Hidden context tree modeling of EEG data

Antonio Galves

joint work with A. Duarte, R. Fraiman, G. Ost and C. Vargas

Universidade de S.Paulo and NeuroMat

MathStatNeuro 2015

Looking for experimental evidence that the brain is a statistician

- Is the brain a statistician?
- Stanislas Dehaene claims that the idea that the brain is a Bayesian statistician is already sketched in von Helmholtz work!
- See for instance the two lessons by Dehaene available on the web:

Looking for experimental evidence that the brain is a statistician

- Is the brain a statistician?
- Stanislas Dehaene claims that the idea that the brain is a Bayesian statistician is already sketched in von Helmholtz work!
- See for instance the two lessons by Dehaene available on the web:
    - Le cerveau statisticien

Looking for experimental evidence that the brain is a statistician

- Is the brain a statistician?
- Stanislas Dehaene claims that the idea that the brain is a Bayesian statistician is already sketched in von Helmholtz work!
- See for instance the two lessons by Dehaene available on the web:
  - Le cerveau statisticien
  - Le bébé statisticien

Is the brain a statistician?

- How to obtain experimental evidence supporting this conjecture?
- Dehaene presents experimental evidence that unexpected occurrences in regular sequences produce characteristic markers in EEG data.
- But we need more than evidences of mismatch negativity to support this conjecture.
- To discuss this issue we need to do statistical model selection in a new class of stochastic processes:

Is the brain a statistician?

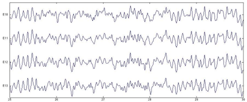
- How to obtain experimental evidence supporting this conjecture?
- Dehaene presents experimental evidence that unexpected occurrences in regular sequences produce characteristic markers in EEG data.
- But we need more than evidences of mismatch negativity to support this conjecture.
- To discuss this issue we need to do statistical model selection in a new class of stochastic processes:

Hidden context tree models.

Neurobiological problem

A random source produces sequences of auditory stimuli.

Random Source



How to retrieve the structure of the source from EEG data?

Example of a random source: samba

- Auditory segments:
  - 2 - strong beat
  - 1 - weak beat
  - 0 - silent event
- Chain 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 probability ϵ.

A typical sample would be

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

A typical sample would be

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

How to define the structure of this source?

A typical sample would be

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

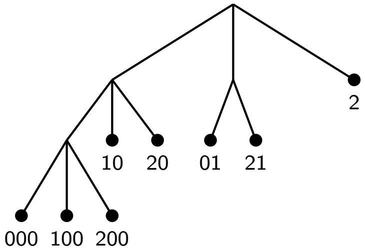
$$\cdots \mathbf{2}\ 1\ \mathbf{0}\ 0\ \mathbf{2}\ 1\ \mathbf{0}\ 1\ \mathbf{2}\ 0\ \mathbf{0}\ 0\ \mathbf{2} \cdots$$

How to define the structure of this source?

- By describing the algorithm producing each next symbol, given the shortest relevant sequence of past symbols.

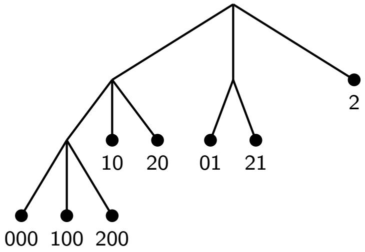
The structure of the random source

The structure of the random source



Contexts(w)	p(0|w)	p(1|w)	p(2|w)
2	ϵ	$1-\epsilon$	0
21	ϵ	$1-\epsilon$	0
01	0	0	1
20	ϵ	$1-\epsilon$	0
10	ϵ	$1-\epsilon$	0
000	0	0	1
100	0	0	1
200	ϵ	$1-\epsilon$	1

The structure of the random source



Contexts(w)	p(0|w)	p(1|w)	p(2|w)
2	ϵ	$1-\epsilon$	0
21	ϵ	$1-\epsilon$	0
01	0	0	1
20	ϵ	$1-\epsilon$	0
10	ϵ	$1-\epsilon$	0
000	0	0	1
100	0	0	1
200	ϵ	$1-\epsilon$	1

The stochastic chain generated by the source samba

$$
\begin{array}{cccccccccc}
\cdots & \mathbf{2} & 1 & 0 & 0 & \mathbf{2} & 1 & 0 & 1 & \cdots \\
 & \uparrow & \uparrow & \uparrow & \uparrow & \uparrow & \uparrow & \uparrow & \uparrow & \\
 & X_{-5} & X_{-4} & X_{-3} & X_{-2} & X_{-1} & X_0 & X_1 & X_2 &
\end{array}
$$

- $X_n \in A = \{0, 1, 2\}$

- $X_n \in A = \{0, 1, 2\}$
- $(X_n)_{n \in \mathbb{Z}}$ is stochastic chain

- $X_n \in A = \{0, 1, 2\}$
- $(X_n)_{n \in \mathbb{Z}}$ is stochastic chain
- with memory of variable length

- $X_n \in A = \{0, 1, 2\}$
- $(X_n)_{n \in \mathbb{Z}}$ is stochastic chain
- with memory of variable length
- generated by the probabilistic context tree

- $X_n \in A = \{0, 1, 2\}$
- $(X_n)_{n \in \mathbb{Z}}$ is stochastic chain
- with memory of variable length
- generated by the probabilistic context tree

Contexts(w)	p(0|w)	p(1|w)	p(2|w)
2	ϵ	$1-\epsilon$	0
21	ϵ	$1-\epsilon$	0
01	0	0	1
20	ϵ	$1-\epsilon$	0
10	ϵ	$1-\epsilon$	0
000	0	0	1
100	0	0	1
200	ϵ	$1-\epsilon$	1

Context tree models

- Introduced by Rissanen

Context tree models

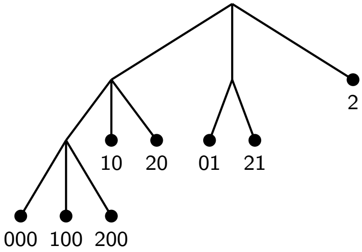
- Introduced by Rissanen
- *A universal data compression system*, IEEE, 1983.

Context tree models

- Introduced by Rissanen
- *A universal data compression system*, IEEE, 1983.
- stochastic chains with memory of variable length

Context tree models

- Introduced by Rissanen
- *A universal data compression system*, IEEE, 1983.
- stochastic chains with memory of variable length
- generated by a probabilistic context tree



The neurobiological question

Is it possible

The neurobiological question

Is it possible

to retrieve the samba context tree

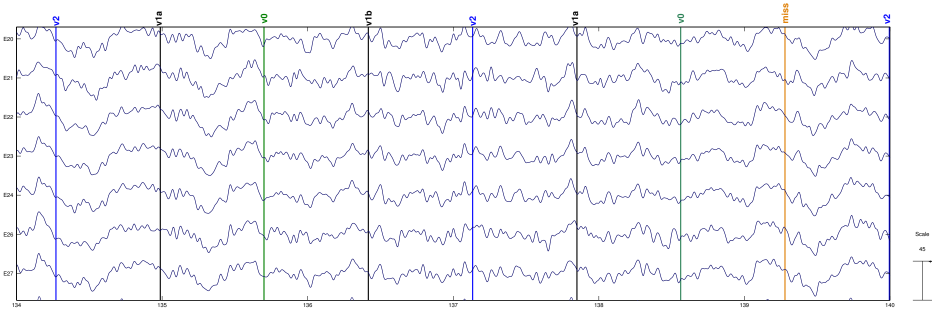
The neurobiological question

Is it possible

to retrieve the samba context tree

from the EEG data recorded during the exposure to the sequence of auditory stimuli generated by the samba source?

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

How to address the identification problem?

We have

How to address the identification problem?

We have

- EEG data recorded with 18 electrodes

How to address the identification problem?

We have

- EEG data recorded with 18 electrodes
- for each electrode e and each step n

How to address the identification problem?

We have

- EEG data recorded with 18 electrodes
- for each electrode e and each step n
- call $Y_n^e = (Y_n^e(t), t \in [0, T])$

How to address the identification problem?

We have

- EEG data recorded with 18 electrodes
- for each electrode e and each step n
- call $Y_n^e = (Y_n^e(t), t \in [0, T])$ the EEG signal recorded at electrode e during the exposure to the auditory stimulus X_n
- $Y_n^e \in L^2([0, T])$, where $T = 450$ms is the time distance between the onsets of two consecutive auditory stimuli

Hidden context tree model (HCTM)

Ingredients:

- finite alphabet A

Hidden context tree model (HCTM)

Ingredients:

- finite alphabet A

  In our example $A = \{0, 1, 2\}$

Hidden context tree model (HCTM)

Ingredients:

- finite alphabet A

  In our example $A = \{0, 1, 2\}$

- measurable space $(F, \mathcal{F})$

Hidden context tree model (HCTM)

Ingredients:

- finite alphabet A

  In our example $A = \{0, 1, 2\}$
- measurable space $(F, \mathcal{F})$

  In our example $F = L^2([0, T])$ and $\mathcal{F}$ is the Borel σ-algebra on F.

Hidden context tree model (HCTM)

Ingredients:

- finite alphabet A

  In our example $A = \{0, 1, 2\}$
- measurable space $(F, \mathcal{F})$

  In our example $F = L^2([0, T])$ and $\mathcal{F}$ is the Borel σ-algebra on F.
- probabilistic context tree (τ, p)

Hidden context tree model (HCTM)

Ingredients:

- finite alphabet A

  In our example $A = \{0, 1, 2\}$
- measurable space $(F, \mathcal{F})$

  In our example $F = L^2([0, T])$ and $\mathcal{F}$ is the Borel σ-algebra on F.
- probabilistic context tree (τ, p)
- family $\{Q_w : w \in \tau\}$ of probabilities on $(F, \mathcal{F})$

Hidden context tree model (HCTM)

Ingredients:

- finite alphabet A

  In our example $A = \{0, 1, 2\}$
- measurable space $(F, \mathcal{F})$

  In our example $F = L^2([0, T])$ and $\mathcal{F}$ is the Borel σ-algebra on F.
- probabilistic context tree (τ, p)
- family $\{Q_w : w \in \tau\}$ of probabilities on $(F, \mathcal{F})$
- stochastic chain $(X_n, Y_n) \in A \times F$.

Hidden context tree model

$(X_n, Y_n)_{n \in \mathbb{Z}}$ HCTM compatible with (τ, p) and $(Q_w : w \in \tau)$ if

Hidden context tree model

$(X_n, Y_n)_{n \in \mathbb{Z}}$ HCTM compatible with (τ, p) and $(Q_w : w \in \tau)$ if

- $(X_n)_{n \in \mathbb{Z}}$ is generated by (τ, p)

Hidden context tree model

$(X_n, Y_n)_{n\in\mathbb{Z}}$ HCTM compatible with (τ, p) and $(Q_w : w \in \tau)$ if

- $(X_n)_{n\in\mathbb{Z}}$ is generated by (τ, p)
- for any $m, n \in \mathbb{Z}$ with $m \leq n$

Hidden context tree model

$(X_n, Y_n)_{n\in\mathbb{Z}}$ HCTM compatible with (τ, p) and $(Q_w : w \in \tau)$ if

- $(X_n)_{n\in\mathbb{Z}}$ is generated by (τ, p)
- for any $m, n \in \mathbb{Z}$ with $m \leq n$
- any string $x_{m-\ell(\tau)+1}^{n} \in A^{n-m+\ell(\tau)}$

Hidden context tree model

$(X_n, Y_n)_{n\in\mathbb{Z}}$ HCTM compatible with (τ, p) and $(Q_w : w \in \tau)$ if

- $(X_n)_{n\in\mathbb{Z}}$ is generated by (τ, p)
- for any $m, n \in \mathbb{Z}$ with $m \leq n$
- any string $x_{m-\ell(\tau)+1}^{n} \in A^{n-m+\ell(\tau)}$
- and any sequence $I_m^n = (I_m, \ldots, I_n)$ of $\mathcal{F}$-measurable sets,

Hidden context tree model

$(X_n, Y_n)_{n\in\mathbb{Z}}$ HCTM compatible with (τ, p) and $(Q_w : w \in \tau)$ if

- $(X_n)_{n\in\mathbb{Z}}$ is generated by (τ, p)
- for any $m, n \in \mathbb{Z}$ with $m \leq n$
- any string $x_{m-\ell(\tau)+1}^{n} \in A^{n-m+\ell(\tau)}$
- and any sequence $I_m^n = (I_m, \ldots, I_n)$ of $\mathcal{F}$-measurable sets,

$$\mathbb{P}\Big(Y_m^n \in I_m^n | X_{m-\ell(\tau)+1}^{n} = x_{m-\ell(\tau)+1}^{n}\Big) = \prod_{k=m}^{n} Q_{c_\tau(x_{k-\ell(\tau)+1}^{k})}(I_k)$$

- $\ell(\tau) =$ height of τ
- $c_\tau(x_{k-\ell(\tau)+1}^{k}) =$ context assigned to $x_{k-\ell(\tau)+1}^{k}$ by τ

Rephrasing our problem

Rephrasing our problem

Taking

Rephrasing our problem

Taking

- $(X_n)_{n \in \mathbb{Z}}$ sequence of auditory stimuli produced by the samba source

Rephrasing our problem

Taking

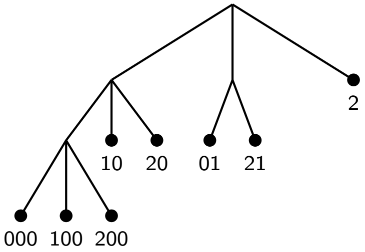
- $(X_n)_{n \in \mathbb{Z}}$ sequence of auditory stimuli produced by the samba source
- $(Y_n^e)_{n \in \mathbb{Z}}$ successive chunks of EEG signals

Rephrasing our problem

Taking

- $(X_n)_{n\in\mathbb{Z}}$ sequence of auditory stimuli produced by the samba source
- $(Y_n^e)_{n\in\mathbb{Z}}$ successive chunks of EEG signals

Question: Is $(X_n, Y_n^e)_{n\in\mathbb{Z}}$ a HCTM compatible with τ?



is $(X_n, Y_n^e)_{n\in\mathbb{Z}}$ a HCTM compatible with τ?

In other terms, for any $w \in \tau$, is it true that

$$\mathcal{L}(Y_n^e | X_{n-\ell(w)+1}^n = w, X_{-\infty}^{-\ell(\tau)} = u) = \mathcal{L}(Y_n^e | X_{n-\ell(w)+1}^n = w, X_{-\infty}^{-\ell(\tau)} = v)$$

for any pair of strings u and v?

Pruning the tree

- A version of Rissanen's algorithm Context will be applied

Pruning the tree

- A version of Rissanen's algorithm Context will be applied
- Start with a maximal admissible candidate tree

Pruning the tree

- A version of Rissanen's algorithm Context will be applied
- Start with a maximal admissible candidate tree
- For any string w and pair of symbols $a, b \in A$ with aw and bw belonging to the candidate tree
- test the equality

$$\mathcal{L}(Y_n^e | X_{n-\ell(w)}^n = aw) = \mathcal{L}(Y_n | X_{n-\ell(w)}^n = bw)$$

Pruning the tree

- If for all pairs of symbols (a, b) the equality is rejected then prune all the leaves aw
- Repeat the pruning procedure until no more pruning is required

How to test the equality

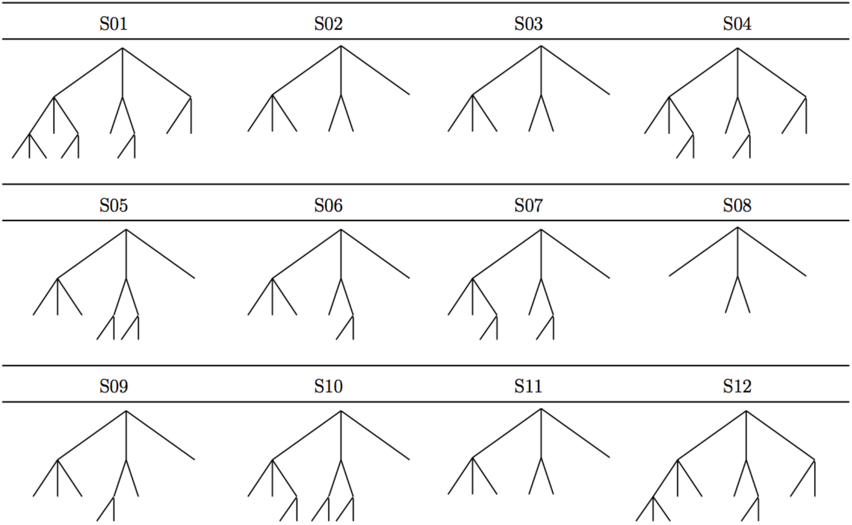
$$\mathcal{L}(Y_n|X_{n-\ell(w)}^n = aw) = \mathcal{L}(Y_n|X_{n-\ell(w)}^n = bw)\,?$$

Apply the projective method introduced by Cuestas-Albertos, Fraiman and Ransford (2006).

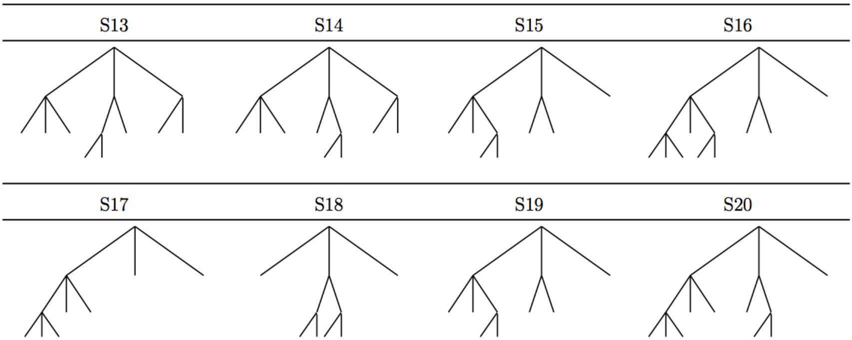
Experimental results

- Context tree selection procedure for the EEG data recorded during the exposure to the sequence of auditory stimuli generated by the samba source
- Sample composed by 20 subjects
- For each subject EEG data from 18 electrodes was recorded

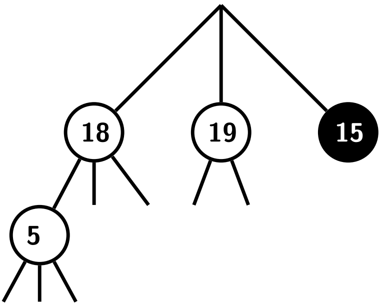
Experimental results



Experimental results



Summary



White nodes indicate the number of subjects which correctly identify the node as not being a context. **Black nodes** indicate the number of subjects which correctly identify the node as a **context**. For instance, 18 subjects correctly identify that the symbol 0 alone is not enough to predict the next symbol. And 15 subjects correctly identify the symbol 2 as **a context**.