Inference on SPMs: Random Field Theory & Alternatives

Thomas Nichols, Ph.D.
Department of Statistics &
Warwick Manufacturing Group
University of Warwick

Vancouver SPM Course
August 2010
realignment & motion correction → smoothing → General Linear Model
→ model fitting → statistic image → Thresholding & Random Field Theory
image data → design matrix → kernel

normalisation → anatomical reference → Statistical Parametric Map

Corrected thresholds & p-values
Assessing Statistic Images...
Assessing Statistic Images

Where’s the 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

...but why threshold?!
Blue-sky inference: What we’d like

• Don’t threshold, model the signal!
  – Signal location?
    • Estimates and CI’s on (x,y,z) location
  – Signal magnitude?
    • CI’s on % change
  – Spatial extent?
    • Estimates and CI’s on activation volume
    • Robust to choice of cluster definition
• ...but this requires an explicit spatial model
Blue-sky inference: What we need

• 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.*
Real-life inference: What we get

• Signal **location**
  – Local maximum  – *no inference*
  – Center-of-mass  – *no inference*
    • Sensitive to blob-defining-threshold

• Signal **magnitude**
  – Local maximum intensity  – P-values (& CI’s)

• Spatial **extent**
  – Cluster volume  – P-value, no CI’s
    • Sensitive to blob-defining-threshold
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

![Diagram showing significant and non-significant voxels with threshold $u_\alpha$]

Significant Voxels

No significant Voxels

space
Cluster-level Inference

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

Cluster not significant

Cluster significant

$u_{\text{clus}}$

space

$k_\alpha$
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

$u_{clus}$

Cluster not significant

$\alpha$

$\alpha$

Cluster significant
Set-level Inference

- Count number of blobs $c$
  - Minimum blob size $k$
- Worst spatial specificity
  - Only can reject global null hypothesis

Here $c = 1$; only 1 cluster larger than $k$
Multiple comparisons...
Hypothesis Testing

- **Null Hypothesis** $H_0$
- **Test statistic** $T$
  - $t$ observed realization of $T$
- **$\alpha$ level**
  - Acceptable false positive rate
  - Level $\alpha = P( T > u_\alpha | H_0 )$
  - Threshold $u_\alpha$ controls false positive rate at level $\alpha$
- **P-value**
  - Assessment of $t$ assuming $H_0$
  - $P( T > t | H_0 )$
    - Prob. of obtaining stat. as large or larger in a new experiment
  - $P(\text{Data}|\text{Null}) \not= P(\text{Null}|\text{Data})$
Multiple Comparisons Problem

• Which of 100,000 voxels are sig.? 
  – $\alpha=0.05 \Rightarrow 5,000$ false positive voxels

• Which of (random number, say) 100 clusters significant? 
  – $\alpha=0.05 \Rightarrow 5$ false positives clusters
MCP Solutions: Measuring False Positives

- Familywise Error Rate (FWER)
  - Familywise Error
    • Existence of one or more false positives
  - FWER is probability of familywise error
- False Discovery Rate (FDR)
  - FDR = E(V/R)
    • R voxels declared active, V falsely so
      • Realized false discovery rate: V/R
MCP Solutions: Measuring False Positives

• Familywise Error Rate (FWER)
  – Familywise Error
    • Existence of one or more false positives
  – FWER is probability of familywise error

• False Discovery Rate (FDR)
  – FDR = E(V/R)
  – R voxels declared active, V falsely so
    • Realized false discovery rate: V/R
Family of hypotheses

- \( H_k \) \( k \in \Omega = \{1, \ldots, K\} \)
- \( H^\Omega = \bigcap H^k \)

Familywise Type I error

- weak control – omnibus test
  - \( \text{Pr}(\text{reject } H^\Omega \mid H^\Omega) \leq \alpha \)
  - “anything, anywhere”?
- strong control – localising test
  - \( \text{Pr}(\text{reject } H^W \mid H^W) \leq \alpha \)
  - \( \forall W: W \subseteq \Omega \& H^W \)
  - “anything, & where”?

Adjusted \( p \)-values

- test level at which reject \( H^k \)
FWE MCP Solutions: Bonferroni

- For a statistic image $T$...
  - $T_i$ $i^{th}$ voxel of statistic image $T$

- ...use $\alpha = \alpha_0 / V$
  - $\alpha_0$ FWER level (e.g. 0.05)
  - $V$ number of voxels
  - $u_\alpha$ $\alpha$-level statistic threshold, $P(T_i \geq u_\alpha) = \alpha$

- By Bonferroni inequality...

\[
\text{FWER} = P(\text{FWE}) = P( \bigcup_i \{T_i \geq u_\alpha\} \mid H_0) \\
\leq \sum_i P( T_i \geq u_\alpha \mid H_0 ) \\
= \sum_i \alpha \\
= \sum_i \alpha_0 / V = \alpha_0
\]

Conservative under correlation

Independent: $V$ tests
Some dep.: $? \text{ tests}$
Total dep.: 1 test
Random field theory...
SPM approach: Random fields...

- Consider statistic image as lattice representation of a continuous random field
- Use results from continuous random field theory
FWER MCP Solutions: Controlling FWER w/ Max

• FWER & distribution of maximum

\[
\text{FWER} = P(\text{FWE}) \\
= P( \bigcup_i \{T_i \geq u\} \mid H_0) \\
= P( \max_i T_i \geq u \mid H_0)
\]

• 100(1-\(\alpha\))%ile of max dist^n controls FWER

\[
\text{FWER} = P( \max_i T_i \geq u_{\alpha} \mid H_0) = \alpha
\]

– where

\[
u_{\alpha} = F^{-1}_{\max}(1-\alpha)
\]
FWER MCP 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
RFT Details: Expected Euler Characteristic

\[ E(\chi_u) \approx \lambda(\Omega) \, |\Lambda|^{1/2} \left( u^2 - 1 \right) \exp\left( -u^2/2 \right) / (2\pi)^2 \]

- \( \Omega \) → Search region \( \Omega \subset \mathbb{R}^3 \)
- \( \lambda(\Omega) \) → volume
- \( |\Lambda|^{1/2} \) → roughness

• Assumptions
  - Multivariate Normal
  - Stationary*
  - ACF twice differentiable at 0

* Stationary
  - Results valid w/out stationary
  - More accurate when stat. holds

Only very upper tail approximates \( 1 - F_{\max}(u) \)
Random Field Theory
Smoothness Parameterization

- $E(\chi_{ul})$ depends on $|\Lambda|^{1/2}$
  - $\Lambda$ roughness matrix:

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

\[
\Lambda = \text{Var} \left( \frac{\partial G}{\partial (x, y, z)} \right) = \begin{pmatrix}
\text{Var} \left( \frac{\partial G}{\partial x} \right) & \text{Cov} \left( \frac{\partial G}{\partial x}, \frac{\partial G}{\partial y} \right) & \text{Cov} \left( \frac{\partial G}{\partial x}, \frac{\partial G}{\partial z} \right) \\
\text{Cov} \left( \frac{\partial G}{\partial y}, \frac{\partial G}{\partial x} \right) & \text{Var} \left( \frac{\partial G}{\partial y} \right) & \text{Cov} \left( \frac{\partial G}{\partial y}, \frac{\partial G}{\partial z} \right) \\
\text{Cov} \left( \frac{\partial G}{\partial z}, \frac{\partial G}{\partial x} \right) & \text{Cov} \left( \frac{\partial G}{\partial z}, \frac{\partial G}{\partial y} \right) & \text{Var} \left( \frac{\partial G}{\partial z} \right)
\end{pmatrix} = \begin{pmatrix}
\lambda_{xx} & \lambda_{xy} & \lambda_{xz} \\
\lambda_{yx} & \lambda_{yy} & \lambda_{yz} \\
\lambda_{zx} & \lambda_{zy} & \lambda_{zz}
\end{pmatrix}
\]

\[
|\Lambda|^{1/2} = \frac{(4 \log 2)^{3/2}}{\text{FWHM}_x \text{FWHM}_y \text{FWHM}_z}.
\]
Random Field Theory

Smoothness Parameterization

• RESELS
  – Resolution Elements
  – 1 RESEL = FWHM_x × FWHM_y × FWHM_z
  – RESEL Count R
    • \( R = \lambda(\Omega) \sqrt{|\Delta|} = (4\log2)^{3/2} \lambda(\Omega) \ / \ (\text{FWHM}_x \times \text{FWHM}_y \times \text{FWHM}_z) \)
    • Volume of search region in units of smoothness
    • Eg: 10 voxels, 2.5 FWHM 4 RESELS

• Beware RESEL misinterpretation
  – RESEL are not “number of independent ‘things’ in the image”
Random Field Theory
Smoothness Estimation

- Smoothness est’d from standardized residuals
  - Variance of gradients
  - Yields resels per voxel (RPV)

- RPV image
  - Local roughness est.
  - Can transform in to local smoothness est.
    - FWHM Img = (RPV Img)^{-1/D}
    - Dimension $D$, e.g. $D=2$ or 3
Random Field Intuition

• Corrected P-value for voxel value $t$
  
  $P^c = P(\text{max} \ T > t)$
  
  $\approx E(\chi_i)$
  
  $\approx \lambda(\Omega) |\Lambda|^{1/2} t^2 \exp(-t^2/2)$

• Statistic value $t$ increases
  – $P^c$ decreases (but only for large $t$)

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

• Roughness increases (Smoothness decreases)
  – $P^c$ increases (more severe MCP)
RFT Details: Unified Formula

- General form for expected Euler characteristic
  - $\chi^2$, $F$, & $t$ fields • restricted search regions • $D$ dimensions

$$E[\chi_u(\Omega)] = \sum_d R_d(\Omega) \rho_d(u)$$

$R_d(\Omega)$: $d$-dimensional Minkowski functional of $\Omega$
- function of dimension, space $\Omega$ and smoothness:

$R_0(\Omega) = \chi(\Omega)$ Euler characteristic of $\Omega$
$R_1(\Omega) =$ resel diameter
$R_2(\Omega) =$ resel surface area
$R_3(\Omega) =$ resel volume

$\rho_d(\Omega)$: $d$-dimensional EC density of $Z(x)$
- function of dimension and threshold, specific for RF type:

E.g. Gaussian RF:

$$\rho_0(u) = 1 - \Phi(u)$$
$$\rho_1(u) = (4 \ln2)^{1/2} \exp(-u^2/2) / (2\pi)$$
$$\rho_2(u) = (4 \ln2) \exp(-u^2/2) / (2\pi)^{3/2}$$
$$\rho_3(u) = (4 \ln2)^{3/2} (u^2 - 1) \exp(-u^2/2) / (2\pi)^2$$
$$\rho_4(u) = (4 \ln2)^2 (u^3 - 3u) \exp(-u^2/2) / (2\pi)^{5/2}$$
Random Field Theory
Cluster Size Tests

• Expected Cluster Size
  – $E(S) = E(N)/E(L)$
  – S cluster size
  – $N$ suprathreshold volume
    $\lambda(\{T > u_{clus}\})$
  – L number of clusters

• $E(N) = \lambda(\Omega) P( T > u_{clus} )$
• $E(L) \approx E(\chi_{ul})$
  – Assuming no holes
Random Field Theory
Cluster Size Distribution

• Gaussian Random Fields (Nosko, 1969)

\[ S^{2/D} \sim \text{Exp} \left( \frac{E(N)}{\Gamma(D/2+1)E(L)} \right)^{-2/D} \]

– D: Dimension of RF

• t Random Fields (Cao, 1999)

– B: Beta dist
– U’s: \( \chi^2 \)’s
– c chosen s.t.
\[ E(S) = \frac{E(N)}{E(L)} \]

\[ S \sim c B^{1/2} \left[ \frac{U_0^D}{\prod_{b=0}^{D} U_b} \right]^{2/D} \]
Random Field Theory
Cluster Size Corrected P-Values

• Previous results give uncorrected P-value
• Corrected P-value
  – Bonferroni
    • Correct for expected number of clusters
    • Corrected $P^c = E(L) \cdot P_{uncorr}$
  – Poisson Clumping Heuristic (Adler, 1980)
    • Corrected $P^c = 1 - \exp(-E(L) \cdot P_{uncorr})$
Review: Levels of inference & power

Set level...
Cluster level...
Voxel level...
Random Field Theory

Limitations

• Sufficient smoothness
  – FWHM smoothness $3-4 \times$ voxel size ($Z$)
  – More like $\sim 10 \times$ for low-df T images

• Smoothness estimation
  – Estimate is biased when images not sufficiently smooth

• Multivariate normality
  – Virtually impossible to check

• Several layers of approximations

• Stationary required for cluster size results
Real Data

• fMRI Study of Working Memory
  – 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 t test
Real Data: RFT Result

- **Threshold**
  - \( S = 110,776 \)
  - \( 2 \times 2 \times 2 \) voxels
  - \( 5.1 \times 5.8 \times 6.9 \) mm
  - FWHM
  - \( u = 9.870 \)

- **Result**
  - 5 voxels above the threshold
  - 0.0063 minimum FWE-corrected p-value
Real Data: SnPM Promotional

- Nonparametric method more powerful than RFT for low DF
- “Variance Smoothing” even more sensitive
- FWE controlled all the while!
False Discovery Rate...
MCP Solutions: Measuring False Positives

• Familywise Error Rate (FWER)
  – Familywise Error
    • Existence of one or more false positives
  – FWER is probability of familywise error

• False Discovery Rate (FDR)
  – FDR = E(V/R)
    – R voxels declared active, V falsely so
      • Realized false discovery rate: V/R
False Discovery Rate

- For any threshold, all voxels can be cross-classified:

- **Realized FDR**

  \[
  rFDR = \frac{V_{0R}}{V_{1R} + V_{0R}} = \frac{V_{0R}}{N_R}
  \]

  - If \( N_R = 0 \), \( rFDR = 0 \)

- But only can observe \( N_R \), don’t know \( V_{1R} \) & \( V_{0R} \)

  - We control the *expected* \( rFDR \)

  \[
  FDR = E(rFDR)
  \]
False Discovery Rate

Illustration:
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

Control of False Discovery Rate at 10%

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

- Select desired limit $q$ on FDR
- Order p-values, $p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(V)}$
- Let $r$ be largest $i$ such that

\[ p_{(i)} \leq \frac{i}{V} \times q \]

- Reject all hypotheses corresponding to $p_{(1)}, \ldots, p_{(r)}$.

Adaptiveness of Benjamini & Hochberg FDR

Ordered p-values $p(i)$

P-value threshold when no signal: $\alpha/V$

P-value threshold when all signal: $\alpha$
Real Data: FDR Example

• Threshold
  – Indep/PosDep
    \( u = 3.83 \)
  – Arb Cov
    \( u = 13.15 \)

• Result
  – 3,073 voxels above Indep/PosDep \( u \)
  – <0.0001 minimum FDR-corrected p-value

FDR Threshold = 3.83
3,073 voxels
FWER Perm. Thresh. = 9.87
7 voxels
FDR Changes

• Before SPM8
  – Only voxel-wise FDR

• SPM8
  – Cluster-wise FDR
  – Peak-wise FDR

Item Recognition data

Cluster-forming threshold P=0.001
  Cluster-wise FDR: 40 voxel cluster, PFDR 0.07
  Peak-wise FDR: t=4.84, PFDR 0.836

Cluster-forming threshold P=0.01
  Cluster-wise FDR: 1250 - 4380 voxel clusters, PFDR <0.001
  Cluster-wise FDR: 80 voxel cluster, PFDR 0.516
  Peak-wise FDR: t=4.84, PFDR 0.027
Benjamini & Hochberg
Procedure Details

• **Standard Result**
  - Positive Regression Dependency on Subsets
    \[ P(X_1 \geq c_1, X_2 \geq c_2, \ldots, X_k \geq c_k \mid X_i = x_i) \text{ is non-decreasing in } x_i \]
  - Only required of null \( x_i \)'s
    - Positive correlation between null voxels
    - Positive correlation between null and signal voxels
  - Special cases include
    - Independence
    - Multivariate Normal with all positive correlations

• **Arbitrary covariance structure**
  - Replace \( q \) by \( q/c(V) \),
    \[ c(V) = \sum_{i=1,\ldots,V} 1/i \approx \log(V) + 0.5772 \]
    - Much more stringent

*Ann. Stat.*
29:1165-1188
Benjamini & Hochberg: Key Properties

- FDR is controlled
  \[ E(\text{rFDR}) \leq q \, m_0/V \]
  - 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 \( i/V \times q/c(V) \)
Controlling FDR: Varying Signal Extent

\[ p = \quad z = \]

Signal Intensity 3.0  
Signal Extent 1.0  
Noise Smoothness 3.0
Controlling FDR: Varying Signal Extent

\[ p = \quad z = \]

Signal Intensity 3.0  Signal Extent 2.0  Noise Smoothness 3.0
Controlling FDR: Varying Signal Extent

\[ p = \quad z = \]

Signal Intensity 3.0  Signal Extent 3.0  Noise Smoothness 3.0
Controlling FDR: Varying Signal Extent

\[ p = 0.000252 \quad z = 3.48 \]
Controlling FDR: Varying Signal Extent

\[ p = 0.001628 \quad z = 2.94 \]

Signal Intensity 3.0  Signal Extent 9.5  Noise Smoothness 3.0
Controlling FDR: Varying Signal Extent

\[ p = 0.007157 \quad z = 2.45 \]
Controlling FDR: Varying SignalExtent

\[ p = 0.019274 \quad z = 2.07 \]
Controlling FDR: Benjamini & Hochberg

- Illustrating BH under dependence
  - Extreme example of positive dependence

8 voxel image

32 voxel image (interpolated from 8 voxel image)
Conclusions

• Must account for multiplicity
  – Otherwise have a fishing expedition

• FWER
  – Very specific, not very sensitive

• FDR
  – Less specific, more sensitive
  – Sociological calibration still underway
References

• Most of this talk covered in these papers

