Neuroinformatics Journal Club
Dynamic Brain Networks (HMM)
Electrophysiology

Horacio Gómez-Acevedo
Department of Biomedical Informatics
University of Arkansas for Medical Sciences

July 29, 2022

UAMS

Today's paper

NeuroImage 126 (2016) 81–95

Contents lists available at ScienceDirect

NeuroImage

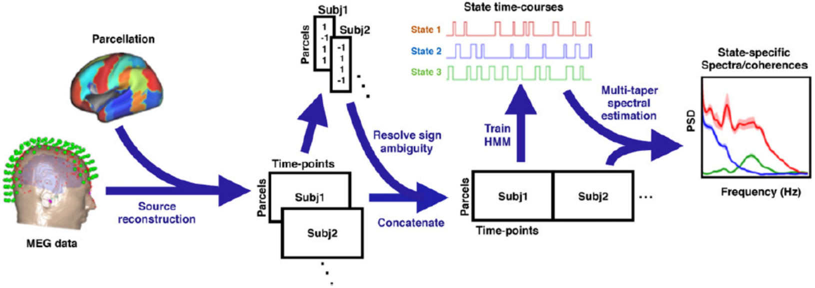
journal homepage: www.elsevier.com/locate/ynimg

Spectrally resolved fast transient brain states in electrophysiological data

Diego Vidaurre [a,*], Andrew J. Quinn [a], Adam P. Baker [a], David Dupret [b], Alvaro Tejero-Cantero [b,c], Mark W. Woolrich [a]

[a] Oxford Centre for Human Brain Activity, Department of Psychiatry, University of Oxford, UK
[b] MRC Brain Network Dynamics Unit, Department of Pharmacology, University of Oxford, UK
[c] Computational Neuroscience, Department Biologie II Ludwig Maximilian, University Munich, Germany

Workflow



Parcellation
MEG data
Source reconstruction
Subj1
Subj2
Parcels
1
-1
1
1
-1
1
1
-1
Time-points
Subj1
Subj2
Resolve sign ambiguity
Concatenate
State time-courses
State 1
State 2
State 3
Train HMM
Parcels
Subj1
Subj2
...
Time-points
Multi-taper spectral estimation
State-specific Spectra/coherences
PSD
Frequency (Hz)

Modeling ingredients

The paper follows some of the similar themes as the previous one

- Hidden Markov Models
- Autoregressive models

To make things "interesting" it adds some more topics

- Bayesian methods
- Some Fourier Analysis

Basic Setup

We consider K hidden states, and T time points.
Let $\mathbf{y}_t^N$ be the multichannel source signal.
The construction of the multivariate autoregressive (MAR) model is different than before.

$$\mathbf{y}'_t | x_t = k \sim \mathcal{N}\left(\sum_{l \in A} \mathbf{y}'_{t-1} \mathbf{W}_l^{(k)}, \Sigma^{(k)}\right) \qquad (1)$$

Don't panic (just yet)

- $\mathbf{y}'_t | x_t = k$ This part refers to as the emission at time t when you are in the hidden state k. It is the transpose of the vector of N entries.
- The $\mathcal{N}$ refers to the multivariate normal distribution. In the 1-dimensional case $\mathcal{N}(\mu, \sigma)$ the μ is transformed into a vector, and the σ by a matrix (variance-covariance).

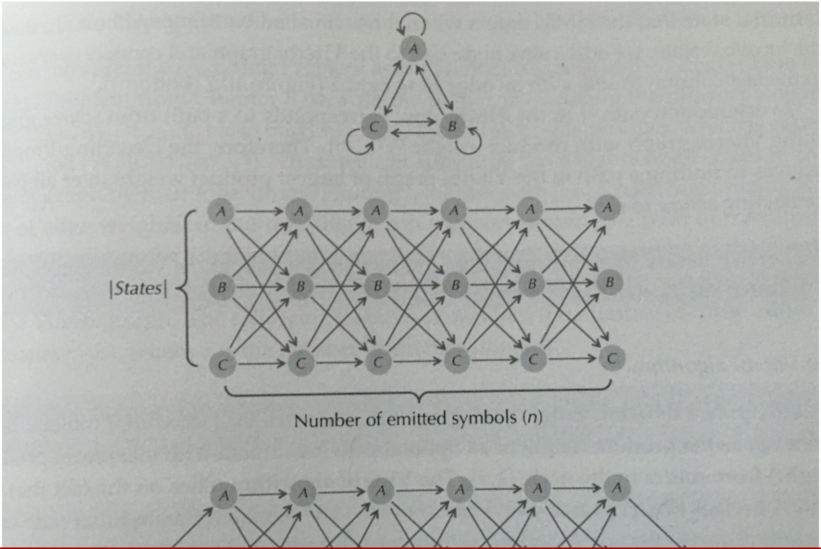
Basic Setup (cont)

- $\mathbf{y}'_{t-1}\mathbf{W}_l^{(k)}$ refers to as the MAR component. The matrix $\mathbf{W}_l^{(k)}$ is of type $N \times N$ for the kth (hidden) state for the lag l. We add all the individual contributions of each lag step.
- $\Sigma^{(k)}$ is the noise covariance matrix.

The Markovian property

$$P(x_t = k_1 | x_{t-1} = k_2) = \Theta_{k_1 k_2}, \quad P(x_t = k) = \eta_k \tag{2}$$

Viterbi's representation HMMs



MAR model

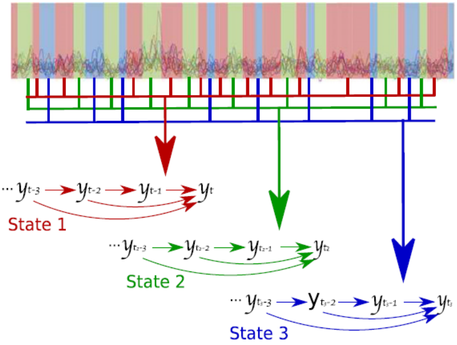


Fig. 2. Graphical representation of the HMM–MAR. The time series (background) is partitioned into three states denoted by the blue, red and green slabs. Each state is characterised by a different set of dynamics, determined by the linear historical interactions between data points $\boldsymbol{y}_t$ (small arrows).

MAR model (cont)

The paper key component for the MAR section is the $\mathcal{A}$ set. It contains the number of previous steps to be considered

$$\mathcal{A} = \{P_0 + 1, P_0 + Q, P_0 + \left\lfloor \frac{1}{2} Q^2 \right\rfloor, \ldots, P\} \tag{3}$$

It calls it exponential due to the Taylor representation of the exponential function

$$\exp(X) = 1 + X + \frac{1}{2!}X^2 + \frac{1}{3!}X^3 + \cdots$$

Oh Bayes where art thou?

Unlike the traditional (frequentist) inference, the Bayesian approach tries to determine conclusions about a parameter θ or unobserved data $\bar{y}$ as probability statements. For example, given the observed data y, Bayesian statements are based on the conditionals such as $P(\bar{y}|y)$ for *new data* $\bar{y}$, or $P(\theta|y)$ for a parameter θ. So following our Bayes theorem,

$$P(\theta|y) = \frac{P(\theta)P(y|\theta)}{P(y)} \tag{4}$$

where $P(y) = \sum_\theta P(\theta)P(y|\theta)$ (or in the continuous case $P(y) = \int P(\theta)P(y|\theta)d\theta$).
The expression $P(\theta)$ is called the **prior**, whereas $P(\theta|y)$ is called **posterior**, and $P(y|\theta)$ is called **likelihood**, and $P(y)$ is called **evidence**.

Oh Bayes where art thou? (cont)

Table 5.5 Applying Bayes' rule to data and parameters

Data	Model parameter			Marginal
	...	θ value	...	
⋮		⋮		⋮
D value	...	$p(D,\theta) = p(D\|\theta)\,p(\theta)$	...	$p(D) = \sum_{\theta^*} p(D\|\theta^*)\,p(\theta^*)$
⋮		⋮		⋮
Marginal	...	$p(\theta)$	...	

When conditionalizing on row value D, the conditional probability $p(\theta|D)$ is the cell probability $p(D,\theta)$ divided by the marginal probability $p(D)$. When these probabilities are algebraically re-expressed as shown in the table, this is Bayes' rule. This table is merely Table 5.1 with its rows and columns re-named.

Oh Bayes where art thou? (cont)

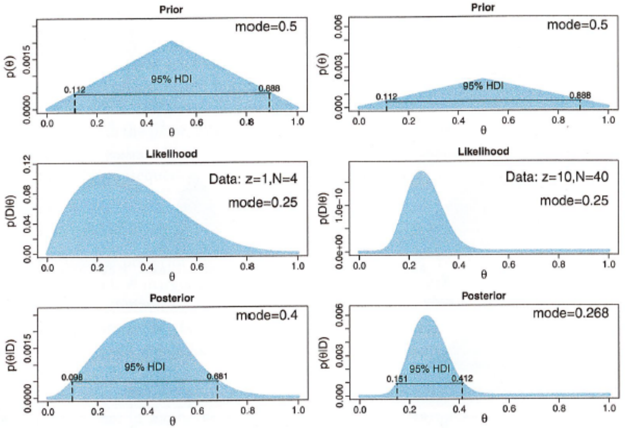


Figure 5.2 The two columns show different sample sizes with the same proportion of heads. The prior is the same in both columns but plotted on a different vertical scale. The influence of the prior is overwhelmed by larger samples, in that the peak of the posterior is closer to the peak of the likelihood function. Notice also that the posterior HDI is narrower for the larger sample.

Oh Bayes where art thou?

The precision matrix is modeled with a Wishart distribution

$$\Omega^{(k)} \sim \mathcal{W}(t_0, \mathbf{B}_0) \tag{5}$$

An example: Sample covariance for idd multivariate normal. You can think about it as a multivariate version of the Gamma distribution.

Priors.

We expect each state to be characterized by a certain set of connections and certain frequency profile... $\sigma_{ij}^{(k)}$ weight the presence of a specific connection between nodes (i, j) when in state k. The precision $\alpha_j^{(k)}$ adaptively weight to the presence of interaction at a certain lag l for all nodes when in state k.

Bayes cont

The authors generate a *Hierarchical Bayes Model*

$$W_{l_{ij}}^{(k)} \sim \mathcal{N}(0, \sigma_{ij}^{(k)} \alpha_j^{(k)}) \tag{6}$$

where

$$\sigma_{ij}^{(k)} \sim \Gamma(\phi_0, d_0) \tag{7}$$

$$\alpha_j^{(k)} \sim \Gamma(\xi_0, d_0) \tag{8}$$

Bayes cont

$$\Theta_k \sim Dir(\nu_0) \quad \eta \sim Dir(\psi_0) \tag{9}$$

The Dirichlet distribution is the conjugate prior for the parameters of the normal multinomial distribution.

Inference of Model Parameters

Bayesian models are difficult to perform since the Bayes' rule involves computing the evidence (nasty math problem). Thus, the models are restricted to ones with *simple* likelihood functions with corresponding formulas for prior distributions (called *conjugate priors*) that play nice with the likelihood function. Another alternative is to use what is called *variational approximation*.

Spectral Properties

The authors use something called *Fourier Transform* to investigate find out the multiregion spectral properties of the state time courses.
In a nutshell this methodology takes a real function f and it produces a continuous function of frequency. So, the expression

$$S(f) = \frac{}{R} \sum_{r=1}^{R} \sum_{t=1}^{T} \delta_t^{(r)} y_t \exp(-2\pi i f t) \tag{10}$$