Bootstrap statistics

Robert Oostenveld
r.oostenveld@fcdonders.ru.nl

Arno Delorme (SCCN, San Diego, USA)
Eric Maris (NICI, Nijmegen, The Netherlands)
Alternative title:

Parametric and non-parametric methods for statistical inference: bootstrapping and randomization
Outline

• Background on statistical inference
• Motivation for bootstrapping and randomization
• Theory behind randomization
• Bootstrapping and randomization applied in EEGLAB
Outline

• **Background on statistical inference**
  • Motivation for bootstrapping and randomization
  • Theory behind randomization
  • Bootstrapping and randomization applied in EEGLAB
What is statistics about?

- Statistics is not about the truth
- Statistics is about dealing with uncertainty
  - Decision making under uncertainty (inferential statistics, statistical testing)
  - Data analysis: methods that reveal patterns in the data which cannot be identified by eyeballing
Statistical decision making

- Based on rational principles
- You make the decision…
  - … that convinces you
  - … that convinces the reviewer
  - … that convinces the audience
- What are the reasons for not being certain?
Implications for Neuroscience

• Inferential statistics is about principles of rational decision making and not about the nature of the biological system
• Therefore it must be boring for a neuroscientist
• However, because there is uncertainty in neuroscience data, inferential statistics cannot be ignored
Statistical inference

• Formulate a so-called null-hypothesis
  – H0: there is no effect
• Formulate an alternative hypothesis
  – H1: there is an effect, describe it

• Determine the likelihood (p-value) of H0
• Reject H0 if it is too unlikely
  e.g. smaller than 5% -> accept H1
Outline

- Background on statistical inference
- *Motivation for bootstrapping and randomization*
- Theory behind randomization
- Bootstrapping and randomization applied in EEGLAB
Conventional statistics

- Parametric statistics
  - t-test
  - ANOVA
  - Correlation

- Parameter
  - t
  - F
  - R

Behaviour of these parameters is known under the null-hypothesis
Parametric statistics

t-test

T-test: Compare paired/unpaired samples for continuous data.

**Paired**

\[ t = \frac{\text{Mean}_{\text{difference}}}{\text{Standard deviation}} \sqrt{N - 1} \]

**Unpaired**

\[ t = \sqrt{N} \frac{\text{Mean}_A - \text{Mean}_B}{\sqrt{(SD_A)^2 - (SD_B)^2}} \]
**ANOVA:** compare several groups (can test interaction between two factors for the repeated measure ANOVA)

\[
F = \frac{\text{Variance}_{\text{interGroup}}}{\text{Variance}_{\text{WithinGroup}}} \times \frac{N_{\text{Group}} - 1}{N - N_{\text{Group}}}
\]

\[\text{F dist. } \begin{cases} \text{df}_n = 5 \\ \text{df}_d = 10 \end{cases}\]

5% of area
Why something else than parametric statistics?

• Problems
  – Assumes normal distribution of the value of interest
  – Assumes that error terms (deviation from the group mean) are normally distributed
  – Assumes known behaviour of the parameters
Why something else than parametric statistics?

• More problems
  – Difficult to solve the multiple comparison problem

• New opportunities
  – Use intuitive explanation of the data -> express parameters the way we want
  – Increased sensitivity for our ‘new parameters’
New approaches for statistics

- Bootstrap
- Jackknife
- Resampling
- Randomization
- Permutation
- Pseudo data
- Surrogate data
New approaches for statistics

• Resampling framework
  – Systematic resampling
  – Random resampling

• Randomization framework
  – Randomization
  – Permutation
Outline

• Background on statistical inference
• Motivation for bootstrapping and randomization
• *Theory behind randomization*
• Bootstrapping and randomization applied in EEGLAB
Baron von Münchhausen
Baron von Münchhausen
New approaches for statistics

• Stuck in the swamp?
• Pull yourself out by your bootstraps!
General principle

• Observed data
  – Some subjects are observed in one condition
  – Some subjects are observed in another condition

  – Independent variable: condition
  – Dependent variable: data
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
Randomization approach
Randomization approach
Randomization approach

- Analyze
- Difference
- $X_{org}$
Randomization approach
Randomization approach
Non-parametric statistics

• Randomization of independent variable
• Hypothesis is about data, not about the specific parameter
• Randomization distribution of the statistic of interest “x” is approximated using Monte-Carlo approach
• H0 is tested by comparing observed statistic v.s. randomization distribution
Outline

• Background on statistical inference
• Motivation for bootstrapping and randomization
• Theory behind randomization
• Bootstraping and randomization applied in EEGLAB
Permutation for comparing between conditions

Help mask cross-coherence difference
Bootstrapping & Jacknife
(parametric testing)

- Estimating variance for non-linear measures
  - Jacknife: leaving one out
  - Bootstrap: taking sub-samples with replacement

E.g. testing for coherence

![Diagram showing mean, standard deviation, leave one out, correction for Jacknifing, and parametric testing for non-0 mean (t-test)]
Correcting for multiple comparisons

• Bonferoni correction: divide by the number of comparisons (Bonferroni CE. Sulle medie multiple di potenze. Bollettino dell'Unione Matematica Italiana, 5 third series, 1950; 267-70.)

• Holms correction: sort all p values. Test the first one against $\alpha/N$, the second one against $\alpha/(N-1)$

• For time-domain: find minimum resolution where effect is visible, then divide significance threshold by the total number of points

• For time-frequency decomposition:
  - minimal resolution in frequency is real wavelets
  - Minimal resolution in time is the time at which neighboring time points are significantly correlated (Approximately 200 in most cases).
  - For 5% significance between 2-100Hz, the new p value should be $0.05 \times 1/200 = 2.5 \times 10^{-4}$ (Bonferroni)