

Non-parametric inferential statistical testing with permutation and clustering

Jan Mathijs Schoffelen

Donders Institute, Radboud University, Nijmegen, NL

janmathijs.schoffelen@donders.ru.nl

Talk outline

Inferential (parametric) statistics

Problems with parametric mass-univariate tests

Solutions to the problems

Non-parametric permutation testing with clustering

Some things to think about

Talk outline

Inferential (parametric) statistics

Problems with parametric mass-univariate tests

Solutions to the problems

Non-parametric permutation testing with clustering

Some things to think about

What types of statistics do we have?

How do large distributions of “something” behave?

Binomial, Gaussian, Poisson

How can I describe (or summarize) a distribution?

Mean, standard deviation, variance, kurtosis, parameter ‘x’

How can I make a decision or draw a conclusion?

Inferential statistics, hypothesis testing



“Data don’t make any sense,
we will have to resort to statistics.”

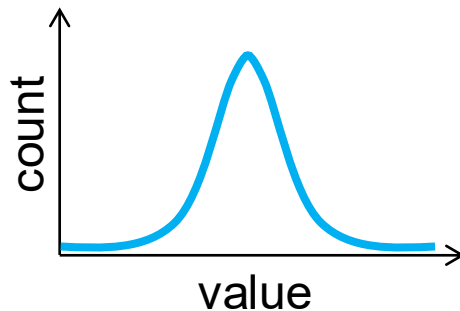
making decisions under uncertainty

Inferential parametric statistics

You make N observations and want to find whether some hypothesis “H1” holds

Step 1: Gather data

Observation	Value
0	1.65
1	1.82
⋮	
N	1.70

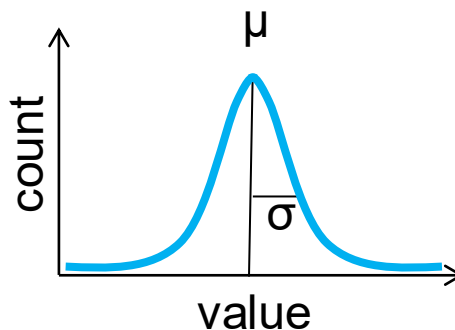


Inferential parametric statistics

You make N observations and want to find whether some hypothesis “H1” holds

Step 2: Compute the parameter(s) of interest

Observation	Value
0	1.65
1	1.82
⋮	
N	1.70

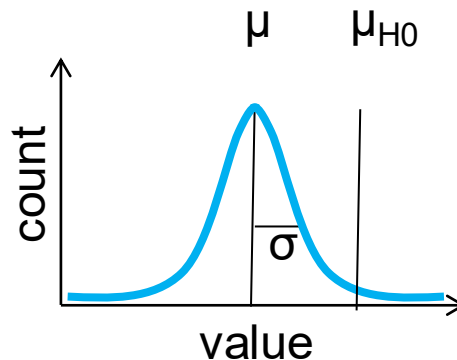


Inferential parametric statistics

You make N observations and want to find whether some hypothesis “H1” holds

Step 3: Statistical testing

Observation	Value
0	1.65
1	1.82
⋮	
N	1.70



Determine probability of t under “H0”

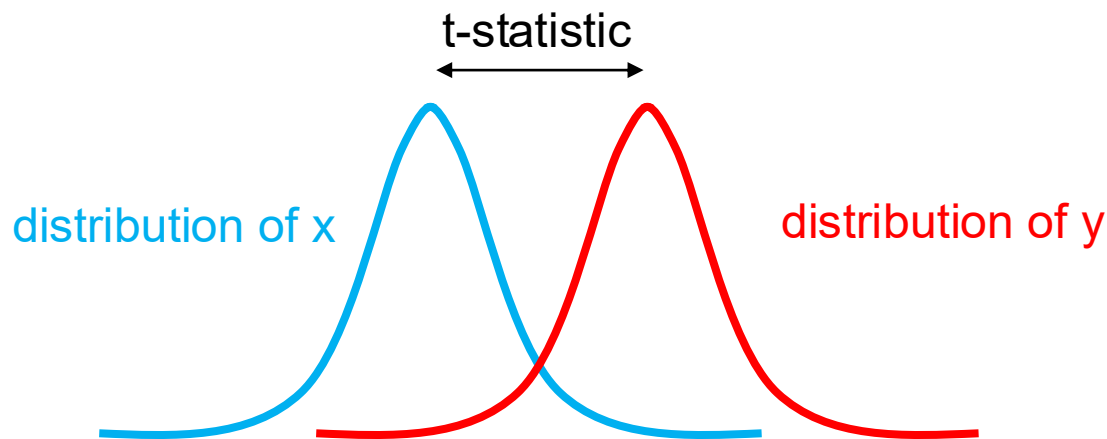
$$t = \frac{\mu - \mu_{H0}}{\sigma / \sqrt{N}}$$

If the observed t sufficiently unlikely, reject H0 in favour of H1

Inferential parametric statistics

Observations
in condition 1:
 $\{x_1, x_2, x_3, x_4, \dots\}$

Observations
in condition 2:
 $\{y_1, y_2, y_3, y_4, \dots\}$



Talk outline

Inferential (parametric) statistics

Problems with parametric mass-univariate tests

Solutions to the problems

Non-parametric permutation testing with clustering

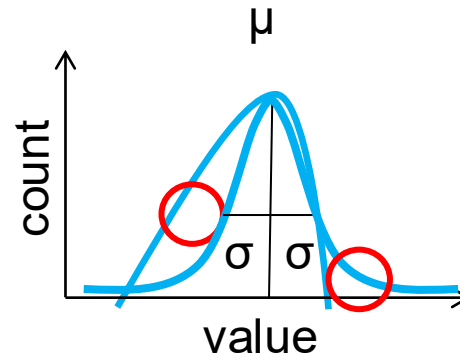
Some things to think about

Parametric statistical testing: problem 1

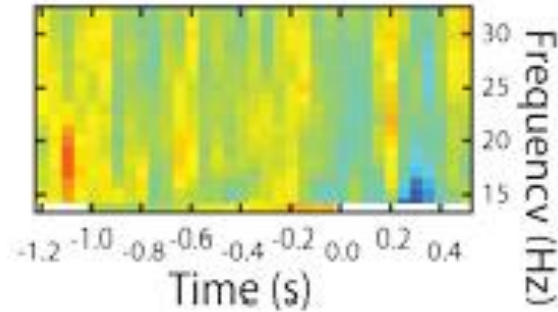
You make N observations and want to find whether some hypothesis H_1 is true.

The first problem is that this requires a *known distribution* of the test statistic.

Observation	Value
0	2.5
1	-3.2
...	
N	2.4



Parametric statistical testing: problem 2



Mass-univariate testing amounts to many decisions

16 * 30 (e.g.) time-frequency tiles, i.e. 480 comparisons.

T-test with $\alpha = 0.05$ (chance of false alarm rate) **for one test**

99.99...% chance of at least one false alarm in 480 tests

$480 * 0.05 = 24$ false alarms expected

And it gets worse with multiple channels, dipoles in the brain

Talk outline

Inferential (parametric) statistics

Problems with parametric mass-univariate tests

Solutions to the problems

Non-parametric permutation testing with clustering

Some things to think about

Solutions to control the family-wise error rate

Bonferroni correction (super conservative)

Do something fancy with Gaussian random field theory (less conservative)

Use the false discovery rate (is not the same as FWER control)

Use a Monte Carlo approximation of the
permutation distribution of some maximum statistic

Reduce the number of formal tests to 1 (or 2, if two-sided)

Talk outline

Inferential (parametric) statistics

Problems with parametric mass-univariate tests

Solutions to the problems

Non-parametric permutation testing with clustering

Some things to think about

Permutation test: general principle

- *Independent variable: condition*
- *Dependent variable: data*

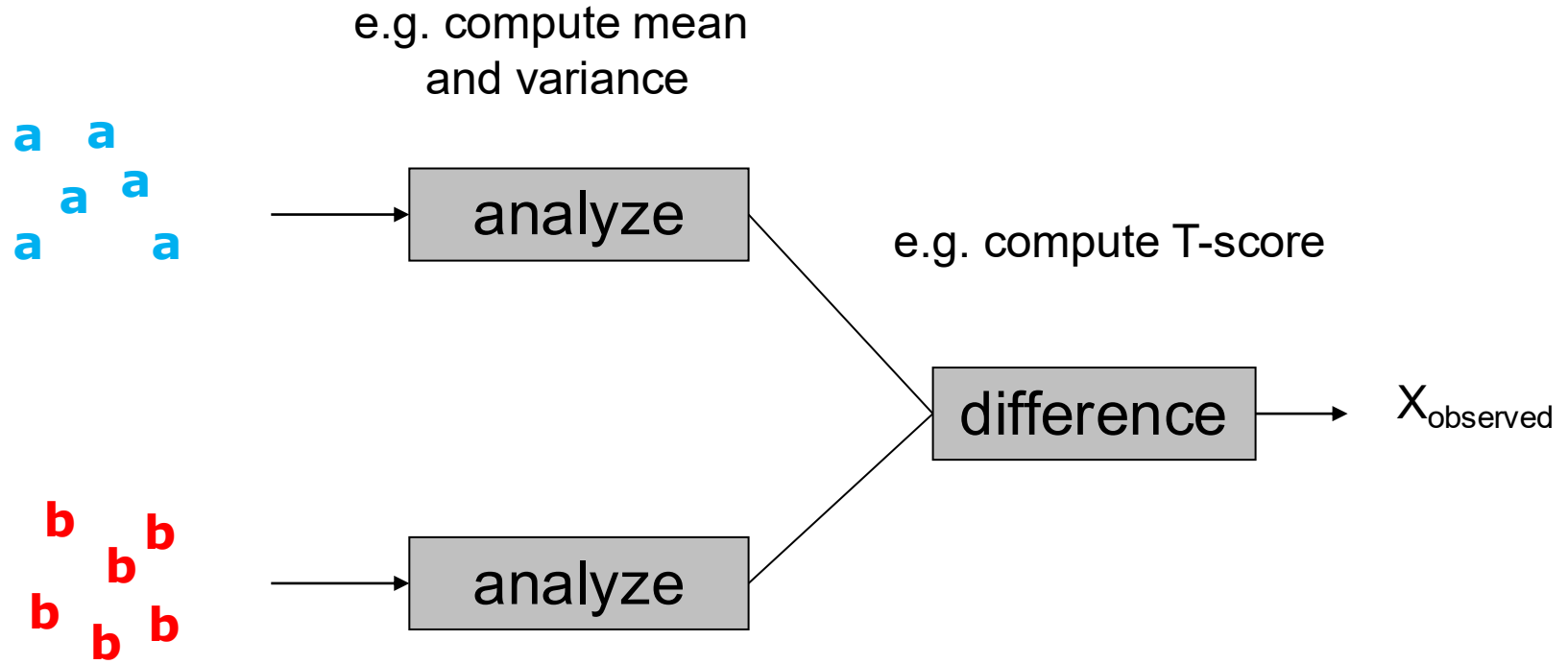
H0: the data is **independent** from the condition in which it was observed

The data in the two conditions is **not** different

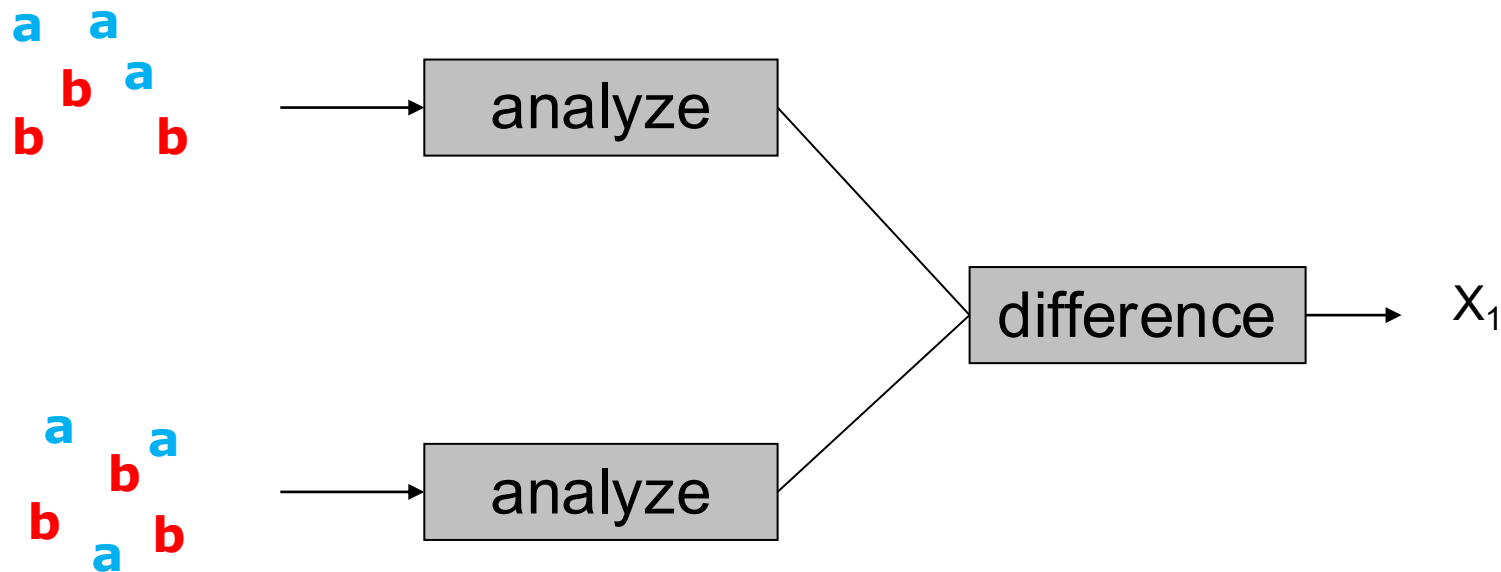
More formally: null hypothesis is about exchangeability between conditions

The test is not about a parameter anymore

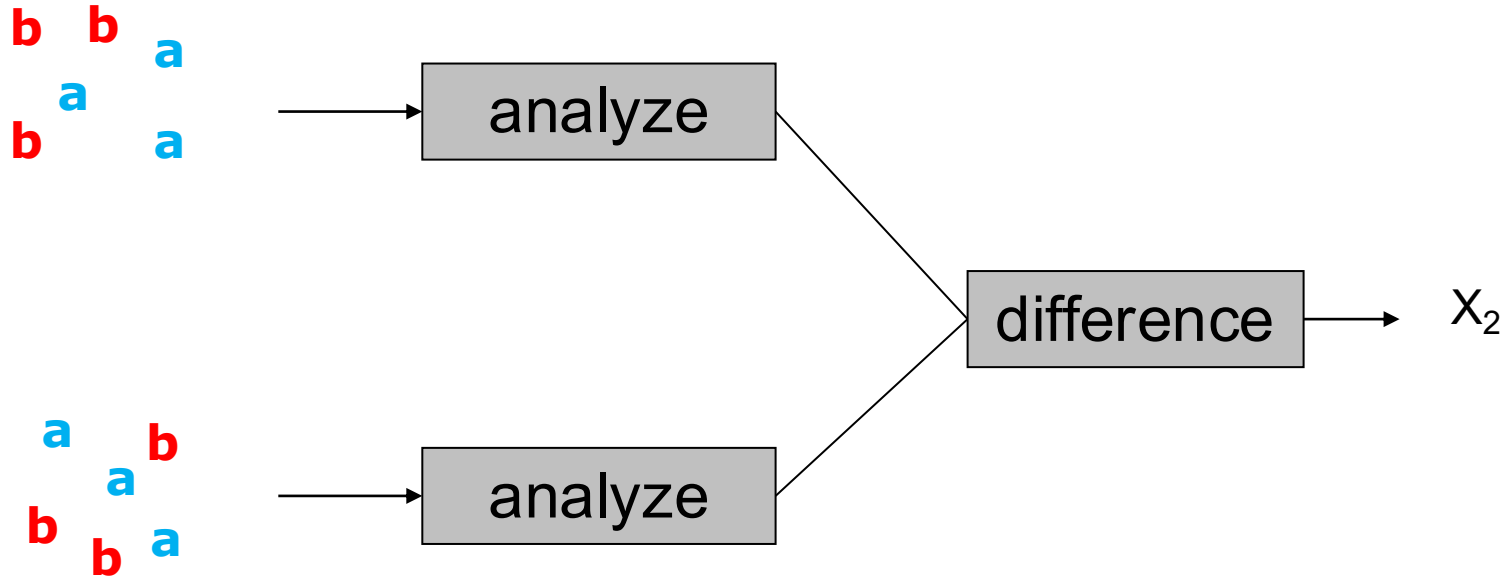
Permutation approach



Permutation approach

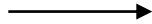


Permutation approach



Permutation approach

b **b** **a**
a **a**
b **a**

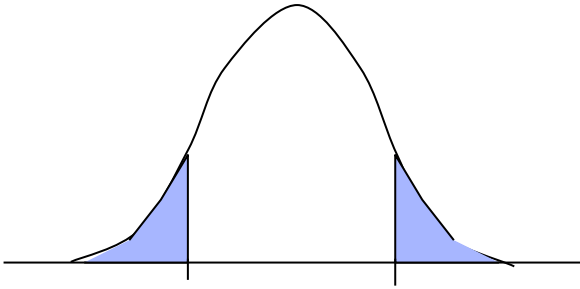


→ X_2

a **b**
a **b**
b **a**



Distribution of “x” can take any shape



Non-parametric statistics

Permutation of independent variable

Hypothesis is about data, not about the specific parameter

The distribution of the statistic of interest “x” is approximated using the Monte-Carlo approach, i.e. by random sampling

H0 is tested by comparing the observed statistic against the randomization distribution

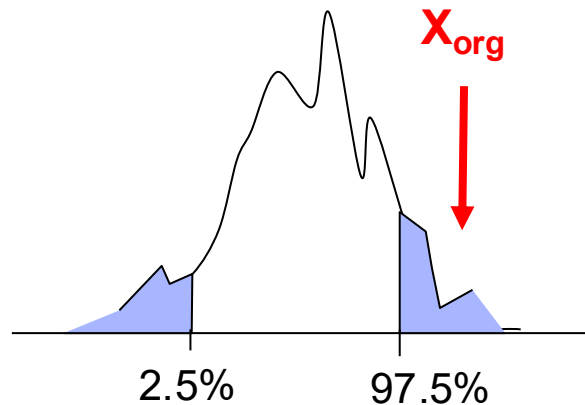
Avoid the multiple comparison problem

The statistic “x” can be anything

Rather than testing everything, only test the most extreme observation (i.e. the max statistic)

Compute the permutation distribution for the most extreme statistic over all channels/times/frequencies

Note that often we compute **two** extrema, one for each tail



Increasing the sensitivity of permutation tests

channel/time/frequency points are not independent

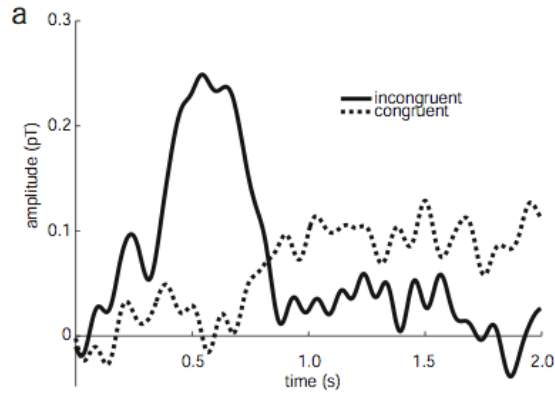
neighbouring channel/time/frequency points are expected to show similar behaviour

combine neighbouring samples into clusters ->

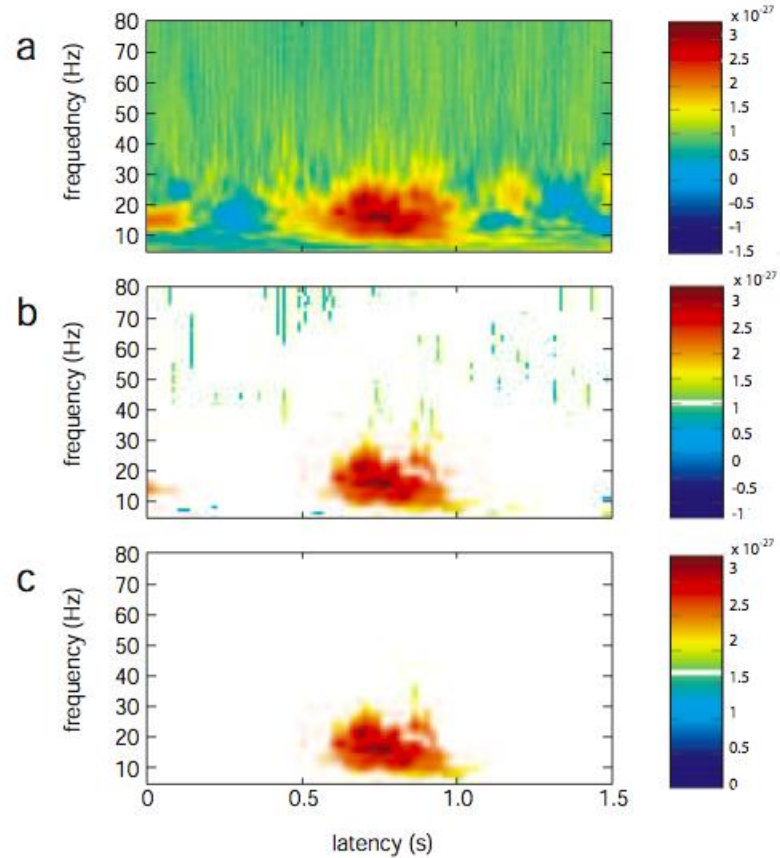
“accumulate the evidence” = cluster-based statistics

avoid the MCP by comparing the largest observed cluster versus the permutation distribution of the largest clusters

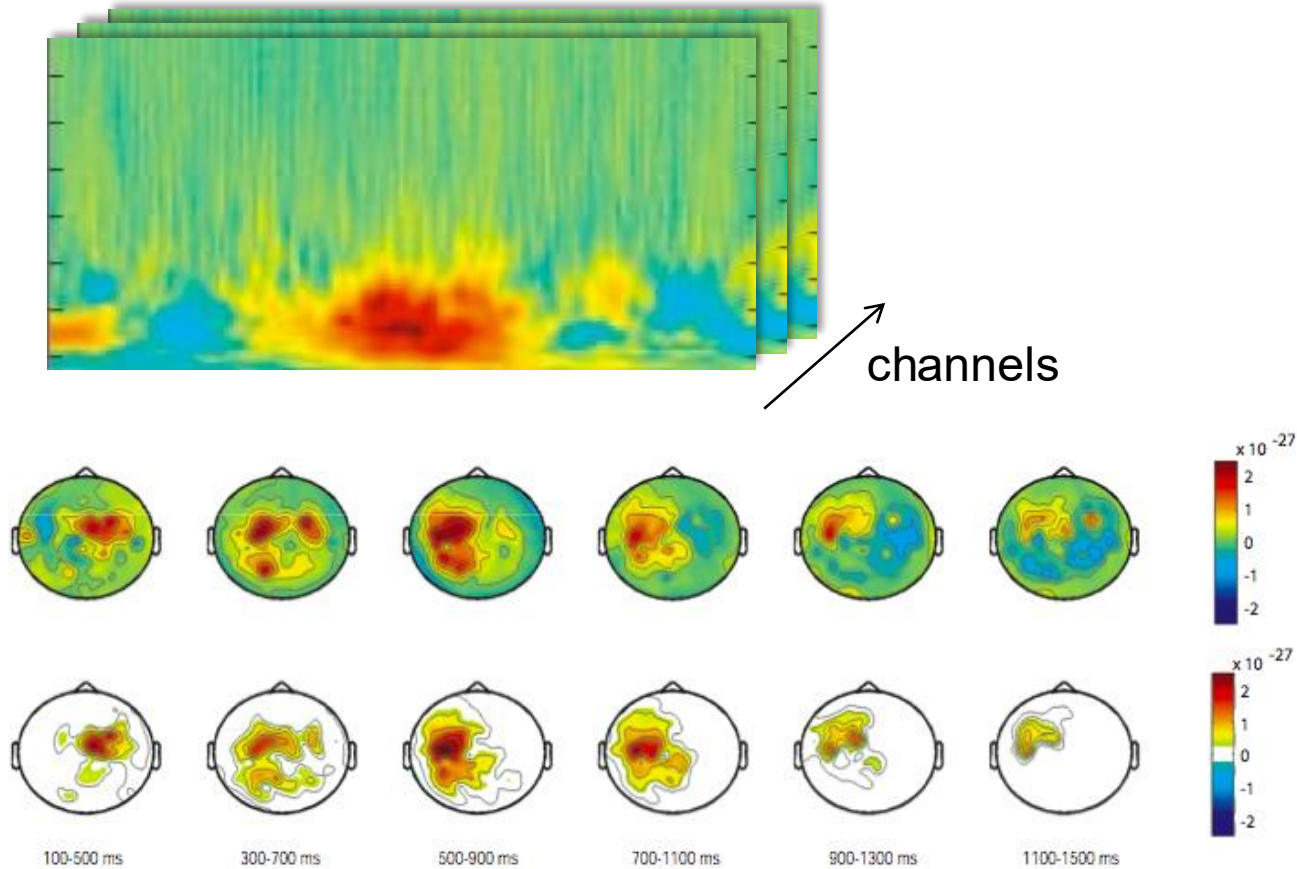
Clustering in time



Clustering in time and frequency



Clustering in time, frequency and space



Summary: mass-univariate statistical inference

Parametric statistical test for all channel-time-frequency points

- probability for H_0

- one H_0 for each channel-time-frequency

- multiple comparison problem

Non-parametric approach for estimating probability

- randomization or permutation

- probability of H_0 (exchangeability) for arbitrary statistic

- incorporate prior knowledge in statistic

- avoid MCP using max statistic

Caveats

The p-value is about the H_0 hypothesis, which is “exchangability of data between conditions”.

The largest cluster is not significant (the test is).

The second-largest cluster does not play a role in the decision

The cluster itself is not always informative.

Following the decision, there is still the need for interpretation.

Talk outline

Inferential (parametric) statistics

Problems with parametric mass-univariate tests

Solutions to the problems

Non-parametric permutation testing with clustering

Some things to think about

Some things to think about

How to interpret the outcome?

Randomization test for data with a bias

How to test parametric effects?

How to test a single condition?

How to choose the thresholds?

Alternatives for clustering?

Some things to think about

How to interpret the outcome?

Randomization test for data with a bias

How to test parametric effects?

How to test a single condition?

How to choose the thresholds?

Alternatives for clustering?

How to interpret the outcome?



“There is no significant cluster”

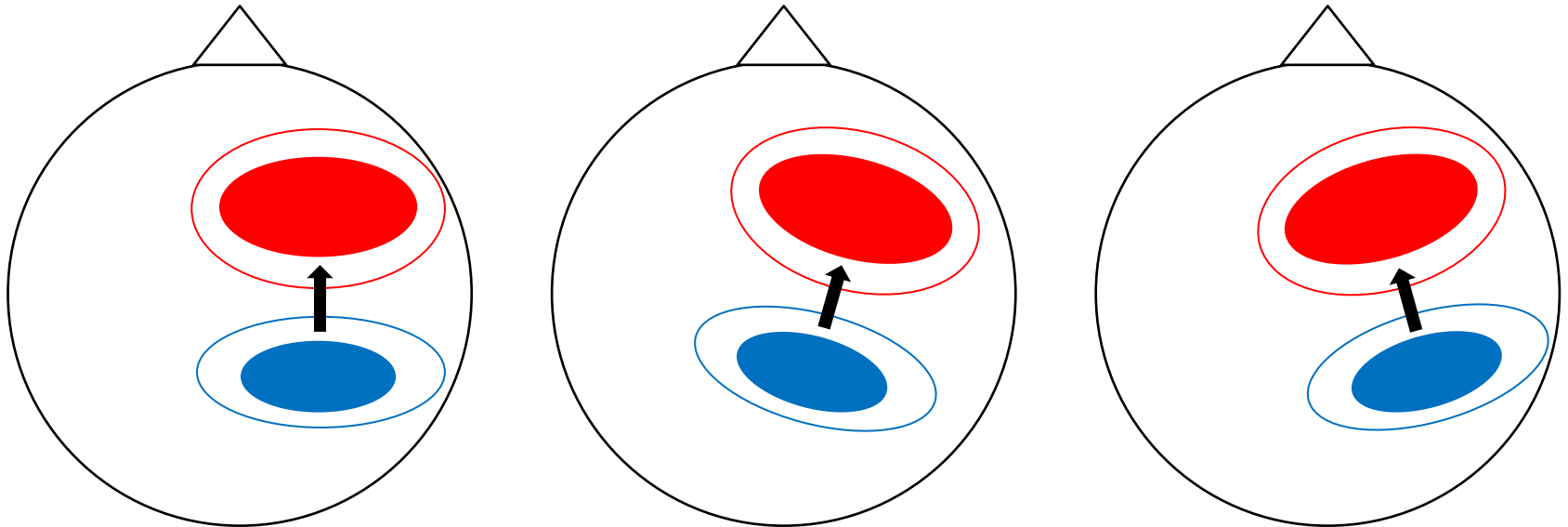
Well, there is. But ...

The computed probability tells us that H_0 is unlikely,
i.e., the decision that you make is that the *data* is are not exchangeable

How to interpret the outcome?

Variability in the orientation of an underlying dipolar source over participants

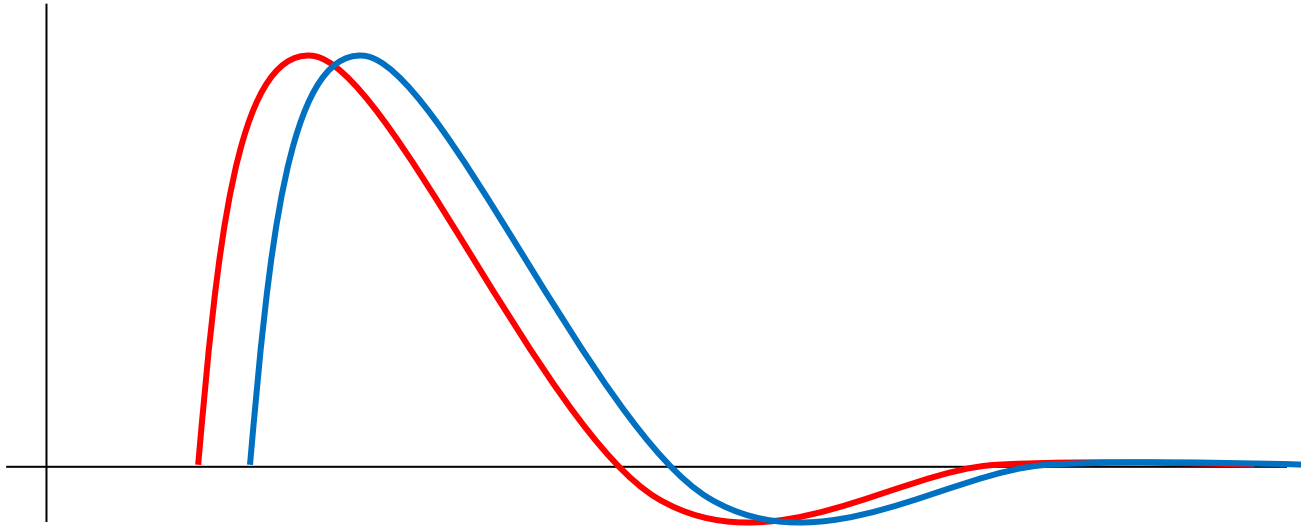
-> only the positive cluster may show up as “significant”



How to interpret the outcome?

Difference in the latency of the ERF component

-> only the steep rising flank may show up as “significant”



How to interpret the outcome?

Inferential statistics is about the ***decision*** making

After the decision that there is a significant difference, you still have to do the ***interpretation***. That may or may not involve the cluster on basis of which you made the decision.

For the interpretation you should use your understanding of the data, the recording system, the analysis methods, the potential confounds, etc.

The largest cluster may not localize the effect properly.

Some things to think about

How to interpret the outcome?

Randomization test for data with a bias

How to test parametric effects?

How to test a single condition?

How to choose the thresholds?

Alternatives for clustering?

Randomization test for data with a bias

80 trials in condition A, 20 trials in condition B

Question: is there a difference in source distribution?

However, for source reconstruction we know

Condition A: sharp (better quality) source reconstruction

Condition B: blurry (lower quality) source reconstruction

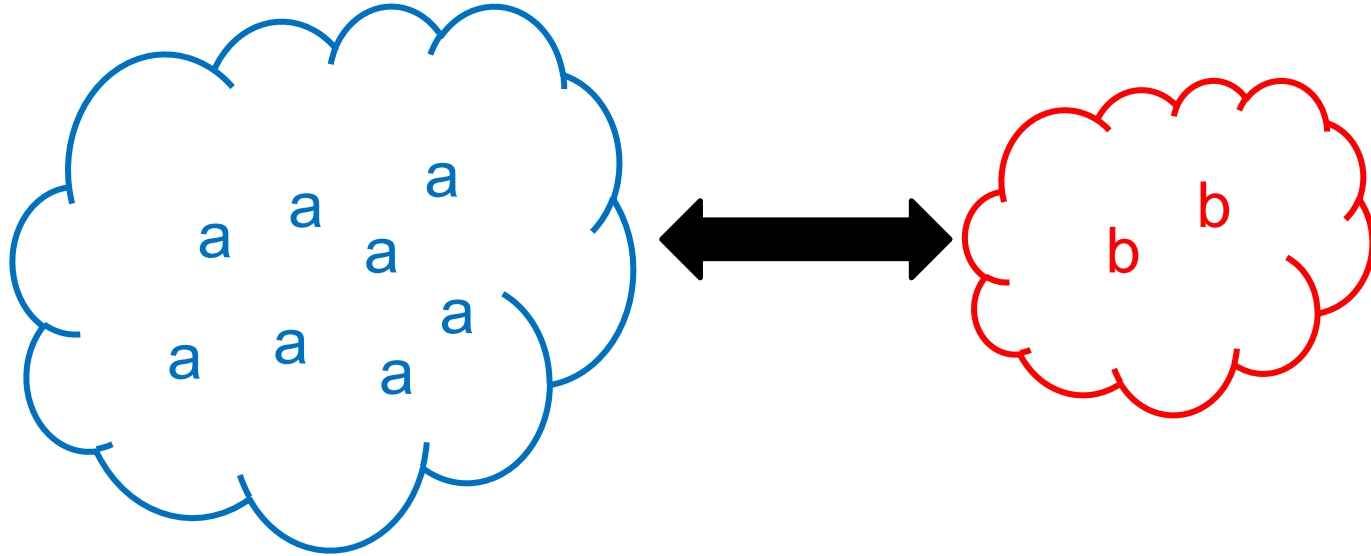
Question: is there a difference in fronto-occipital coherence?

However, for many connectivity estimates we know

Condition A: lower (less biased) coherence

Condition B: higher (more biased) coherence

Randomization test for data with a bias



Keep both conditions the *same size* with each shuffle

Include the bias in the estimate of the randomization distribution

> [J Neurosci Methods](#). 2007 Jun 15;163(1):161-75. doi: 10.1016/j.jneumeth.2007.02.011.
Epub 2007 Feb 20.

Nonparametric statistical testing of coherence differences

Eric Maris¹, Jan-Mathijs Schoffelen, Pascal Fries

Some things to think about

How to interpret the outcome?

Randomization test for data with a bias

How to test parametric effects?

How to test a single condition?

How to choose the thresholds?

Alternatives for clustering?

How to test parametric effects?

Unfamiliar or familiar or scrambled faces

Each stimulus was shown twice

	1st	2nd
unfamiliar	100	100
familiar	100	100
scrambled	100	100

faces vs scrambled

100	100
100	100
100	100

unfamiliar vs familiar

100	100
100	100
100	100

1st vs 2nd presentation

Wakeman and Henson (2015) Sci Data
Frontiers (2021), "From raw MEG/EEG to publication"

100	100
100	100
100	100

interaction (ANOVA)

How to test parametric effects?

Compute the difference between the 2nd presentation of a face minus the 1st presentation

Use this condition difference instead of the original data and shuffle that between unfamiliar and familiar faces

1st and 2nd presentation

100	100
100	100
100	100

difference between 1st and 2nd

100	unfamiliar familiar
100	
100	

Some things to think about

How to interpret the outcome?

Randomization test for data with a bias

How to test parametric effects?

How to test a single condition?

How to choose the thresholds?

Alternatives for clustering?

How to test a single condition?

We create “dummy” data in the unobserved condition,
i.e., we make it all zero.

This corresponds to flipping the +/- sign of the original data.

H0: the data can be swapped with zeros (thin ice!)

H0: the sign of the data can be swapped

How to test a single condition?

Two conditions: shuffle between conditions

Condition A

Condition B

S1_a

S1_b

S2_a



S2_b

S3_a

S3_b

S4_a

S4_b

S5_a



S5_b

S6_a



S6_b

S7_a

S7_b

S8_a

S8_b

S9_a

S9_b

S10_a



S10_b

How to test a single condition?

Condition A

Condition B

S1_a		0
S2_a		0
S3_a		0
S4_a		0
S5_a		0
S6_a		0
S7_a		0
S8_a		0
S9_a		0
S10_a		0

How to test a single condition?

Condition A

S1_a

0

S3_a

S4_a

0

0

S7_a

S8_a

S9_a

0

Condition B

0

S2_a

0

0

S5_a

S6_a

0

0

0

S10_a



Some things to think about

How to interpret the outcome?

Randomization test for data with a bias

How to test parametric effects?

How to test a single condition?

How to choose the thresholds?

Alternatives for clustering?

How to choose the thresholds?

There is one threshold for the clustering and another for the inferential decision

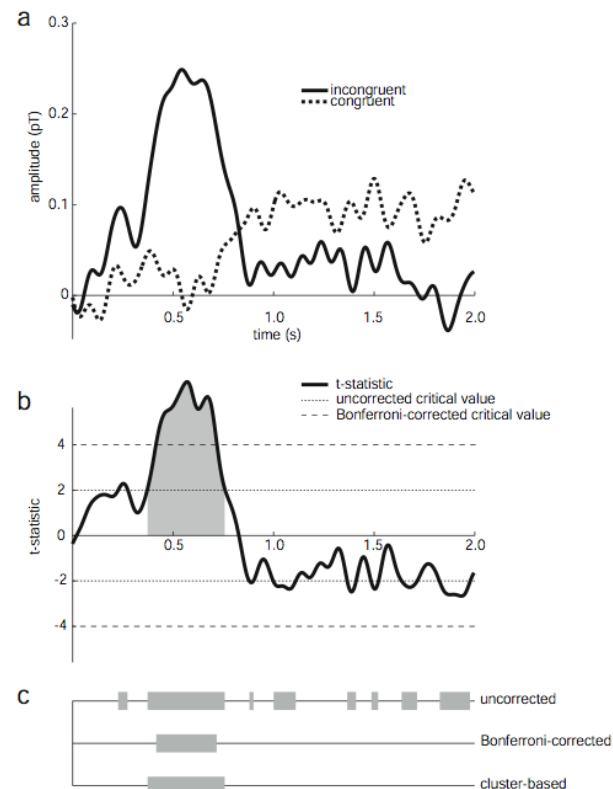
The cluster threshold does not affect the **validity** of the test, but it does affect the **sensitivity** of the test.

Too low, you'll get very large clusters

Too high, you won't get any clusters

By default we set it to the uncorrected 5% parametric threshold

Alternative: TFCE (also in FieldTrip, not discussed here)



Some things to think about

How to interpret the outcome?

Randomization test for data with a bias

How to test parametric effects?

How to test a single condition?

How to choose the thresholds?

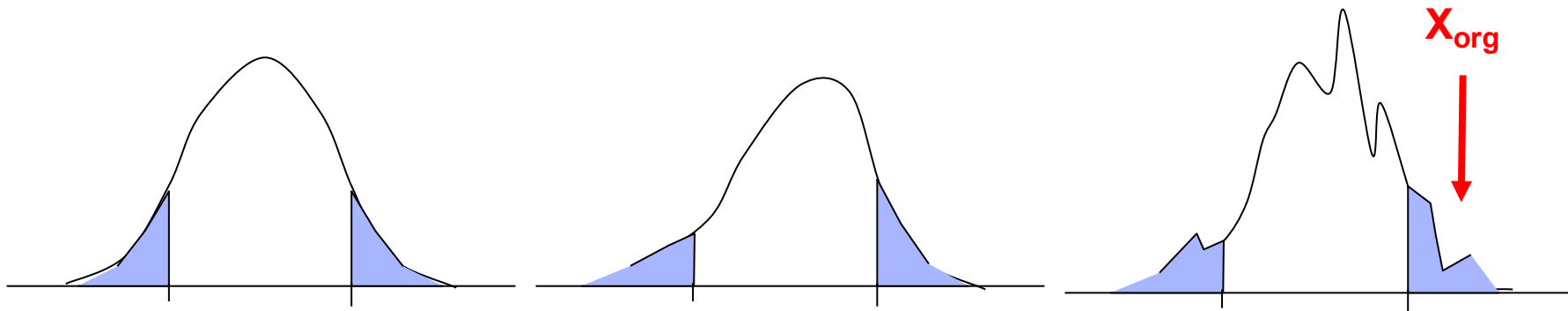
Alternatives for clustering?

Alternatives for clustering?

Clustering is used to convert the high-dimensional multivariate observations into a single statistical test

You can also use any other multivariate analysis that results in a scalar condition difference

For example: “the difference in classifier performance on data in condition A and condition B”



Rausschmeisser

Anything to keep in mind when applying
all this to OPM data?

Non-parametric inferential statistical testing with permutation and clustering

Jan Mathijs Schoffelen

Donders Institute, Radboud University, Nijmegen, NL

janmathijs.schoffelen@donders.ru.nl