Exploratory Inference for Brain Imaging

International Seminar on Selective Inference
7 May 2020

Aldo Solari
DEMS
University of Milano-Bicocca

Joint work with

Jelle Goeman	Leiden University Medical Centre
Jonathan Rosenblatt	Ben Gurion University of the Negev
Livio Finos	University of Padova
Rosa Meijer	Antes PG
Thijmen Krebs	Delft University
Wouter Weeda	Leiden University

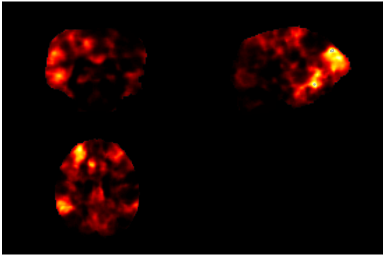
Post-selection inference

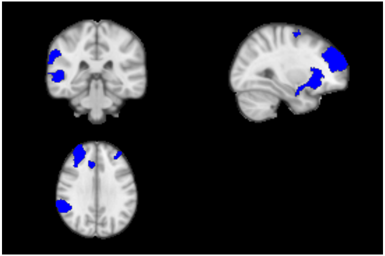
Examining the data to *select* interesting patterns,
then carrying out *inference* about the selection with the same data

Question

How to correct for overoptimism in inference due to data-driven selection?

fMRI data





Brain activity map

Selection

Outline

1. Cluster-Size Inference
2. All-Resolutions Inference
3. Closed Testing
4. Relationship to Hommel and Benjamini-Hochberg
5. Conclusions

fMRI experiment

Subjects perform mental tasks in MRI scanner
MRI measures oxygenated blood flow in brain (brain activity)

Brain activity map

Significance (p-value) for brain activity at each location (*voxel*)

Goal

Find *regions* of brain activity

Aggregation

Micro-inferences (voxels) $\rightarrow$ larger-scale inferences (regions)

Go/NoGo data

Lee, Weeda, Somerville, Insel, Krabbendam, Huizinga (2018)

34 subjects performing an emotional Go/NoGo task

- Go: press button when seeing happy face
- NoGo: hold when face is not happy

225212 voxels

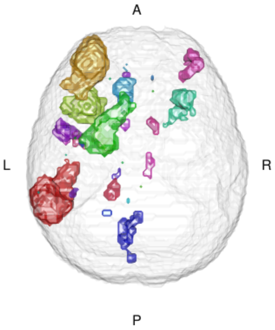
Cluster-size inference

Poline and Mazoyer (1993)

- A cluster is ‘significant’ if its size is larger than ‘chance’
- Size threshold: $1 - \alpha$ quantile of the null distribution of the maximum size of clusters
- Cluster-size inference controls the familywise error rate at α

Clusters

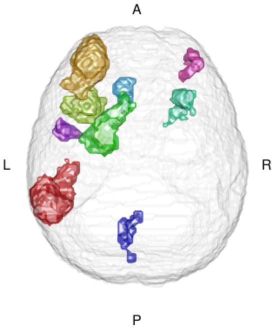
cluster $\equiv$ contiguous voxels with $p < 0.0007$



32 clusters of size 2191, 1835, 1400, 698, 421, 304, 245, 232, 187, 82, 69, 43, 43, 28, 14, 12, 12, 10, 10, 6, 5, 4, 3, 2, 1, 1, 1, 1, 1, 1, 1, 1

Significant clusters

cluster size > 161



9 significant clusters of size 2191, 1835, 1400, 698, 421, 304, 245, 232, 187

Cluster null hypothesis

- Cluster-size inference tests a (random) number of cluster null hypotheses
- Cluster null hypothesis: ‘*all the voxels in the cluster are null*’
- Its rejection implies ‘*at least one voxel in the cluster is active*’

Spatial specificity paradox

- The most we can say is that ‘*an activation has occurred somewhere inside the cluster*’
- The larger the cluster, the weaker the finding

1. Cluster-Size Inference
2. All-Resolutions Inference
3. Closed Testing
4. Relationship to Hommel and Benjamini-Hochberg
5. Conclusions

$M = \{1, \dots, m\}$	collection of $m = \lvert M \rvert$ voxels
$M_1 \subset M$	active voxels with $m_1 = \lvert M_1 \rvert$ and $\pi_1 = m_1/m$
$M_0 = M_1 \setminus M_1$	null voxels with $m_0 = \lvert M_0 \rvert$ and $\pi_0 = m_0/m$
$H_i : i \in M_0$	voxel null hypothesis with p-value $p_i, \quad i \in M$

$M = \{1, \ldots, m\}$	collection of $m = \lvert M\rvert$ voxels
$M_1 \subset M$	active voxels with $m_1 = \lvert M_1\rvert$ and $\pi_1 = m_1/m$
$M_0 = M_1 \setminus M_1$	null voxels with $m_0 = \lvert M_0\rvert$ and $\pi_0 = m_0/m$
$H_i : i \in M_0$	voxel null hypothesis with p-value p_i, $\quad i \in M$

Selection

$S \subseteq M$	selected voxels
$m_1(S) = \lvert M_1 \cap S\rvert$	number of true discoveries in the selection
$\pi_1(S) = m_1(S)/\lvert S\rvert$	true discoveries proportion in S
$m_0(S) = \lvert S\rvert - m_1(S)$	number of false discoveries in S
$\pi_0(S) = 1 - \pi_1(S)$	false discovery proportion in S, i.e. $\mathrm{FDP}(S)$

All-Resolutions Inference

Goeman and Solari (2011); Rosenblatt, Finos, Weeda, Solari, Goeman (2018)

Lower confidence bound for the number of true discoveries in the selection, simultaneously valid for all possible selections

$$\mathrm{P}\Big(\forall\, S \subseteq M :\ \underbrace{\underline{m}_1(S)}_{\text{lower bound}} \ \leq\ \underbrace{m_1(S)}_{\text{parameter}}\Big) \geq 1 - \alpha$$

First proposed by Genovese and Wasserman (2004)

Cluster-size inference

cluster S	*size* $\lvert S \rvert$	*# active* $m_1(S)$
C_1	2191	≥ 1
C_2	1835	≥ 1
C_3	1400	≥ 1
C_4	698	≥ 1
C_5	421	≥ 1
C_6	304	≥ 1
C_7	245	≥ 1
C_8	232	≥ 1
C_9	187	≥ 1

ARI

cluster S	*size* $\lvert S \rvert$	*# active* $m_1(S)$	*% active* $\pi_1(S)$
C_1	2191	$\geq$ 624	$\geq$ 29 %
C_2	1835	$\geq$ 847	$\geq$ 46 %
C_3	1400	$\geq$ 454	$\geq$ 32 %
C_4	698	$\geq$ 0	$\geq$ 0 %
C_5	421	$\geq$ 25	$\geq$ 6 %
C_6	304	$\geq$ 33	$\geq$ 11 %
C_7	245	$\geq$ 0	$\geq$ 0 %
C_8	232	$\geq$ 0	$\geq$ 0 %
C_9	187	$\geq$ 0	$\geq$ 0 %

Bonferroni inference

cluster S	*size* $\lvert S \rvert$	*# active* $m_1(S)$	*% active* $\pi_1(S)$
C_1	2191	≥ 7	≥ 0.3 %
C_2	1835	≥ 86	≥ 4 %
C_3	1400	≥ 82	≥ 6 %
C_4	698	≥ 0	≥ 0 %
C_5	421	≥ 0	≥ 0 %
C_6	304	≥ 0	≥ 0 %
C_7	245	≥ 0	≥ 0 %
C_8	232	≥ 0	≥ 0 %
C_9	187	≥ 0	≥ 0 %

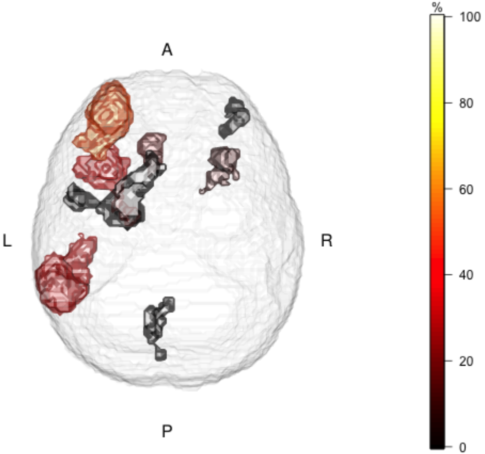
Interactive (exploratory) inference

ARI legitimates post-selection inference with full flexibility

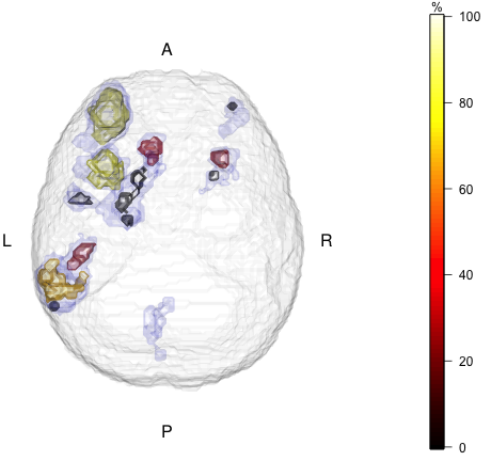
- the user looks at the data and selects interesting $S_1, S_2, \ldots$
- ARI informs the user about $\underline{m}_1(S_1), \underline{m}_1(S_2), \ldots$
- then the user may consider others $S'_1, S'_2, \ldots$
- ARI informs the user about $\underline{m}_1(S'_1), \underline{m}_1(S'_2), \ldots$
- $\ldots$

All ARI's statements are simultaneously correct with high prob.

$p < t$



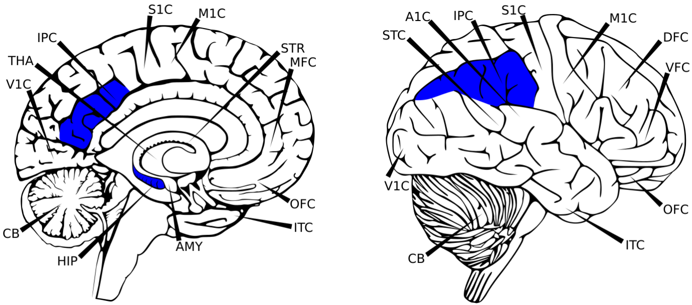
$$p < t' < t$$



Sub-clusters

cluster	*threshold*	*size*	*# active*	*% active*
C_1	$p < t$	2191	624	29 %
1	$p < t'$	405	267	66 %
2	$p < t'$	133	31	23 %
3	$p < t'$	6	0	0 %
C_2	$p < t$	1835	847	46 %
1	$p < t'$	963	826	86 %
$\vdots$				

Anatomical regions



1. Cluster-Size Inference
2. All-Resolutions Inference
3. Closed Testing
4. Relationship to Hommel and Benjamini-Hochberg
5. Conclusions

Closed testing

Marcus, Peritz, Gabriel (1976)

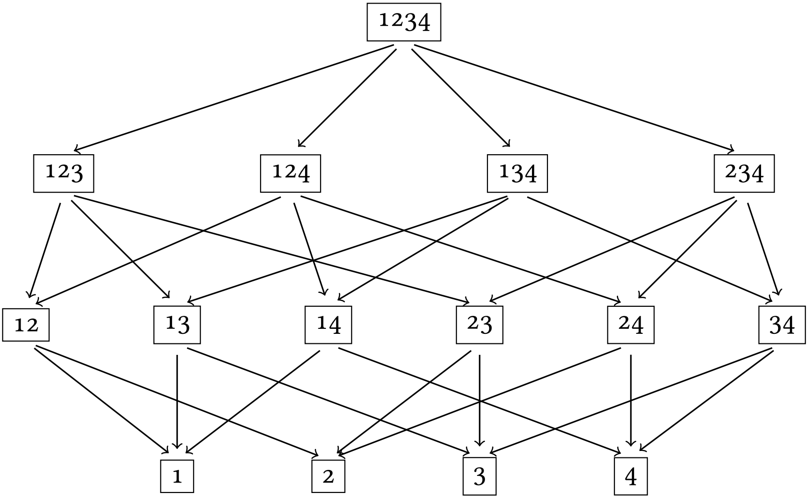
$H_1, \ldots, H_m$	elementary hypotheses
$H_S = \bigcap_{i \in S} H_i \qquad \forall\ S \subseteq M$	intersection hypotheses
$\phi_S = \mathbb{1}\{H_S \text{ rejected at level } \alpha\}$	local tests
$\tilde{\phi}_S = \min\left\{\phi_K : S \subseteq K \subseteq M\right\}$	closed testing adjusted tests

Closed testing guarantees familywise error rate control at α over all intersection hypotheses

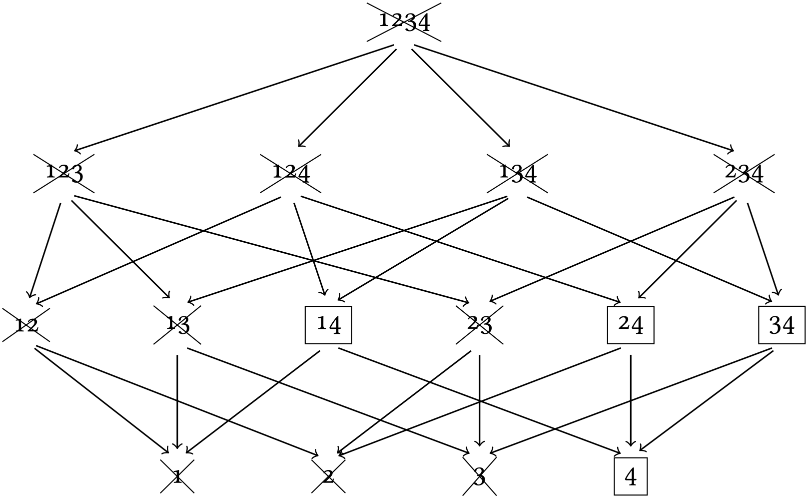
Four-pixels brain

1	2
3	4

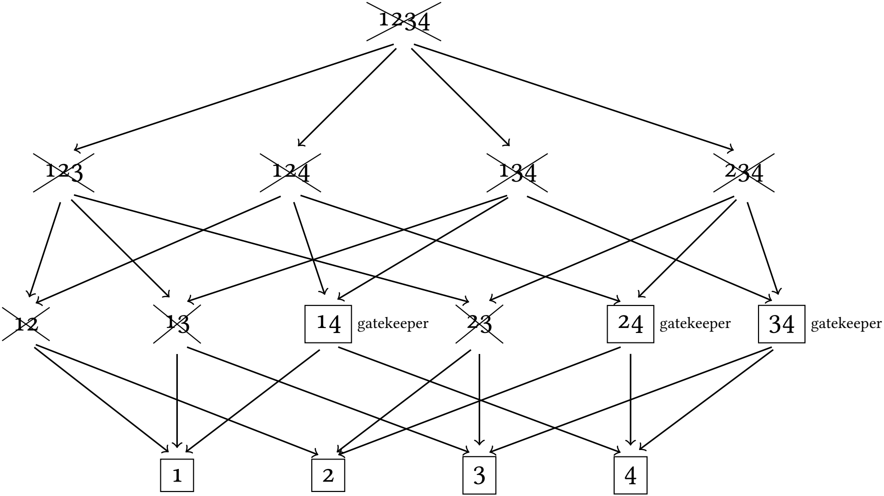
Intersection hypotheses



Rejections



Closed testing rejections



Confidence bound

Goeman and Solari (2011)

$$\bar{m}_0(S) = \max_{K \subseteq S} \left\{ |K| : \tilde{\phi}_K = 0 \right\}$$

The size of the largest subset of S for which the corresponding intersection hypothesis is not rejected by closed testing

$$\underline{m}_1(S) = |S| - \bar{m}_0(S)$$

S	$m_1(S)$	$\pi_1(S)$
$\{1\}$	≥ 0	≥ 0 %
$\{2\}$	≥ 0	≥ 0 %
$\{3\}$	≥ 0	≥ 0 %
$\{4\}$	≥ 0	≥ 0 %
$\{1,2\}$	≥ 1	≥ 50 %
$\{1,3\}$	≥ 1	≥ 50 %
$\{1,4\}$	≥ 0	≥ 0 %
$\{2,3\}$	≥ 1	≥ 50 %
$\{2,4\}$	≥ 0	≥ 0 %
$\{3,4\}$	≥ 0	≥ 0 %
$\{1,2,3\}$	≥ 2	≥ 66.6 %
$\{1,2,4\}$	≥ 1	≥ 33.3 %
$\{1,3,4\}$	≥ 1	≥ 33.3 %
$\{2,3,4\}$	≥ 1	≥ 33.3 %
$\{1,2,3,4\}$	≥ 2	≥ 50 %

Closed testing bottleneck

The required number of tests is 2^m

Shortcut

Computation time can be reduced to polynomial time by specific choice of local tests

Simes test

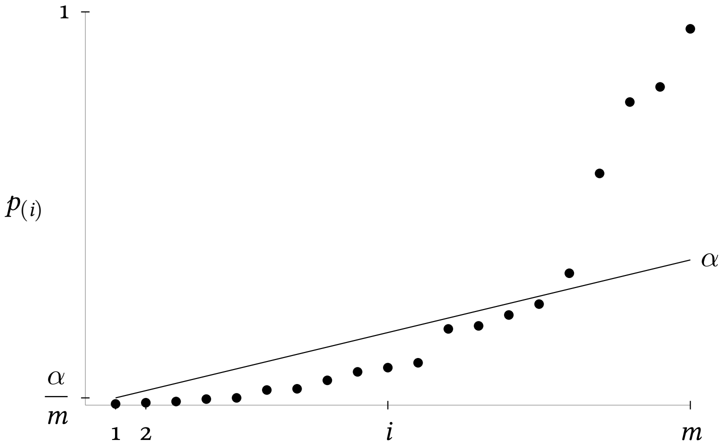
Simes test for H_S

$$\phi_S = \mathbb{1}\left\{ \bigcup_{i \in S} \left\{ p_{(i:S)} \leq \frac{i\alpha}{|S|} \right\} \right\}$$

where $p_{(i:S)}$ is the ith smallest p-value in $\{p_i : i \in S\}$

Assumption
Simes inequality (1986) holds for null p-values

$$\mathrm{P}\left(\bigcap_{i=1}^{m_0} \left\{ p_{(i:M_0)} > \frac{i\alpha}{m_0} \right\} \right) \geq 1 - \alpha$$

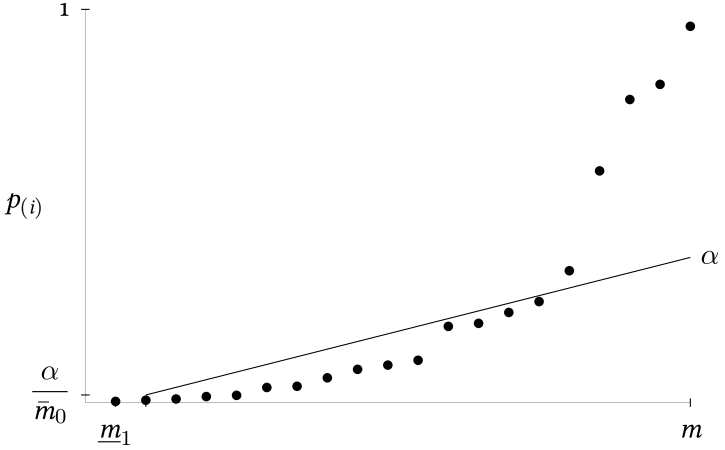


$$m_0 = m\,?\ \text{No} \ \rightarrow m_0 \leq \bar{m}_0 = m - 1$$

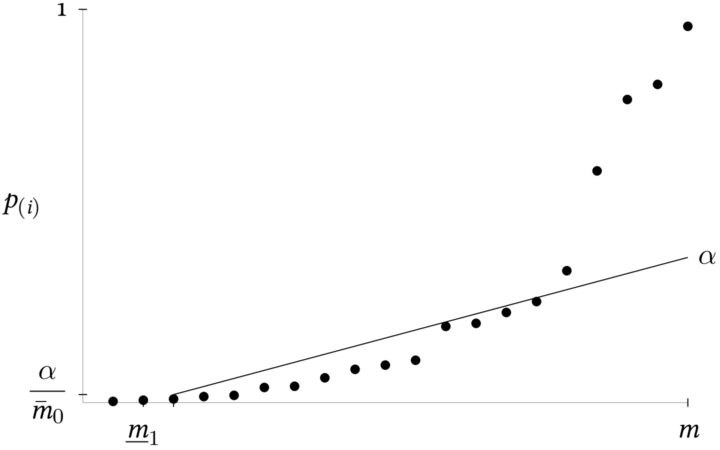
Upper bound for m_0

Find the upper confidence bound for m_0

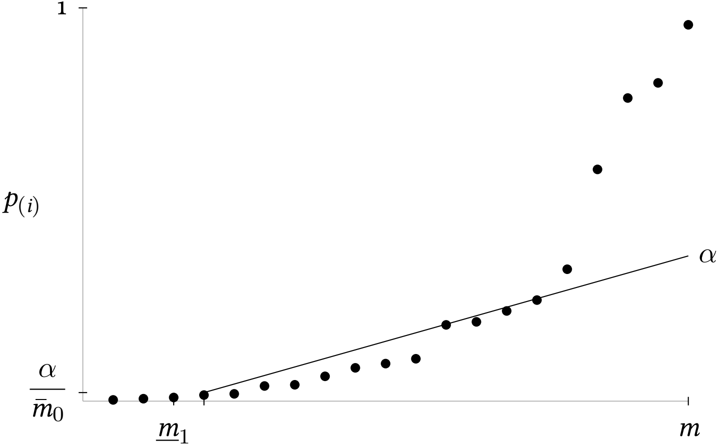
$$\bar{m}_0 = \max\left\{0 \le k \le m : \bigcap_{i=1}^{k} \left\{p_{(m-k+i)} > \frac{i\alpha}{k}\right\}\right\}$$

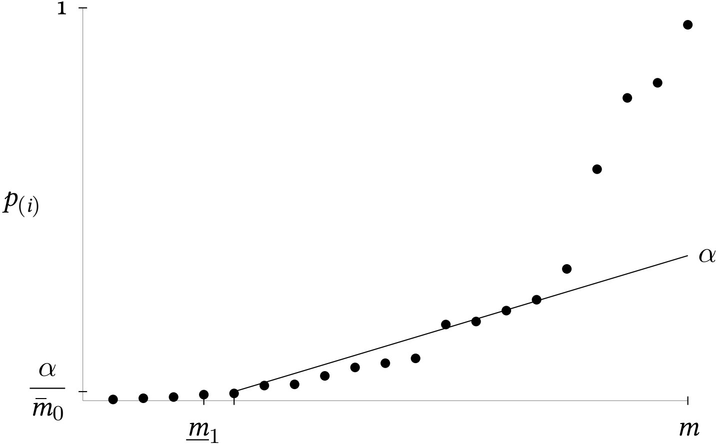


1
$p_{(i)}$
α
$\frac{\alpha}{\bar{m}_0}$
$\underline{m}_1$
m

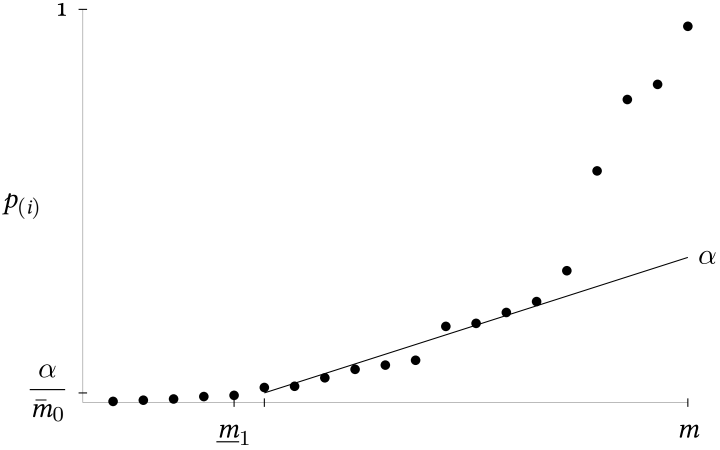


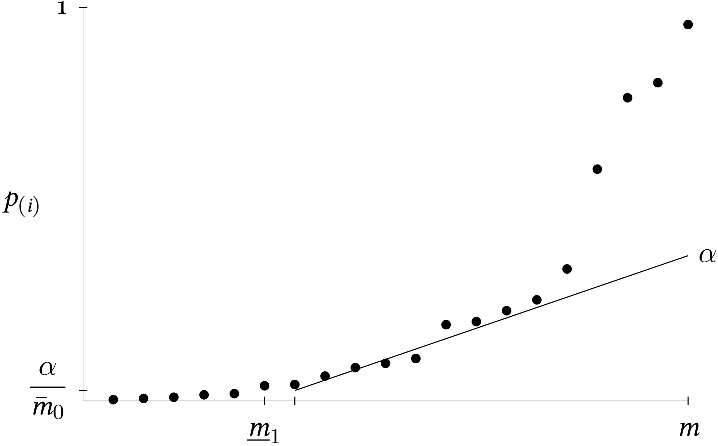
1
$p_{(i)}$
α
$\frac{\alpha}{\bar{m}_0}$
$\underline{m}_1$
m


1
$p_{(i)}$
α
$\frac{\alpha}{\bar{m}_0}$
$\underline{m}_1$
m

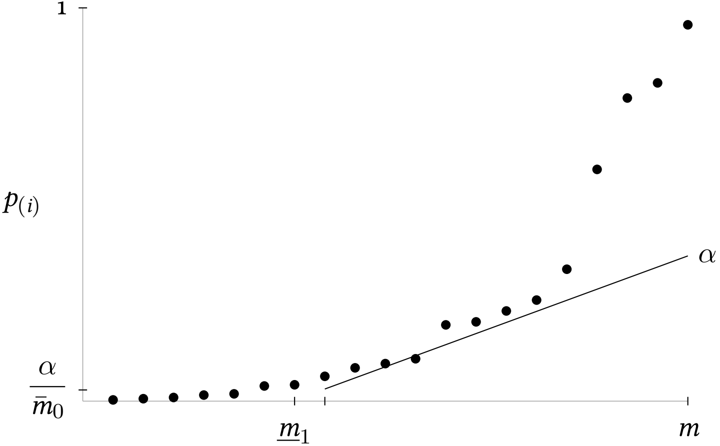


1
$p_{(i)}$
α
$\frac{\alpha}{\bar{m}_0}$
$\underline{m}_1$
m

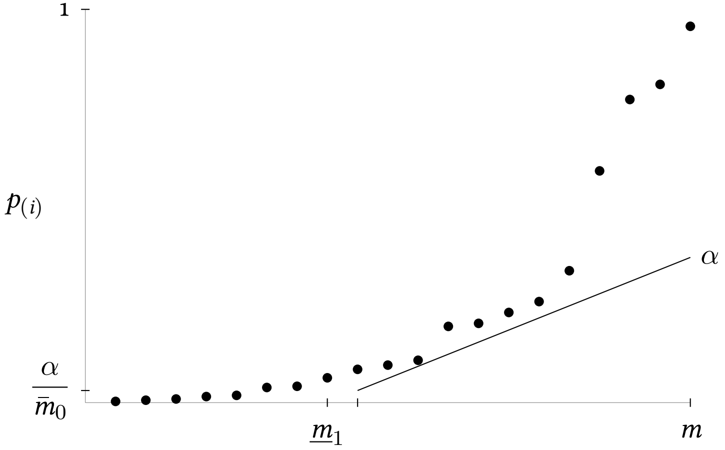

1
$p_{(i)}$
α
$\frac{\alpha}{\bar{m}_0}$
$\underline{m}_1$
m



1
$p_{(i)}$
α
$\frac{\alpha}{\bar{m}_0}$
$\underline{m}_1$
m

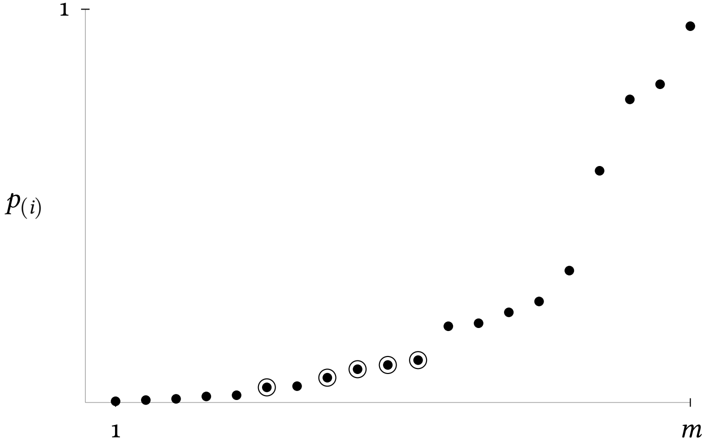


1
$p_{(i)}$
α
$\frac{\alpha}{\bar{m}_0}$
$\underline{m}_1$
m



1
$p_{(i)}$
α
$\frac{\alpha}{\bar{m}_0}$
$\underline{m}_1$
m

Arbitrary selection



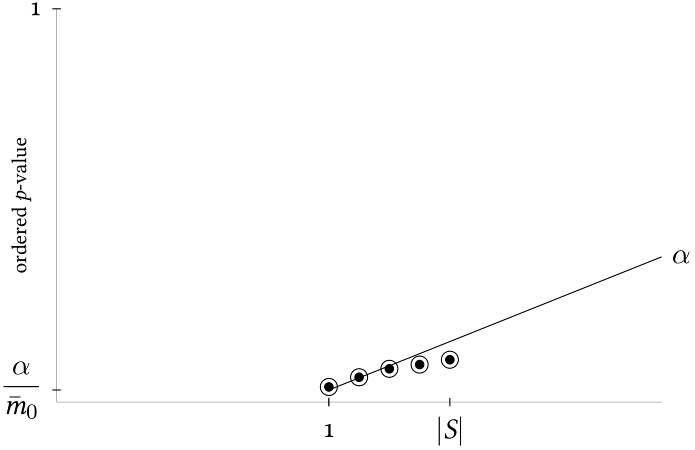
$S \subseteq M$

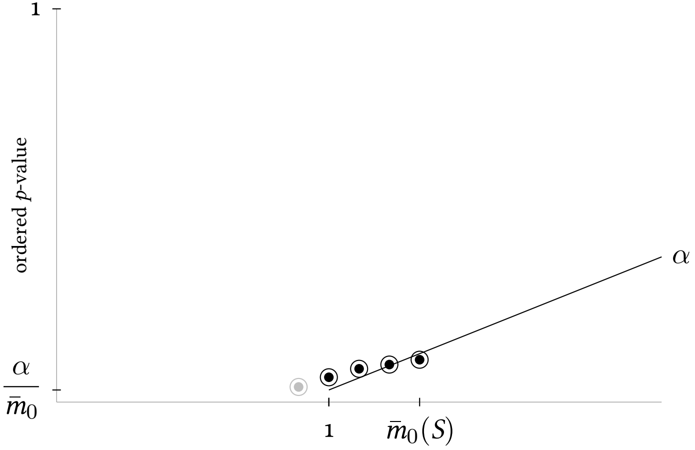
Confidence bound

Goeman, Meijer, Krebs, Solari (2019)

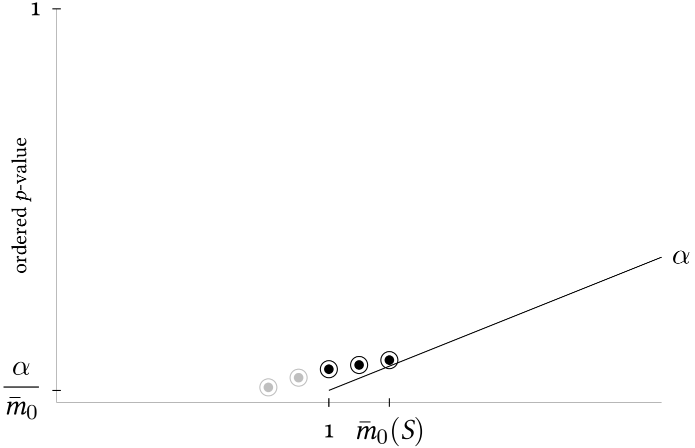
Theorem

$$\underline{m}_1(S) = \min\left\{0 \leq k \leq |S| : \bigcap_{i=1}^{|S|-k} \left\{p_{(k+i:S)} > \frac{i\alpha}{\bar{m}_0}\right\}\right\}$$


1
ordered p-value
α
$\frac{\alpha}{\bar{m}_0}$
1
$|S|$



1
ordered p-value
α
$\frac{\alpha}{\bar{m}_0}$
1
$\bar{m}_0(S)$



1
ordered p-value
α
$\frac{\alpha}{\bar{m}_0}$
1
$\bar{m}_0(S)$

Algorithm

	Operation	*Complexity*
1	Sort the p-values	$O(m \log m)$
2	Compute $\bar{m}_0$	$O(m)$
3	For each S, compute $\underline{m}_1(S)$	$O(\lvert S \rvert)$

- $\bar{m}_0$ in linear time

  Meijer, Krebs, Goeman (2019)
- Implemented in the R package `hommel`

  Goeman, Meijer, Krebs

1. Cluster-Size Inference
2. All-Resolutions Inference
3. Closed Testing
4. Relationship to Hommel and Benjamini-Hochberg
5. Conclusions

Relationship to Hommel (FWER)

Hommel (1988)

- Reject the hypotheses with indexes in

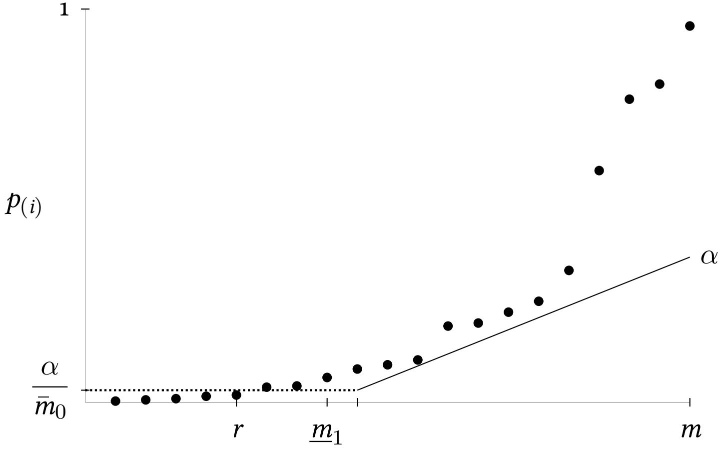
$$R = \left\{ i \in M : p_i \leq \frac{\alpha}{\bar{m}_0} \right\}$$

  with familywise error rate control at α

- Voxels in R represent *localized activations*

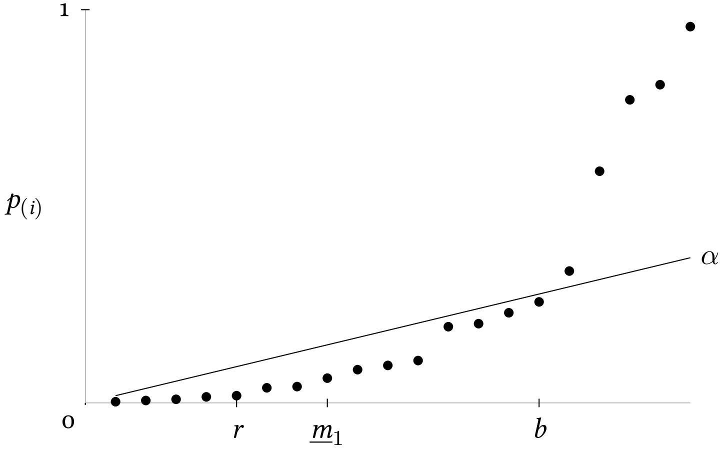
$$\underline{m}_1(R) = |R| = r$$

Hommel rejections



Relationship to Benjamini-Hochberg (FDR)

Benjamini and Hochberg (1995)



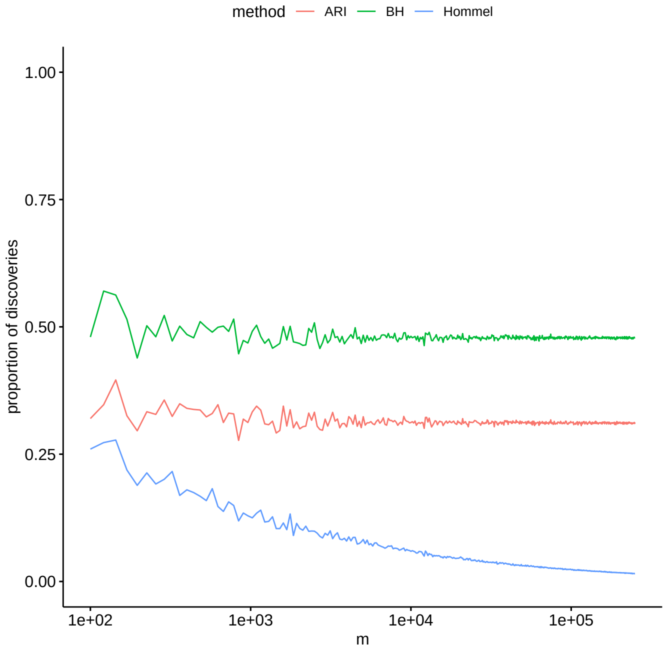
Large-scale testing

Assume $p_1, \ldots, p_m \overset{i.i.d.}{\sim} F$
with a mixture distribution $F(u) = \pi_0 u + \pi_1 F_1(u)$

Lemma
Fix $\alpha \in (0, 1)$. As the number of hypotheses $m \to \infty$

$$\operatorname*{plim}_{m\to\infty} \frac{r}{m} = 0 \qquad \operatorname*{plim}_{m\to\infty} \frac{m_1}{m} = k > 0 \qquad \operatorname*{plim}_{m\to\infty} \frac{b}{m} = k' > 0$$

if a minimal level of signal is present



method
ARI
BH
Hommel
1.00
0.75
0.50
0.25
0.00
proportion of discoveries
1e+02
1e+03
1e+04
1e+05
m

From FDR to FDP confidence

Let

$$b_q = \max\left\{1 \le i \le m : p_{(i)} \le \frac{iq}{m}\right\}$$

and $B_q = \{i \in M : p_i \le p_{(b_q)}\}$ the index set of BH rejections at level q

$$\mathrm{FDR}(B_q) \le q$$

Lemma

With probability $\ge 1 - \alpha$

$$\mathrm{FDP}(B_{\tilde{q}}) \le q$$

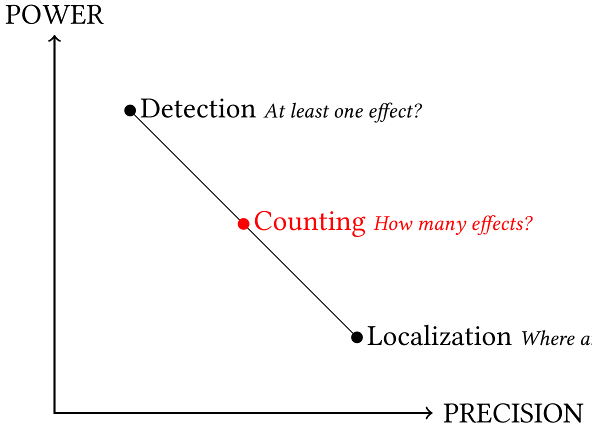
where

$$\tilde{q} = q \cdot \frac{\alpha}{\bar{\pi}_0}$$

1. Cluster-Size Inference
2. All-Resolutions Inference
3. Closed Testing
4. Relationship to Hommel and Benjamini-Hochberg
5. Conclusions

Conclusions

ARI is a flexible approach to large-scale post-selection inference



Ongoing work

Closed testing is admissible for FDP control

Goeman, Hemerik, Solari (2020+)

ARI re-analysis of ≈ 2000 fMRI data sets

Weeda, van Kempen, Chen, Goeman (2020+)

ARI with permutations

Hemerik, Solari, Goeman (2019); Andreella, Hemerik, Finos, Goeman (2020+)

ARI with different local tests

- *Higher criticism*
  Goeman, Hemerik, Solari (2020+)
- *Global test*
  Xu, Solari, Goeman (2020+)
- *Harmonic mean p-value*
  Goeman, Rosenblatt, Nichols (2019); Tian, Goeman, Ramdas, Katsevich (2020+)
- *Sum-type tests*
  Vesely, Finos, Goeman (2020+)
- ...

References

Multiple Testing for Exploratory Research
Goeman, Solari
Statistical Science 2011 26:584-597

All-Resolutions Inference for Brain Imaging
Rosenblatt, Finos, Weeda, Solari, Goeman
NeuroImage 2018 181:786-796

Simultaneous Control of All False Discovery Proportions in Large-Scale Multiple Hypothesis Testing
Goeman, Meijer, Krebs, Solari
Biometrika 2019 106:841-856

Hommel R package
Goeman, Meijer, Krebs
cran.r-project.org