Toolkit of Cognitive Neuroscience

Statistical Testing of Electrophysiological Data

Eric Maris

Radboud University Nijmegen

Outline

1. Why do we need statistics in neuroscience?
   - Inferential statistics and data analysis
   - The interocular traumatic test (ITT) and why it fails so often
2. Inferential statistics is principled decision making under uncertainty
3. The Neyman-Pearson approach
   - Its principles and how it deals with the multiple-comparison problem
   - Between- and within-units-of-observation designs
4. The Permutation-based approach
   - Comparison with the Neyman-Pearson approach
   - Home-made test statistics
5. To remember

Why Do We Need Statistics in Neuroscience?

1. Statistics helps in making decisions under uncertainty. This branch of statistics is called *inferential statistics*, and its tools are *statistical tests*.
2. Statistics provides methods that reveal patterns in the data which cannot be identified by eyeballing. This branch of statistics is called *data analysis*.

This talk is about inferential statistics.

Why Do We Need Statistics in Neuroscience?

- Because the *interocular traumatic test* (Berkson, 1950) very often fails.
- The interocular traumatic test (ITT):

  "When you look at the data, the conclusion hits you right between the eyes!"

- A study in which the ITT does work:

  In a sample of 20 participants, *all* show the same difference pattern (across space, frequency, and time) between condition A and condition B.

- Very often, the ITT fails because the observer is uncertain about the conclusion that can be drawn, or because different observers disagree with respect to the conclusion.

Why Does the ITT often Fail in Neuroscience?

1. The signal-to-noise ratio of many measurements (EEG, MEG, fMRI-measured BOLD-response) is small, especially at the single-trial level.
2. The measurements are high-dimensional data structures that cannot be compared on the basis of visual inspection alone.
   - MEEG: one comparison for every [sensor,time]-pair or [sensors,frequency,time]-triplet (called *samples* in the following)
   - fMRI: one comparison for every voxel

Is There a "Statistics for Neuroscience"?

- *No*, the general principles of statistical decision making also apply to neuroscience
- But neuroscience has its own statistical problems that require specialized methods
- These statistical problems are mainly the result of the high dimensionality of typical neuroscience data
- The *multiple comparisons problem* (MCP): the false alarm (type I error) rate increases with the number of comparisons

Inferential Statistics is Principled Decision Making

- Inferential statistics is decision making based on rational principles
- The neuroscientist can choose between multiple rational principles
- One can arrive at different conclusions depending on the principle on which the decision is based

In this sense, *statistics is not about the truth*.

Which Rational Decision Making Principles are Used?

1. The Neyman-Pearson approach
2. False discovery rate control
3. The permutation-based approach
4. The Bayesian approach

- I will discuss the Neyman-Pearson approach (the leader in the field) and the permutation-based approach (my personal favorite)
- I will focus on how these approaches deal with the MCP

Outline of the Discussion

For every statistical decision making principle, we will do the following:

1. Illustrate the statistical testing procedure by means of an example (i.c., the difference between two means)
2. Explain the rationale behind the procedure
3. Show how the MCP is solved
4. Evaluate this solution

The Neyman-Pearson Approach

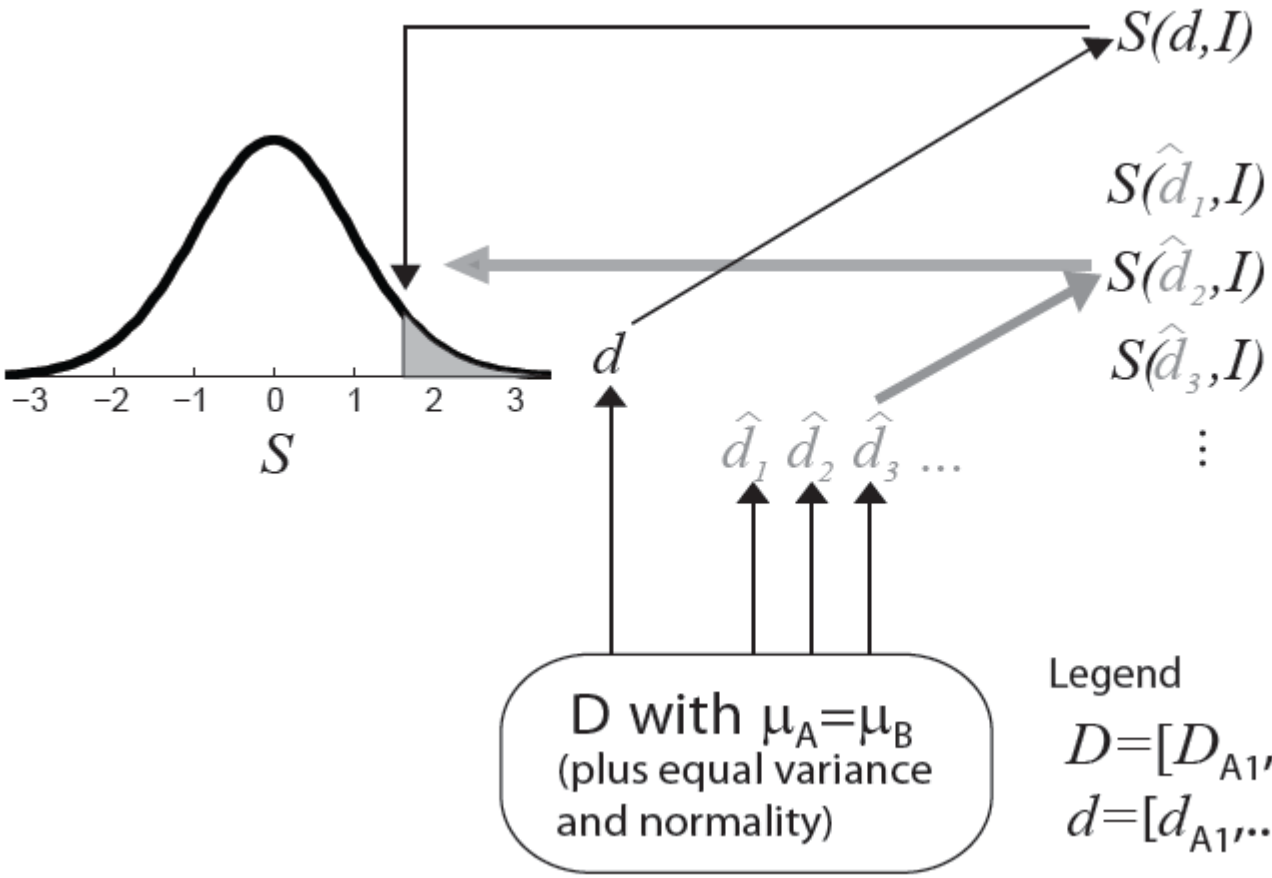
1. Formulate a so-called *null hypothesis* in terms of unknown population parameters. For example: the difference between the expected values in two populations (A and B).
2. Take some *test statistic*. For example: the two-sample T-statistic.
3. Find a *critical value* for the test statistic such that, under the null hypothesis, the probability of exceeding this critical value is controlled (e.g., at 0.05). This critical value is defined by some critical p-value under the *sampling distribution*.

The Neyman-Pearson Approach

Constraints:

1. The probability distribution of the test statistic under the null hypothesis has to be known. This requires *auxiliary assumptions* about the data (normality, equal variance, independence), which may be false.
2. Under the *alternative hypothesis*, the probability of exceeding the critical value must be large. In other words, the test must be powerful.

The Neyman-Pearson Approach



The Neyman-Pearson Approach

How is the MCP solved?

1. Formulate a null hypothesis in terms of unknown population parameters at *all* elements of the multivariate random variable *D* (electrophys: all [channel,frequency,time]-triplets; fMRI: all voxels)
2. Take a test statistic that depends on the data at all elements jointly. For example: the maximum (over all elements) T-statistic, the size of the largest cluster exceeding some threshold
3. Find a critical value for this test statistic such that, under the null hypothesis, the probability of exceeding this critical value is controlled

The Neyman-Pearson Approach

Evaluation

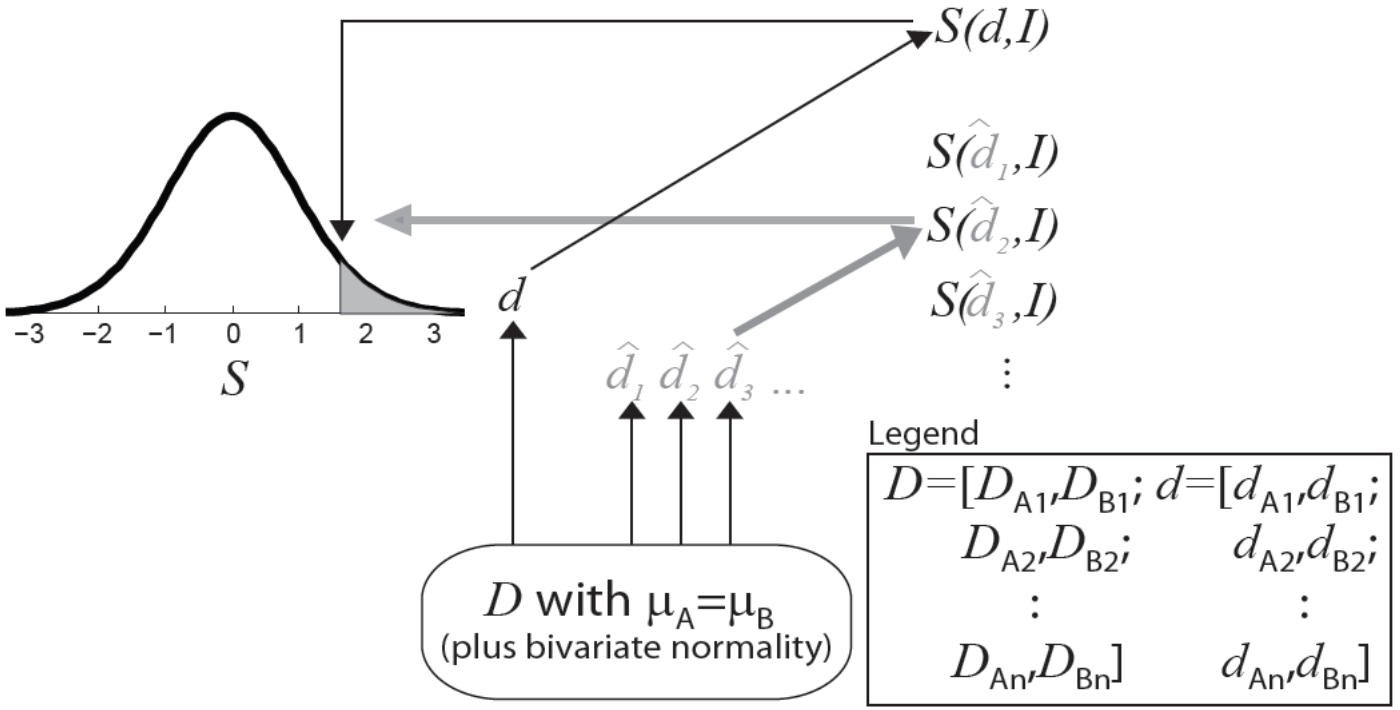
1. The distribution under the null hypothesis is only known under auxiliary assumptions (e.g., a Gaussian Random Field for the T-statistics), and these may be false.
2. The test statistic for which the distribution is known may have a very low power.

Terminology: Between- and Within UO-designs

The observed data: a collection of trials (single-subject study) or a collection of subjects (multi-subject study). Trials and subjects are called *units of observation* (UO).

Two possible experimental designs: a *between-UO* design and a *within-UO* design. In a between-UO design, every UO (trial or subject) is observed in only a single experimental condition, and in a within-UO design, every UO is observed in ALL experimental conditions.

The Neyman-Pearson Approach for a Within-UO Design



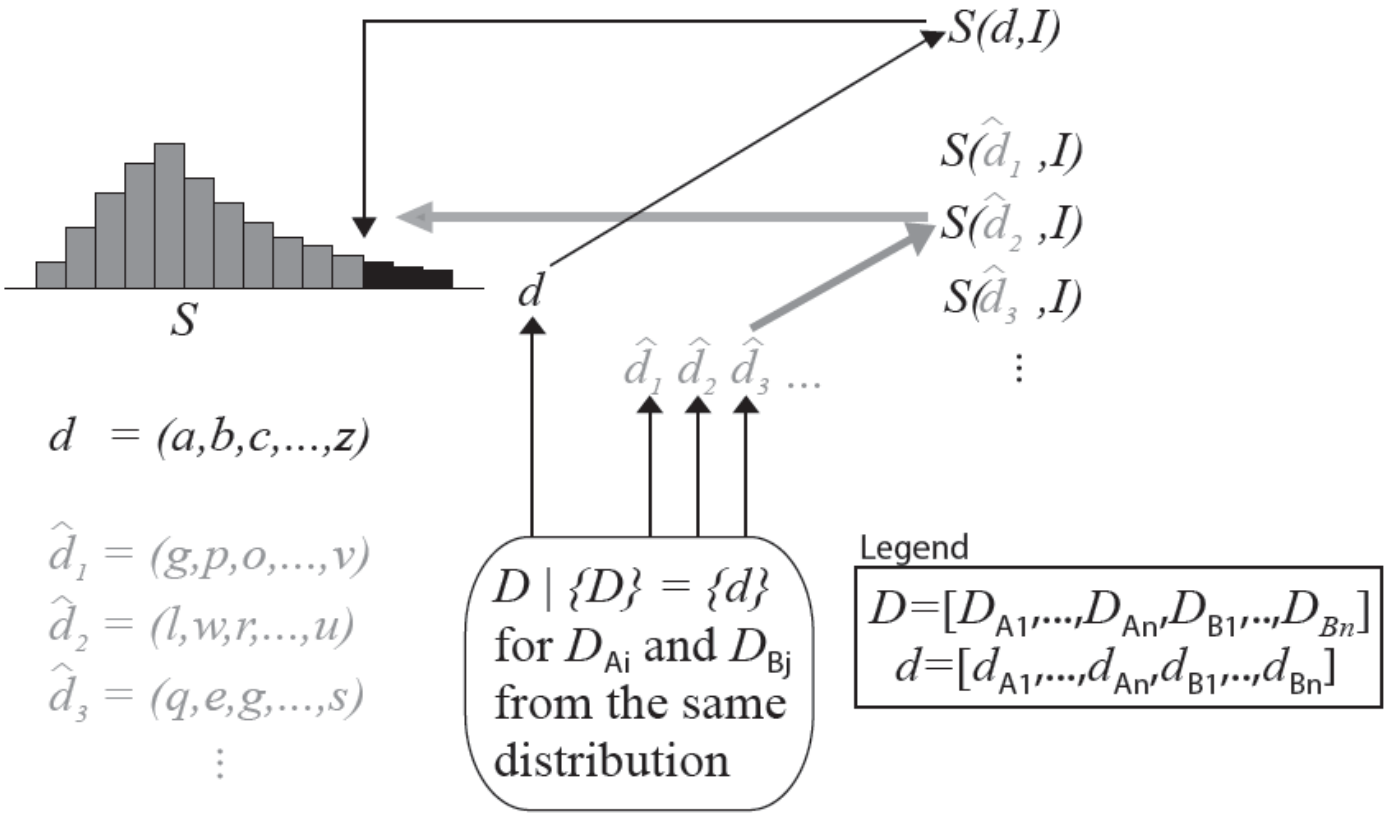
The Permutation-based Approach

1. Formulate a null hypothesis in terms of the *probability distributions* of the observations: the probability distributions of the observations in the different experimental conditions are identical.
2. Take some test statistic.
3. Find a critical value for the test statistic such that, under the null hypothesis, the probability of exceeding this critical value is controlled (e.g., at 0.05). This critical value is defined by some critical p-value under the *permutation distribution*.

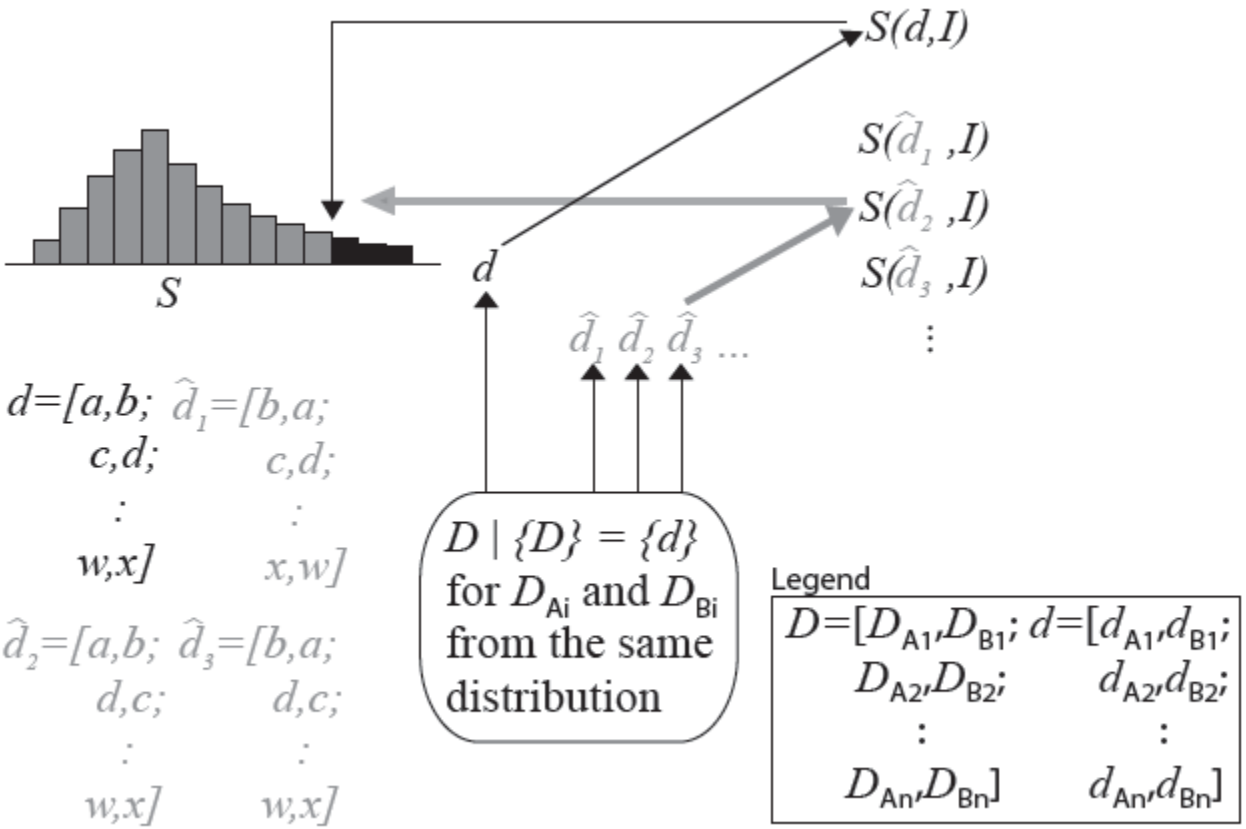
Differences with the Neyman-Pearson Approach

1. The null hypothesis is about the whole probability distributions of the observations in the experimental conditions, and not about some unknown population parameters.
2. The p-value under the permutation distribution can be calculated for *every* test statistic without any auxiliary assumption
3. If you want your test to be sensitive for a particular aspect of the data, you must choose your test statistic accordingly.

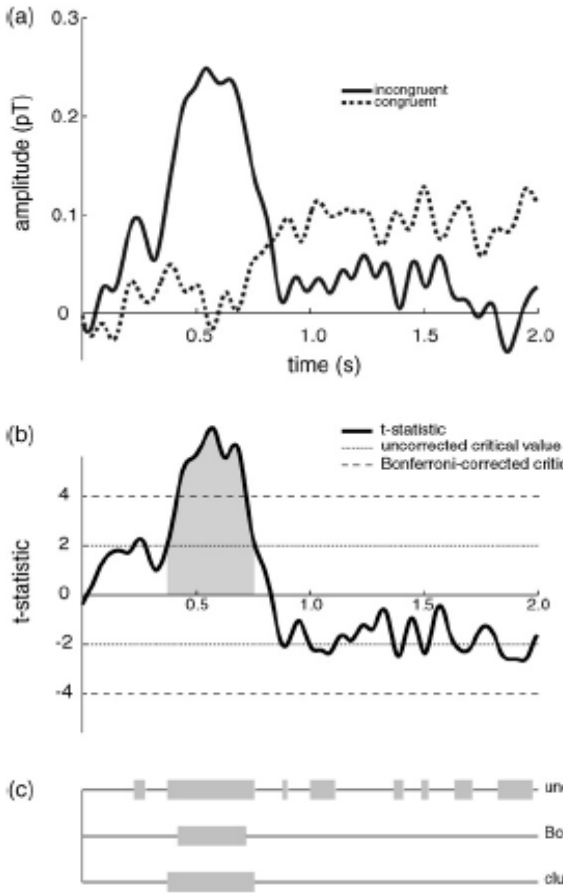
The Permutation-based Approach for a Between-UO Design



The Permutation-based Approach for a Within-UO Design



The Permutation-based Approach: A *Home-Made* Test Statistic



After randomly permuting congruent and incongruent trials:

Random differences between the dashed (randomly assigned *congruent* trials) and the solid (randomly assigned *incongruent* trials) lines.

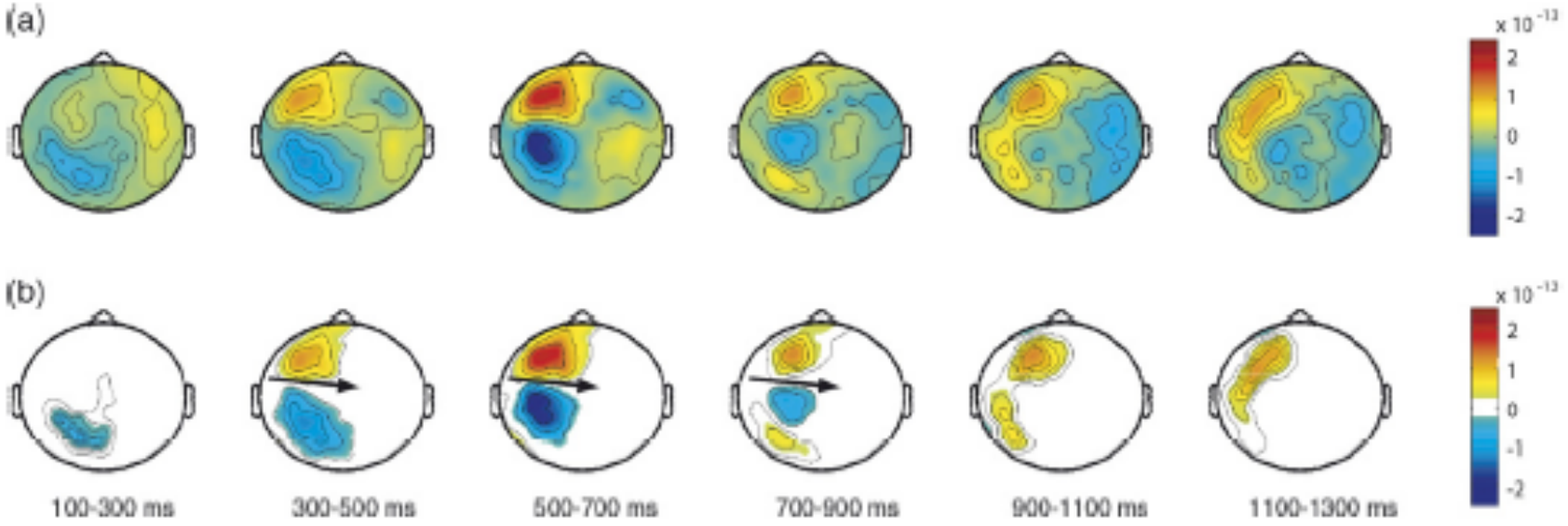
Random fluctuations of the t-statistic signal around 0, exceeding the 5% univariate critical values (-1.96 and 1.96) only on 5% of the samples.

Based on the permutation distribution of the *maximum cluster-level sum*

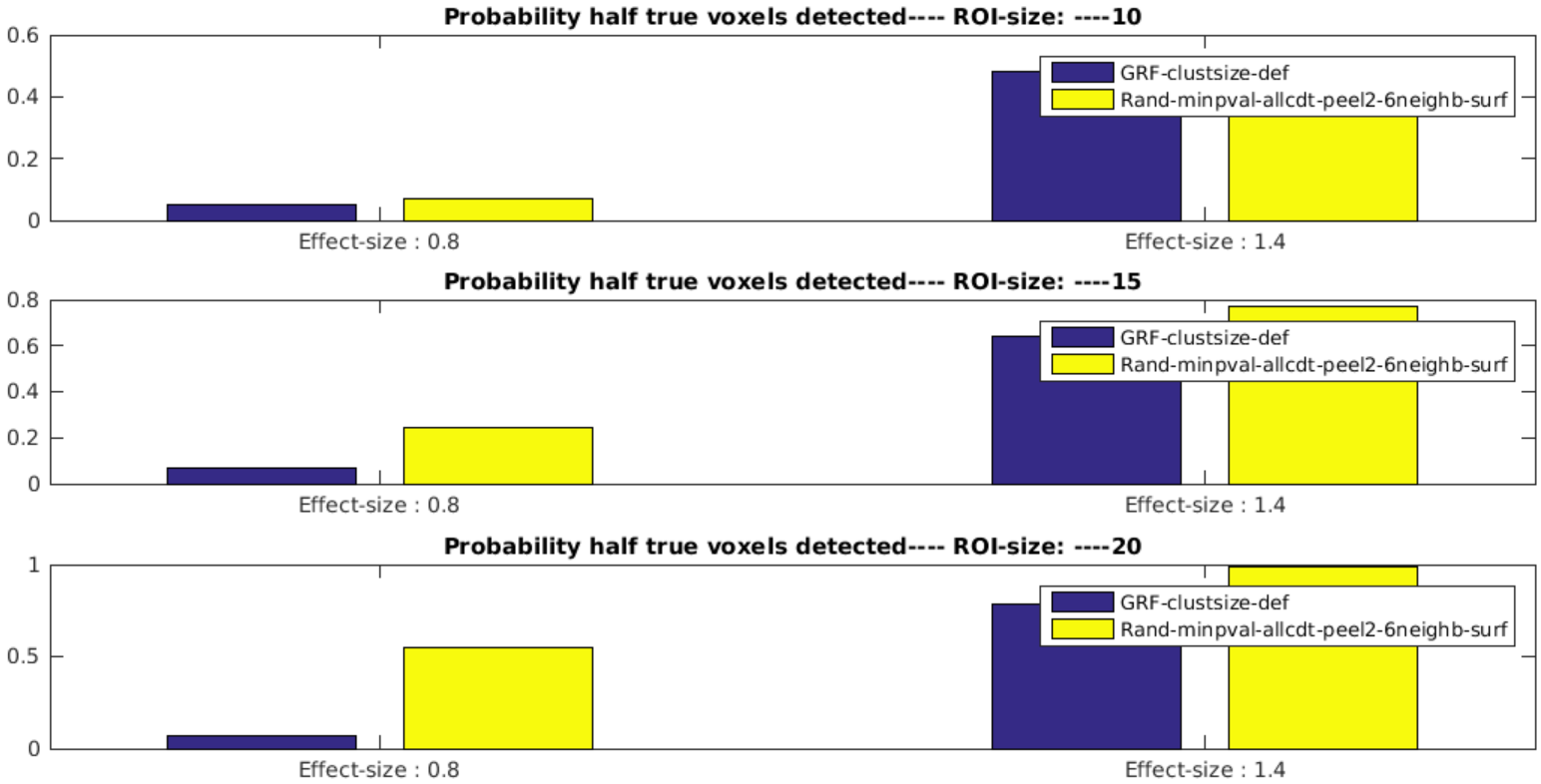
The Permutation-based Approach: A *Home-Made* Test Statistic

1. Calculate a T-statistic for each of the samples in the multidimensional data structure (time-samples, [channel,time]-pairs, [channel,frequency,time]-triplets).
2. Threshold these sample-specific statistics.
3. Construct connected clusters of samples that (1) exceed the threshold and (2) have the same sign.
4. Calculate the cluster-level statistics by taking the sum of the sample-specific T-statisics.
5. Take the maximum of the cluster-level statistics.
6. Evaluate this maximum under its *permutation distribution*.

The Permutation-based Approach: Another *Home-Made* Test Statistic



The Randomization Approach for fMRI data: Comparison with GRF-based Inference



Permutation-based and the Neyman-Pearson Approach to the MCP

The MCP is solved in the same way as in the Neyman-Pearson theory, but now the evaluation is much more positive:

1. The distribution under the null hypothesis does not depend on auxiliary assumptions
2. The test statistic, which depends on the data at all samples jointly, can be chosen such that it is maximally sensitive to effects that are biophysically plausible.

To Remember

- Statistics is about decision making under uncertainty.
- Due to the low signal-to-noise ratio of most biological signals, and the dimensionality of its data structures, neuroscience cannot do without statistics.
- Every decision making principle that is applied in neuroscience must be able to solve the MCP.
- Permutation tests are ideally suited for neuroscience:
  1. They solve the MCP without making auxiliairy assumptions.
  2. They can increase sensitivity by incorporating biophysically plausible constraints in the test statistic.

References

- Maris E. (2004) Randomization tests for ERP topographies and whole spatiotemporal data matrices, *Psychophysiology*, 41 (1): 142-151
- Maris E., Oostenveld R. (2007) Nonparametric statistical testing of EEG- and MEG-data**.** *J Neurosci Methods, 164*, 177-190.
- Maris E., Schoffelen J.M., Fries P. (2007) Nonparametric statistical testing of coherence differences**.** *J Neurosci Methods*, 163(1):161-75.
- Maris E. (2012) Statistical testing in electrophysiological studies, *Psychophysiology,* 49, 549–565.

References

- Applications in Science, Nature, Neuron, Journal of Neuroscience, PNAS, NeuroImage, Human Brain Mapping, etc.
- http://www.ru.nl/fcdonders/fieldtrip