Multiple Testing and Thresholding

NITP, 2009

Thanks for the slides Tom Nichols!
Overview

• Multiple Testing Problem
  – Which of my 100,000 voxels are “active”?

• Two methods for controlling false positives
  – Familywise Error Rate
    • Controlling the chance of any false positives
    • Bonferroni, Random Field and Nonparametric Methods
  – False Discovery Rate
    • Controlling the fraction of false positives
Overview

• Multiple Testing Problem
  – Which of my 100,000 voxels are “active”?

• Two methods for controlling false positives
  – Familywise Error Rate
    • Controlling the chance of any false positives
    • Bonferroni, Random Field and Nonparametric Methods
  – False Discovery Rate
    • Controlling the fraction of false positives
Hypothesis Testing Review

- Establish $H_0$: no activation in voxel $i$
- Establish significance level $\alpha$
  - Derive threshold $u_\alpha$
Hypothesis Testing Review

- Establish $H_0$: no activation in voxel $i$
- Establish significance level $\alpha$
  - Derive threshold $u_\alpha$
- Calculate test statistic $t$
- P-value
  - $P(T > t|H_0)$
- Decision: Reject or accept $H_0$
Hypothesis Testing in fMRI

• Mass Univariate Modeling
  – Fit a separate model for each voxel
  – Look at images of statistics

  – Apply Threshold…
Assessing Statistic Images

• What threshold will show us signal?

High Threshold
\[ t > 5.5 \]
Good Specificity
Poor Power (risk of false negatives)

Med. Threshold
\[ t > 3.5 \]

Low Threshold
\[ t > 0.5 \]
Poor Specificity (risk of false positives)
Good Power
Voxel-level Inference

• Retain voxels above $\alpha$-level threshold $u_\alpha$
• Gives best spatial specificity
  – The null hyp. at a single voxel can be rejected

Statistic values
Voxel-level Inference

• Retain voxels above $\alpha$-level threshold $u_\alpha$
• Gives best spatial specificity
  – The null hyp. at a single voxel can be rejected
Voxel-level Inference

- Retain voxels above $\alpha$-level threshold $u_\alpha$
- Gives best spatial specificity
  - The null hyp. at a single voxel can be rejected
Cluster-level Inference

• Two step-process
  – Define clusters by arbitrary threshold $u_{\text{clus}}$
Cluster-level Inference

• Two step-process
  – Define clusters by arbitrary threshold $u_{\text{clus}}$
  – Retain clusters larger than $\alpha$-level threshold $k_\alpha$
Cluster-level Inference

- Typically better sensitivity
- Worse spatial specificity
  - The null hyp. of entire cluster is rejected
  - Only means that *one or more* of voxels in cluster active

\[ u_{\text{clus}} \]

Cluster not significant

\[ k_{\alpha} \]

Cluster significant
Voxel-wise Inference & Multiple Testing Problem (MTP)

• Standard Hypothesis Test
  – Controls Type I error of each test, at say 5%
  – But what if I have 100,000 voxels?
    • 5,000 false positives on average!

• Must control false positive rate
  – What false positive rate?
  – Chance of 1 or more Type I errors?
  – Proportion of Type I errors?
Overview

• Multiple Testing Problem
  – Which of my 100,000 voxels are “active”?

• Two methods for controlling false positives
  – Familywise Error Rate
    • Controlling the chance of any false positives
    • Bonferroni, Random Field and Nonparametric Methods
  – False Discovery Rate
    • Controlling the fraction of false positives
FWER MTP Solutions

- Bonferroni
- Maximum Distribution Methods
  - Random Field Theory
  - Permutation
Bonferroni

- Based on the Bonferroni inequality
  - For uncorrelated events $E_i$
    - $P(E_1 \text{ or } E_2 \text{ or } \ldots E_n) \leq \sum_{i=1}^{n} P(E_i)$
- If $P(Y_i \text{ passes}|H_0) \leq \alpha/n$ then
  - $P(\text{some } Y_i \text{ passes}|H_0) \leq \sum P(Y_i \text{ passes}|H_0) \leq \alpha$
- For 100,000 voxels
  - $\alpha = 0.05/100,000 = 0.0000005$
Bonferroni

- Can be too conservative
- Bonferroni assumes all tests are independent
- fMRI data tend to be spatially correlated
  - # of independent tests < # voxels
Bonferroni

• Where does spatial correlation come from?
  – How images are constructed from the scanner
  – Physiologic signal
  – Preprocessing steps (realignment, smoothing, etc.)
Why not use a spatial model

- If we can model temporal correlation, why not spatial?
- Need an explicit spatial model
- No routine spatial modeling methods exist
  - High-dimensional mixture modeling problem
  - Activations don’t look like Gaussian blobs
  - Need realistic shapes, sparse representation
    - Some work by Hartvig et al., Penny et al.
FWER MTP Solutions: Controlling FWER w/ Max

- FWER & distribution of maximum
  \[ \text{FWER} = P(\text{FWE}) = P(\text{One or more voxels } \geq u \mid H_0) = P(\text{Max voxel } \geq u \mid H_0) \]

- \(100(1-\alpha)\)%ile of max dist\(^n\) controls FWER
  \[ \text{FWER} = P(\text{Max voxel } \geq u_\alpha \mid H_0) \leq \alpha \]
FWER MTP Solutions: Random Field Theory

- Euler Characteristic $\chi_u$
  - Topological Measure
    - #blobs - #holes
    - At high thresholds, just counts blobs
  - FWER $= P(\text{Max voxel} \geq u \mid H_o)$
    $= P(\text{One or more blobs} \mid H_o)$
    $\approx P(\chi_u \geq 1 \mid H_o)$
    $\approx E(\chi_u \mid H_o)$

No holes
Never more than 1 blob

Suprathreshold Sets
RFT Details: Expected Euler Characteristic

- \( E(\chi_u) \approx V \sqrt{|\Lambda|} (u^2 - 1) \exp(-u^2/2) / (2\pi)^2 \)
  - \( V \rightarrow \) volume
  - \( \sqrt{|\Lambda|} \rightarrow \) roughness, measuring the covariance of the gradient of GRF, \( G \)

\[
\Lambda = \text{Var} \left( \frac{\delta G}{\delta (x, y, z)} \right) = \begin{pmatrix}
\text{Var} \left( \frac{\delta G}{\delta x} \right) & \text{Cov} \left( \frac{\delta G}{\delta x}, \frac{\delta G}{\delta y} \right) & \text{Cov} \left( \frac{\delta G}{\delta x}, \frac{\delta G}{\delta z} \right) \\
\text{Cov} \left( \frac{\delta G}{\delta y}, \frac{\delta G}{\delta x} \right) & \text{Var} \left( \frac{\delta G}{\delta y} \right) & \text{Cov} \left( \frac{\delta G}{\delta y}, \frac{\delta G}{\delta z} \right) \\
\text{Cov} \left( \frac{\delta G}{\delta z}, \frac{\delta G}{\delta x} \right) & \text{Cov} \left( \frac{\delta G}{\delta z}, \frac{\delta G}{\delta y} \right) & \text{Var} \left( \frac{\delta G}{\delta z} \right)
\end{pmatrix}
\]
Random Field Theory
Smoothness Parameterization

- Smoothness parameterized as Full Width at Half Maximum
  - FWHM of Gaussian kernel needed to smooth a white noise random field to roughness $\Lambda$

- Parameterize $\sqrt{|\Lambda|}$ in terms of FWHM

$$\sqrt{|\Lambda|} = \frac{(4\log2)^{1/2}}{FWHM_x \ FWHM_y \ FWHM_z}$$
Random Field Theory
Smoothness Parameterization

- RESELS – Resolution Elements
  - 1 RESEL = FWHM$_x \times$ FWHM$_y \times$ FWHM$_z$
  - RESEL Count $R$
    - $R = \sqrt{\sqrt{|\Lambda|}} \leftarrow \text{The only data-dependent part of } E(\chi_u)$
    - Volume of search region in units of smoothness
    - Eg: 10 voxels, 2.5 voxel FWHM smoothness, 4 RESELS

- RESELS not # of independent ‘things’ in the image
Random Field Intuition

• Corrected P-value for voxel value $t$

$$P^c = P(\max T > t) \approx E(\chi_t) \approx V \sqrt{\Lambda} t^2 \exp(-t^2/2)$$

• Statistic value $t$ increases
  – $P^c$ decreases (of course!)

• Search volume $V$ increases
  – $P^c$ increases (*more* severe MCP)

• Smoothness increases ($\sqrt{\Lambda}$ smaller)
  – $P^c$ decreases (*less* severe MCP)
Nonparametric Permutation Test

- **Parametric methods**
  - Assume distribution of statistic under null hypothesis

- **Nonparametric methods**
  - Use *data* to find distribution of statistic under null hypothesis
  - Any statistic!
Permutation Test
Toy Example

• Data from voxel in visual stim. experiment
  A: Active, flashing checkerboard    B: Baseline, fixation
  6 blocks, ABABAB    Just consider block averages...

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>A</th>
<th>B</th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>103.00</td>
<td>90.48</td>
<td>99.93</td>
<td>87.83</td>
<td>99.76</td>
<td>96.06</td>
</tr>
</tbody>
</table>

• Null hypothesis $H_0$
  – No experimental effect, A & B labels arbitrary

• Statistic
  – Mean difference
Permutation Test
Toy Example

• Under $H_o$
  – Consider all equivalent relabelings

<table>
<thead>
<tr>
<th>AAABBB</th>
<th>ABABAB</th>
<th>BAAABB</th>
<th>BABBA</th>
<th>BABBA</th>
</tr>
</thead>
<tbody>
<tr>
<td>AABABB</td>
<td>ABABBA</td>
<td>BAABAB</td>
<td>BBAAB</td>
<td>BBAAAB</td>
</tr>
<tr>
<td>ABBBAB</td>
<td>ABBAAB</td>
<td>BAABBA</td>
<td>BBAABA</td>
<td>BBABA</td>
</tr>
<tr>
<td>AABBBA</td>
<td>ABBABA</td>
<td>BABABA</td>
<td>BBABAA</td>
<td>BBBAAA</td>
</tr>
<tr>
<td>ABBABB</td>
<td>ABBAAB</td>
<td>BABABA</td>
<td>BBABAA</td>
<td>BBBAAA</td>
</tr>
</tbody>
</table>
Permutation Test
Toy Example

• Under $H_o$
  – Consider all equivalent relabelings
  – Compute all possible statistic values

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>AAABBB</td>
<td>4.82</td>
<td>ABABAB</td>
<td>9.45</td>
<td>BAAABB</td>
</tr>
<tr>
<td>AABABB</td>
<td>-3.25</td>
<td>ABABBA</td>
<td>6.97</td>
<td>BAABAB</td>
</tr>
<tr>
<td>AABBBAB</td>
<td>-0.67</td>
<td>ABBAAB</td>
<td>1.38</td>
<td>BAABBA</td>
</tr>
<tr>
<td>AABBBBA</td>
<td>-3.15</td>
<td>ABBABA</td>
<td>-1.10</td>
<td>BABAAB</td>
</tr>
<tr>
<td>ABAABB</td>
<td>6.86</td>
<td>BBBAAA</td>
<td>1.48</td>
<td>BABABA</td>
</tr>
<tr>
<td>BABBBAA</td>
<td>-6.86</td>
<td>BAAABA</td>
<td>3.15</td>
<td>BBAABA</td>
</tr>
<tr>
<td>BABBBAB</td>
<td>3.25</td>
<td>BBBAAAB</td>
<td>4.82</td>
<td>BBABAA</td>
</tr>
</tbody>
</table>
Permutation Test
Toy Example

• Under $H_o$
  – Consider all equivalent relabelings
  – Compute all possible statistic values
  – Find 95%ile of permutation distribution

<table>
<thead>
<tr>
<th>AABBBB</th>
<th>4.82</th>
<th>ABABAB</th>
<th>9.45</th>
<th>BAAABB</th>
<th>-1.48</th>
<th>BABBA</th>
<th>-6.86</th>
</tr>
</thead>
<tbody>
<tr>
<td>AABABB</td>
<td>-3.25</td>
<td>ABABBA</td>
<td>6.97</td>
<td>BAABAB</td>
<td>1.10</td>
<td>BBAABA</td>
<td>3.15</td>
</tr>
<tr>
<td>ABBBAB</td>
<td>-0.67</td>
<td>ABBAAB</td>
<td>1.38</td>
<td>BAABBA</td>
<td>-1.38</td>
<td>BBAABA</td>
<td>0.67</td>
</tr>
<tr>
<td>AABBBA</td>
<td>-3.15</td>
<td>ABBA</td>
<td>-1.10</td>
<td>BABABA</td>
<td>-6.97</td>
<td>BBABAA</td>
<td>3.25</td>
</tr>
<tr>
<td>ABAABB</td>
<td>6.86</td>
<td>ABBBA</td>
<td>1.48</td>
<td>BABABA</td>
<td>-9.45</td>
<td>BBBAAA</td>
<td>-4.82</td>
</tr>
</tbody>
</table>
Permutation Test
Toy Example

• Under $H_0$
  – Consider all equivalent relabelings
  – Compute all possible statistic values
  – Find 95%ile of permutation distribution

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAABBB</td>
<td>4.82</td>
</tr>
<tr>
<td>AABABB</td>
<td>-3.25</td>
</tr>
<tr>
<td>AABBBAB</td>
<td>-0.67</td>
</tr>
<tr>
<td>AABBBBA</td>
<td>-3.15</td>
</tr>
<tr>
<td>ABAABB</td>
<td>6.86</td>
</tr>
<tr>
<td>ABABABA</td>
<td>9.45</td>
</tr>
<tr>
<td>BABABB</td>
<td>-1.48</td>
</tr>
<tr>
<td>BBAAAB</td>
<td>3.15</td>
</tr>
<tr>
<td>BAABBA</td>
<td>1.10</td>
</tr>
<tr>
<td>ABBAAB</td>
<td>1.38</td>
</tr>
<tr>
<td>BBABBA</td>
<td>-1.38</td>
</tr>
<tr>
<td>AABABB</td>
<td>1.48</td>
</tr>
<tr>
<td>BABBAA</td>
<td>1.48</td>
</tr>
</tbody>
</table>
Permutation Test

Toy Example

• Under $H_0$
  – Consider all equivalent relabelings
  – Compute all possible statistic values
  – Find 95\%ile of permutation distribution
Small Sample Sizes

• Permutation test doesn’t work well with small sample sizes
  – Possible p-values for previous example:
    • 0.05, 0.1, 0.15, 0.2, etc
  – Tends to be conservative for small sample sizes
Controlling FWER: Permutation Test

• Parametric methods
  – Assume distribution of max statistic under null hypothesis

• Nonparametric methods
  – Use data to find distribution of max statistic under null hypothesis
  – Again, any max statistic!
Permutation Test & Exchangeability

• Exchangeability is fundamental
  – Def: Distribution of the data unperturbed by permutation
  – Under H₀, exchangeability justifies permuting data
  – Allows us to build permutation distribution

• Subjects are exchangeable
  – Under Ho, each subject’s A/B labels can be flipped

• fMRI scans are not exchangeable under Ho
  – If no signal, can we permute over time?
  – No, permuting disrupts order, temporal autocorrelation
Permutation Test & Exchangeability

• fMRI scans are not exchangeable
  – Permuting disrupts order, temporal autocorrelation

• Intra-subject fMRI permutation test
  – Must decorrelate data, model before permuting
  – What is correlation structure?
    • Usually must use parametric model of correlation
  – E.g. Use wavelets to decorrelate
    • Bullmore et al 2001, HBM 12:61-78

• Inter-subject fMRI permutation test
  – Create difference image for each subject
  – For each permutation, flip sign of some subjects
Permutation Test
Other Statistics

• Collect max distribution
  – To find threshold that controls FWER

• Consider smoothed variance $t$ statistic
  – To regularize low-df variance estimate
Permutation Test
Smoothed Variance $t$

- Collect max distribution
  - To find threshold that controls FWER
- Consider smoothed variance $t$ statistic
Permutation Test
Smoothed Variance $t$

- Collect max distribution
  - To find threshold that controls FWER
- Consider smoothed variance $t$ statistic
Permutation Test Example

- fMRI Study of Working Memory
  - 12 subjects, block design  Marshuetz et al (2000)
  - Item Recognition
    - Active: View five letters, 2s pause, view probe letter, respond
    - Baseline: View XXXXX, 2s pause, view Y or N, respond

- Second Level RFX
  - Difference image, A-B constructed for each subject
  - One sample, smoothed variance $t$ test
Permutation Test
Example

• Permute!
  – \(2^{12} = 4,096\) ways to flip 12 A/B labels
  – For each, note maximum of \(t\) image
Permutation Test

Example

• Compare with Bonferroni
  – \( \alpha = \frac{0.05}{110,776} \)

• Compare with parametric RFT
  – 110,776 2×2×2mm voxels
  – 5.1×5.8×6.9mm FWHM smoothness
  – 462.9 RESELs
\[ t_{11} \text{ Statistic, Nonparametric Threshold} \]

- \( u_{\text{Perm}} = 7.67 \)
- 58 sig. vox.

\[ t_{11} \text{ Statistic, RF & Bonf. Threshold} \]

- \( u_{\text{RF}} = 9.87 \)
- \( u_{\text{Bonf}} = 9.80 \)
- 5 sig. vox.

- RFT threshold is conservative (not smooth enough, d.f. too small)
- Permutation test is more efficient than Bonferroni since it accounts for smoothness
- Smooth variance is more efficient for small d.f.

\[ \text{Smoothed Variance } t \text{ Statistic, Nonparametric Threshold} \]

- 378 sig. vox.
Does this Generalize?

RFT vs Bonf. vs Perm.

<table>
<thead>
<tr>
<th>Category</th>
<th>df</th>
<th>RF</th>
<th>Bonf</th>
<th>Perm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal Fluency</td>
<td>4</td>
<td>4701.32</td>
<td>42.59</td>
<td>10.14</td>
</tr>
<tr>
<td>Location Switching</td>
<td>9</td>
<td>11.17</td>
<td>9.07</td>
<td>5.83</td>
</tr>
<tr>
<td>Task Switching</td>
<td>9</td>
<td>10.79</td>
<td>10.35</td>
<td>5.10</td>
</tr>
<tr>
<td>Faces: Main Effect</td>
<td>11</td>
<td>10.43</td>
<td>9.07</td>
<td>7.92</td>
</tr>
<tr>
<td>Faces: Interaction</td>
<td>11</td>
<td>10.70</td>
<td>9.07</td>
<td>8.26</td>
</tr>
<tr>
<td>Item Recognition</td>
<td>11</td>
<td>9.87</td>
<td>9.80</td>
<td>7.67</td>
</tr>
<tr>
<td>Visual Motion</td>
<td>11</td>
<td>11.07</td>
<td>8.92</td>
<td>8.40</td>
</tr>
<tr>
<td>Emotional Pictures</td>
<td>12</td>
<td>8.48</td>
<td>8.41</td>
<td>7.15</td>
</tr>
<tr>
<td>Pain: Warning</td>
<td>22</td>
<td>5.93</td>
<td>6.05</td>
<td>4.99</td>
</tr>
<tr>
<td>Pain: Anticipation</td>
<td>22</td>
<td>5.87</td>
<td>6.05</td>
<td>5.05</td>
</tr>
</tbody>
</table>
Does this Generalize?
RFT vs Bonf. vs Perm.

<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>RF</th>
<th>Bonf</th>
<th>Perm</th>
<th>SmVar t Perm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal Fluency</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Location Switching</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>158</td>
<td>354</td>
</tr>
<tr>
<td>Task Switching</td>
<td>9</td>
<td>4</td>
<td>6</td>
<td>2241</td>
<td>3447</td>
</tr>
<tr>
<td>Faces: Main Effect</td>
<td>11</td>
<td>127</td>
<td>371</td>
<td>917</td>
<td>4088</td>
</tr>
<tr>
<td>Faces: Interaction</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Item Recognition</td>
<td>11</td>
<td>5</td>
<td>5</td>
<td>58</td>
<td>378</td>
</tr>
<tr>
<td>Visual Motion</td>
<td>11</td>
<td>626</td>
<td>1260</td>
<td>1480</td>
<td>4064</td>
</tr>
<tr>
<td>Emotional Pictures</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>Pain: Warning</td>
<td>22</td>
<td>127</td>
<td>116</td>
<td>221</td>
<td>347</td>
</tr>
<tr>
<td>Pain: Anticipation</td>
<td>22</td>
<td>74</td>
<td>55</td>
<td>182</td>
<td>402</td>
</tr>
</tbody>
</table>
Overview

• Multiple Testing Problem
  – Which of my 100,000 voxels are “active”?

• Two methods for controlling false positives
  – Familywise Error Rate
    • Controlling the chance of any false positives
    • Bonferroni, Random Field and Nonparametric Methods
  – False Discovery Rate
    • Controlling the fraction of false positives
False Discovery Rate

- For any threshold, all voxels can be cross-classified:
  - **Null True (no effect)**
    - Accept Null ("Negative")
      - $V_{0N}$
    - Reject Null ("Positive")
      - $V_{0P}$
  - **Null False (true effect)**
    - $V_{1N}$
    - $V_{1P}$

  - **False Discovery Proportion**
    - $FDP = \frac{V_{0P}}{V_{P}}$ (FDP=0 if $V_{P}=0$)

- But only can observe $V_{P}$, don’t know $V_{0P}$
  - We control the expected FDP
    - $FDR = E(FDP)$
False Discovery Rate
Illustration:

Noise

Signal

Signal+Noise
Control of Per Comparison Rate at 10%

Percentage of Null Pixels that are False Positives

Control of Familywise Error Rate at 10%

Occurrence of Familywise Error

FWE

Control of False Discovery Rate at 10%

Percentage of Activated Pixels that are False Positives
Benjamini & Hochberg Procedure

- Select desired limit $\alpha$ on FDR
- Order p-values, $p(1) \leq p(2) \leq \ldots \leq p(v)$
- Let $r$ be largest $i$ such that
  \[ p(i) \leq \frac{i}{V} \times \alpha \]
- Reject all hypotheses corresponding to $p(1), \ldots, p(r)$. 

![Diagram showing the Benjamini-Hochberg procedure with a dotted line representing the inequality $p(i) \leq i/V \times \alpha$.]
Adaptiveness of Benjamini & Hochberg FDR

When no signal: P-value threshold $\alpha/v$

When all signal: P-value threshold $\alpha$

...FDR adapts to the amount of signal in the data
Benjamini & Hochberg: Key Properties

• FDR is controlled
  \[ \mathbb{E}(\text{FDP}) \leq \alpha \frac{m_0}{v} \]
  – Conservative, if large fraction of nulls false

• Adaptive
  – Threshold depends on amount of signal
    • More signal, More small p-values,
      More \( p(i) \) less than \( \frac{i}{v} \times \frac{\alpha}{c(v)} \)
FDR Example

FWER Perm. Thresh. = 7.67
58 voxels

FDR Threshold = 3.83
3,073 voxels
Conclusions for voxelwise tests

• Multiple Testing Problem
  – Choose a MTP metric (FDR, FWE)
  – Use a powerful method that controls the metric

• Nonparametric Inference
  – More power for small group FWE inferences

• References
  – Permutation: Nichols & Holmes, HBM, 2001: 1-20
Cluster-based inference

• We use RFT all the time, so it can’t be as bad as the RFT results we just saw
• Use cluster size as the test statistic for RFT
• Permutation tests use cluster size or cluster mass
Cluster-level Inference

- Two step-process
  - Define clusters by arbitrary threshold $u_{\text{clus}}$
Cluster-level Inference

- Typically better sensitivity
- Worse spatial specificity
  - The null hyp. of entire cluster is rejected
  - Only means that one or more of voxels in cluster active
Extent vs Mass

- Cluster extent
  - How many voxels are in cluster
  - Sensitive to spatially extended signals

- Cluster mass
  - Combines signal extent and intensity
  - Can be done with FSL’s randomise and SnPM
  - Generally works better, but RFT-based distribution is difficult
RF vs Perm: cluster mass

Most popular threshold

Hayasaka, et al, NI 2003
Conclusions

• Cluster extent RF test
  – Generally conservative (especially for low smoothness)
  – Only close to 0.05 for high threshold (0.0001) and smooth data
  – In some cases extremely anticonservative
  – Results seem to worsen with larger sample sizes (not sure why)
Conclusions

• Cluster extent permutation test
  – In general works well for smooth data with sufficient DF
  – Generally conservative due to discreteness of the test
What to do?

• Start with fast RFT-based approaches
• If you think you have something use longer permutation-based thresholding
• Also check out new threshold free cluster enhancement (TFCE) option in FSL
  – No need to choose 2 thresholds!