$\rho_i(z) = \mathbb{E}\left[(Z - z)^+ \det(-\ddot{Z}_i) | \dot{Z}_i = 0\right] f_{\dot{Z}_i}(0)$$

$\phi_i(\Omega)$ Minkowski functional of Ω

$$\phi_i(\Omega) = \frac{1}{2\pi^{\frac{N-i}{2}} \Gamma(\frac{N-i}{2})} \int_{\partial\Omega} \text{detr}_{N-1-i}(K_{\partial\Omega}) dA$$

$K_{\partial\Omega}$ curvature matrix of $\partial\Omega$

$\text{detr}_i(K_{\partial\Omega})$ sum of $i \times i$ principal minors of $K_{\partial\Omega}$

Expected EC for a stationary Gaussian field

Smoothness of random field.

Proportional to 1/FWHM

$$\begin{aligned}
 E(\text{EC}) &= \frac{\text{Volume} \lambda^3}{(2\pi)^2} (t^2 - 1) e^{-t^2/2} \\
 &+ \frac{(1/2) \text{Area} \lambda^2}{(2\pi)^{3/2}} t e^{-t^2/2} \\
 &+ \frac{2 \text{Diameter} \lambda}{(2\pi)} e^{-t^2/2} \\
 &+ \frac{\text{EC}}{(2\pi)^{1/2}} \int_t^\infty e^{-z^2/2} dz.
 \end{aligned}$$

Unified Resel Approach

(works for irregular search volume, small ROI, surface)

$$\mathbb{E}\chi(A_h) = \sum_{i=1}^N R_i \rho_i(h)$$

RESEL

$$R_i(\Omega) = \frac{V_i(\Omega)}{FWHM^i}$$

EC-density:

differently defined
from one paper to
another: the main
cause of confusion.

Note: Do not interpret it as the number of independent clusters.

$V_i(\Omega)$

Geometric measure of search region (volume, area etc)

T-stat result for 3D convex search volume

$Resels_d(S)$	$EC_d(t)$
1	$\int_t^\infty \frac{\Gamma(\frac{m+1}{2})}{(m\pi)^{1/2}\Gamma(\frac{m}{2})} \left(1 + \frac{u^2}{m}\right)^{-(m+1)/2} du$
$\frac{2 \text{ Diameter}(S)}{FWHM}$	$\frac{(4 \log_e 2)^{1/2}}{2\pi} \left(1 + \frac{t^2}{m}\right)^{-(m-1)/2}$
$\frac{\frac{1}{2} \text{ Surface area}(S)}{FWHM^2}$	$\frac{(4 \log_e 2)}{(2\pi)^{3/2}} \frac{\Gamma(\frac{m+1}{2})}{(\frac{m}{2})^{1/2}\Gamma(\frac{m}{2})} \left(1 + \frac{t^2}{m}\right)^{-(m-1)/2} t$
$\frac{\text{Volume}(S)}{FWHM^3}$	$\frac{(4 \log_e 2)^{3/2}}{(2\pi)^2} \left(1 + \frac{t^2}{m}\right)^{-(m-1)/2} \left(\frac{m-1}{m} t^2 - 1\right)$

These terms are small compared to the 3D Resel term.

1st order approximation

$$\mathbb{E}\chi(A_h) = \sum_{i=1}^N R_i \rho_i(h)$$

The last term dominates the expansion. In practice, it is sufficient to use the last term only.

$$\mathbb{E}\chi(A_h) \approx R_N \rho_N(h)$$

Smoothness in random field theory

Gaussian random field

$$P\left(\sup_{t \in \Omega} Z(t) > h\right) \approx \frac{V_3(\Omega)|\Lambda|^{1/2}}{(2\pi)^2} h^2 \exp\left(-\frac{h^2}{2}\right)$$

Arbitrary random field

$$P\left(\sup_{t \in \Omega} T(t) > h\right) \approx V_3(\Omega)|\Lambda|^{1/2} \dots$$

Determinant of the covariance matrix of derivative of field

Smoothness Estimation

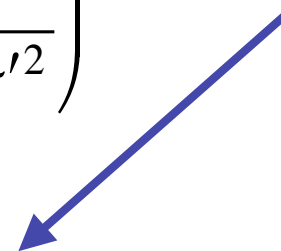
$$\eta = K_{\sigma'} * \epsilon \quad \Lambda = \text{Cov}\left(\frac{d\eta}{dt}\right)$$



For anisotropic kernel, we obtain symmetric positive definite covariance matrix.

$$\Lambda = \begin{pmatrix} \frac{1}{2\sigma'^2} & 0 & \dots & 0 \\ 0 & \frac{1}{2\sigma'^2} & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \dots & 0 & \frac{1}{2\sigma'^2} \end{pmatrix}$$

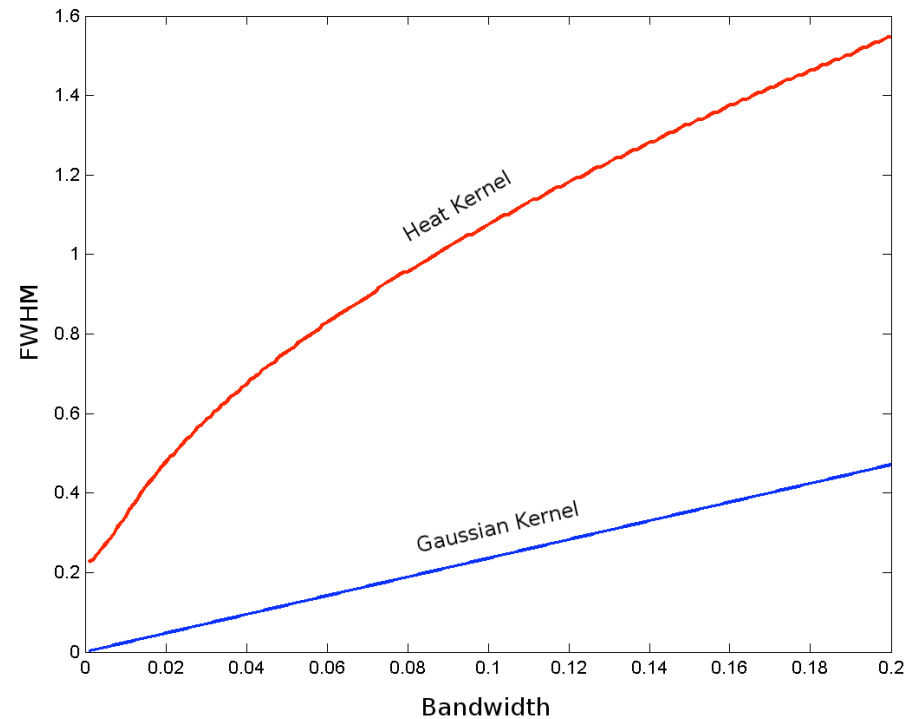
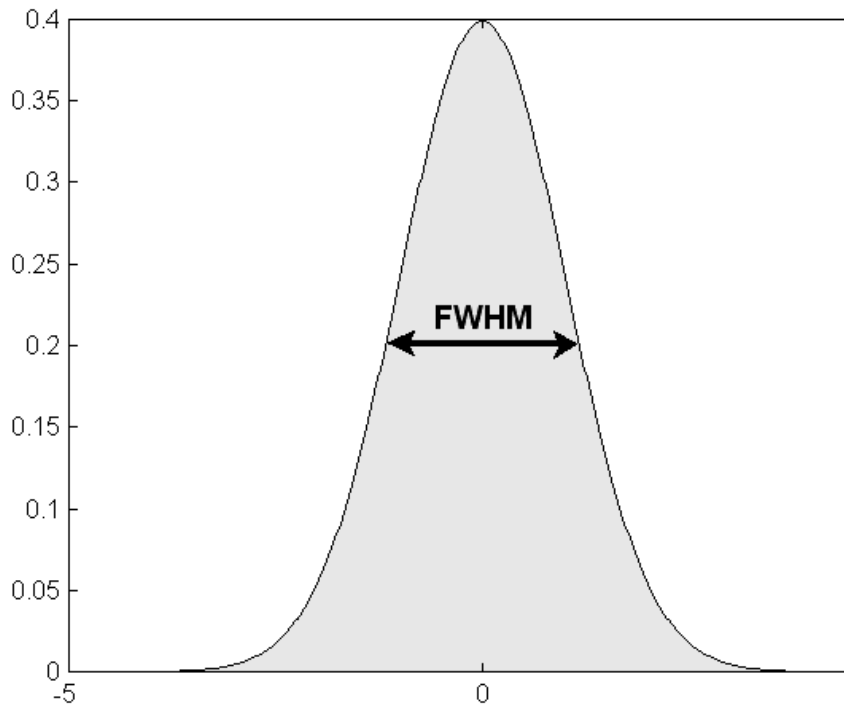
$$|\Lambda|^{1/2} = \frac{1}{2^{N/2} (\sigma')^N}$$



Small $\det(\Lambda)$ --> more smooth

Large $\det(\Lambda)$ --> less smooth

FWHM (full width at half maximum)



Linear relationship between FWHM and bandwidth

Smoothness term

$$FWHM = \sigma' \sqrt{8 \ln 2}$$

$$|\Lambda|^{1/2} = \frac{1}{2^{N/2} (\sigma')^N} = \frac{(4 \log 2)^{N/2}}{FWHM^N}$$

Main issue in estimating smoothness of noise component

- Unbiased estimate of FWHM is important for P-value computation
- Enable us to compare results across different studies properly.

Main issue on FWHM estimation

Linear model at each voxel t : $Y(t) = X\beta(t) + \epsilon(t)$

Gaussian white noise

Spatially varying parameter: $\beta(t) = (\beta_1, \dots, \beta_p)'$

After smoothing we have the following model:

$$K_\sigma * Y(t) = X\beta(t) + K_{\sigma'} * \epsilon(t)$$

Smoothness of kernel and smoothness of residual may not match.

Large kernel bandwidth

- For large kernel smoothing bandwidth, $\sigma = \sigma'$
- No need for estimating FWHM from the residual of a linear model.

Small kernel bandwidth

$$\sigma \neq \sigma'$$

- It provides more accurate P-value computation.
- Any nonstandard smoothing approach surface-based smoothing and anisotropic smoothing require this approach.

$$\sigma \neq \sigma'$$

Smoothness of linear model

1. Estimate the residual at each voxel:

$$R(t) = K_{\sigma} * Y(t) - X\beta(t)$$

2. Compute the partial derivatives using the finite difference

$$\frac{\partial R(t)}{\partial t_i} \approx \frac{R(t + \Delta t_i) - R(t)}{\|\Delta t_i\|}$$

For 3D image, 3 partial derivatives are required.
For 2D image (or surface), 2 partial derivatives.

3. At each voxel, construct the partial derivative vector

$$\frac{\partial R}{\partial t} = \begin{pmatrix} \frac{\partial R(t)}{\partial t_1} \\ \frac{\partial R(t)}{\partial t_2} \\ \frac{\partial R(t)}{\partial t_3} \end{pmatrix}$$

4. Compute the covariance matrix using the sample covariance matrix

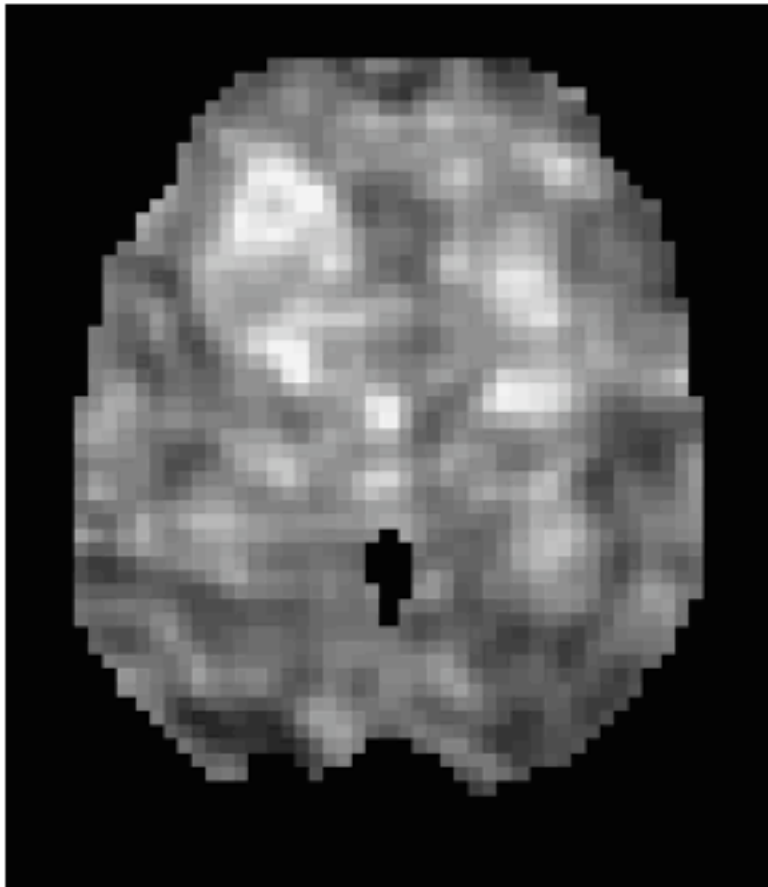
$$\Lambda = \text{cov}\left(\frac{\partial R}{\partial t}\right)$$

5. Compute the determinant
(It will not be uniform in majority of studies.
Weakness of this approach).

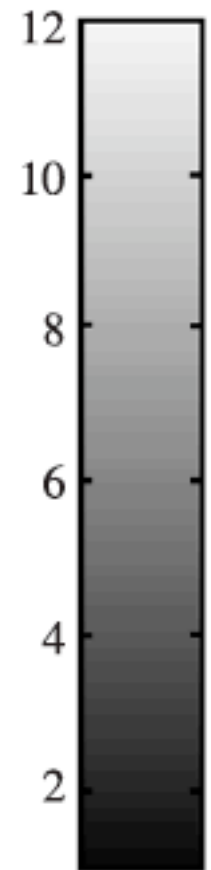
6. Average the determinant across voxel.

Nonuniformity of FWHM

FWHM image
(fMRI)



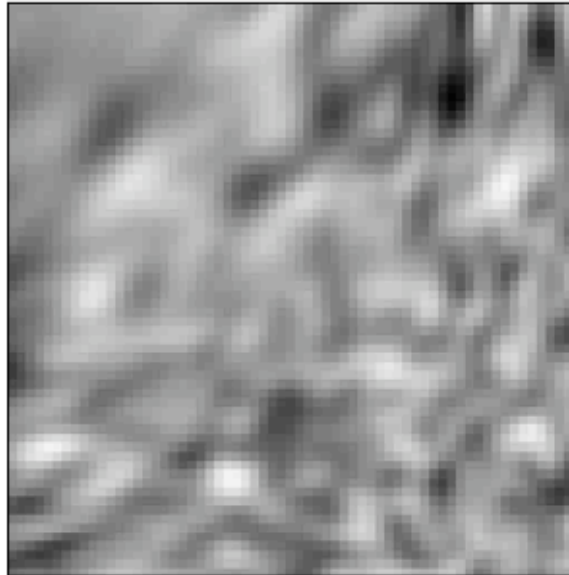
FWHM image
(VBM)



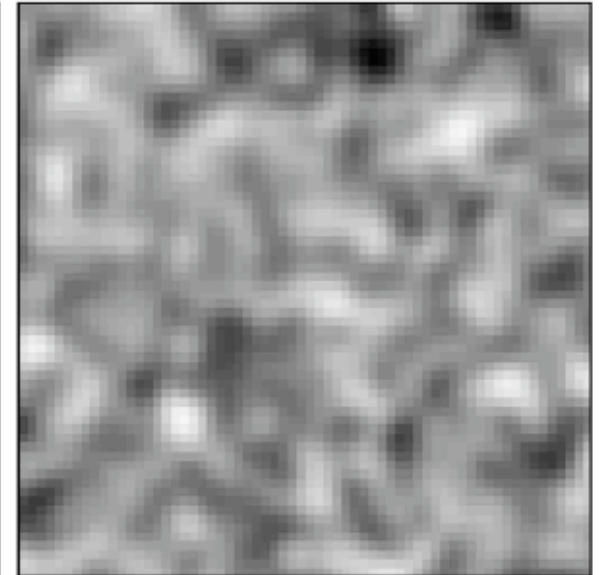
For nonstationary images,
we figure out the
deformation that makes
image stationary and
estimate the smoothness.

See Hayasaka et al.
NeuroImage 22:676-687

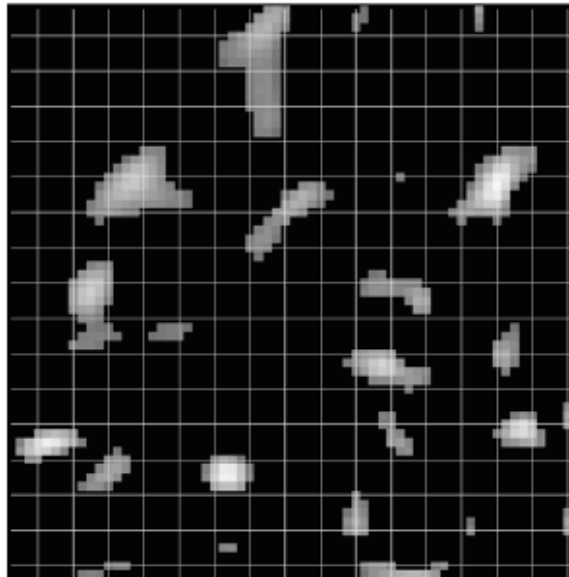
Non-stationary Image



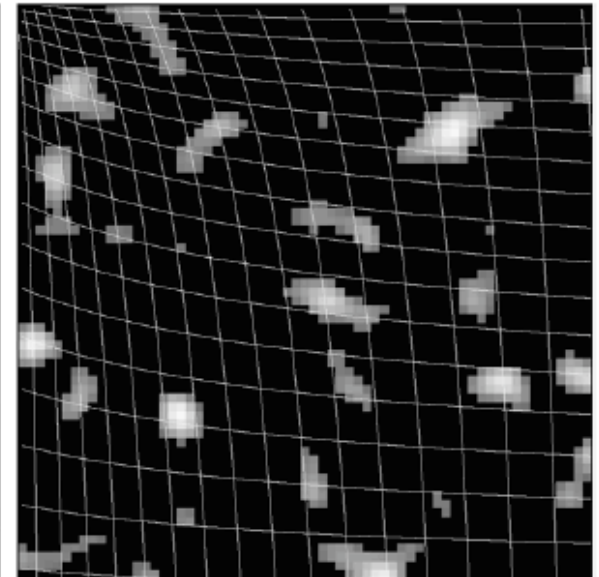
Stationary Image



Clusters, Non-stationary



Clusters, Stationary



Permutation Method

- Simulation based approach to numerically estimate the distribution of the $\sup T(t)$ under the null hypothesis.
- Find units exchangeable under the null hypothesis.
- There may be a situation we may not be able to exchange.

Two sample example

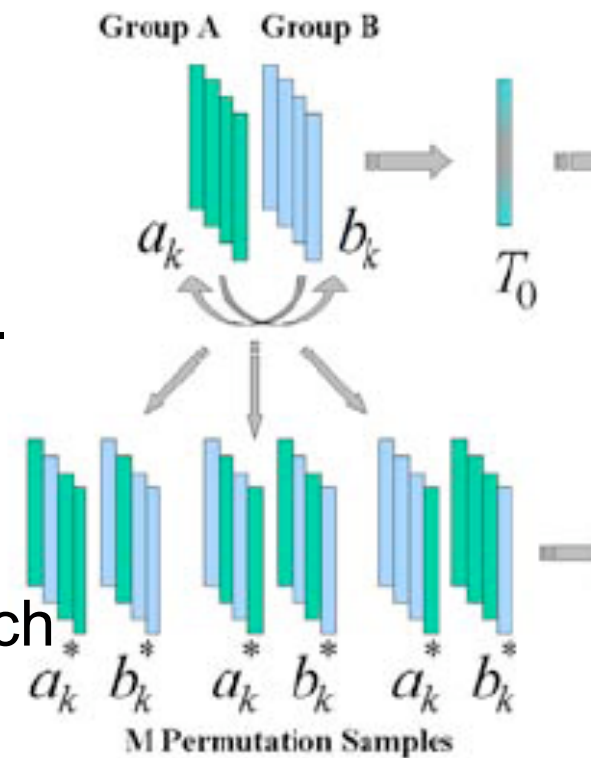
- Group A (n subjects) vs. Group B (m subjects)
Null hypothesis: equal distribution between groups.

- Mix all subjects together.

- Separate the mixed subjects into n subjects and m subjects.

There are $\binom{n+m}{n}$ possible permutations.

- Need to assume the joint distribution of each permutation is identical.



Example

Group A (AAA) Group B (BBB)

$$\# \text{ permutation} = \binom{3+3}{3} = 20$$

1. AAABBB

2. AABABB

3. AABBAB

4. AABBBB

5. ABAABB

6. ABABAB

7. ABABBA

8. ABBAAB

9. ABBABA

10. ABBBAA

11. BAAABB

12. BAABAB

13. BAABBA

14. BABAAB

15. BABABA

16. BABBAA

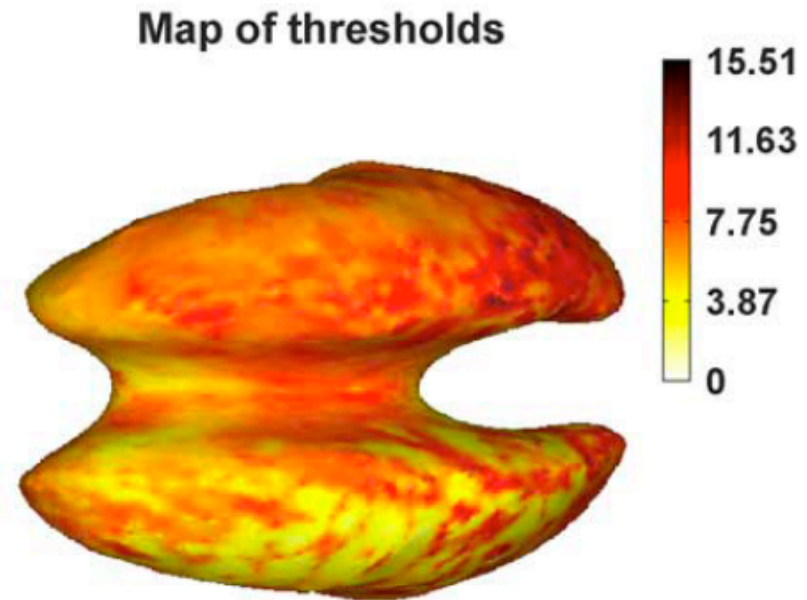
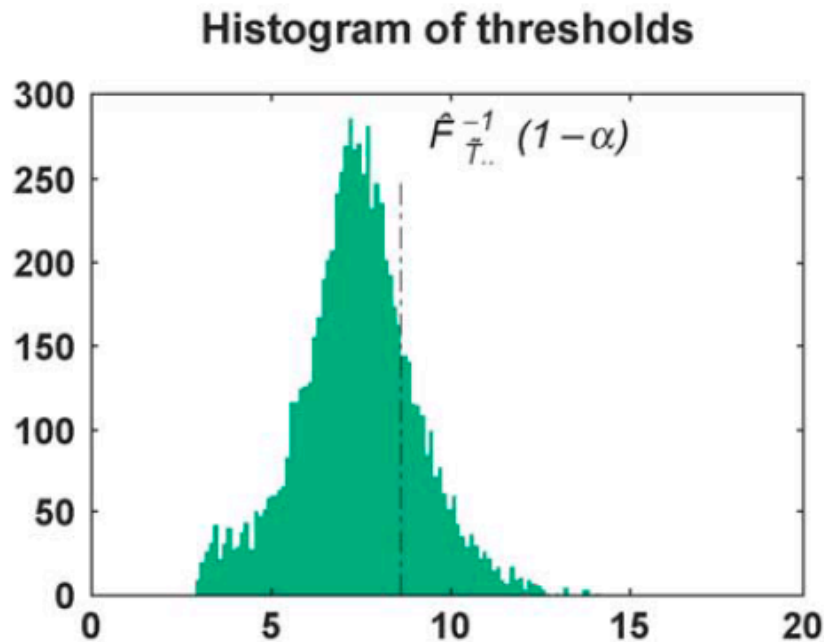
17. BBAAAB

18. BBABAB

19. BBABAA

20. BBBAAA

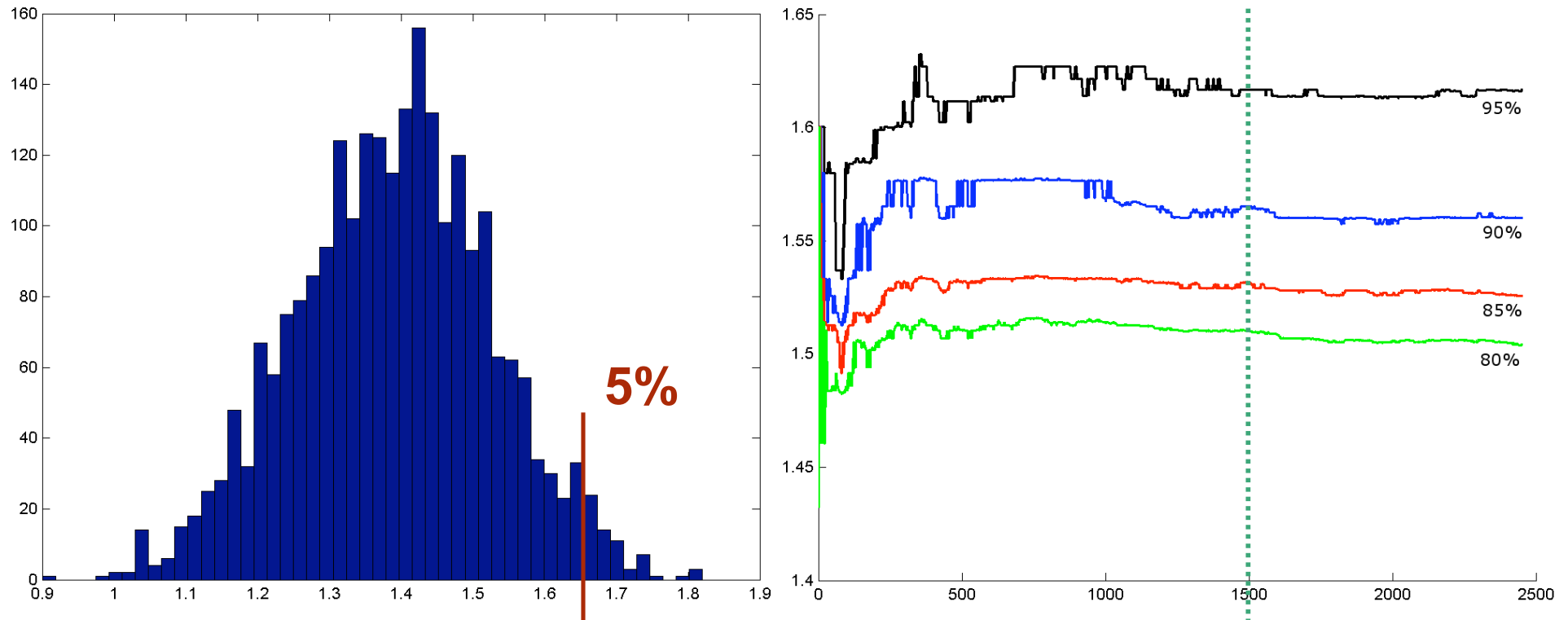
- Compute the $\sup T(t)$ for each permutation.
- Construct histogram of $\sup T(t)$ and threshold at 5% level.



Pantazis et al. NeuroImage 2005.

Permutation test - model free statistical inference

Not recommended for most of studies.



More than 1500 permutations are needed to guarantee the convergence of the thresholding.
8 hours of running time in MATLAB.

Lecture 11 Topics

**Anisotropic Smoothing &
Anisotropic Diffusion**

Lecture 12 Topics

Guest Lecture by Prof. Li Shen