Alternative thresholding procedure with application to pre-surgical fMRI

Joke Durnez[a], Beatrijs Moerkerke[a], Andreas Bartsch[b] and Thomas Nichols[c]

[a] Department of Data Analysis, Ghent University, Belgium
[b] Department of Neuroradiology, University of Würzburg, Germany
[c] Department of Statistics & Warwick Manufacturing Group, University of Warwick, United Kingdom

1. Power concerns in pre-surgical fMRI

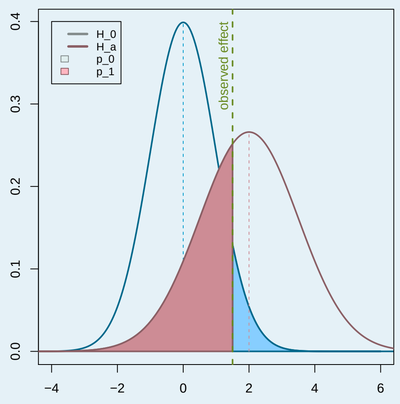
- When surgically resecting brain tumors, one wants to minimize risk of resecting brain tissue involved in important human tasks .
- **Pre-surgical fMRI** uses these tasks to localize eloquent (necessary) brain tissue
- Statistical inference in cognitive neuroscience focuses on **control of false positives**.
- The scientific discipline deems stringent control of false positives necessary, accepting the concomitant sacrifices in sensitivity.
- In a clinical setting, a loss in power means true activation is not discovered, and this might result in the **resection of vital brain tissue**.
- This asymmetrical way of penalising errors in statistical inference is undesirable in this context.
- We therefore present a **new hypothesis thresholding procedure** that incorporates information on both false positives and false negatives and thus is ideally suited for pre-surgical fMRI.

2. Measures of evidence against the H_0 and H_a

- $H_0 : \Delta = 0$
  (Δ: BOLD effect of interest in units of percent BOLD change)
- $H_a : \Delta = \Delta_1$
  (Δ_1: the non-zero effect magnitude expected under activation)
- In classical hypothesis testing, the evidence against H_0 is measured with the p-value, the null hypothesis probability of data as or more extreme than that observed.
- Thresholding a p-value at α produces a statistical test that **controls the false positive rate at** α.
- To allow direct control of false negative risk, we present a symmetrical measure p_1 which quantifies evidence against the H_a.
- Thresholding this probability measure at β ensures **control of the false negative rate at** β.

3. p_0 and p_1

- $t = \widehat{\Delta}/\mathrm{SE}(\widehat{\Delta})$
- **Evidence against** H_0:
  $p_0 = P(T \geq t|H_0)$
  $T_i \sim \mathcal{N}(0,1)$ under H_0
- **Evidence against** H_a:
  $p_1 = P(T \leq t|H_a)$
  - We don't expect a single magnitude of true activation:
    $\Delta_1 \sim \mathcal{N}(\mu, \tau^2)$
  - Consequently:
    $T_i \sim \mathcal{N}\left(\frac{\mu}{\mathrm{SE}(\widehat{\Delta}_i)}, \frac{\mathrm{SE}(\widehat{\Delta}_i)^2+\tau^2}{\mathrm{SE}(\widehat{\Delta}_i)^2},\right)|H_a$



4. Combining measures of significance

For given value of μ (expected activation) and τ (its uncertainty):
When thresholding p_0 and p_1 at respectively level α and β:

1st layer of activation $\mathbf{p_0 \leq \alpha = 0.001}$:
Evidence against the null of no activation.

2nd layer of activation $\mathbf{p_1 \geq \beta = 0.20}$:
Voxels cannot be confidently declared inactive.

1st layer ∩ 2nd layer $\mathbf{p_0 \leq \alpha \cap p_1 \geq \beta}$:
Evidence against the null of no activation and no evidence against activation ⇒ NO GO for surgical resection.

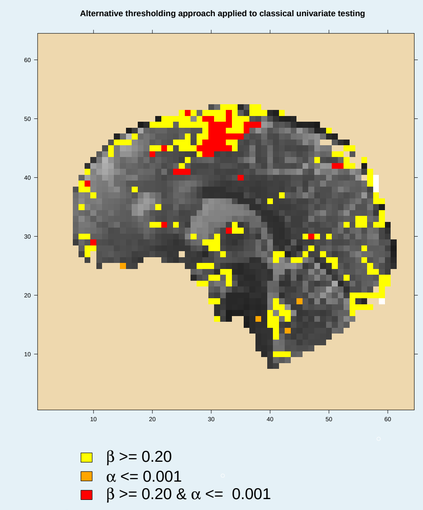
3rd layer $\mathbf{p_0 > \alpha \cap p_1 < \beta}$:
Activation can be confidently excluded.

5. Data

- Data: patient with left prefrontal brain tumor.
- Box-car design: alternate between recitation of tongue-twisters and quiescence.
- For analysis: FSL used.
- $\theta = 0.7\%$ BOLD change
- $\tau = 0.3\%$ BOLD change

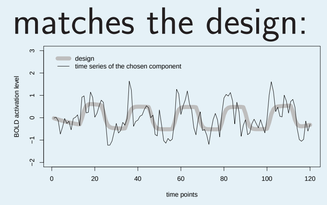
6. Alternative Thresholding (GLM)

- Red: H_0 rejected and H_a not rejected: strong evidence for the effect.
- Yellow: neither H_0 nor H_a can be rejected: lack of confidence in ruling out activation.
- No color: H_0 cannot be rejected but H_a can: good evidence for a lack of activation.
- Key strength of the procedure: among voxels traditionally classified as nonactive: procedure distinguishes between voxels with compelling evidence for non-activation and voxels where we cannot rule out the possibility of activation.

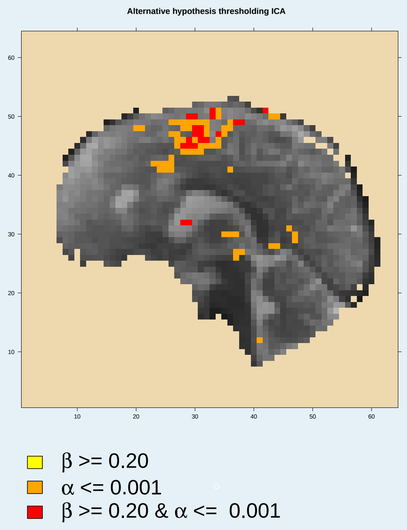


7. Alternative Thresholding (ICA)

- Define a meaningful BOLD effect sizes with ICA: see Durnez, Moerkerke, Bartsch and Nichols. (in press).
- For a given component who's temporal mode matches the design:



- **Differences GLM vs. ICA:** Smaller voxelwise variances.



8. Discussion and conclusion

- We argue for a control of false negatives while preserving control of the false positive rate.
- This results in multiple layers of significance:
  1. The first layer represents voxels exhibiting strong evidence against the null of no activation.
  2. The second layer is formed by voxels for which activation cannot be confidently excluded.
  3. The third layer then consists of voxels for which the presence of activation can be confidently rejected.
- We use voxelwise inference: maximal spatial precision is needed for pre-surgical fMRI.
- Main limitation of the procedure: Need to define the expected effect size and variance: literature on power studies available to guide estimation.
- The procedure is presented for both GLM and ICA.
- We would like to stress that this procedure does not abandon null hypothesis significance testing. The classical significance testing framework is still included in the procedure, represented by one layer of significance.