# Predicting neuro-developmental scores from electroencephalogram data: Stochastic modelling and machine learning

M. Veretennikova    A. Sikorskii    M. Boivin

Snegiri conference, NRU HSE, 4th - 7th of December, 2017

7th of December, 2017

# Outline

- Cerebral malaria
- Coma
- EEG
- Methods
- Results
- Further work
- References

# Cerebral malaria (CM)

- Cerebral malaria (CM) affects over half a million people annually.
- High prevalence in sub-Saharian Africa.
- Complications may include coma, seizures, convulsions, neurodevelopmental impairments, and metabolic disturbances.
- The mortality rate among children is over 40 percent.
- A large proportion of children who recover from CM have neurological consequences, and some develop difficulties in cognition and behaviour.

# Cerebral malaria (CM)

- During CM the red blood cells are parasitized, most often by *Plasmodium Falciparum*.
- During CM, the blood-brain barrier is broken.
- Brain capillaries are blocked and brain swelling occurs.

# Coma

- Coma is defined as prolonged unconsciousness and unresponsiveness.
- It helps the body to heal from the damage before awaking.
- Neurologists differentiate between the alpha coma, theta coma, high-voltage delta coma, low-voltage delta coma, burst suppression, spindle coma and electro-cerebral inactivity.

# Electroencephalogram recordings

- Electroencephalography allows to monitor electrical activity within the brain.
- The electrodes are connected to the head skin with a gel.
- Usually they are evenly spaced on the scalp.
- A common choice is 32 EEG channels, but modern devices use many more electrodes.
- EEG measures the changes in electric potentials with a common reference to a quiet electrode.

# Brain activity and EEG

- Neuron cells are interconnected by synapses: special connections between cells.
- Synapses are gateways for passing or retaining information to be passed from one neuron to another.
- Each synaptic activity generates an electrical impulse.
- For EEG to detect electric activity within the brain, thousands of neurons must fire in sync,
- because to be detected the collective impulse must get through the head tissues and the skull.

# Within the brain

- Na+ (sodium), K+ (potassium), Ca++ (calcium) and Cl- (chloride) ions play the key role during neuronal firing.
- Also there are neurotransmitters, such as glutamate and GABA, which can amplify/relay signals between neurons and other cells.
- Astrocytes (glial cells) play a role in supporting neurons.

# Traditional EEG signal decomposition by frequency ranges

- Gamma ($> 30$ Hz): (Role not certain) Several hypotheses: Epileptic activity, pathologies; learning, info processing, sensory uptake, data exchange between different brain regions
- Beta ($13 - 30$ Hz): Mental activity, survival thought processes, dreaming (paradoxal) sleep
- Alpha ($7.5 - 13$ Hz): Quiet, awake resting state, closed eyes, unfocused relaxation
- Theta ($3.5 - 7.5$ Hz): Falling asleep
- Delta ($< 3.5$ Hz): Deep dream-free sleep, childhood, serious organic brain diseases

# Data

- Where: Mulago National Referral and Teaching Hospital in Kampala, Uganda, 2008– 2015
- Our dataset consists of 54 medical features and approx. 30 minutes of EEG recordings for 78 children, some of the medical data is missing
- 362 statistical features were extracted from the standard 10–20 EEG recordings with the sampling frequency of 500 Hz (approx 500 by 60 by 30 observations for each person).

# Data preprocessing

- The EEG recordings were de-noised using Persyst software, which is based on neural networks.
- The noise was due to breath, heartbeat and possible muscle movement.
- It also allowed us to obtain the binary variable indicating seizure occurrence.

# Methods - LASSO regression

The Least Absolute Shrinkage and Selection Operator (LASSO) is a regression method, which focuses on the following optimization problem:

$$\min_{\bar{\beta}} \left( \sum_{i=1}^{n} \left( y_i - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \alpha \sum_{j=1}^{p} |\beta_j| \right), \qquad (1)$$

for $p >> n$, i.e. there are many more features than observations, where $\alpha$ is a tuning parameter, controlling the penalty for having large parameters $|\beta_j|$.
LASSO regression performs feature subset selection.

# Methods -Elastic Net (EN)

The Elastic Net (EN) technique solves the following optimization problem:

$$\min_{\bar{\beta}} \left( \sum_{i=1}^{n} \left( y_i - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \text{penalty term} \right), \tag{2}$$

where the penalty term is

$$+\alpha \times l_1 \sum_{j=1}^{p} |\beta_j| + 0.5 \times \alpha \times (1 - l_1) \sum_{j=1}^{p} \beta_j^2.$$

EN more suitable in face of multicollinearity than the Lasso when the sample size is large.

# Methods - Group Lasso

The Group Lasso optimization problem:

$$\min_{\beta} \left( \sum_{i=1}^{n} \left( y_i - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \alpha \sum_{k=1}^{L} \sqrt{p_k} \|\beta_k\|_2 \right), \tag{3}$$

where $p_l$ is the size of the $l$-th collinear cluster, and $\|\beta_k\|_2^2$ is the sum of the squares of the coefficients $\beta_j$ within a cluster.

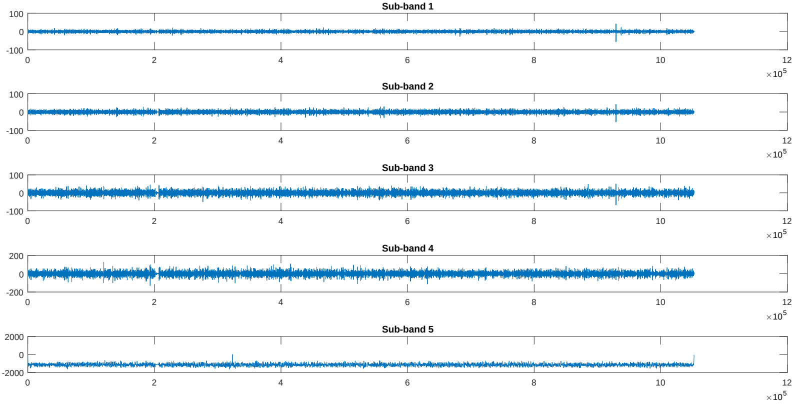
A data-driven approach yielded 28 groups of variables for our dataset, and the variable group assignment was implemented using the k-means algorithm and synthetic cluster principal component construction for correlation analysis.

# Signal processing

Daubechies' (Db4) wavelets for signal separation into frequency bands:

| Traditional | Db4 band's central frequency |
|---|---:|
| Delta 0 – 3.5 Hz | 2.7 Hz |
| Theta 3.5 – 7.5 Hz | 5.57 Hz |
| Alpha 7.5 – 13 Hz | 11 Hz |
| Beta 13 – 30 Hz | 22.3 Hz |
| Gamma $>$ 30 Hz | four subsequent bands |

# Daubechies wavelet signal decomposition



# Feature selection

The EEG features:

- Proportion of the flat line signal
- Variance and first four moments
- Features for delta, alpha, beta and three gamma frequency bands: variance, Shannon entropy, relative frequency range energy
- Frequency of peaks higher than their neighbours by 1/3, 1, 2, 3 standard deviations (FP 1/3, 1, 2, 3 in our notation)
- Seizure presence (binary)

## Medical features

- Height, weight, age
- Blood properties, e.g. platelet count, hemoglobin base level
- The Socio-Economic-Score, Blantyre Coma Score
- Plasma and cerebrospinal fluid features

The Mullen Scales of Early Learning (MSEL) and Kauffman Assessment Battery for Children KABC-2 were used for quantification of the neurodevelopment level, which was the response variable in this prediction problem.

# Features from stochastic modelling

Consider the SDE

$$dX_t = -\theta(X_t - \mu)dt + \sqrt{\frac{2\theta\delta^2}{\nu - 1}\left(1 + \left(\frac{X_t - \mu}{\delta}\right)^2\right)}dW_t,$$

with $\nu > 1$, $\delta > 0$, $\mu \in \mathbb{R}$, $\theta > 0$.
This SDE has a unique ergodic Markovian weak solution, which is a diffusion process $X_t$ with the symmetric scaled Student stationary distribution with parameters $\mu$, $\sigma$ and $\nu$, appearing the SDE above.

# Stochastic modelling

There are other stochastic processes which have the stationary distribution as a Student's t-distribution, for example the Student Ornstein-Uhlenbeck process

$$dX_t = -\lambda X_t dt + dY_t, \tag{4}$$

where $Y_t$ is the corresponding background Levy process (BDLP) process.

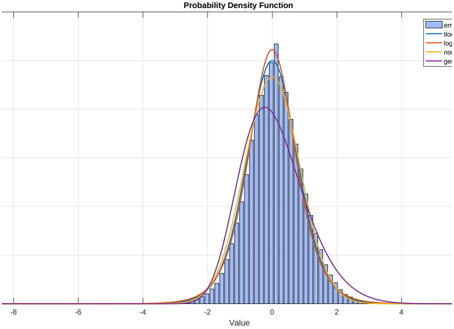
For this process it is not necessary to have the parameter $\nu > 1$.

Our data suggests that $\nu$ isn't definitely above 1 for every frequency band, although for the frequency band D5 it seems to be so for every person for two channels.

Stationarity: we use the rule of thumb for the ratio of variances approx. between 0.25 and 4 (see the references on list 3)

# Features from stochastic modelling

We have considered the increment process for the signal split into frequency bands, and we observed that the t-distribution is a particularly good fit for the histograms:



# Features from stochastic modelling

- So far we have only added the quasi-MLE estimates of $\sigma$'s and $\nu$'s for delta, theta, alpha and beta freq. bands for one channel: F7-CZ.
- We have used the rule of thumb to check stationarity of the increment process heuristically: the ratio of variances in consecutive time intervals is always roughly between 0.25 and 4, and the mean for the process is constant.

# Methods - matrix completion

We had to complete 504 missing matrix entries out of $78 \times 416$

## Lemma

*Suppose the matrix $W_{m\times n}$ has rank $r$. The solution to the optimization problem*

$$\min_{Z}\left(\frac{1}{2}\|W-Z\|_F^2+\lambda\|Z\|_*\right) \tag{5}$$

*is given by $\hat{Z}=S_\lambda(W)$, where $S_\lambda(W)=UD_\lambda V^T$, with $D_\lambda = diag[(d_1-\lambda,\cdots,(d_r-\lambda)_+]$, where $UDV^T$ is the SVD of $W$, $D=diag[d_1,\cdots,d_r]$, and $t_+=\max(t,0)$.*

Here $\|A\|_F$ is the Frobenius norm of a matrix A, whilst $\|A\|_*$ is the sum of the singular values of the matrix A.

# Results

The regularization parameter $\lambda$ was chosen by the following criteria: the MSE after applying the Elastic Net to the completed matrix, and the number of features left in the model by the Elastic Net.

| $\lambda$ | $\alpha$ | l1 ratio | MSE | Non-zero coeff. |
|---|---|---|---|---|
| **75** | 0.01 | 0.9 | **0.367** | **25** |
| **100** | 0.01 | 0.9 | **0.379** | **21** |

The parameters $\alpha$ and l1 ratio were chosen by leave-one-out cross-validation. The value $\lambda = 100$ yields a near-best and the simplest model, so it was selected for further calculations.

# The best results among different models

| Method | MSE | Nonzero coeff. |
| --- | --- | --- |
| **EN, 424 features with 8 from stochastic modelling** | **0.073** | 68 |
| **EN for 416 features, 54 of these are non-EEG** | **0.379** | 21 |
| **Lasso for 416 features** | **0.3893** | 15 |

Accounting for multicollinearity didn't help in any way, and made the predictions worse in all cases.

# Results for the two models:

- EN: Medical: weight, height, BCS score, platelet count, interleukin-1 receptor alpha (plasma) and range activated clotting time.
- EN: Ten of the 21 coefficients have EEG nature: from the FP 2 and 1 groups, 7 come from the Daubechies wavelet analysis: the variance variables for *delta* and *alpha* bands as well as the entropy variable for the *theta* band
- EN with 424 features: Student's t-distribution fit parameters from each of delta,theta, alpha, beta frequency bands are present among the 68 features selected by the Elastic Net
- Lasso: 9 of the 15 are of EEG nature: FP2, FP1, variance in the *delta* band, Shannon entropy for the *theta* band.

# Multicollinearity - the Elastic Net

- Among the remaining variables only variance in the *theta* band for channels 3 and 12 had a high absolute value of the correlation: 0.949175.
- All the other correlations were lower than 0.66.
- This indicates that the Elastic Net has dealt with multicollinearity very well.

# Other results

| Method | MSE | Nonzero coeff. |
| --- | --- | --- |
| EN, exclude non-EEG features with over 6 missing values for the feature, feature group repr. | 0.5531 | 65 |
| EN, only non-EEG features | 0.6396 | 54 |
| EN, only EEG features | 0.59214 | 55 |

# Conclusion

- It is beneficial to use features of EEG nature in predicting neuro-developmental scores of children after coma due to cerebral malaria.
- Features from stochastic modelling dramatically improve performance of the regression models
- Useful EEG biomarkers: variance in the delta and alpha frequency bands, entropy in the theta band, FP2 and FP1, $\sigma$ and $\nu$ parameters for the Student's t distribution fit for the increments in delta, theta, alpha and beta frequency band signals

# References (Machine learning and EEG)


- R. Mazumder, T. Hastie, R. Tibshirani, Spectral Regularization Algorithms for Learning Large Incomplete Matrices, Journal of Machine Learning Research 11, pp 2287-2322, 2010
- R. Idro et al., Cerebral Malaria; Mechanisms of Brain Injury And Strategies For Improved Neuro-Cognitive Outcome, Pediatr. Res. 2010
- N. K. Smolentsev, Main wavelet theories. Wavelets in Matlab, (in Russian), DMK Press, 4th edition, 2014, ISBN 5-978-5-94074-955-4
- S. Z. M. Tumari, R. Sudirman, A. H. Ahmad, Selection of a Suitable Wavelet for Cognitive Memory Using Electroencephalograph Signal, Scientific Research, Engineering (5), pp 15-19, 2013

# References (Stochastic modelling and diffusion)


- F. Avram, N. N. Leonenko, N. Suvak, On spectral analysis of heavy-tailed Kolmogorov-Pearson diffusions, Markov Processes Relat. Fields, 19, pp 249-298, 2013
- C. C. Heyde, N. N. Leonenko, Student processes, Adv. Appl. Prob. 37, pp 342-365, 2005
- N. N. Leonenko, Statistical inference for student diffusion processes, Stochastic analysis and applications, Taylor and Francis, 2010
- N. N. Leonenko, A. Sikorskii, The Student subordinator model with dependence for risky asset returns, Communications in Statistics - Theory and Methods, 40, pp 3509-3523, 2011

# Stationarity - rules of thumb

- D. C. Montgomery, Design of Experiments, 8th edition, Wiley, 2012
- http://data.library.virginia.edu/a-rule-of-thumb-for-unequal-variances/

# Thanks

Thank you for attention!