O'Bayes 2022 - UCSC

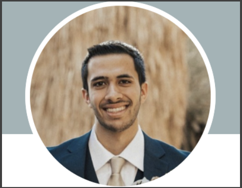
Functional Partial/Mixed Membership Models

Donatello Telesca
UCLA, Fielding School of Public Health
Department of Biostatistics

Joint Work with
Nicholas Marco, Damla Şentürk, Shafali Jeste, and Abigail Dickinson

The Team at UCLA

UCLA Biostatistics



Nicholas Marco

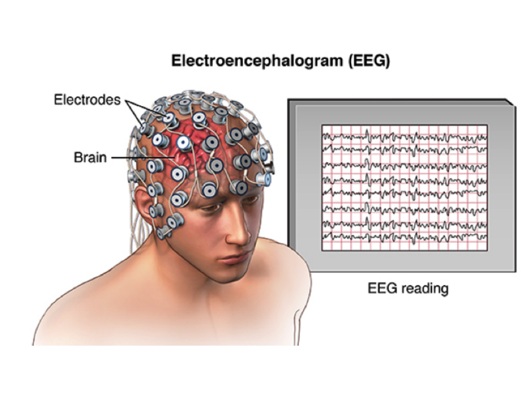
Damla Şentürk

UCLA Semel Institute

Shafali Jeste

Abigail Dickinson

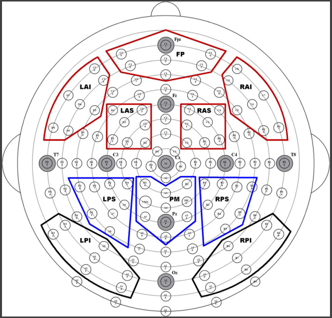
Motivation: EEG as a Functional Brain Imaging Modality

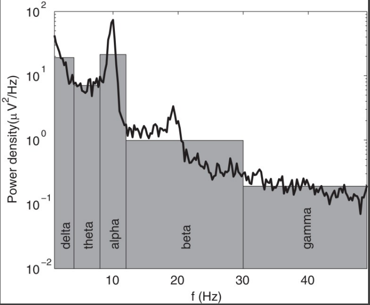


◦ EEG sensors measure distributed neuronal activity on cortical patches perpendicular to the sensors

◦ We study the response of a population of neurons – [Learning, memory formation, task execution, ...]

Resting State EEG and Spectral Features

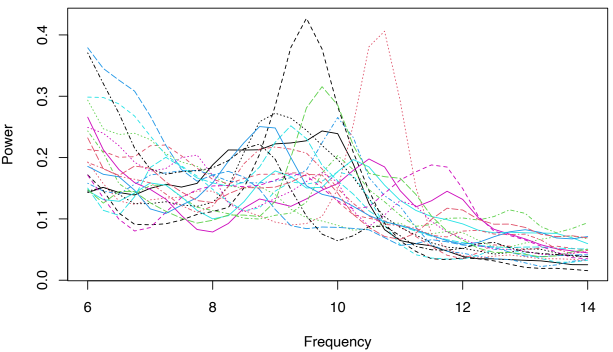




- Power spectrum analysis associates spectral features in a specific frequency range with bio-behavioral characterizations of brain activity.

- We focus on the alpha frequency range whose patterns at rest are thought to play a role in neural coordination and communication between distributed brain regions.

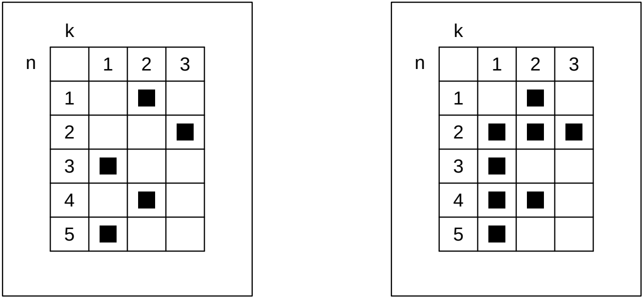
EEG Spectral Power (ASD + TD)



- Can we use spectral power dynamics to identify latent neuro-developmental classes?

- Is the uncertain membership (clustering) framework appropriate for this application?

Uncertain Membership (Clustering) Vs. Mixed Membership



Mixed Membership and Feature Allocation Models

$$\underbrace{p(Z)}_{\text{Allocation Matrix}} \quad \underbrace{p(Y \mid Z, \boldsymbol{\theta}_z)}_{\text{Sampling Model}}$$

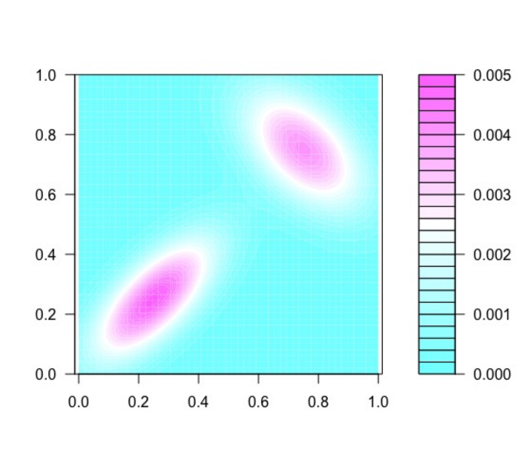
Probability Models over Multisets - $p(Z)$

- IBP - (Ghahramani and Griffiths, 2006)
- Exchangeable feature probability function - (Broderick et al., 2013)
- Non-exchangeable schemes - (Benedetto et al., 2020)

Applications

- Topic modeling (Williamson et al., 2010),
- Image analysis (Zhou et al., 2011),
- Inference in tumor heterogeneity (Lee et al., 2015; Xu et al., 2015).

Uncertain Membership (Clustering) vs. Mixed Membership



Functional Data Analysis

- Functional Data Analysis (FDA) focuses on methods used to analyze sample paths of an underlying continuous stochastic process Y
- Typically we consider:

$$Y_i(t) = f_i(t) + \epsilon_i(t); \quad f_i(t) \sim GP\{\mu(t), C(\cdot,\cdot)\}; \quad \epsilon_i(t) \sim N(0, \sigma_\epsilon^2)$$

Note: Often the literature on GP focuses on direct (parametrized) modeling of the covariance function $C(\cdot,\cdot)$

Example: $C(s,t) = a^2 \exp\{-0.5||s-t||^2/\ell^2\}$

FDA: Estimation of $C(s,t)$ from random samples $[Y_1(t), \ldots, Y_n(t)]$

- Established literature on flexible priors for $C(\cdot,\cdot)$ [Yang et al., 2017; Montagna et al., 2012; Shamshoian et al., 2022]

Functional Clustering (GP Mixtures)

- The FDA literature on clustering is very mature (James and Sugar, 2003; Chiu and Li, 2007) .
- From a Bayesian perspective, assuming there exist K latent GPs

$$f^{(k)} \sim \mathcal{GP}\left(\mu^{(k)}, C^{(k)}\right), \quad k = 1, 2, \ldots, K$$

Each sample paths f_i, (i=1,2,. . . , N), follows a finite mixture of GPs:

$$p\left(f_i \mid \rho^{(1:K)}, \mu^{(1:K)}, C^{(1:K)}\right) = \sum_{k=1}^{K} \rho^{(k)} \, \mathcal{GP}\left(f_i \mid \mu^{(k)}, C^{(k)}\right);$$

where $\rho^{(k)} \in [0, 1]$ is the mixing proportion quantifying uncertain membership to GP (k).

Mixed Membership Functions

- Allowing sample paths $(f_1, f_2, \ldots, f_N)$ mixed membership to K underlying GPs is conceptually straightforward.

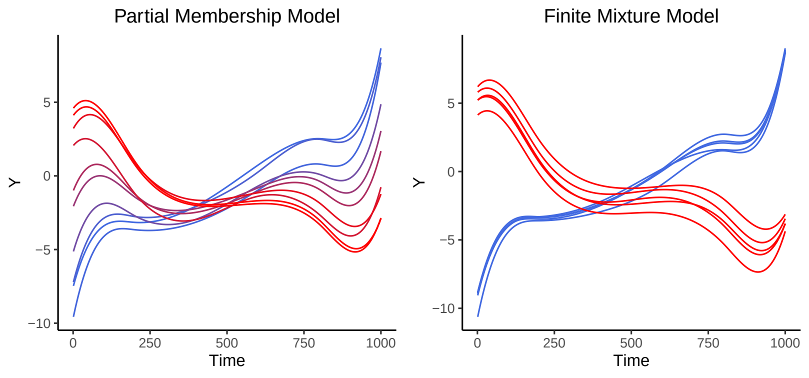
- Introducing path-specific mixed membership probabilities $\mathbf{z}_i = [Z_{i1} \cdots Z_{iK}]$, with $Z_{ik} \in (0,1)$, we define a mixed membership process as follows

$$f_i \mid \mathbf{z}_i = \sum_{k=1}^{K} Z_{ik} f^{(k)}$$

where $\sum_{k=1}^{K} Z_{ik} = 1$,

- Typically, we cannot assume that the latent $\mathrm{GP}(f^{(k)})$ are mutually independent

Functional Clustering vs. Functional Mixed Membership



Mixed Membership Vs. Local Functional Clustering

Related approaches in the literature:

- Local clustering on random partitions of the evaluation domain - (Petrone et al., 2009)

- Clustering based on local functional features - (Suarez and Ghosal, 2016)

Mixed Membership Functions

- The proposed sampling model assumes

$$f_i \mid \cdots \sim GP\left(\sum_k Z_{ik}\mu^{(k)}, \sum_k Z_{ik}^2 C^{(k)} + \sum_k \sum_{k' \neq k} Z_{ik} Z_{ik'} C^{(k,k')}\right)$$

- Model K Gaussian Processes (GPs), $f^{(k)}$
  - K mean functions, $\mu^{(k)}(t)$
  - K covariance functions, $C^{(k,k)}(s,t)$
  - $\frac{K(K-1)}{2}$ cross-covariance functions, $C^{(k,j)}(t_k, t_j)$
- These functions are infinite dimensional and computationally intractable
- We desire a concise and efficient finite representation of the K Gaussian Processes

Joint Representation of K Gaussian Processes

- We assume $f^{(k)}$ can be represented by a set of **uniformly continuous** basis functions.
- Let $\mathbf{B}(t)$ is a vector of the P basis functions evaluated at t
- The Multivariate Karhunen-Loève theorem (Happ and Greven, 2018) jointly decomposes K GPs:

$$f^{(k)}(t) = \boldsymbol{\nu}_k' \mathbf{B}(t) + \sum_{m=1}^{KP} \chi_m \boldsymbol{\phi}_{km}' \mathbf{B}(t), \tag{1}$$

  where $\boldsymbol{\nu}_k \in \mathbb{R}^P$, $\boldsymbol{\phi}_{km} \in \mathbb{R}^P$, and $\chi_m \sim \mathcal{N}(0,1)$

- Using this decomposition, we have:
  - $\mu^{(k)}(t) = \boldsymbol{\nu}_k' \mathbf{B}(t)$
  - $C^{(k,j)}(t_k, t_j) = \sum_{m=1}^{KP} \boldsymbol{\phi}_{km}' \mathbf{B}(t_k) \boldsymbol{\phi}_{jm}' \mathbf{B}(t_j)$

Multivariate Karhunen-Loève Theorem (cont.)

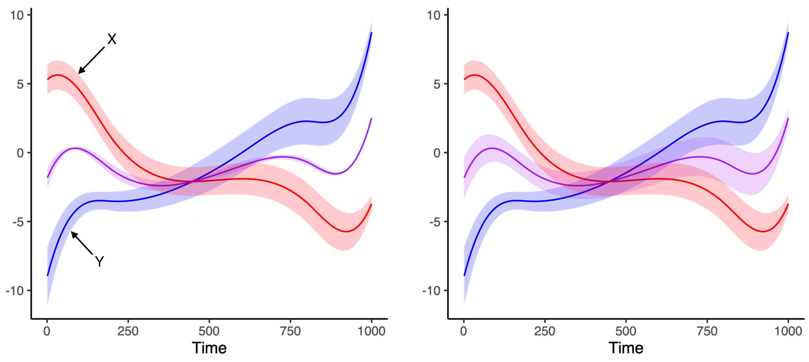
- The Karhunen-Loève theorem typically allows for a reduced dimensional representation with $M \leq KP$ components, s.t.

$$f^{(k)}(t) \approx \boldsymbol{\nu}_k' \mathbf{B}(t) + \sum_{m=1}^{M} \chi_m \boldsymbol{\phi}_{km}' \mathbf{B}(t), \qquad (2)$$

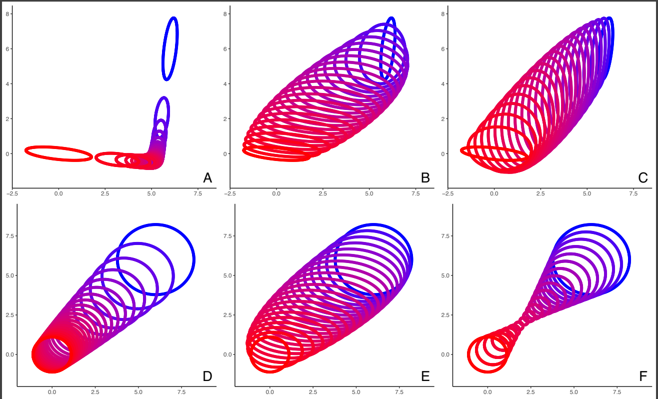
- Number of parameters needed to model the covariance structure:
  - Multivariate Karhunen-Loève: $\mathcal{O}(KPM)$
  - Naïve : $\mathcal{O}(K^2P^2)$

Effects of the Cross-Covariance Function

$$\mathrm{Cov}^{(X,Y)}(s,t) = \mathrm{Cov}\left(X(s), Y(t)\right)$$



Relation to Other PMMs (Multivariate Analysis)



- The proposed representation is not unique and builds on several related ideas in multivariate analysis (Heller et Al., 2008), (Griffiths and Ghahramani, 2011), (Broderick et Al. 2013).

Finite Dimensional Margins

- $Z_{ik} \in (0,1) \longrightarrow$ mixed membership probability of path i to GP (k).
- Using the multivariate KL construction, we obtain:

$$y_i(t)|\boldsymbol{\Theta} \sim \mathcal{N}\left(\sum_{k=1}^{K} Z_{ik} \underbrace{\left(\boldsymbol{\nu}_k' \mathbf{B}(t) + \sum_{m=1}^{M} \chi_{im} \boldsymbol{\phi}_{km}' \mathbf{B}(t)\right)}_{f^{(k)}(t)},\ \sigma^2\right) \tag{3}$$

- Integrating over χ_i yields

$$y_i(\mathbf{t}_i)|\boldsymbol{\Theta}_{-\chi} \sim \mathcal{N}\left(\sum_{k=1}^{K} Z_{ik} \underbrace{\mathbf{S}'(\mathbf{t}_i)\boldsymbol{\nu}_k}_{\boldsymbol{\mu}^{(k)}(\mathbf{t}_i)},\ \sum_{k=1}^{K}\sum_{j=1}^{K} Z_{ik} Z_{ij} \underbrace{\left(\mathbf{S}'(\mathbf{t}_i) \sum_{m=1}^{M} \left(\boldsymbol{\phi}_{km} \boldsymbol{\phi}_{jm}'\right) \mathbf{S}(\mathbf{t}_i)\right)}_{C^{(k,j)}(\mathbf{t}_i, \mathbf{t}_i)} + \sigma^2 \mathbf{I}_{n_i}\right) \tag{4}$$

Prior Distributions

- The ϕ parameters construct scaled eigenfunctions of the covariance operator
  - Mutually orthogonal
  - Magnitude of the scaled eigenfunctions should decrease
    - Multiplicative gamma process shrinkage prior (Bhattacharya and Dunson, 2011)

$$\phi_{kpm}|\gamma_{kpm}, \tilde{\tau}_{mk} \sim \mathcal{N}\left(0, \gamma_{kpm}^{-1}\tilde{\tau}_{mk}^{-1}\right),$$

$$\gamma_{kpm} \sim \Gamma\left(\nu_\gamma/2, \nu_\gamma/2\right), \quad \tilde{\tau}_{mk} = \prod_{n=1}^{m} \delta_{nk},$$

$$\delta_{1k} \sim \Gamma(a_{1k}, 1), \quad \delta_{jk} \sim \Gamma(a_{2k}, 1), \quad a_{1k} \sim \Gamma(\alpha_1, \beta_1), \quad a_{2k} \sim \Gamma(\alpha_2, \beta_2)$$

Posterior Distributions

- Let $\mathbf{\Sigma}_{jk} := \sum_{p=1}^{KP}\left(\boldsymbol{\phi}_{jp}\boldsymbol{\phi}'_{kp}\right)$ and

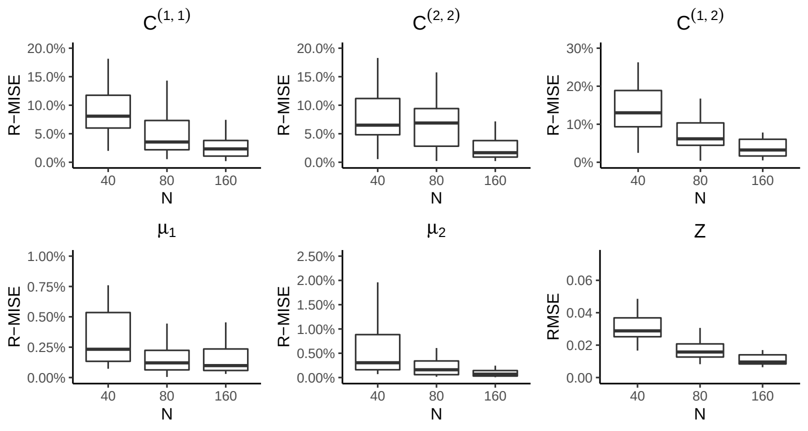
$$\boldsymbol{\omega} := \left\{\boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_K, \mathbf{\Sigma}_{11}, \ldots, \mathbf{\Sigma}_{1K}, \ldots, \mathbf{\Sigma}_{KK}, \sigma^2\right\}.$$

- The parameters in $\boldsymbol{\omega} \in \mathbf{\Omega}$ completely specify the mean and covariance structure of our model. We will denote the true set of parameters as $\boldsymbol{\omega}_0$
- Assumptions:
  1. $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ are observed on a grid of R points in the domain, $\{t_1, \ldots, t_R\}$
  2. The variables Z_{ik} are fixed and known (not-random)
  3. $\sigma_0^2 > 0$
- Consider the fully saturated model (M = KP). Under these assumptions, the posterior distribution is weakly consistent at $\boldsymbol{\omega}_0 \in \mathbf{\Omega}$

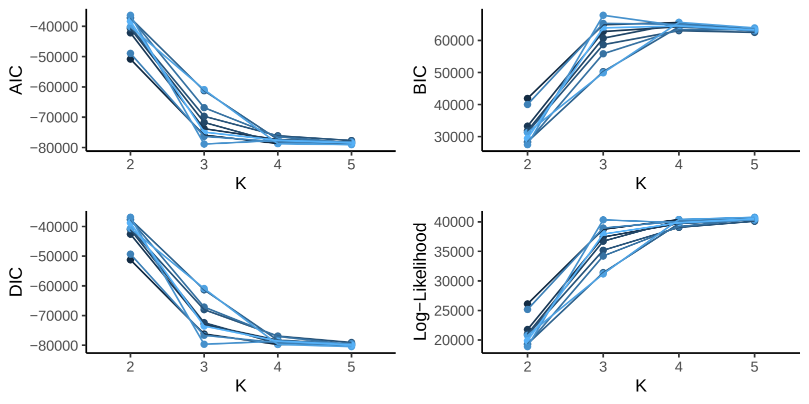
Posterior Simulation

- Posterior simulation is conducted by using Metropolis-within-Gibbs sampling
  - Implemented in RCPP
- Tempered transitions are utilized to move across areas of low posterior probability
- Post-processing is conducted to help with identifiability and the interpretability of the parameters

Operating Characteristics on Engineered Data

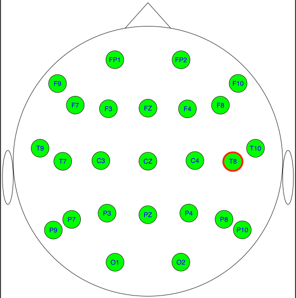


Selecting the Number of Features



Case Study: Peak Alpha Frequency (TD and ASD)

- Autism spectrum disorder (ASD) is a term used to describe individuals with a collection of social communication deficits and restricted or repetitive sensory-motor behaviors
- This case study contains electroencephalogram (EEG) data for 39 typically developing (TD) children and 58 children with ASD between the ages of 2 and 12 years old
- We fit a 2 functional feature mixed membership model on data from the T8 electrode



EEG Case Study Data

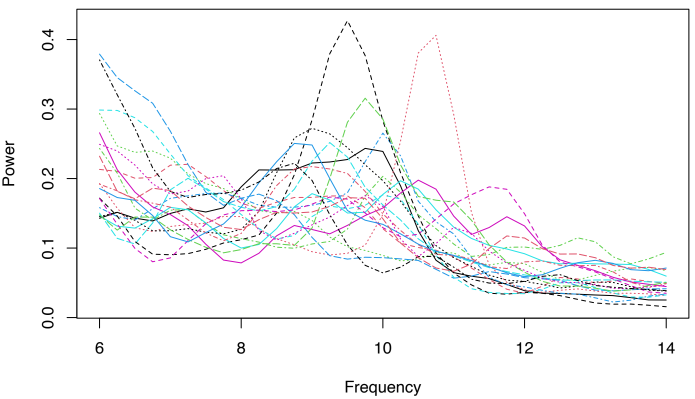


Figure: EEG data from the T8 electrode for 20 individuals (ASD and TD)

EEG Case Study Data (cont.)

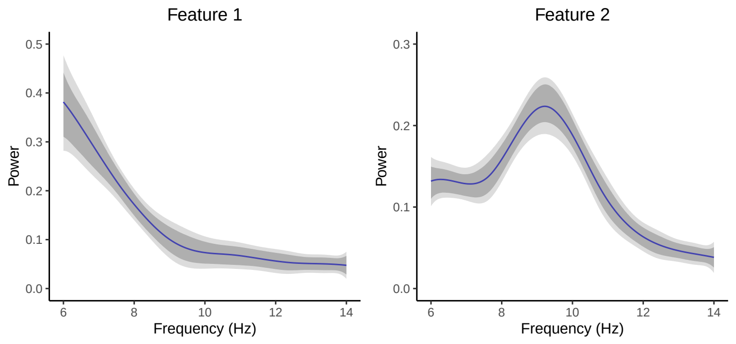


Figure: Posterior median and 95% credible (pointwise credible interval in dark gray and simultaneous credible interval in light gray) of the mean function for each latent functional feature.

EEG Case Study (cont.)

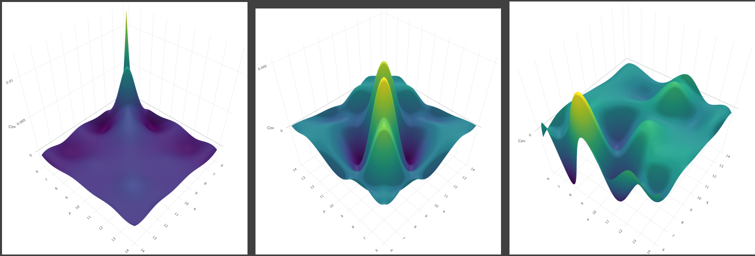
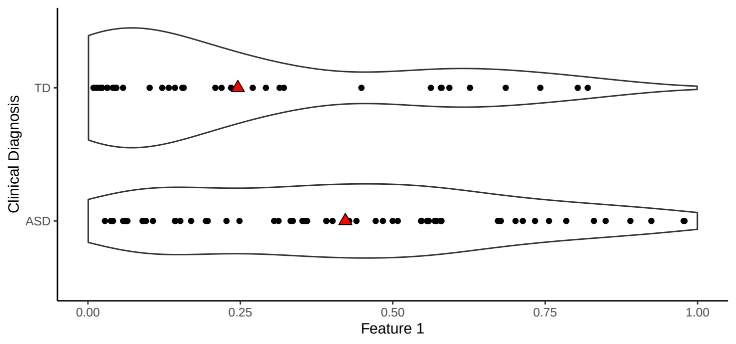


Figure: Posterior estimates of the covariance functions (From left to right: covariance of feature 1, covariance of feature 2, cross-covariance between features 1 and 2)

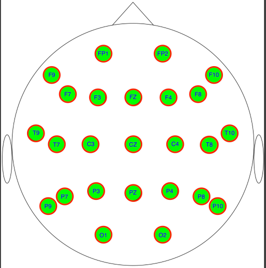
EEG Case Study (cont.)



- Children with an TD clinical diagnosis are highly likely to load on the second functional feature, whereas children with ASD exhibit a higher level of heterogeneity

Analysis of Multi-Channel EEG Data

- In the previous case study, we only used the T8 electrode and discarded the information from the 24 other electrodes
- For this case study, we will model all electrodes using a functional model, assuming $\mathcal{T} \subset \mathbb{R}^3$
  - Two of the indices will contain the spatial location of the electrodes
  - The third index will contain the frequency domain



Analysis of Multi-Channel EEG Data (cont.)

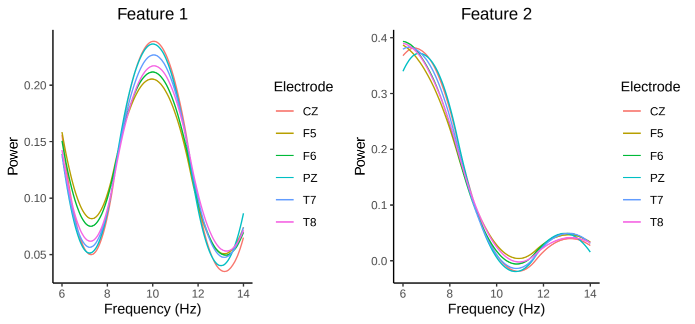


Figure: Posterior estimates of the means of the two functional features viewed at specific electrodes of interest.

Analysis of Multi-Channel EEG Data (cont.)

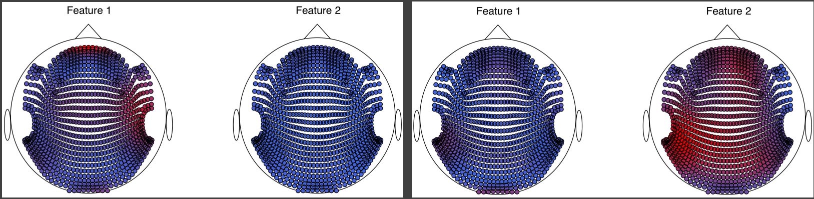


Figure: Variance of electrodes at 6 Hz (left) and 10 Hz (right)

- For the second functional feature, we can see that there is high heterogeneity around the T8 electrode at 6 Hz

Summary

- Functional Mixed Membership Models are likely important in BioX applications
- Multivariate KL constructions allow for efficient representation and dimension reduction of multivariate GPs
- Some work is needed for dimension selection and theoretical guarantees on latent membership
- Some work is needed to account for covariate information
- In our applications, results are robust to increasing dimensionality (multi-channel analyses)

Thank You!

R Packages

HFM	Multivariate FDA	https://github.com/Qian-Li/HFM
BayesFMMM	Funct. Mixed Membership Models	https://github.com/ndmarco/BayesFMMM

Manuscripts

◦ Marco N, Senturk D, Jeste S, Dickinson A and D. Telesca D (2022) *Functional Partial Membership Models.* (arXiv:2206.12084).

◦ Shamshoian J, Senturk D, Jeste S, Telesca D, (2022) *Bayesian Analysis of Multidimensional and Longitudinal Functional Data.* (Biostatistics)

◦ Li Q, Senturk D, Sugar C, DiStefano C, Jeste S, Telesca D, (2020) *Region-referenced Spectral Power Dynamics of EEG Signals: a Hierarchical Modeling Approach.* (AOAS)

Funding

R01 MH122428-01 (DS, DT) from the NIH/NIMH