Seleção estatística de árvores de contextos a partir de dados de EEG

Aline Duarte

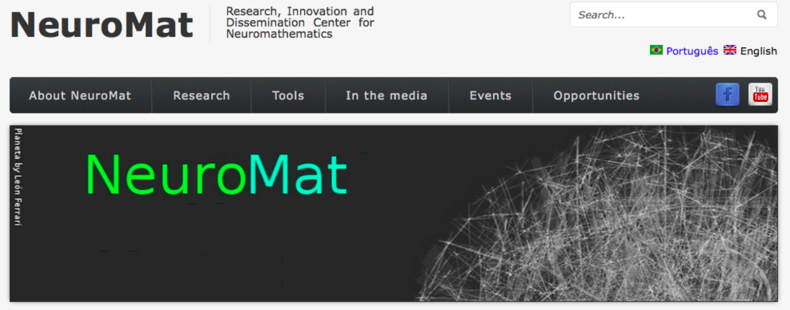
trabalho em colaboração com R. Fraiman, A. Galves, G. Ost and C. Vargas

Universidade de São Paulo/ CEPID NeuroMat

Universidade Federal Fluminense

25 de outubro de 2017

CEPID NeuroMat



Conjectura: O cérebro é um estatístico

Conjectura: O cérebro é um estatístico

- Como obter evidências experimentais?

Conjectura: O cérebro é um estatístico

- Como obter evidências experimentais?

R.: Novo quadro experimental

Conjectura: O cérebro é um estatístico

- Como obter evidências experimentais?

Protocolo experimental + Classe de mod-els estocásticos

Conjectura: O cérebro é um estatístico

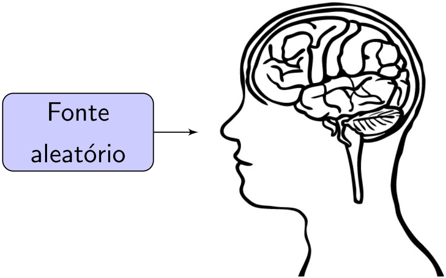
- Como obter evidências experimentais?

Protocolo experimental + Classe de modelos estocásticos

Processos estocásticos conduzidos por árvores de contextos

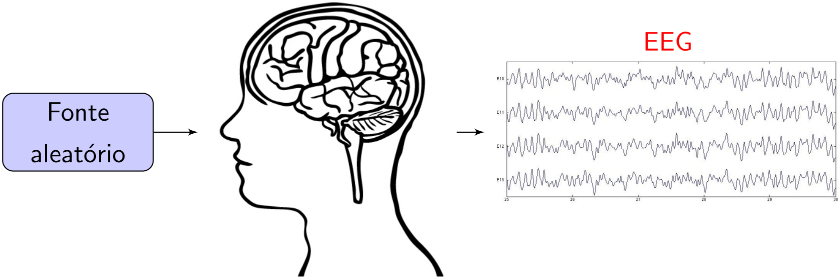
Protocolo experimental

- Uma fonte aleatória produz uma sequência de estímulos auditivos



Protocolo experimental

- Uma fonte aleatória produz uma sequência de estímulos auditivos



- Como recuperar dos dados de EEG alguma assinatura da fonte aleatória?

Exemplo de fonte aleatória: Valsa

Unidades de estímulo:

- 2 - batida forte
- 1 - batida fraca
- 0 - unidade de silêncio

Exemplo de fonte aleatória: Valsa

Unidades de estímulo:

- 2 - batida forte
- 1 - batida fraca
- 0 - unidade de silêncio

Sequência estímulo:

- Comece com a sequência determinística

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

Exemplo de fonte aleatória: Valsa

Unidades de estímulo:

- 2 - batida forte
- 1 - batida fraca
- 0 - unidade de silêncio

Sequência estímulo:

- Comece com a sequência determinística

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

- Substitua cada batida fraca, de maneira independente, por uma unidade de silêncio com probabilidade ϵ

Exemplo de fonte aleatória: Valsa

Unidades de estímulo:

- 2 - batida forte
- 1 - batida fraca
- 0 - unidade de silêncio

Sequência estímulo:

- Comece com a sequência determinística

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

- Substitua cada batida fraca, de maneira independente, por uma unidade de silêncio com probabilidade ϵ

Um exemplo de sequência seria

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

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

- Denote $X_0, X_1, X_2, \ldots$ a cadeia estocástica gerada dessa maneira

Essa cadeia pode ser gerada passo por passo como segue:

- Denote $X_0, X_1, X_2, \ldots$ a cadeia estocástica gerada dessa maneira

Essa cadeia pode ser gerada passo por passo como segue:

- Se $X_{n-1} = 2$ então

$$X_n = \begin{cases} 1, & \text{com probabilidade } 1 - \epsilon, \\ 0, & \text{com probabilidade } \epsilon. \end{cases}$$

- Denote $X_0, X_1, X_2, \ldots$ a cadeia estocástica gerada dessa maneira

Essa cadeia pode ser gerada passo por passo como segue:

- Se $X_{n-1} = 2$ então

$$X_n = \begin{cases} 1, & \text{com probabilidade } 1 - \epsilon, \\ 0, & \text{com probabilidade } \epsilon. \end{cases}$$

- Se $X_{n-1} = 1$ ou $X_{n-1} = 0$ então precisamos checar um passo a mais no passado
  - se $X_{n-2} = 2$ então

$$X_n = \begin{cases} 1, & \text{com probabilidade } 1 - \epsilon, \\ 0, & \text{com probabilidade } \epsilon; \end{cases}$$

- Denote $X_0, X_1, X_2, \ldots$ a cadeia estocástica gerada dessa maneira

Essa cadeia pode ser gerada passo por passo como segue:

- Se $X_{n-1} = 2$ então

$$X_n = \begin{cases} 1, & \text{com probabilidade } 1-\epsilon, \\ 0, & \text{com probabilidade } \epsilon. \end{cases}$$

- Se $X_{n-1} = 1$ ou $X_{n-1} = 0$ então precisamos checar um passo a mais no passado
  - se $X_{n-2} = 2$ então

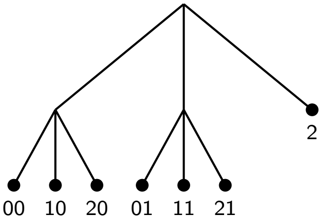
$$X_n = \begin{cases} 1, & \text{com probabilidade } 1-\epsilon, \\ 0, & \text{com probabilidade } \epsilon; \end{cases}$$

  - se $X_{n-2} = 1$ or $X_{n-2} = 0$ então $X_n = 2$ com probabilidade 1.

Caracterizando a fonte aleatória

2 1 1 **2** 1 1 **2** 1 1 **2** $\cdots$

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



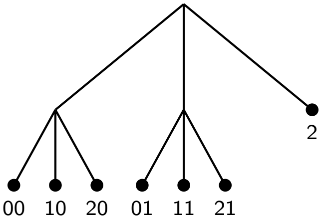
Ternary CTM			
context w	$\mathbf{p(0\|w)}$	$\mathbf{p(1\|w)}$	$\mathbf{p(2\|w)}$
2	ϵ	$1-\epsilon$	0
21	ϵ	$1-\epsilon$	0
20	ϵ	$1-\epsilon$	0
11	0	0	1
10	0	0	1
01	0	0	1
00	0	0	1

$\tau = \{2, 21, 11, 01, 20, 10, 00\}$: Uma partição dos passados possíveis

Caracterizando a fonte aleatória

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

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



Ternary CTM

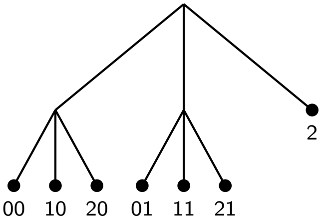
context w	p(0|w)	p(1|w)	p(2|w)
2	ϵ	$1-\epsilon$	0
21	ϵ	$1-\epsilon$	0
20	ϵ	$1-\epsilon$	0
11	0	0	1
10	0	0	1
01	0	0	1
00	0	0	1

$\tau = \{2, 21, 11, 01, 20, 10, 00\}$: Uma partição dos passados possíveis

Caracterizando a fonte aleatória

2 1 1 **2** 1 1 **2** 1 1 **2** $\cdots$

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



Ternary CTM

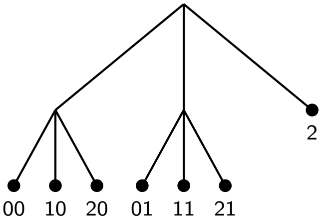
context w	p(0|w)	p(1|w)	p(2|w)
2	ϵ	$1-\epsilon$	0
21	ϵ	$1-\epsilon$	0
20	ϵ	$1-\epsilon$	0
11	0	0	1
10	0	0	1
01	0	0	1
00	0	0	1

$\tau = \{2, 21, 11, 01, 20, 10, 00\}$: Uma partição dos passados possíveis

Caracterizando a fonte aleatória

2 1 1 **2** 1 1 **2** 1 1 **2** · · ·

2 1 0 **2** 1 1 **2** 0 0 **2** · · ·



Ternary CTM			
context w	$\mathbf{p(0\|w)}$	$\mathbf{p(1\|w)}$	$\mathbf{p(2\|w)}$
2	ϵ	$1-\epsilon$	0
21	ϵ	$1-\epsilon$	0
20	ϵ	$1-\epsilon$	0
11	0	0	1
10	0	0	1
01	0	0	1
00	0	0	1

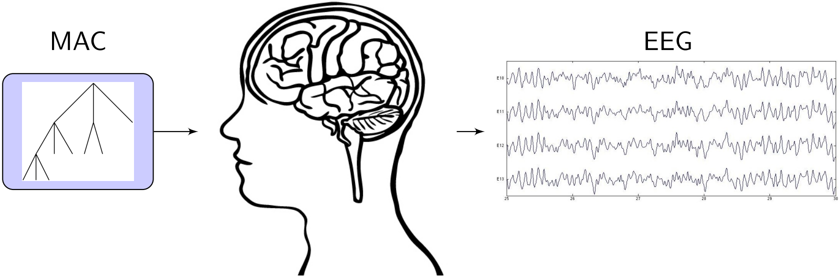
$\tau = \{2, 21, 11, 01, 20, 10, 00\}$: Uma partição dos passados possíveis

- Esses dois objetos caracterizam o algorítimo de simulação

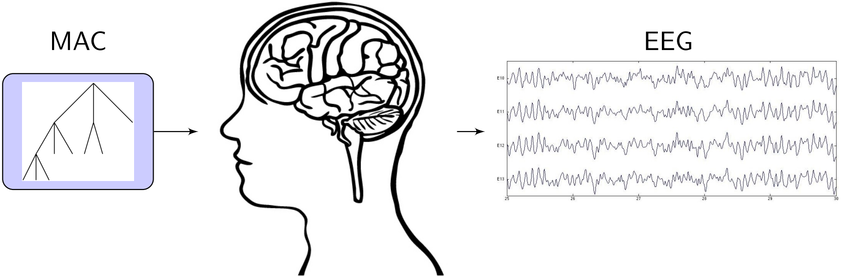
Modelos de árvores de contextos

- Introduzidos por Rissanen como um sistema universal de compressão de dados.
- Cadeias estocásticas com memória de alcance variável
- *Modelos de árvore de contextos* (MAC)

Reescrevendo a conjectura

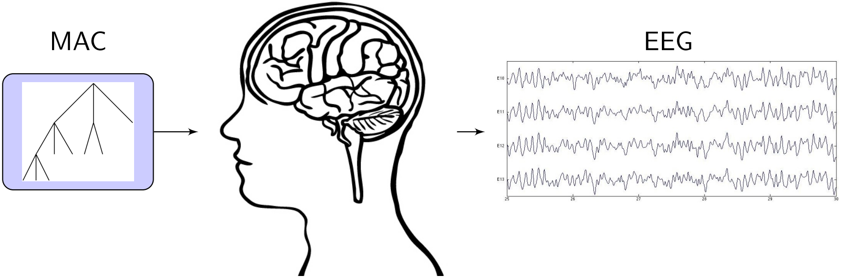


Reescrevendo a conjectura



Pergunta: É possível identificar alguma característica da árvore nos dados de EEG.

Reescrevendo a conjectura



Pergunta: É possível identificar alguma característica da árvore nos dados de EEG.

Problema estatístico: É possível extrair dos dados de EEG a árvore de contextos gerando o estímulo?

Fato: Dada uma sequência de estímulos gerada por um MAC é possível recuperar a árvore por trás do algorítimo (Rissanen 1983)

A Universal Data Compression System

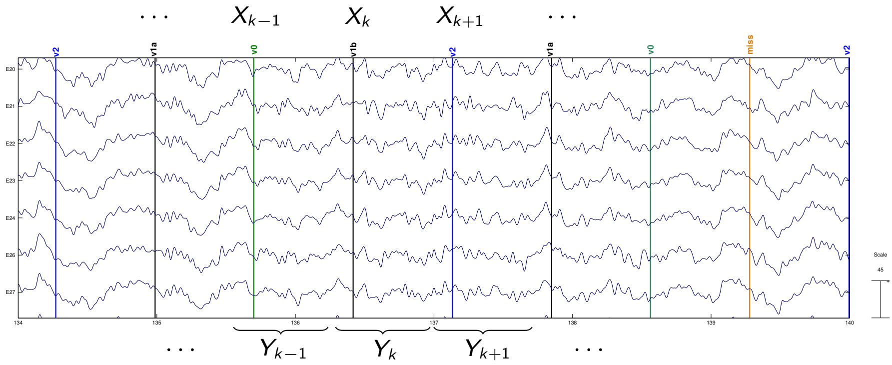
JORMA RISSANEN

Abstract—**A universal data compression algorithm is described which is capable of compressing long strings generated by a "finitely generated" source, with a near optimum per symbol length without prior knowledge of the source. This class of sources may be viewed as a generalization of Markov sources to random fields. Moreover, the algorithm does not require a working storage much larger than that needed to describe the source generating parameters.**

data compression system and the associated notion of compressibility, the real power of the algorithm is its convenient data gathering capability. In order to see this as well as the limitations of the approach, we reinterpret their algorithm in a natural statistical framework of the type discussed in Rissanen and Langdon, [8]. This is done in Section II

Dada a sequência estímulo, é possível recuperar dos dados de EEG a árvore que caracteriza o estímulo?

Processos estocásticos conduzidos por MAC



- Para cada unidade de estímulo aleatório $X_k \in \{0, 1, 2\}$, considere o pedaço de EEG Y_k correspondente a esse estímulo
- A sequência de pares ordenados $(X_k, Y_k)_{k \geq 0}$ define *um modelo estocástico dirigido por um MAC*

Formalmente

- A alfabeto finito
- (τ, p) MAC em A,
- $(F, \mathcal{F})$ espaço mensurável
- $(Q^w : w \in \tau)$ família de medida de probabilidade em $(F, \mathcal{F})$.

A cadeia $(X_n, Y_n)_{n\in\mathbb{Z}}$ tomando valores em $A \times F$ é um *modelo dirigido por uma árvore de contextos* compatível com (τ, p) e $(Q^w : w \in \tau)$ se o seguinte vale

Formalmente

- A alfabeto finito
- (τ, p) MAC em A,
- $(F, \mathcal{F})$ espaço mensurável
- $(Q^w : w \in \tau)$ família de medida de probabilidade em $(F, \mathcal{F})$.

A cadeia $(X_n, Y_n)_{n \in \mathbb{Z}}$ tomando valores em $A \times F$ é um *modelo dirigido por uma árvore de contextos* compatível com (τ, p) e $(Q^w : w \in \tau)$ se o seguinte vale

- $(X_n)_{n \in \mathbb{Z}}$ é um MAC compatível com (τ, p)

Formalmente

- A alfabeto finito
- (τ, p) MAC em A,
- $(F, \mathcal{F})$ espaço mensurável
- $(Q^w : w \in \tau)$ família de medida de probabilidade em $(F, \mathcal{F})$.

A cadeia $(X_n, Y_n)_{n\in\mathbb{Z}}$ tomando valores em $A \times F$ é um *modelo dirigido por uma árvore de contextos* compatível com (τ, p) e $(Q^w : w \in \tau)$ se o seguinte vale

- $(X_n)_{n\in\mathbb{Z}}$ é um MAC compatível com (τ, p)
- Para quaisquer $0 < m < n < \infty$, $J_m, \ldots, J_n \in \mathcal{F}$ e sequência $\mathbf{x} = (x_m, \ldots, x_n) \in A^{n-m+1}$

$$P\big(Y_m \in J_m, \ldots, Y_n \in J_n | X_m = x_m, \ldots, X_n = x_n\big) = \prod_{k=m}^{n} Q^{c_\tau(\mathbf{x})}(J_k),$$

onde $c_\tau(x_{k-\ell(\tau)+1}^{k})$ é o único contexto em τ atribuído a sequência de símbolos do passado $\mathbf{x}$.

- A pergunta que se coloca é como Y_k depende de

$$\ldots, \ X_{k-3}, \ X_{k-2}, \ X_{k-1}, \ X_k$$

- Se a conjectura for verdadeira a distribuição de Y_k deveria depender apenas do contexto associado a sequência de estímulos passados

- A pergunta que se coloca é como Y_k depende de

$$\ldots,\ X_{k-3},\ X_{k-2},\ X_{k-1},\ X_k$$

- Se a conjectura for verdadeira a distribuição de Y_k deveria depender apenas do contexto associado a sequência de estímulos passados
- Por exemplo, se

$$\ldots, X_{n-3} = 1, X_{n-2} = 2, X_{n-1} = 0, X_n = 1,$$

$01 \in \tau \Rightarrow \mathcal{L}(Y_n)$ depende apenas $X_n = 1$ e $X_{n-1} = 0$

- A pergunta que se coloca é como Y_k depende de

$$\dots, \; X_{k-3}, \; X_{k-2}, \; X_{k-1}, \; X_k$$

- Se a conjectura for verdadeira a distribuição de Y_k deveria depender apenas do contexto associado a sequência de estímulos passados
- Por exemplo, se

$$\dots, X_{n-3} = 1, X_{n-2} = 2, X_{n-1} = 0, X_n = 1,$$

$01 \in \tau \Rightarrow \mathcal{L}(Y_n)$ depende apenas $X_n = 1$ e $X_{n-1} = 0$

- Em particular, se para algum $m \neq n$

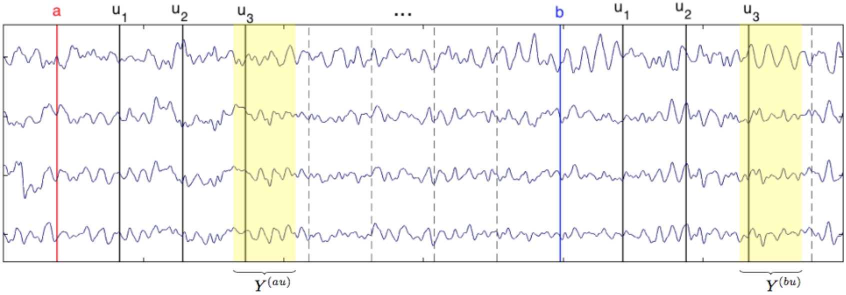
$$\dots, X_{m-3} = 0, X_{m-2} = 2, X_{m-1} = 0, X_m = 1$$

essa realização também é associada ao contexto $01 \Rightarrow \mathcal{L}(Y_n) = \mathcal{L}(Y_m)$

- Como verificar tal hipótese?

Método Projetivo + Algorítimo contexto

Algorítimo contexto de Rissanen



Pergunta: Como testar a igualdade

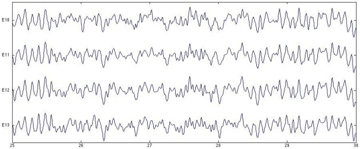
$$
\begin{aligned}
&\mathcal{L}(Y_k \mid X_k = u_3, X_{k-1} = u_2, X_{k-2} = u_1, X_{k-3} = a) \\
&\qquad = \mathcal{L}(Y_k \mid X_k = u_3, X_{k-1} = u_2, X_{k-2} = u_1, X_{k-3} = b)
\end{aligned}
$$

Bull Braz Math Soc, New Series 37(4), 477-501

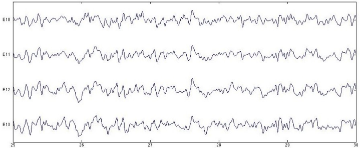
Random projections and goodness-of-fit tests in infinite-dimensional spaces

Juan Antonio Cuesta-Albertos*, Ricardo Fraiman and Thomas Ransford**

Método projetivo

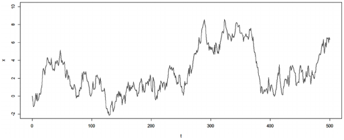


$Y_k^{(au)}$



$Y_k^{(bu)}$

Projeção em um Movimento Browniano



$R_k^{(bu)} \in \mathbb{R}$ $\quad$ $R_k^{(au)} \in \mathbb{R}$

Cuesta-Albertos (2006): Se a lei de $R_k^{(au)}$ e $R_k^{(bu)}$ forem diferentes, então as leis de $Y_k^{(au)}$ e $Y_k^{(bu)}$ também são.

Método projetivo

Fixadas duas sequências $au = au_1u_2u_3$ e $bu = bu_1u_2u_3$

- Selecione seus respectivos pedaços de EEG
  $\mathbf{Y}^{(au)} = \{Y_1^{(au)}, Y_2^{(au)}, \ldots, Y_{n_1}^{(au)}\}$
  $\mathbf{Y}^{(bu)} = \{Y_1^{(bu)}, Y_2^{(bu)}, \ldots, Y_{n_2}^{(bu)}\}$

Método projetivo

Fixadas duas sequências $au = au_1u_2u_3$ e $bu = bu_1u_2u_3$

- Selecione seus respectivos pedaços de EEG
  $\mathbf{Y}^{(au)} = \{Y_1^{(au)}, Y_2^{(au)}, \ldots, Y_{n_1}^{(au)}\}$
  $\mathbf{Y}^{(bu)} = \{Y_1^{(bu)}, Y_2^{(bu)}, \ldots, Y_{n_2}^{(bu)}\}$

$*$ Cada $Y_k^{(\cdot)}$ é um função em $L^2([0, T])$

Método projetivo

Fixadas duas sequências $au = au_1u_2u_3$ e $bu = bu_1u_2u_3$

- Selecione seus respectivos pedaços de EEG
  $\mathbf{Y}^{(au)} = \{Y_1^{(au)}, Y_2^{(au)}, \ldots, Y_{n_1}^{(au)}\}$
  $\mathbf{Y}^{(bu)} = \{Y_1^{(bu)}, Y_2^{(bu)}, \ldots, Y_{n_2}^{(bu)}\}$

∗ Cada $Y_k^{(\cdot)}$ é um função em $L^2([0, T])$

- Gere uma movimento browniano B em $[0, T]$

Movimento Browniano

Método projetivo

Fixadas duas sequências $au = au_1u_2u_3$ e $bu = bu_1u_2u_3$

- Selecione seus respectivos pedaços de EEG
  $\mathbf{Y}^{(au)} = \{Y_1^{(au)}, Y_2^{(au)}, \ldots, Y_{n_1}^{(au)}\}$
  $\mathbf{Y}^{(bu)} = \{Y_1^{(bu)}, Y_2^{(bu)}, \ldots, Y_{n_2}^{(bu)}\}$

* Cada $Y_k^{(\cdot)}$ é um função em $L^2([0, T])$

- Gere uma movimento browniano B em $[0, T]$

  Movimento Browniano

- Para cada $Y_k^{(\cdot)}$, projete $Y_k^{(\cdot)}$ em B
  $R_k^{(\cdot)} = \int_0^T Y_k^{(\cdot)}(t)B(t)dt \in \mathbb{R}$

Método projetivo

Fixadas duas sequências $au = au_1u_2u_3$ e $bu = bu_1u_2u_3$

- Selecione seus respectivos pedaços de EEG
  $\mathbf{Y}^{(au)} = \{Y_1^{(au)}, Y_2^{(au)}, \ldots, Y_{n_1}^{(au)}\}$
  $\mathbf{Y}^{(bu)} = \{Y_1^{(bu)}, Y_2^{(bu)}, \ldots, Y_{n_2}^{(bu)}\}$

* Cada $Y_k^{(\cdot)}$ é um função em $L^2([0, T])$

- Gere uma movimento browniano B em $[0, T]$

  Movimento Browniano

- Para cada $Y_k^{(\cdot)}$, projete $Y_k^{(\cdot)}$ em B
  $R_k^{(\cdot)} = \int_0^T Y_k^{(\cdot)}(t)B(t)dt \in \mathbb{R}$

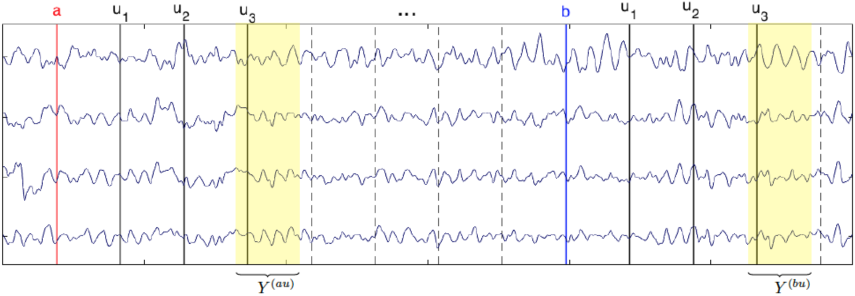
- Com isso, construímos dois conjuntos de números reias
  $\mathbf{R}^{(au)} = \{R_1^{(au)}, R_2^{(au)}, \ldots, R_{n_1}^{(au)}\}$
  $\mathbf{R}^{(bu)} = \{R_1^{(bu)}, R_2^{(bu)}, \ldots, R_{n_2}^{(bu)}\}$

Método projetivo

$$\mathcal{L}(\mathbf{Y}_k^{(au)}) \neq \mathcal{L}(\mathbf{Y}_k^{(bu)})$$

$$\Updownarrow$$

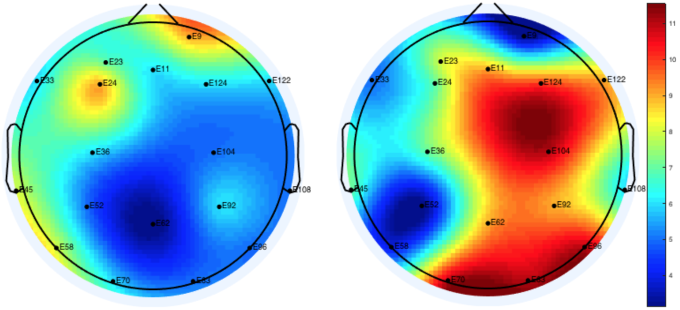
$$\mathcal{L}(\mathbf{R}_k^{(au)}) \neq \mathcal{L}(\mathbf{R}_k^{(bu)})$$



- Se $\mathcal{L}(\mathbf{R}^{(au_1u_2u_3)}) \neq \mathcal{L}(\mathbf{R}^{(bu_1u_2u_3)})$, então $au_1u_2u_3,\ bu_1u_2u_3 \in \hat{\tau}$
- Se $\mathcal{L}(\mathbf{R}^{(au_1u_2u_3)}) = \mathcal{L}(\mathbf{R}^{(bu_1u_2u_3)})$, então $au_1u_2u_3,\ bu_1u_2u_3 \notin \hat{\tau}$
  - Recomeçamos testando as sequências au_2u_3 e bu_2u_3

Resultados experimentais

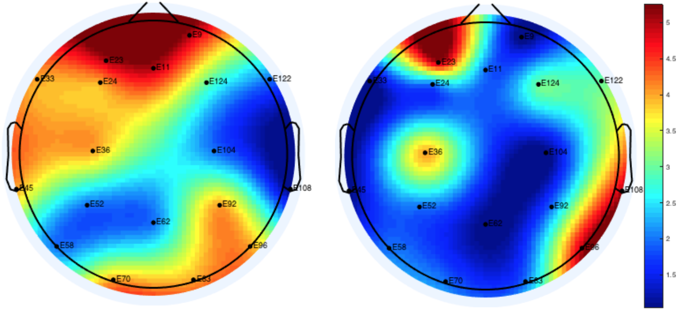
Resultados experimentais - 2 é contexto



Valsa

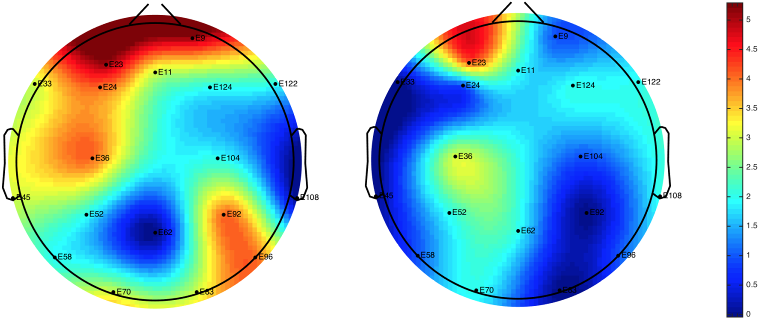
Samba

Resultados experimentais - 1 **não** é contexto



Valsa　　　　Samba

Resultados - 2 é contexto and 1 não é um contexto



Valsa

Samba