Acknowledgement / Disclaimer

Many of the slides in this lecture have been adapted from slides available in talks available on the SPM web site.
Overview

• Motivation
• Linear model formulation
• Region of interest analyses
• Pixel/voxel based analyses
• Multiple comparisons for images
• Bayesian image analysis methods
Motivation

- Imaging data – statistical methods to look for “regional effects”
- Tissue differences between groups or over time – VBM, TBM (voxel/tensor-based morphometry)
- PET (positron emission tomography), fMRI (functional MRI) – determine “activation” in the brain due to thought, stimulus or task
- Diffusion (DWI, DTI, tractography), Bone mineral density etc. etc.
FMRI Data:
Set of Volumes (over time) or
Set of Time-Series (over space)
Software etc.

SPM – PET, fMRI, VBM and TBM, EEG/MEG (http://www.fil.ion.ucl.uk/spm/ needs Matlab)

FSL – fMRI primarily + DTI (http://www.fmrib.ox.ac.uk/fsl/)

R – AnalyzeFMRI package + linear models in general (http://www.r-project.org/ and then go to your nearest CRAN mirror)
Also, check “Venables and Ripley” Splus book + many R books (see R web site) + online tutorials
Challenges

- Generating suitable (statistical) imaging models
- Dealing with highly multivariate responses (curse of dimensionality)
- Defining imaging "hypotheses"
- Creating computationally efficient analysis procedures
Aims of Statistical Modeling

• Summarize data
• Estimation: point and interval estimates
• Inference: hypotheses / relationships
• Prediction
Aims of Statistical Modeling

- Summarize data
- Estimation: point and interval estimates
- Inference: hypotheses / relationships
- Prediction
Statistical Modeling Strategy

• Propose a model for the data
• Fit the model
• Assess the model’s adequacy
• Fit other plausible models
• Compare all fitted models
• Interpret the best model
Statistical Models: Definitions

• Univariate response variable $y_i$ (for exp. unit $i$)

• Covariates $(x_{i1}, x_{i2}, ..., x_{ik}) = \mathbf{x}_i^T$
  (variables of interest and “nuisance” variables)

• Data is: $\left\{y_i, \mathbf{x}_i^T; i = 1, ..., n\right\}$, $n$ experimental units

*Continuous covariates*: e.g. age, blood pressure etc., (random or controlled)

*Factors*: e.g. diagnosis, gender, drinking level (low, medium, high) etc.
The (General) Linear Model

A simple \textit{linear model} might take the form:

\[ y_i = \beta_1 + x_{i2}\beta_2 + x_{i3}\beta_3 + \ldots + x_{im}\beta_m + \epsilon_i \]

\text{e.g.}

\[ y_i = \beta_{\text{mean}} + x_{i,\text{age}}\beta_{\text{age}} + x_{i,\text{gender}}\beta_{\text{gender}} + \ldots + x_{i,\text{diagnosis}}\beta_{\text{diagnosis}} + \epsilon_i \]

\[ \epsilon_i \sim N(0, \sigma^2), \quad i.i.d. \quad i = 1, \ldots, n \]

\textit{i.i.d.} = \text{independently and identically distributed}
The (General) Linear Model

For univariate data:
\[ y_i = x_i^T \beta + \varepsilon_i, \quad i = 1, \ldots, n \]
\[ \beta = (\beta_1, \ldots, \beta_m)^T \] is a set of unknown parameters

or in matrix notation
\[ y = X^T \beta + \varepsilon \]

This can be extended to a multivariate response
\[ Y = X^T B + E \]
Ex. Hippocampal Volume

HCV ~ Age + Diagnosis

(Wilkinson notation)

Diagnosis can be normal control (NC) or Alzheimer’s disease (AD)
Ex. Hippocampal Volume

HCV ~ Age + Diagnosis + Age*Diagnosis

(Wilkinson notation)

Diagnosis can be normal control (NC) or Alzheimer’s disease (AD)
Structural T1 weighted MRI’s

Hippocampal volumes manually traced

Volume measure = response for each subject

Disease status encoded 1 for AD and 0 for NC

(the \( x_{\text{diag.}} \) term)
\[ y_i = \beta_1 + x_{i,\text{age}} \beta_{\text{age}} + x_{i,\text{diag.}} \beta_{\text{diag.}} + x_{i,\text{age}} x_{i,\text{diag.}} \beta_{\text{inter}} + \epsilon_i \]

**Case 1**

\[ \beta_{\text{age}} = 0, \beta_{\text{diag.}} = 0, \beta_{\text{inter}} = 0 \]
\[ y_i = \beta_1 + x_{i,\text{age}} \beta_{\text{age}} + x_{i,\text{diag.}} \beta_{\text{diag.}} + x_{i,\text{age}} x_{i,\text{diag.}} \beta_{\text{inter}} + \varepsilon_i \]

**Case 2**

\[ \beta_{\text{age}} \neq 0, \beta_{\text{diag.}} = 0, \beta_{\text{inter}} = 0 \]
\[ y_i = \beta_0 + x_{i,\text{age}} \beta_{\text{age}} + x_{i,\text{diag.}} \beta_{\text{diag.}} + x_{i,\text{age}} x_{i,\text{diag.}} \beta_{\text{inter}} + \epsilon_i \]

Case 3

\[ \beta_{\text{age}} \neq 0, \beta_{\text{diag.}} \neq 0, \beta_{\text{inter}} = 0 \]
\[ y_i = \beta_0 + x_{i,\text{age}} \beta_{\text{age}} + x_{i,\text{diag.}} \beta_{\text{diag.}} + x_{i,\text{age}} x_{i,\text{diag.}} \beta_{\text{inter}} + \epsilon_i \]
\[ y_i = \beta_1 + x_{i,\text{age}} \beta_{\text{age}} + x_{i,\text{diag.}} \beta_{\text{diag.}} + x_{i,\text{age}} x_{i,\text{diag.}} \beta_{\text{inter}} + \varepsilon_i \]

Case 4

\[ \beta_{\text{age}} \neq 0, \beta_{\text{diag.}} \neq 0, \beta_{\text{inter}} \neq 0 \]
Linear models can be more general - only needs to be linear in the parameters: $\beta$

We can have:

$$y_i = x_{\text{age}} \beta_1 + x_{\text{age}}^2 \beta_2 + \exp(x_{\text{height}}) \beta_3 + x_{\text{age}}^{\pi} x_{\text{height}} \beta_4 + \epsilon_i$$

$$i = 1, \ldots, n$$

But not

$$y_i = x_{\text{age}} \exp(\beta_1 + x_{\text{height}} \beta_2)$$

$$i = 1, \ldots, n$$
Estimation

Minimize squared error (Least Squares Error) = Maximum Likelihood Estimation for linear model

\[ \hat{\beta} = (X^TX)^{-1}X^Ty \]
\[ E(\hat{\beta}) = \beta \]
\[ V(\hat{\beta}) = \sigma^2 (X^TX)^{-1} \]

Estimate \( \sigma^2 \) by

\[ \hat{\sigma}^2 = \frac{\text{sum of squares error}}{n} \]

or divide by \( n-1 \) for unbiased estimate
Inference – Model Comparison

Take linear model

\[ y = X^T \beta + \epsilon \]

And add constraint \[ A\beta = c \]

this defines a new model that is a simplification of the previous one
E.g., cf. model $y_i = \beta_1 + \beta_2 x_{i1} + \beta_3 x_{i2} + \epsilon_i$
to simplification with $\beta_3 = 0$

i.e. $y_i = \beta_1 + \beta_2 x_i + \epsilon_i$

$(0,0,1)\begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = 0$

i.e. $A\beta = c$
What about $\beta_2 = 0$ & $\beta_3 = 0$?

$$A\beta = c \quad \Rightarrow \quad \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$
And what about $\beta_2 = \beta_3$?

$$A\beta = c \implies \begin{pmatrix} 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = 0$$

Are 2 different conditions equivalent? E.g. is the activation effect: reading a word vs imagining the object equal?
Definition: Linear model nested in another if 1\textsuperscript{st} model can be obtained by linear constraint on the 2\textsuperscript{nd}

Nesting tree:
F-test for General Linear Hypothesis

\[ y = X^T \beta + \varepsilon \quad \varepsilon \sim N_n \left( 0, \sigma^2 I_n \right) \]

Consider

\[ H_0 : A\beta = c \]

This is the General Linear Hypothesis
Under $H_0$, i.e., $A\beta = c$

$$F = \frac{(SSE_{\text{nested}} - SSE_{\text{larger}}) / (p_{\text{larger}} - p_{\text{nested}})}{(SSE_{\text{larger}}) / (n - p_{\text{larger}})} \sim F^{p_{\text{larger}} - p_{\text{nested}}, n - p_{\text{larger}}}_{}$$

$p$ denotes the number of model parameters
$n$ denotes the number of data points
$SSE = \text{Deviance} = \text{sum of squares of residuals}$

Tests whether ratio of variances = 1
FMRI Data:
Set of Volumes (over time) or
Set of Time-Series (over space)
realignment & motion correction

normalisation

smoothing

Linear Model
- model fitting
- statistic image

Thresholding & Random Field Theory

Statistical Parametric Map (test statistics)

Corrected thresholds & p-values
Estimation

The estimation entails finding the parameter values such that the linear combination best fits the data.
Parameter Estimates

- Same model for all voxels
- Different parameters for each voxel

\[
\hat{\beta} = \begin{bmatrix}
0.83 \\
0.16 \\
2.98 \\
\end{bmatrix}
\]

\[
\hat{\beta} = \begin{bmatrix}
0.03 \\
0.06 \\
2.04 \\
\end{bmatrix}
\]

\[
\hat{\beta} = \begin{bmatrix}
0.68 \\
0.82 \\
2.17 \\
\end{bmatrix}
\]
We trust: Long series with large effects and small error
Spatial Modeling
Spatial Hypotheses

Question - how do we extend from standard univariate hypotheses to answering spatially motivated questions?

Not easy - curse of dimensionality (millions of voxels)

e.g. in 1D it makes sense to infer A is less than B, but what is the equivalent in 2D?
Spatial Testing Solutions

• Summarize the image into one dimensional quantities for testing (e.g. region of interest analysis)

• Consider the overall test as a combination of individual voxel tests (voxel based analysis)

• Perform shape/object analysis on objects defined via landmarks

• Build Bayesian image analysis models
Spatial Testing Solutions

• Summarize the image into one dimensional quantities for testing (e.g. region of interest analysis)

• Consider the overall test as a combination of individual voxel tests (voxel based analysis)

• Perform shape/object analysis on objects defined via landmarks

• Build Bayesian image analysis models
Voxel based analysis

Each voxel obtains a test statistic from the linear model, e.g. $t$ or $F$

Forms statistical maps of the statistics
realignment & motion correction → normalisation

smoothing

Linear Model
→ model fitting
→ statistic image

Thresholding & Random Field Theory

parameter estimates

Statistical Parametric Map (test statistics)

Corrected thresholds & p-values

image data

design matrix

kernel

anatomical reference

parameter estimates
Hypothesis Testing

- **Null Hypothesis** $H_0$
- **Test statistic** $T$
  - $t$ observed realization of $T$
- **$\alpha$-level**
  - Acceptable false positive risk
  - Level $\alpha = \Pr(T > u_\alpha \mid H_0)$
  - Threshold $u_\alpha$ controls false positive risk at level $\alpha$
Multiple Comparisons Problem

Which of 100,000 voxels are significant?

- $\alpha = 0.05 \Rightarrow 5,000$ false positive voxels
Assessing Statistic Images

Where’s the signal or change?

High Threshold
Good Specificity
Poor Power (risk of false negatives)

Med. Threshold

Low Threshold
Poor Specificity (risk of false positives)
Good Power

How can we determine a sensible threshold level?
Multiple Comparison Solutions: Measuring False Positives

• Familywise Error Rate (FWER)
  – Familywise Error
    • Existence of one or more false positives

• False Discovery Rate (FDR)
  – $\text{FDR} = \mathbb{E}[\frac{\text{FP}}{\text{TP}+\text{FP}}]$ 
  – $\text{TP}+\text{FP}$ voxels declared active, $\text{FP}$ falsely so

Realized false discovery rate: $\frac{\text{FP}}{\text{TP}+\text{FP}}$
Bonferroni Correction

FWE, $\alpha$, for $N$ independent voxels is approximately $\alpha = Nv$ ($v =$ voxel-wise error rate)

To control FWE set $\nu = \alpha / N$

Bonferroni is too conservative for brain images
FWER MCP Solutions: Random Field Theory

- **Euler Characteristic** $\chi_u$
  - Topological Measure
    - $\#\text{blobs} - \#\text{holes}$
  - At high thresholds, just counts blobs

- **FWER** = $\Pr(\text{Max voxel} \geq u \mid H_0)$
  - No holes
    - $= \Pr(\text{One or more blobs} \mid H_0)$
    - $\approx \Pr(\chi_u \geq 1 \mid H_0)$
    - $\approx \mathbb{E}(\chi_u \mid H_0)$
  - Never more than 1 blob

See description at [http://imaging.mrc-cbu.cam.ac.uk/imaging/PrinciplesRandomFields](http://imaging.mrc-cbu.cam.ac.uk/imaging/PrinciplesRandomFields)
Random Field Theory
Limitations

- Multivariate normality (Gaussianity)
  - Virtually impossible to check
- Sufficient smoothness
  - FWHM smoothness $3-4 \times$ voxel size
- Smoothness estimation
  - Estimate is biased when images not sufficiently smooth
- Several layers of approximations
Multiple Comparison Solutions: Measuring False Positives

- **Familywise Error Rate (FWER)**
  - Familywise Error
    - Existence of one or more false positives

- **False Discovery Rate (FDR)**
  - \( \text{FDR} = \mathbb{E}[\frac{\text{FP}}{\text{TP} + \text{FP}}] \)
  - \( \text{TP} + \text{FP} \) voxels declared active, \( \text{FP} \) falsely so
  - Realized false discovery rate: \( \frac{\text{FP}}{\text{TP} + \text{FP}} \)
False Discovery Rate

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

<table>
<thead>
<tr>
<th></th>
<th>Accept Null</th>
<th>Reject Null</th>
</tr>
</thead>
<tbody>
<tr>
<td>Null True</td>
<td>TN</td>
<td>FP</td>
</tr>
<tr>
<td>Null False</td>
<td>FN</td>
<td>TP</td>
</tr>
<tr>
<td>$N_A$</td>
<td></td>
<td>$N_R$</td>
</tr>
</tbody>
</table>

• Realized FDR

$$rFDR = \frac{FP}{(TP+FP)} = \frac{FP}{N_R}$$

– Special case: if $N_R = 0$, $rFDR = 0$

• But only can observe $N_R$, don’t know TP & FP

– We therefore control the expected $rFDR$
False Discovery Rate

Illustration:

Noise

Signal

Signal+Noise
Control of Per Comparison Rate at 10%

Percentage of Null Pixels that are False Positives

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

Control of Familywise Error Rate at 10%

Occurrence of Familywise Error

FWE

Control of False Discovery Rate at 10%

Percentage of Observed “Above Threshold” Pixels that are False Positives

6.7% 10.4% 14.9% 9.3% 16.2% 13.8% 14.0% 10.5% 12.2% 8.7%
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 $p_{(i)} \leq i/V \times q/c(V)$
• Reject all hypotheses corresponding to $p_{(1)}, \ldots, p_{(r)}$

NB, no spatial consideration
Also, Non-Parametric Testing

- If $H_0$ is true then time order irrelevant (if noise really white)
- Therefore permute the timepoints and obtain test statistics
- If true test statistic is extreme compared to others then reject $H_0$
Types of Spatial Inference

- Individual voxel level
- Cluster level
- Set level
- Bayesian model based
Voxel-level Inference

- Retain voxels above $\alpha$-level threshold $u_\alpha$
- Gives best spatial specificity
  - $H_0$ at a single voxel can be rejected

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

![Cluster-level Inference Diagram](image)
Set-level Inference

- Count number of blobs $c$
  - uses minimum blob size $k$ to count
  - significant activity if number of blobs $> n(k, u_{clus})$

- Worst spatial specificity

Here $c = 1$; only 1 cluster larger than $k$
A flexible Bayesian Approach

- Model the form of activity
- Provides an “adaptive thresholding” approach

Active voxels
Bayesian Model

\[ y = z x + \varepsilon \]

- \( y \) = data, parameter estimates of statistics
- \( z \) = binary activation map – modeled as a MRF
- \( x \) = activation level field – modeled as a MRF
- \( \varepsilon \) = residual error

**MRF** = Markov Random Field (similar random field but defined on a lattice)
Other Topics and Omissions

- Hemodynamic response function
- Multiple subjects (random and mixed effects models)
- PCA, ICA
- Multivariate analysis with variogram modeling
- Space-time modeling
Plug for: February 29th SFASA Seminar

Speaker: Yoav Benjamini, PhD., Professor of Statistics
Time: Wednesday, 5pm - 6pm (4:30-5pm pre-seminar social)
Location: UCSF China Basin Landing, Room 6702 (specific directions to classrooms: http://www.epibiostat.ucsf.edu/general/cbl.html)

Title: Hierarchical Testing of Families of Hypotheses

Abstract: As the size of large testing problems encountered in genomic research keeps increasing, more of these problems have further structure where the set of hypotheses can be partitioned into families of the hypotheses, and the true state of the tested signals tends to be more similar within these subsets than across the subsets. Moreover, interest may lie with a discovery of a family with some signal in it, on top of the discovery of a signal in each of the many hypotheses on its own. The challenges in the analysis of such multiple testing problems will be discussed. We then present the concept of the control on the average over the selected families of a desired error-rate, be it the family-wise error rate, the False Discovery Rate, or their generalizations. We discuss the various considerations involved using the genomic part of a Norwegian epidemiological study of breast cancer, and a study involving genomics and brain imaging.

Transportation/Parking: 1) Right at the Caltrain station. 2) If driving, 2-hour parking in the neighborhood as well as reasonably priced parking in lot A for the Giants. The building parking is expensive. 3) If by BART, transfer at Embarcadero to the N-Judah or T-train (both inbound)