Classifying EEG data driven by rhythmic stimuli using a projective test

A. Duarte, R. Fraiman, A. Galves, **G. Ost**, C. Vargas

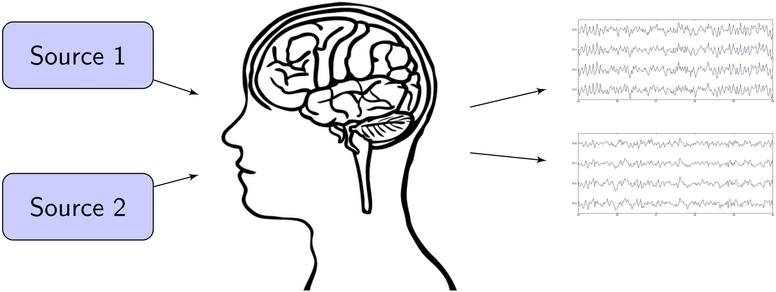
Work in progress

FAPESP - CEPID **NeuroMat**

January 23, 2014

Neurobiological problem

Do stimuli of different sources produce distinct brain processes?



Can we classify them?

- The stimuli consist of independent samples produced by different stochastic rhythmic sources.
- Each sample is a sequence of strong and weak beats, and silent units generated by a probabilistic source

More Precisely

- Each stochastic source is characterized by a probabilistic context tree.
- Statistical fact: each of them can be estimated consistently.
- Question: Can we distinguish samples produced by different sources?

Goal

- Our goal is to classify EEG signals driven by rhythmic stimuli.
- This is a problem of functional random data classification.
- Model selection in Electroencephalographic (EEG) data is a challenging task.

- First rhythm: Waltz (Ternary).
-
- Symbols:
  - 2 - strong beat.
  - 1 - weak beat.
  - 0 - silence unit.
- Stochastic rhythm generation:
  - start with a deterministic sequence

$$\cdots \mathbf{2}\ 1\ 1\ \mathbf{2}\ 1\ 1\ \mathbf{2}\ 1\ 1\ \mathbf{2} \cdots$$

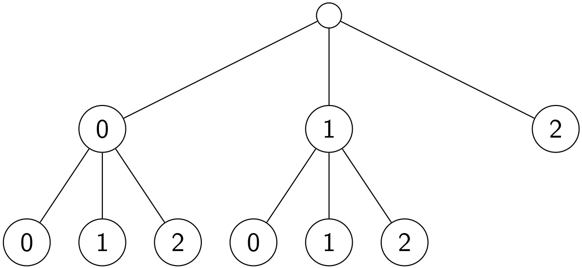
  - replace in a iid way each symbol 1 by 0 with a probability (say 20%).

A typical sample would be

$$\cdots \mathbf{2}\ 1\ 1\ \mathbf{2}\ 1\ 1\ \mathbf{2}\ 1\ 1\ \mathbf{2} \cdots$$

$$\cdots \mathbf{2}\ 1\ 1\ \mathbf{2}\ 1\ 0\ \mathbf{2}\ 0\ 1\ \mathbf{2} \cdots$$

The correspondent context tree is



- Second rhythm: simplified Samba (Quaternary).
-
- Symbols:
  - 2 - strong beat.
  - 1 - weak beat.
  - 0 - constitutive silence unit or omitted weak beat.
- Stochastic rhythm generation:
  - start with a deterministic sequence

$$\cdots \mathbf{2}\ 1\ \mathbf{0}\ 1\ \mathbf{2}\ 1\ \mathbf{0}\ 1\ \mathbf{2}\ 1\ \mathbf{0}\ 1\ \mathbf{2} \cdots$$

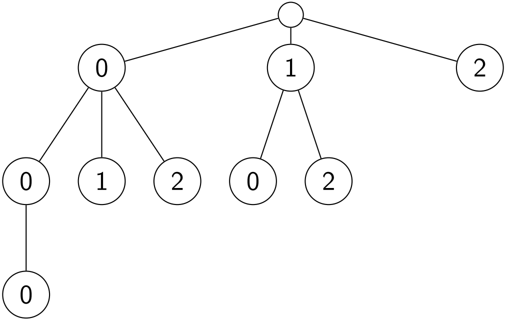
  - replace in a iid way each symbol 1 by 0 with a probability

A typical sample would be

$$\cdots \mathbf{2}\ 1\ \mathbf{0}\ 1\ \mathbf{2}\ 1\ \mathbf{0}\ 1\ \mathbf{2}\ 1\ \mathbf{0}\ 1\ \mathbf{2} \cdots$$

$$\cdots \mathbf{2}\ 1\ \mathbf{0}\ 0\ \mathbf{2}\ 1\ \mathbf{0}\ 1\ \mathbf{2}\ 0\ \mathbf{0}\ 0\ \mathbf{2} \cdots$$

The correspondent context tree is



- Third rhythm: Independent rhythmic units.
-
- Symbols:
  - 2 - strong beat.
  - 1 - weak beat.
  - 0 - silence unit.
- Chain generation:
  - choose any symbol in a iid way with probability $1/3$.

A typical sample would be

$$\cdots 2\ 1\ 0\ 1\ 1\ 2\ 2\ 0\ 1\ 0\ 2 \cdots$$

The correspondent context tree is reduced to the root. Why?

Acquisition

- Each volunteer was exposed to two rhythmic blocks of 12 min each.
- Each rhythmic block is a concatenation of three rhythms:
  - $B_{WIS} = \{\text{Waltz}, \text{Independent}, \text{Samba}\}$
  - $B_{SIW} = \{\text{Samba}, \text{Independent}, \text{Waltz}\}$
- Each sample corresponding to a given rhythm lasts for 3 min and is preceded by a one minute interval of silence.

We mark each stimulus onset:

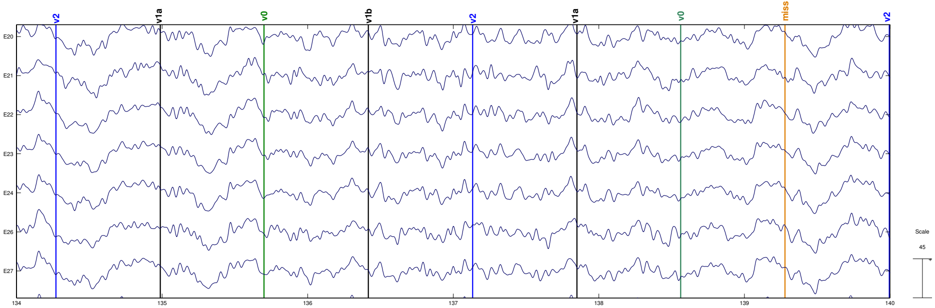
Constitutive silence unit $\longrightarrow$ V_0

Weak beat $\longrightarrow$ V_1

Strong beat $\longrightarrow$ V_2

Omitted weak beat $\longrightarrow$ Miss

EEG data


v2
v1a
v0
v1b
v2
v1a
v0
miss
v2
E20
E21
E22
E23
E24
E26
E27
134
135
136
137
138
139
140
Scale
45

Summarizing

- Stochastic Sources modeled by Probabilistic Context Trees.
- Each source can be statistically retrieved from a sample.
- EEG samples associated to each context tree rhythmic source.

Are the EEG samples statistically different?

How to tackle this question?

Projective Method

- Given two random samples of functional data, we want to test if these two samples came from the same source.
- Projective method: choose a randomly direction and perform a one dimensional statistical test for the projected data.
- This method was introduced in Cuesta-Albertos, Fraiman and Ransford (2006).
- This approach was successfully employed in the classification of linguistic sonority data in Cuesta-Albertos, Fraiman, Galves, Garcia and Svarc (2007).

Projective Method: groundwork

- If the laws of two random mechanisms are such that:
    - one of them is not “heavy-tailed”.
    - the set of the directions in which the laws are the same has positive probability.
- Then: the laws are equals!

How to apply it?

- Consider each EEG signals collected from each electrode as an outcome of suitable random mechanisms.
- Given the *Samba* and *Waltz* EEG signals, we want to test
  - $H_0 = \{P_{Samba} = P_{Waltz}\}$ (null hypothesis)
  - $H_1 = \{P_{Samba} \neq P_{Waltz}\}$ (alternative hypothesis)

Under $H_0 = \{P_{Samba} = P_{Waltz}\}$, for each direction the laws are different.

Algorithm:

- Choose N independent directions $W_i, i = 1, \cdots N$. (Brownian motions)
- For each i:
  - Test the null at level η by projecting *Samba* and *Waltz* on W_i, using Kolmogorov-Smirnov test.
  - Define

$$Z_i = \begin{cases} 1, & \text{if we rejected } H_0 \\ 0, & \text{if we do not rejected } H_0. \end{cases}$$

- Define the average value

$$\bar{Z} = \frac{1}{N}\sum_{i=1}^{N} Z_i$$

and reject H_0 if $\bar{Z} \geq c_\alpha$.

- Question: what should be the value of c_α to have a test of level α?

- To answer this question we use a bootstrap procedure.

Planning:

- To use the protective method to classify EEG samples driven by Samba, Waltz and IID stimuli.
- This has not been done yet: our data sample is not big enough.

Data is being collected!

But something can be done immediately with the small data set of the pilot study.

Preliminary question: for the EEG data driven by Samba

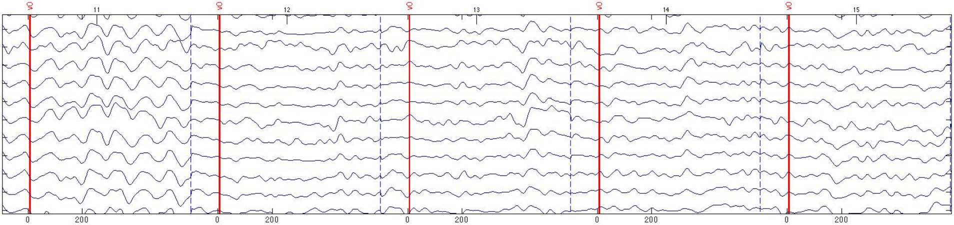
- Both Miss and V_0 time intervals correspond to silence units.
- However, from a structural point of view Miss and V_0 are entirely different.

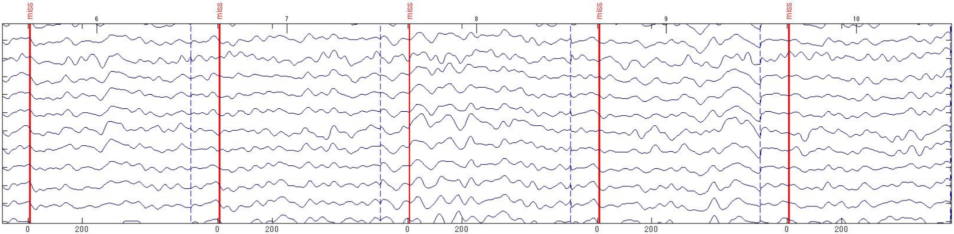
Remember:

- Miss is an omitted weak beat.
- V_0 is a constitutive silence unit.

- From a structural point of view Miss and V_0 are entirely different.
- Is the brain "aware" of this distinction?
- *More pragmatically: is our statistical tool able to catch this difference in the EEG signal?*

V0 and Miss samples

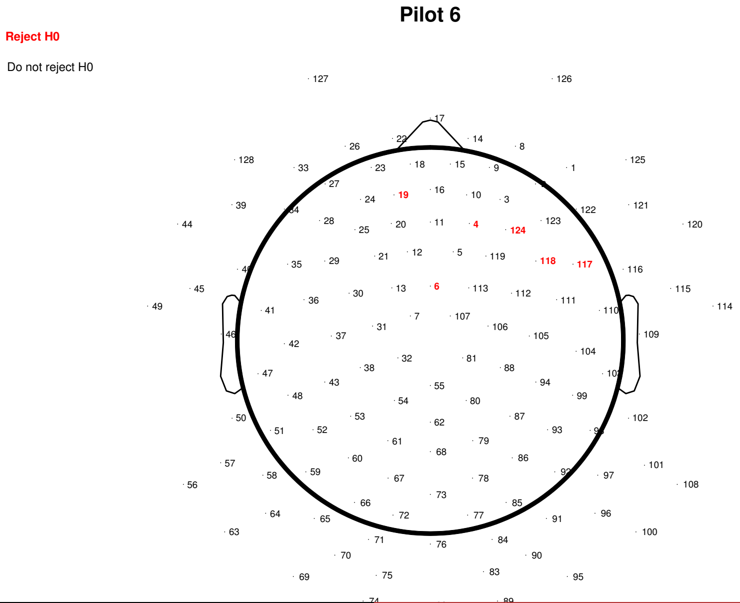




Results

- The number of elements in $\mathcal{M}$ and $\mathcal{V}$ are 54 and 126 respectively.
- We applied the test with $N = 1000$, $B = 100$, $\eta = \alpha = 0.1$.
- Pilot 6:

Elect	p-value
4	0.04
6	0.01
19	0.01
117	0.02
118	0.03
124	0.04


Pilot 6
Reject H0
Do not reject H0
19
4
124
118
117
6

To be continued....