Generalized Gaussian model for EEG data

25th Young Statisticians Meeting

Željka Salinger[1]

School of Mathematics, Cardiff University

YSM 2021

[1] joint work with N.N. Leonenko (Cardiff University, UK), N. Šuvak (J.J. Strossmayer University of Osijek, Croatia), A. Sikorskii and M.J. Boivin (Michigan State University, USA)

Electroencephalogram (EEG)

- Electroencephalogram (EEG) registers electrical neural activity of the brain
- Signals are captured by multiple electrodes called *channels* located over the scalp

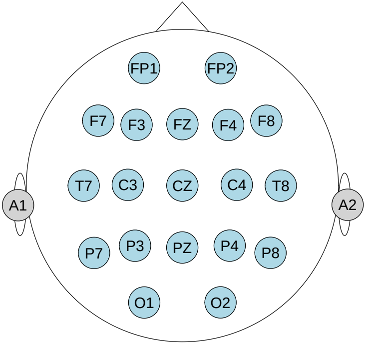


Figure: International 10-20 system

Electroencephalogram (EEG)

- EEG signals observed as realisations of a stochastic process
- Signals nonlinear and nonstationary

Figure: Example of an electroencephalogram[2]

[2]image source: Wikipedia, distributed under a CC-BY 4.0 license

Dataset used in the analysis

- Data were collected during the observational study of severe malaria in Uganda between 2008 and 2015
- **EEG data** was recorded using 19 channels with an average record duration of 30 minutes, obtaining EEG signal for 78 children
- **Non-EEG data** included
  - **neurodevelopmental score** - single measure of neurodevelopment and cognition regardless of age (z-scores) taken at 3 time points
  - **demographic and anthropometric characteristics** - age, sex, height-for-age and weight-for-age-z-score, socioeconomic status, home environment quality...
  - **biomarkers** - panels from plasma and cerebrospinal fluid
- The analysis builds upon previous work by Veretennikova et al.[3]

[3] Veretennikova, Sikorskii, and Boivin, "Parameters of stochastic models for electroencephalogram data as biomarkers for child's neurodevelopment after cerebral malaria".

Main goal

...in short

Model the EEG increments "in some way" and use the obtained information (in addition to non-EEG data) to predict neurodevelopmental and cognitive development of children who were in a coma from cerebral malaria.

Diffusion process

- Model for EEG signal with a stochastic component described using a stochastic differential equation (SDE)

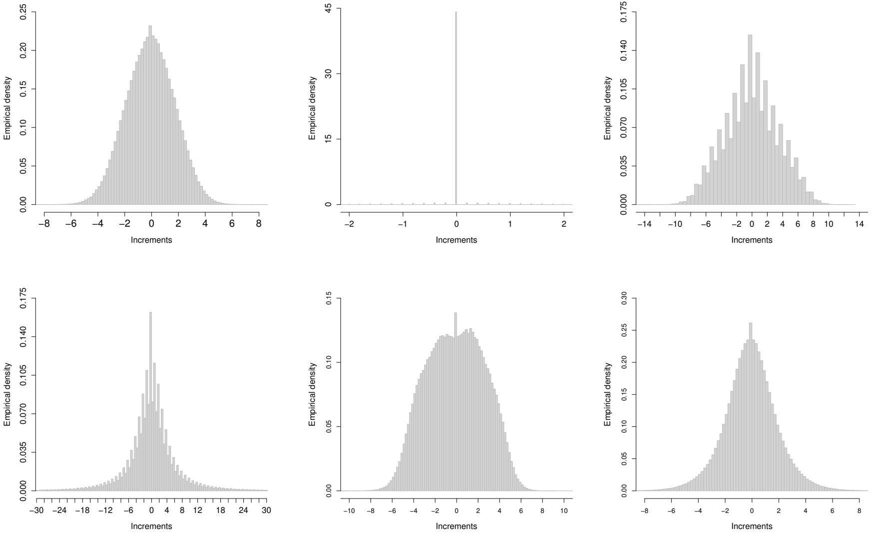
$$dX_t = -\theta X_t \, dt + v(X_t) \, dB_t, \quad \theta > 0, \quad t \geq 0, \tag{1}$$

driven by the standard Brownian motion $(B_t, t \geq 0)$

- Bibby et al.[4] describe the construction of a diffusion process with a stationary probability density function (PDF)
- If the stationary PDF is continuous, bounded, and strictly positive on the whole $\mathbb{R}$, SDE (1) admits the unique weak ergodic solution and defines the diffusion with chosen stationary distribution

[4] Bibby, Skovgaard, and Sørensen, "Diffusion-type models with given marginal distribution and autocorrelation function".

Examples of histograms of EEG increments



Generalized Gaussian distribution (GGD)

- The choice for the stationary distribution - **Generalized Gaussian distribution (GGD)** using the parametrization from Lutwak et al.[5] for $\mu = 0$

$$f_{s,b}(x) = \begin{cases} \dfrac{1}{2(s\sigma^2)^{1/s}\Gamma\left(1+\frac{1}{s}\right)} e^{-\frac{|x|^s}{s\sigma^2}} & , \quad b = 0 \\ \dfrac{bs}{2\sigma^2}\left(\dfrac{s\sigma^2}{b}\right)^{-1/s} \dfrac{\Gamma\left(1+\frac{1}{s}+\frac{\sigma^2}{b}\right)}{\Gamma\left(\frac{1}{s}\right)\Gamma\left(\frac{\sigma^2}{b}\right)} \left(1+\dfrac{b}{s\sigma^2}|x|^s\right)^{-\frac{\sigma^2}{b}-\frac{1}{s}-1} & , \quad b > 0, \end{cases} \tag{2}$$

- Normal distribution with mean 0 and variance σ^2 for $b = 0$ and $s = 2$
- Student-type distribution for $b > 0$ and $s = 2$

[5] Lutwak, Yang, and Zhang, "Moment-entropy inequalities".

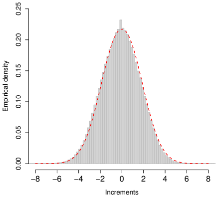
Fitting of light-tailed GGD to EEG increments

- In the light-tailed case ($b = 0$), the two-dimensional parameter $\zeta = (s, \sigma^2)$ of the stationary distribution GGD (2) is estimated by the quasi-likelihood method
- For the purpose of estimation of parameter ζ we disregard the existing exponentially decaying autocorrelation structure of the diffusion and define the quasi log-likelihood function as

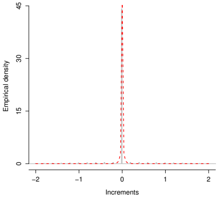
$$l_n(\zeta) = \sum_i^n \ln \left(\frac{1}{2(s\sigma^2)^{1/s}\Gamma\left(1+\frac{1}{s}\right)} e^{-\frac{|X_i|^s}{s\sigma^2}} \right). \qquad (3)$$

- The estimate $\hat{\zeta} = (\hat{s}, \hat{\sigma^2})$ of the parameter $\zeta = (s, \sigma^2)$ is then obtained by maximising (3), which can be performed using existing non-linear optimization methods.

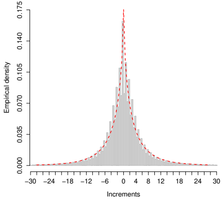
Examples of obtained fit



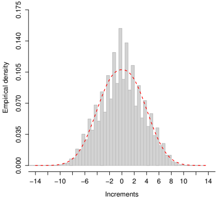
(a) $\text{GGD}(2.09, 3.52)$



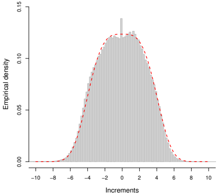
(b) $\text{GGD}(0.02, 0.12)$



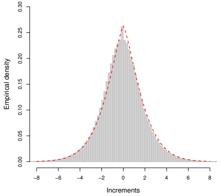
(c) $\text{GGD}(0.68, 2.50)$



(d) $\text{GGD}(2.32, 20.10)$



(e) $\text{GGD}(2.49, 15.72)$



(f) $\text{GGD}(1.35, 1.98)$

Generalized Gaussian distribution (GGD)

- The choice for the stationary distribution - **Generalized Gaussian distribution (GGD)** using the parametrization from Lutwak et al. for $\mu = 0$

$$f_{s,b}(x) = \begin{cases} \frac{1}{2(s\sigma^2)^{1/s}\Gamma\left(1+\frac{1}{s}\right)} e^{-\frac{|x|^s}{s\sigma^2}} & , \quad b = 0 \\ \frac{bs}{2\sigma^2}\left(\frac{s\sigma^2}{b}\right)^{-1/s} \frac{\Gamma\left(1+\frac{1}{s}+\frac{\sigma^2}{b}\right)}{\Gamma\left(\frac{1}{s}\right)\Gamma\left(\frac{\sigma^2}{b}\right)} \left(1+\frac{b}{s\sigma^2}|x|^s\right)^{-\frac{\sigma^2}{b}-\frac{1}{s}-1} & , \quad b > 0, \end{cases} \tag{4}$$

- Normal distribution with mean 0 and variance σ^2 for $b = 0$ and $s = 2$
- Student-type distribution for $b > 0$ and $s = 2$

Empirical scaling function

- The shape of the scaling function is strongly influenced by the tail index
- Tail index α was estimated based on empirical scaling function introduced by Grahovac et al.[6]

$$\hat{\tau}_{N,n}(q) = \frac{\sum_{i=1}^{N} s_i \frac{\ln S_q(n,n^{s_i})}{\ln n} - \frac{1}{N} \sum_{i=1}^{N} s_i \sum_{j=1}^{N} \frac{\ln S_q(n,n^{s_j})}{\ln n}}{\sum_{i=1}^{N} (s_i)^2 - \frac{1}{N} \left(\sum_{i=1}^{N} s_i \right)^2}$$

where S_q is the partition function of the sample $X_1, X_2, \ldots, X_n$

$$S_q(n,t) = \frac{1}{\lfloor n/t \rfloor} \sum_{i=1}^{\lfloor n/t \rfloor} \left| \sum_{j=1}^{\lfloor t \rfloor} X_{(i-1)\lfloor t \rfloor + j} \right|^q ,$$

with $q > 0, \quad 1 \le t \le n$ and $s_i \in (0,1)$, $i = 1, \ldots, N$

[6] Grahovac et al., "Asymptotic properties of the partition function and applications in tail index inference of heavy-tailed data".

Asymptotic form of scaling function

- Estimation can be done by fitting the empirical scaling function to its asymptotic form

$$\tau_\alpha^\infty(q) = \begin{cases} \frac{q}{\alpha}, & \text{if } q \leq \alpha \text{ and } \alpha \leq 2, \\ 1, & \text{if } q > \alpha \text{ and } \alpha \leq 2, \\ \frac{q}{2}, & \text{if } 0 < q \leq \alpha \text{ and } \alpha > 2, \\ \frac{q}{2} + \frac{2(\alpha-q)^2(2\alpha+4q-3\alpha q)}{\alpha^3(2-q)^2}, & \text{if } q > \alpha \text{ and } \alpha > 2 \end{cases}$$

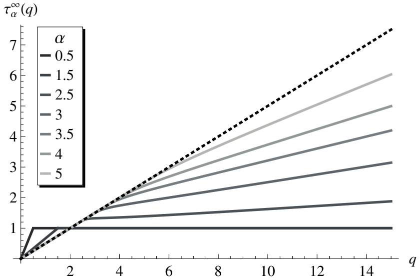
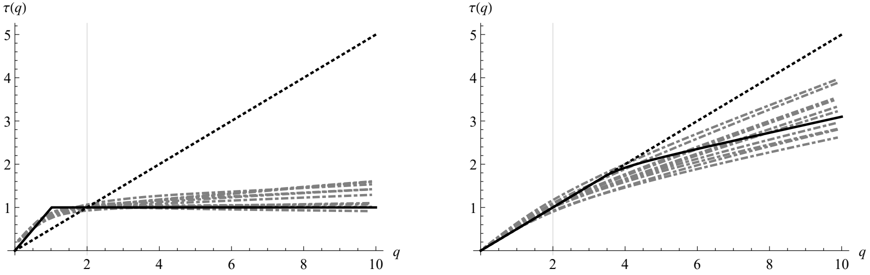


Figure: Asymptotic form of scaling function

Estimation of tail index on EEG increments

- Estimation performed on 10 random samples of size 10000 obtaining estimates of tail index $\hat{\alpha}_i$
- A single value of tail index estimate $\hat{\alpha}$ was chosen to be the median of values $\hat{\alpha}_i$, $i = 1, \ldots, 10$



(a) $\hat{\alpha} = 1.10$ (b) $\hat{\alpha} = 3.49$

Figure: Tail index estimates of EEG increments

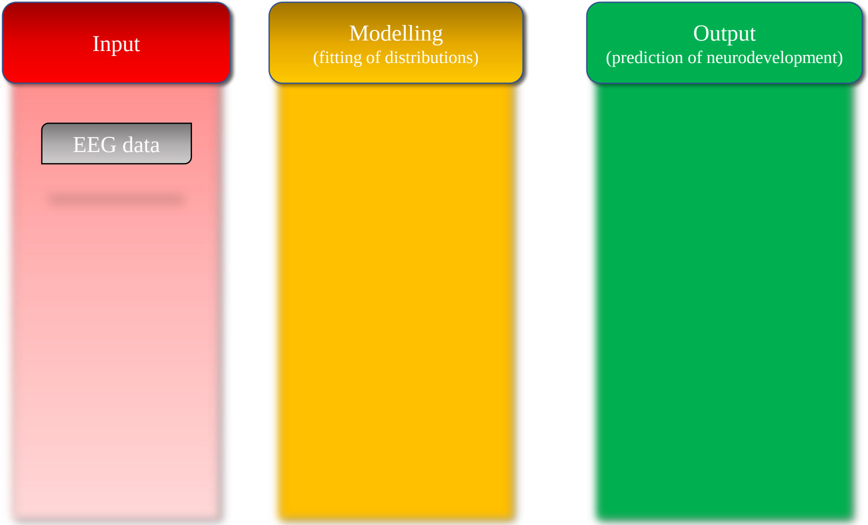
Flowchart of the process

Input

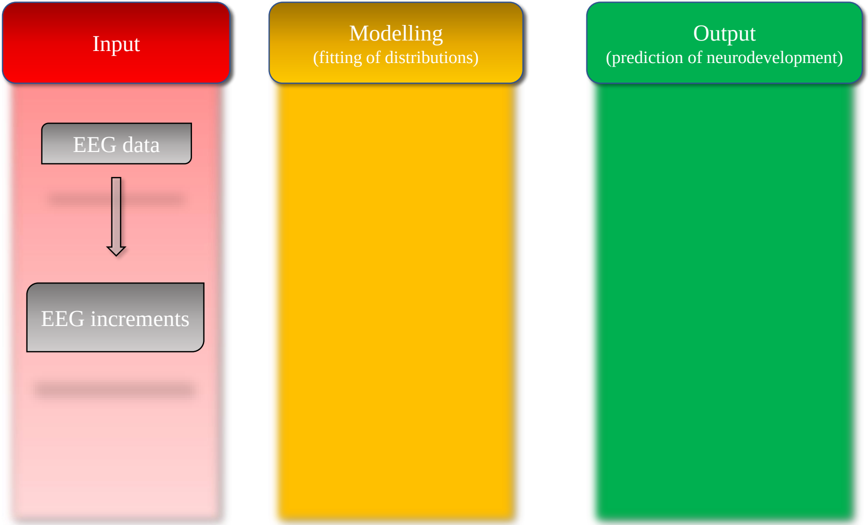
Modelling
(fitting of distributions)

Output
(prediction of neurodevelopment)

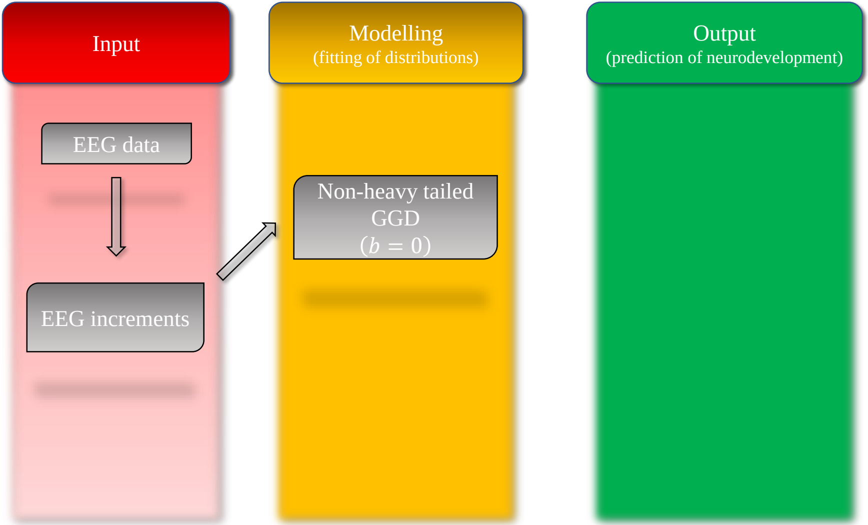
Flowchart of the process



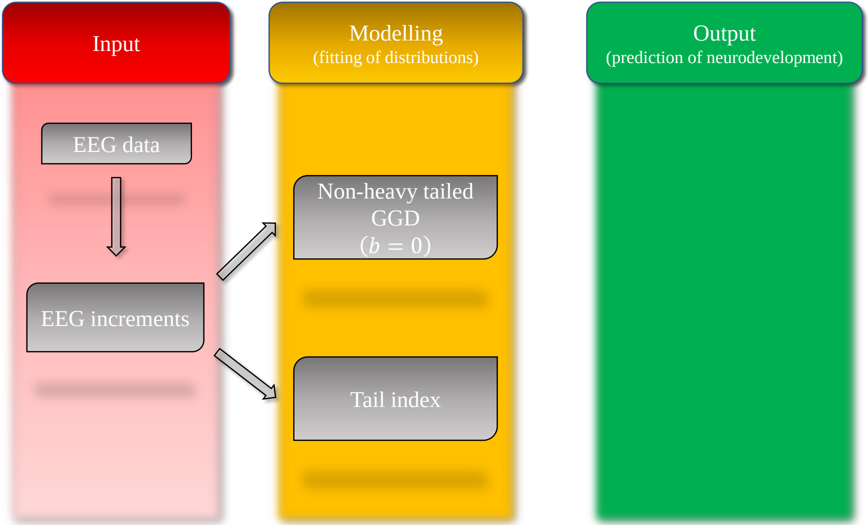
Flowchart of the process



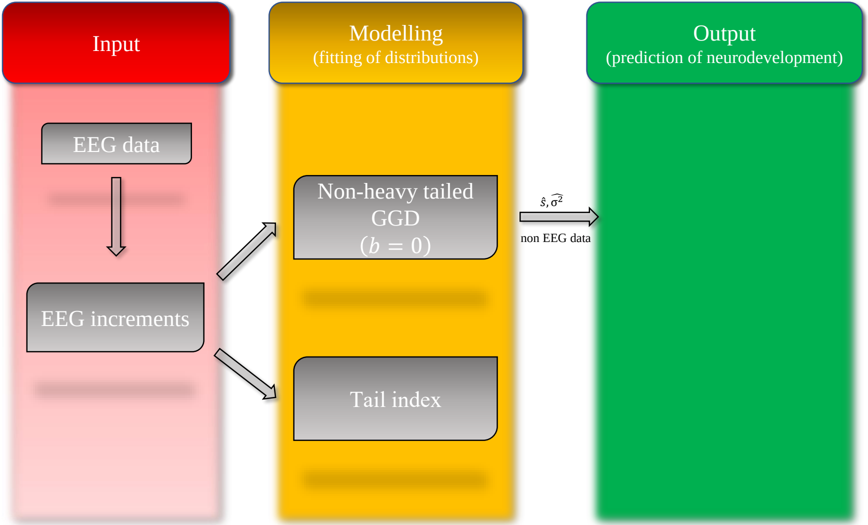
Flowchart of the process



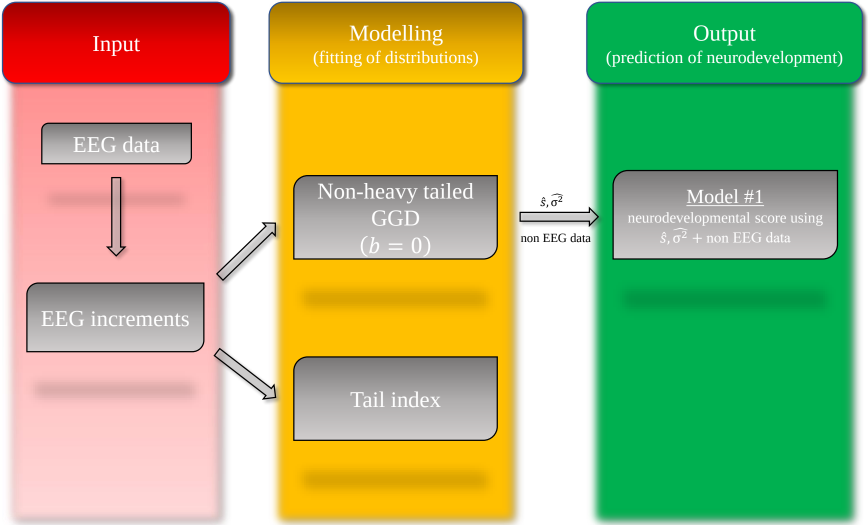
Flowchart of the process



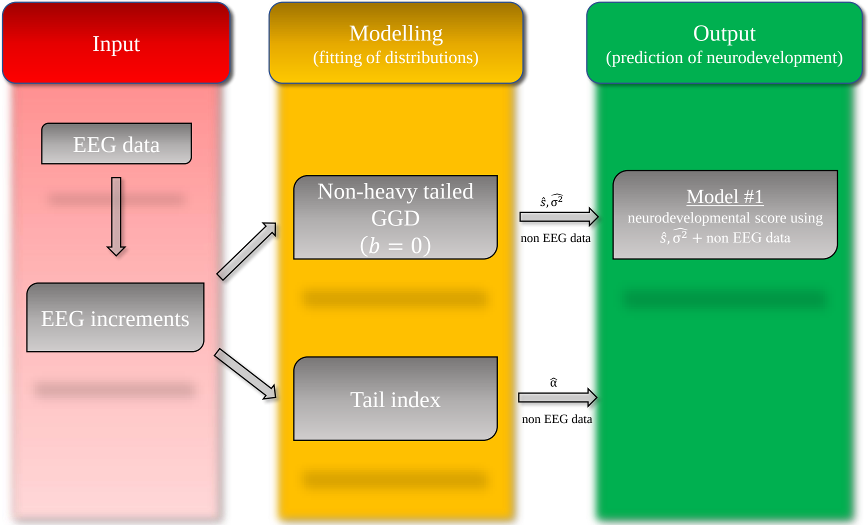
Flowchart of the process



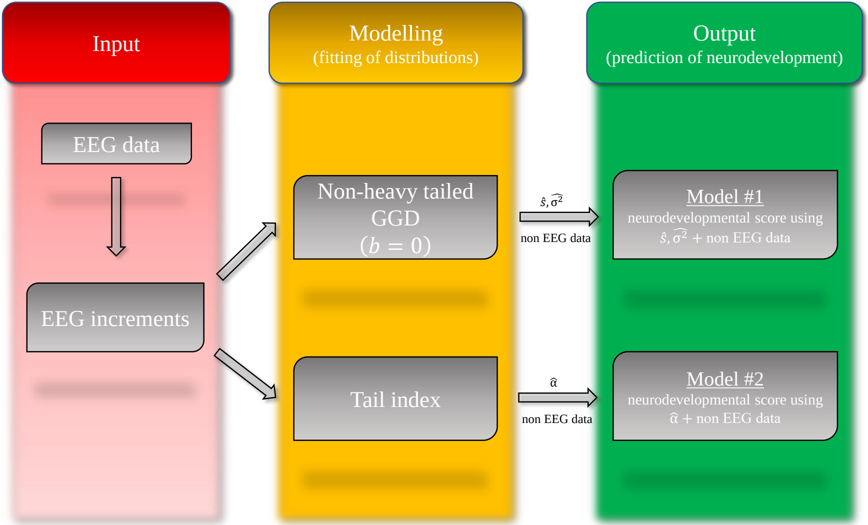
Flowchart of the process



Flowchart of the process



Flowchart of the process



Elastic net regression

- Elastic net regression was used to identify important predictors of neurodevelopment and cognition
- Elastic net regression controls for correlations among predictors and deals with the case where the number of predictors is much bigger than the number of observations
- Elastic net regression can be viewed as a penalized least squares method which minimizes the loss function[7] defined by

$$L(\alpha, \lambda, \boldsymbol{\beta}) = |\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}|^2 + \lambda \left(\frac{1-\alpha}{2} |\boldsymbol{\beta}|^2 + \alpha |\boldsymbol{\beta}|_1 \right),$$

- Hyperparameter α can be seen as a mixing parameter between ridge $(\alpha = 0)$ and LASSO $(\alpha = 1)$ regression

[7] Zou and Hastie, "Regularization and variable selection via the elastic net".

Models used

- Response variable was the standardized neurocognitive score taken 6 months after the discharge from the hospital
- Models investigated based on feature matrix:
  - **non-EEG features model** included just the non-EEG features (baseline neurodevelopmental score, demographic and anthropometric characteristics, biomarkers)
  - **combined non-EEG and GGD features model** - non-EEG features and estimates of s and σ^2 obtained from fitting light-tailed GGD
  - **combined non-EEG and tail index features model** - non-EEG features and median values of estimates $\hat{\alpha}$ of tail index (as continuous and categorical variable)

Comparison of models

Table: Model comparison based on elastic net regression results

Model features included (number of features)	RMSE	Number of non-zero coefficients	Number of non-zero coefficients from EEG features subset
Non-EEG features (54)	0.5670	12	N/A
Non-EEG (54) and GGD (38) features	0.5655	13	1
Non-EEG (54) and continuous tail index features (19)	0.5670	12	0
Non-EEG (54) and categorical tail index features (38 dummy variables)	**0.5499**	**10**	1

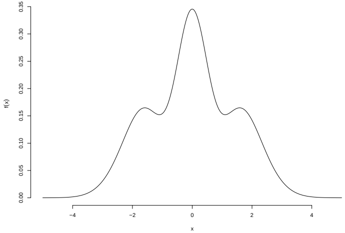
Conclusion

Conclusion

Addition of information obtained from EEG data can improve the prediction of neurodevelopment and cognition in children who recovered from a coma.

Future research

- Investigate possible multimodal distributions, e.g 3-peak distribution from Cammarota et al.[8]



- Machine learning approach, e.g. convolutional neural networks

[8] Cammarota, Marinucci, and Wigman, "On the distribution of the critical values of random spherical harmonics".

References

- Bo Martin Bibby, Ib Michael Skovgaard, and Michael Sørensen. "Diffusion-type models with given marginal distribution and autocorrelation function". In: *Bernoulli* 11.2 (2005), pp. 191–220.
- J.P.N. Bishwal. *Parameter Estimation in Stochastic Differential Equations*. Springer-Verlag, Berlin, Heidelberg, 2007.
- Valentina Cammarota, Domenico Marinucci, and Igor Wigman. "On the distribution of the critical values of random spherical harmonics". In: *J. Geom. Anal.* 26.4 (2016), pp. 3252–3324. arXiv: 1409.1364.
- Alex Dytso, Ronit Bustin, and Harold Vincent Poor. "Analytical properties of generalized Gaussian distributions". In: *J. Stat. Distrib. Appl.* 5.1 (2018), pp. 2195–5832.
- Danijel Grahovac et al. "Asymptotic properties of the partition function and applications in tail index inference of heavy-tailed data". In: *Statistics (Ber).* 49.6 (2015), pp. 1221–1242.
- Bronius Grigelionis. *Student's t-Distribution and Related Stochastic Processes*. Springer-Verlag, Berlin, Heidelberg, 2013.
- Nikolai N. Leonenko et al. "Generalized Gaussian time series model for increments of EEG data". In: *submitted* (2021).
- Erwin Lutwak, Deane Yang, and Gaoyong Zhang. "Moment-entropy inequalities". In: *Ann. Probab.* 32.1B (2004), pp. 757–774.
- Saralees Nadarajah. "A generalized normal distribution". In: *J. Appl. Stat.* 32.7 (2005), pp. 685–694.
- Maria A. Veretennikova, Alla Sikorskii, and Michael J. Boivin. "Parameters of stochastic models for electroencephalogram data as biomarkers for child's neurodevelopment after cerebral malaria". In: *J. Stat. Distrib. Appl.* 5.1 (2018).
- Hui Zou and Trevor Hastie. "Regularization and variable selection via the elastic net". In: *J. R. Stat. Soc. Ser. B (statistical Methodol.* 67.2 (2005), pp. 301–320.