Advances in False Discovery Rate control applied in Neuroimaging

Glenn Lawyer*, Egil Ferkingstad

Psykiatrisk institutt, Vinderen
University of Oslo

May 16, 2007

Outline

Outline

Voxel-based morphometry

1995: running a statistical test at every voxel in an fMRI image.

Voxel-Based Morphometry (VBM).

A typical fMRI image has a resolution of 2-7 mm^2,
i.e a 64x64x7 image. 30000+ simultaneous tests.

Voxel-based morphometry

1995: running a statistical test at every voxel in an fMRI image.

Voxel-Based Morphometry (VBM).

A typical fMRI image has a resolution of 2-7 mm^2,
i.e a 64x64x7 image. 30000+ simultaneous tests.

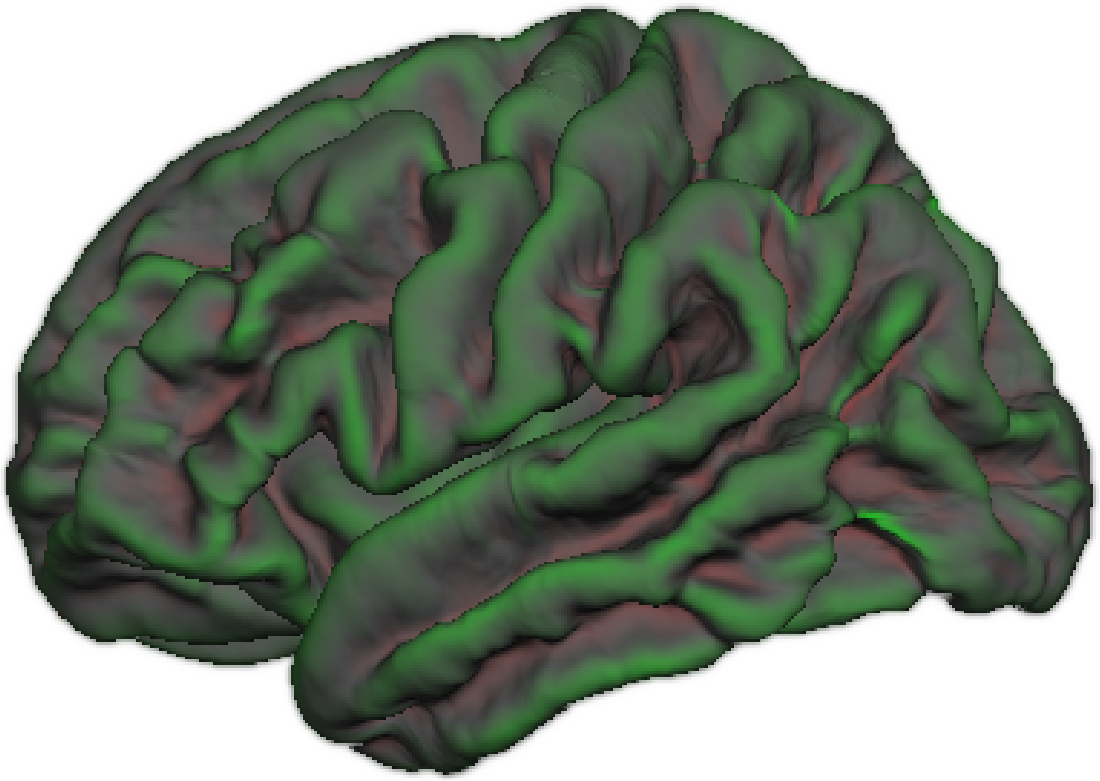
Vertex-based morphometry

Cognition takes place, to a great extent, in the cortex.

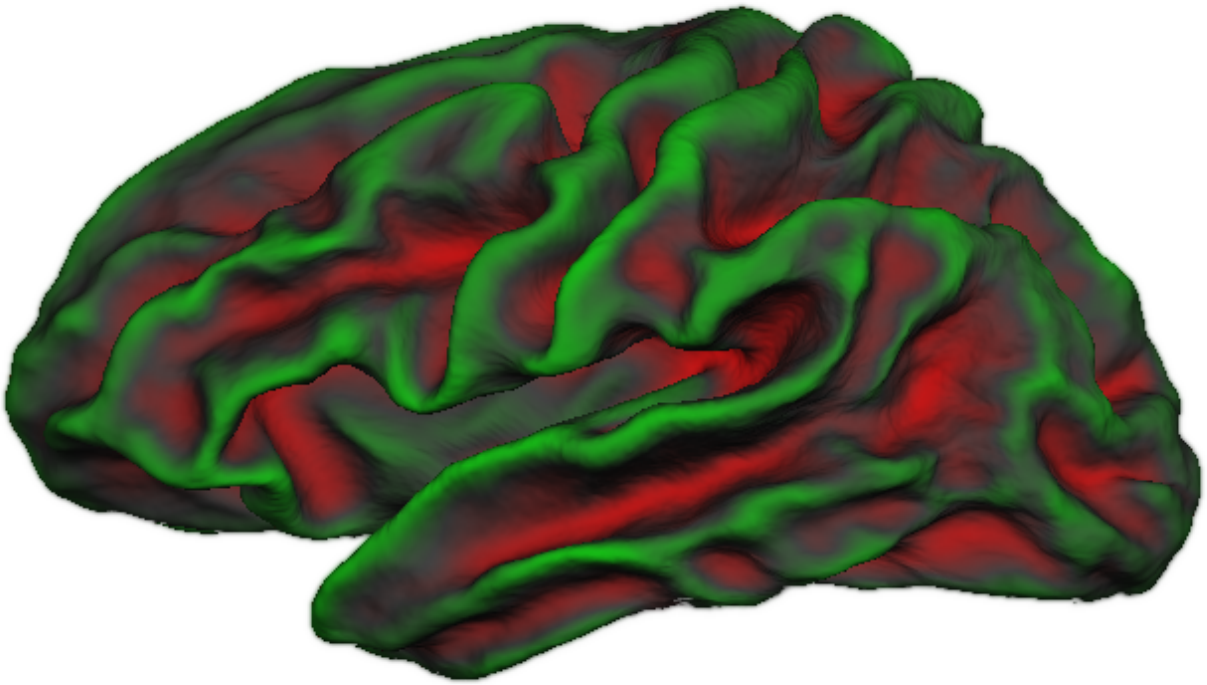
Convoluted sheet of grey matter – the outer layer of the brain.

approx yr. 2000: Measure the thickness of the cortex.

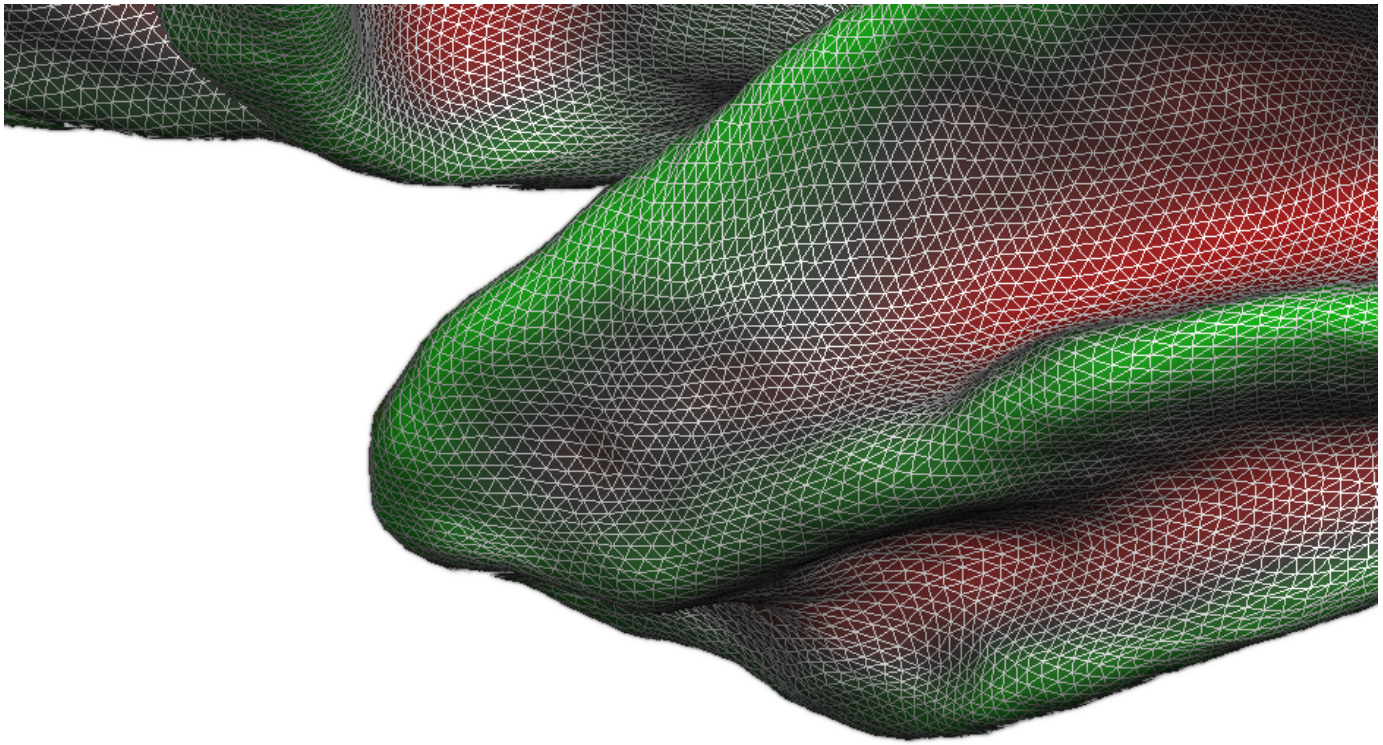
Pial (outside) surface



White matter surface



Wireframe



Multiple comparisons

Test hypothesis at each vertex in the mesh.

Vertex-Based Morphology (VBM).

test over 300,000 hypothesis.

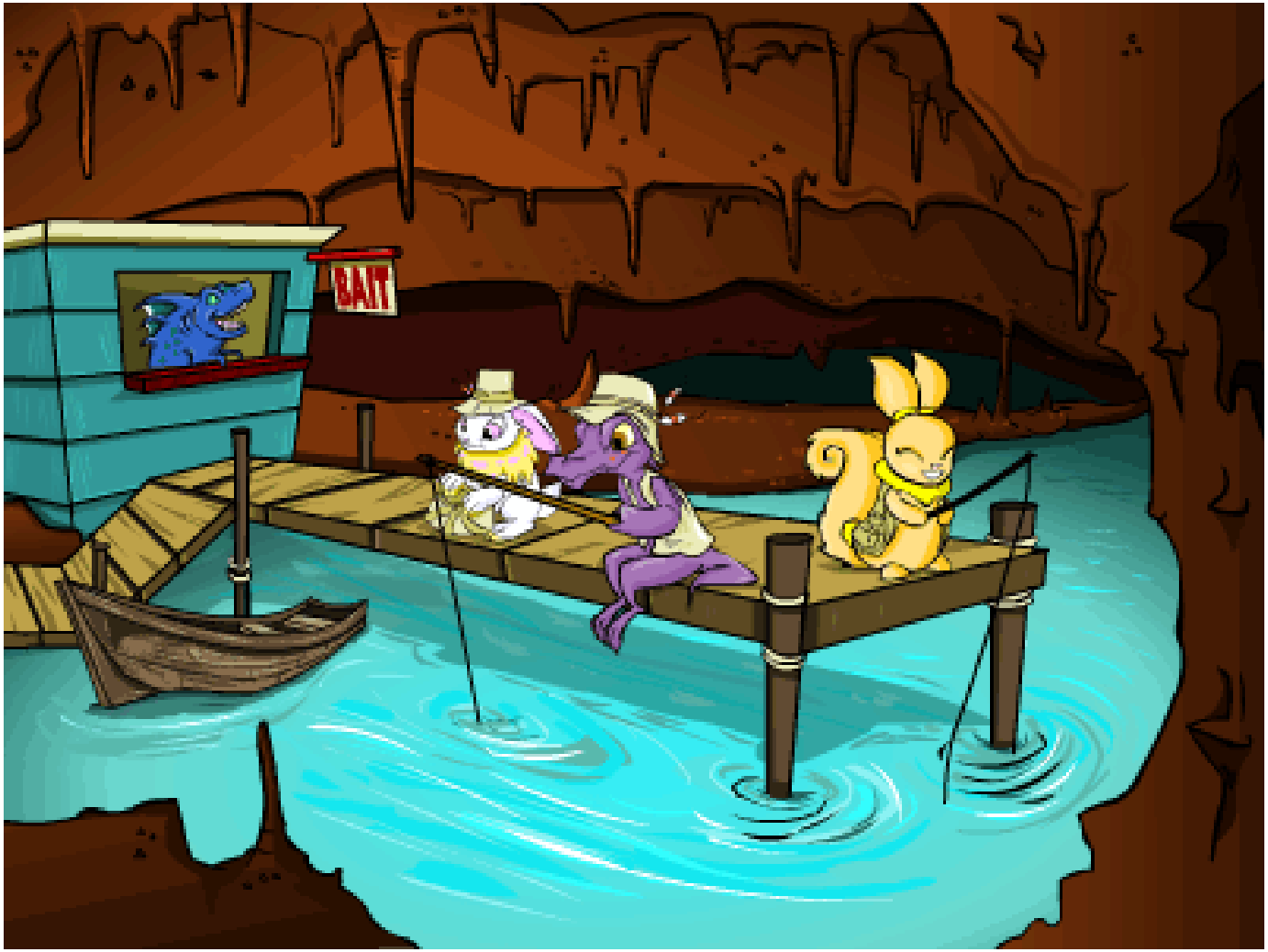
Multiple comparisons

Test hypothesis at each vertex in the mesh.

Vertex-Based Morphology (VBM).

test over 300,000 hypothesis.

Searching for interesting results



Outline

Definition

A cluster is a large, connected set of extreme p-values.

If the null hypothesis is true everywhere, then we expect 5% of the values to be in the range that is 5% likely, but these should be evenly distributed over the image.

It is rather unlikely that we would see a large cluster of vertices all of which were significant unless the null hypothesis were false in that region.

Markov Random Fields

Treat statistical output as a random field. Use characteristics of field to estimate probability of cluster size.

Prof. Keith Worsley, McGill University, Quebec.

While some of the early assumptions were perhaps questionable, the idea is under active development.

- Taylor, J.E. & Worsley, K.J. (2007). Random fields of multivariate test statistics, with applications to shape analysis and fMRI.*Annals of Statistics*, accepted.
- Taylor, J.E. & Worsley, K.J. (2007). Detecting sparse signal in random fields, with an application to brain mapping. *Journal of the American Statistical Association*, accepted.
- Taylor, J.E. & Worsley, K.J. (2007). Maxima of discretely sampled random fields, with an application to ‘bubbles’. *Biometrika*, 94:1:18.

Permutation tests.

presented yesterday by S. Baillet.

Re-label the subjects in the experiment, compute the parameter maps, and measure the size of the largest cluster of significant results. Repeat 10,000 times.

Gives a non-parametric estimate of the distribution of cluster sizes under the null hypothesis.

Strengths and weaknesses

Strength: Nonparametric estimate can provide better fit to the actual distribution when it doesn't fit model assumptions.

Weaknesses:

- Only measures largest cluster at each permutation.
- Assumption of interchangeability (generally holds for neuroimaging).
- Time consuming.

Weakness with blob-based methods

No localization

You know the cluster contains significant findings,

but not where in the cluster.

Weakness with blob-based methods

No localization

You know the cluster contains significant findings,

but not where in the cluster.

Outline

False Discovery Rate control

Control the expected proportion of false positives – weak error control.

Applies when goal is to sift through a mountain of significance tests and report those which might be worth following up on.

examples: microarray, geological data (oil prospecting), neuroimaging.

The original concept

False Discovery Rate (FDR) control,
1995 Benjamini and Hochberg [1].

Find a threshold for the p-values at which the expected proportion of false discoveries is below a user-specified fraction.

The model

A number of hypothesis, $H_1, H_2, \ldots, H_N$

with associated p-values $P_1, P_2, \ldots, P_N$

Desired FDR threshold is q, say 5%,
i.e. of the results reported as significant,
the expected proportion of false reports will be 5%.

The method

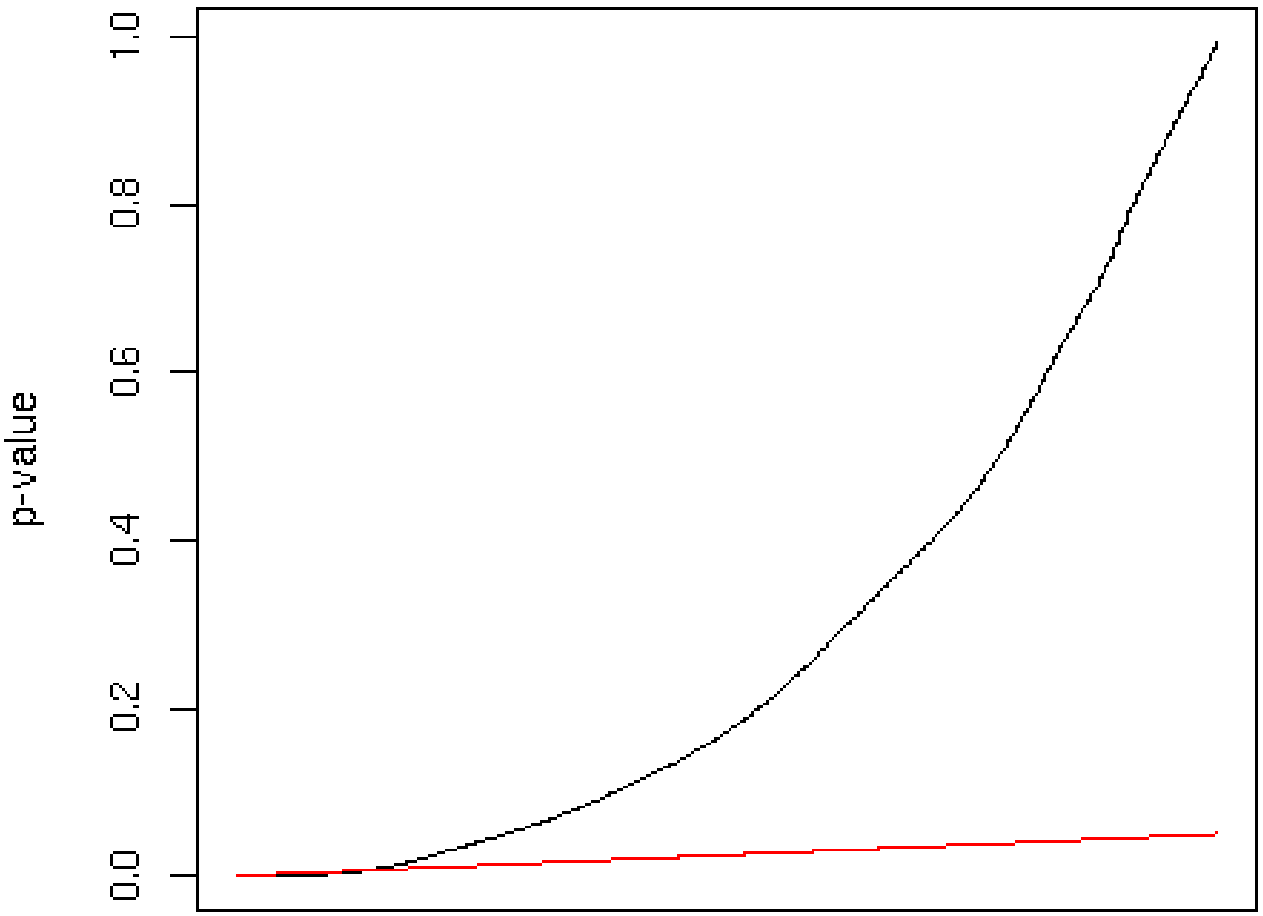
Order the p-values so that $P_{(1)} \leq P_{(2)} \leq \ldots \leq P_{(N)}$.

Let k be the largest i for which

$$P_{(i)} \leq \frac{i}{k} q \tag{1}$$

Reject every null hypothesis for which $P \leq P_{(k)}$.

Graphical



Relies on tail values

Null hypothesis true everywhere $\Rightarrow$ short tails.

Alternate hypothesis frequently true $\Rightarrow$ long (or fat) tails.

FDR in practice

Become widely popular.

No agreed on level for q.

Default in neuroimaging seems to be 0.05 (habit?)

Outline

Local FDR

Place Benjamini's FDR in an empirical Bayesian context.

Makes fewer assumptions – more statistical power.

Developed by Brad Efron Stanford University, USA. 2001 [2]

The model

N hypothesis $H_1, H_2, \ldots, H_N$
with Z-values $Z_1, Z_2, \ldots, Z_N$

Assume the N tests are null or non-null:

$p_0 = Pr\{null\}$	$f_0(z)$ density if null
$p_1 = Pr\{non-null\}$	$f_1(z)$ density if not null

The model, cont.

The z_i have the following mixture distribution:

$$f(z) = p_0 f_0(z) + p_1 f_1(z) \tag{2}$$

Definition

Define the local FDR as

$$loc\_fdr \equiv Pr\{Null \mid z\} = p_0 f_0(z)/f(z) \tag{3}$$

the Bayes posterior probability that a case is null given z.

Definition

Define the local FDR as

$$loc\_fdr \equiv Pr\{Null \mid z\} = p_0 f_0(z)/f(z) \tag{3}$$

the Bayes posterior probability that a case is null given z.

Why local?

The probability $Pr\{Null \mid z\}$ will vary over the range of z-values.

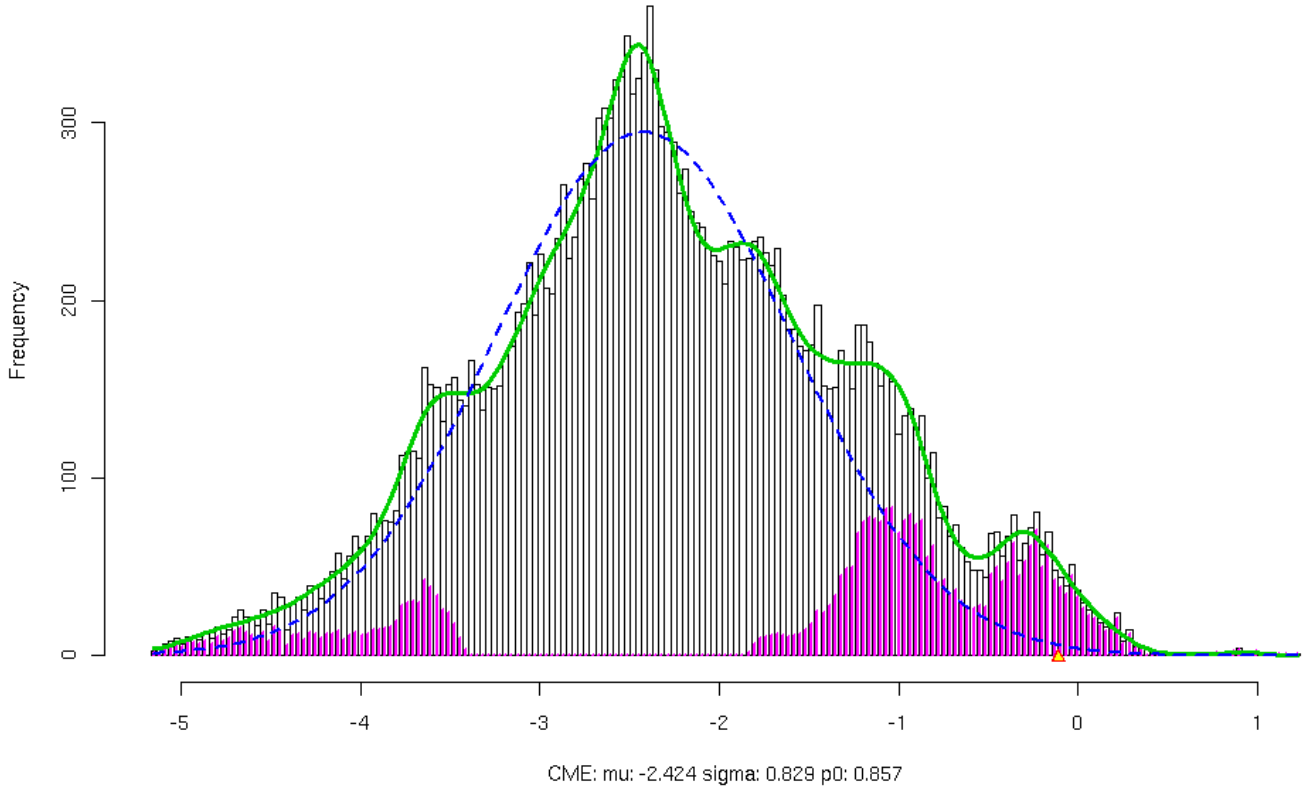
local in terms of z values.

Specifically, each test statistic is examined individually.

Compared to Benjamini's FDR

By contrast, Benjamini's FDR relies strictly on tail areas, and treats extremely high z-scores the same as extremely low z-scores (after converting to p-values).

Histogram of Z-scores



Requirements

Independence of the Z_i is not required,

only a reasonable estimate of their marginal distribution.

Calculating the local FDR

The goal:

$$loc\_fdr \equiv Pr\{Null \mid z\} = p_0 f_0(z)/f(z)$$

need $p_0 f_0(z)$ and $f(z)$.

The mixture $f(z)$

A number of ways to estimate $f(z)$.

As a Poisson regression (Lindsey's method).

The mixture $f(z)$

Place the N z-values in κ bins, with counts $y_1, y_2, \ldots, y_\kappa$.

Assume the counts y_κ are independent Poisson counts, i.e.

$$y_\kappa \overset{ind}{\sim} Po(\nu_\kappa), \quad \kappa \in [1, K]$$

with ν_κ proportional to $f(z)$ at the midpoint of bin κ,

The mixture $f(z)$

Model $\log(\nu_\kappa)$ as a pth degree polynomial

Estimate $f(z)$ empirically by maximum likelihood

$$f(z) = exp\left\{\sum_{j=0}^{p} \beta_j z^j\right\} \tag{4}$$

(Alternately, a natural spline function with p degrees freedom could be fit to the data.)

The numerator $p_0 f_0(z)$

It is assumed that $p_0 f_0(z)$ is a scaled normal density.

$$p_0 \varphi_{\mu,\sigma}(z) \tag{5}$$

where

$$\varphi_{\mu,\sigma}(z) = f_0(z) = \exp\left\{\frac{-1}{2}\left(\frac{z-\mu}{\sigma}\right)^2\right\} \Big/ \sqrt{2\pi\sigma^2} \tag{6}$$

The numerator $p_0 f_0(z)$

Fit a quadratic curve to $\log(\hat{f}(z))$ around $z = 0$.

The central peak of the histogram of z-values consists mainly of null cases.

The numerator $p_0 f_0(z)$

Alternative:

constrain the fitting such that $f_0(z)$ is the theoretical null, $f_0(z) \sim N(0, 1)$.

Does not always fit the data well
(especially if there is dependence between the z_i).

Local FDR in practice

Available as an R package called “locfdr” (CRAN).

Running time is seconds.

Works with Z values, not P values.

Outline

Covariate Modulated FDR

Incorporate prior information into the FDR estimation.

A covariate, measured at each point at which we have a hypothesis, relates to the probability that the null hypothesis is true.

2007: Developed by Egil Ferkingstad, Arnoldo Frigessi, Gudmar Thorleifsson and Augustine Kong [3]

Microarray example

Which genetic variants affect expression levels in the immediate area of the gene (cis-variants from an eQTL study)?

Data:

- p-values (from linkage analysis of the microarray).
- estimate of heritability for each location tested

The heritability of the gene can reasonably be expected to affect the probability of the null hypothesis being true.

The model

N hypothesis H_{0i} *vs* H_{1i}, $\quad i \in [1, N]$
with p-values $P_1, P_2, \ldots, P_N$.
and non-random covariate x_i for each hypothesis H_i.

It is believed that x_i affects the prior probability that null hypothesis i is true.

The model

N hypothesis H_{0i} *vs* H_{1i}, $\quad i \in [1, N]$
with p-values $P_1, P_2, \ldots, P_N$.
and non-random covariate x_i for each hypothesis H_i.

It is believed that x_i affects the prior probability that null hypothesis i is true.

Definition

The covariate-modulated FDR is defined as

$$cmFDR(H_{0i}|p_i, x_i) \equiv P(H_{0i}|p_i, x_i) \tag{7}$$

the posterior probability that the the null hypothesis is true given the p-value and x.

Definition

Use Baye's law

$$P(H_{0i}|p_i, x_i) = \frac{f(p_i, H_{0i}|x_i)}{f(p_i|x_i)} \tag{8}$$

(recall x is not random.)

Definition

$$= \frac{f(p_i, H_{0i}|x_i)}{f(p_i|x_i)} \tag{9}$$

$$= \frac{f(p_i|H_{0i}, x_i)f(H_{0i}|x_i)}{f(p_i|x_i)} \tag{10}$$

$$= \frac{f(p_i|H_{0i})f(H_{0i}|x_i)}{f(p_i|x_i)} \tag{11}$$

Compare to Local FDR

Similar to local FDR (with dependence on x_i).

$$f(H_{0i}|x_i) \equiv \pi_0(x_i) \approx p_0$$

$$f(p_i|x_i) \approx f(p)$$

The model, II

Factor $f(p_i|x_i)$ as

$$f(p_i|x_i) = \pi_0(x_i) + (1 - \pi_0(x_i))f(p_i|H_{1i}, x_i) \tag{12}$$

compare to local FDR:

$$f(z) = p_0 f_0(z) + p_1 f_1(z) \tag{13}$$

Calculating cmFDR

The goal:

Estimate $\pi_0(x_i)$ and $f(p_i|H_{1i}, x_i)$ from data (p_i, x_i).

Observation

Any distribution on $[0, 1]$ can be well approximated by a mixture of beta distributions. [Diaconis and Ylvisaker, 1985.]

A mixture:
uniform distribution $U[0, 1]$ – true null hypotheses
beta distribution $\beta(\xi, \theta)$ – true alternative hypothesis

well approximates the density f underlying a set of p-values [Allison et al., 2002]

Bin the covariates

The dependence on x is dealt with by dividing the (p_i, x_i) into $j = 1, \ldots, B$ bins increasing in x.

Within each bin the mixture is

$$f_j(p_i) = \pi_{0j} + (1 - \pi_{0j}) \frac{\Gamma(\xi_j + \theta_j)}{\Gamma(\xi_j)\Gamma(\theta_j)} p_i^{\xi_j - 1} (1 - p_i)^{\theta_j - 1} \tag{14}$$

Hyperpriors

Since the bins are small, the parameters π_{0j}, ξ_j, and θ_j should be smooth between bins.

$$p(\pi) \propto \exp\left(-\lambda_\pi \sum_{j=2}^{B} (\pi_{0j} - \pi_{0(j-i)})^2\right) \tag{15}$$

with similar priors for λ_ξ and λ_θ.

posterior density for π_{0j}, ξ_j, and θ_j

Bin j contains m_j p-values, $p_{j_1}, p_{j_2}, \ldots, p_{j_{m_j}}$.

The simultaneous posterior density for π_{0j}, ξ_j, and θ_j is ($\propto$) the mixture model times the smoothing parameters.

$$\prod_{h=1}^{m_j} \left(\pi_{0j} + (1 - \pi_{0j}) \frac{\Gamma(\xi_j + \theta_j)}{\Gamma(\xi_j)\Gamma(\theta_j)} p_{j_h}^{\xi_j - 1} (1 - p_{j_h})^{\theta_j - 1} \right) \times S \qquad (16)$$

where S is the smoothing priors

posterior density, cont.

$$s = \exp\left(\begin{array}{ll} -\lambda_\pi \sum_{i=j}^{j+1} \left(\pi_{0i} - \pi_{0(i-1)}\right)^2 & + \\ -\lambda_\xi \sum_{i=j}^{j+1} \left(\xi_i - \xi_{i-1}\right)^2 & + \\ -\lambda_\theta \sum_{i=j}^{j+1} \left(\theta_i - \theta_{i-1}\right)^2 & \end{array}\right) \tag{17}$$

Approximate

π_{0j}, ξ_j, and θ_j can be approximated using Markov Chain Monte Carlo.

takes time, ...

Approximate in minutes using Laplace – Håvard Rue. *submitted*

Approximate

π_{0j}, ξ_j, and θ_j can be approximated using Markov Chain Monte Carlo.

takes time, . . .

Approximate in minutes using Laplace – Håvard Rue. *submitted*

cmFDR with one bin

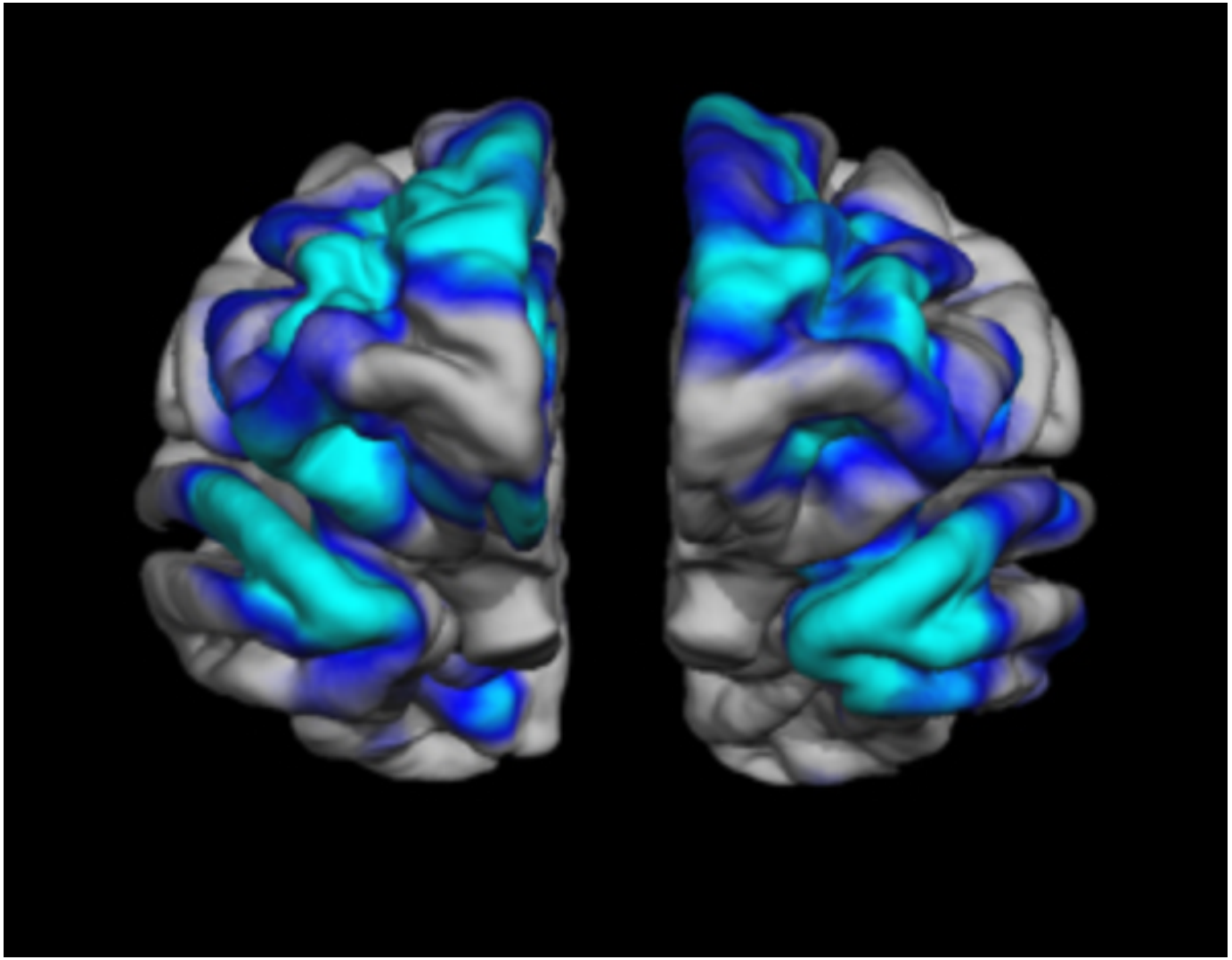
Running cmFDR with one bin effectively discards all information in the covariate. It is thus a variant of the local FDR.

cmFDR assumes parametric model for the p-value density.

Efron estimates density by parametric smoothing of the p-value histogram.

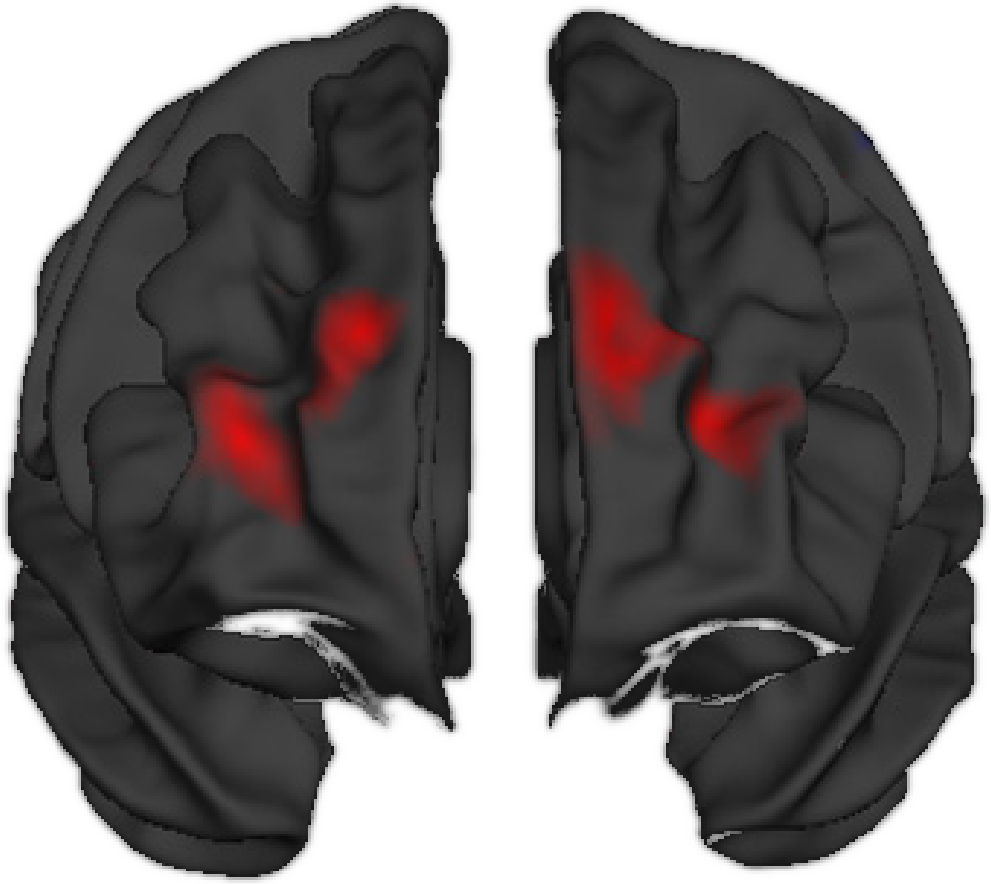
Outline

Genetic variation and cortical thickness



Preliminary indications of 12% increase in findings using local FDR compared to FDR.

BDNF -663



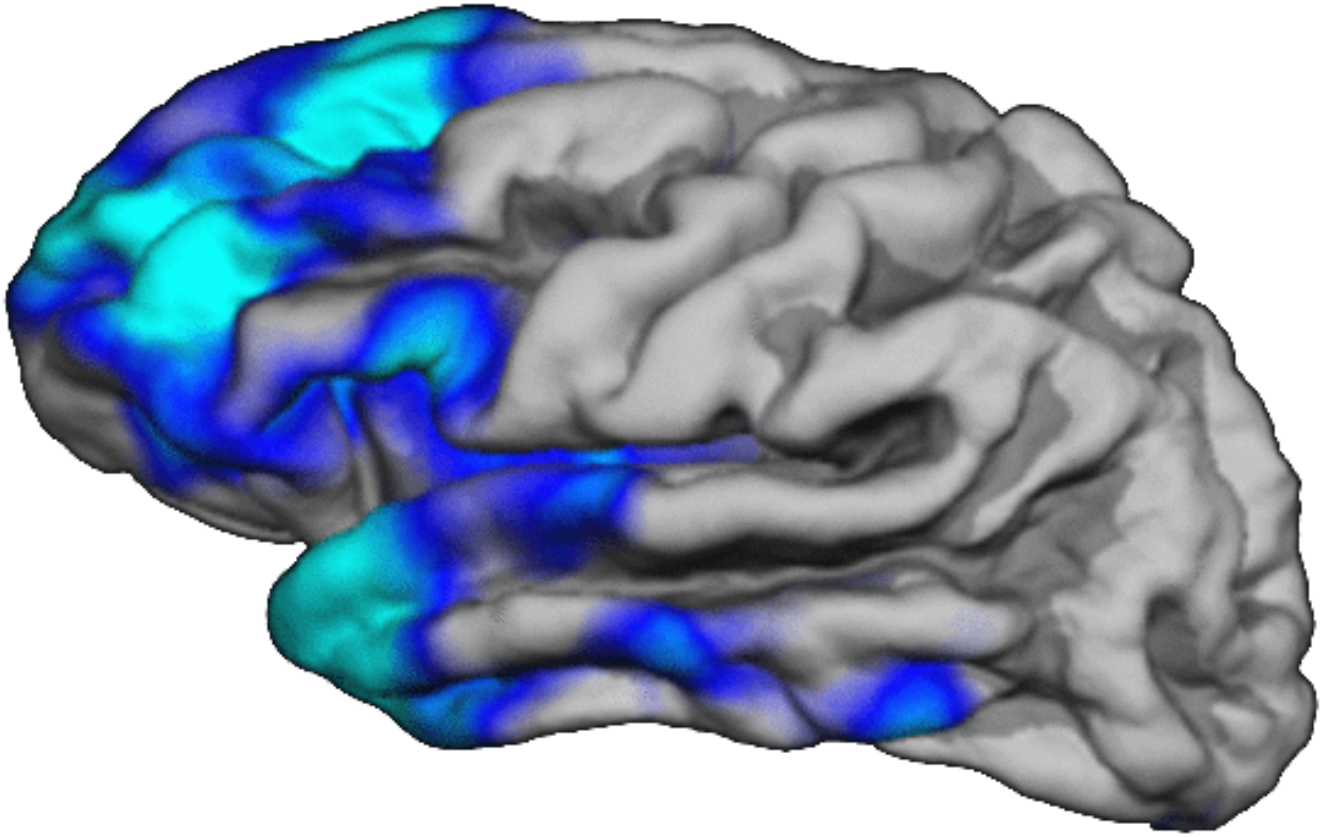
AA polymorphism thinner than TT

cmFDR reasoning

- phenotype suspected to be associated with an increased risk for schizophrenia
- having phenotype *might* be part of the biological cause underlying the thinning in schizophrenia

cortical thickness differences in A/A-subjects compared to T/T-subjects are more believable in regions where patients are thinner than controls.

Comparison



Restricted to ROI (7% of total surface).

Comparison

Num tests	$p < 0.001$	FDR	cmFDR*	cmFDR
22199	504	834	2282	6682

*2.75 increase with one-bin cmFDR.
8-fold increase when covariate information included.

cmFDR with no bins $\Rightarrow \alpha \approx 0.012$
cmFDR with 10 bins $\Rightarrow \alpha \approx 0.080$

Bibliography

- Benjamini Y., Hochberg Y.. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing *J. of the Royal Stat. Soc. B.* 1995;57:289-300.
- Efron Bradley. Local False Discovery rates *authors web page.* 2003.
- Ferkingstad Egil, Frigessi Arnoldo, Thorleifsson Gudmar, Kong Augustine. Covariate-modulated false discovery rates *Annals of applied Statistics.* 2007;submitted.