Stochastic models for increments of EEG recordings using heavy-tailed and multimodal diffusions

Željka Salinger

School of Mathematics

Cardiff University

A thesis submitted for the degree of

Doctor of Philosophy

December 2023

To my beloved parents, Margita and Zdravko ...

Acknowledgements

"Nothing truly valuable can be achieved except by the unselfish cooperation of many individuals."

— Albert Einstein

Doing a PhD is no small feat, and I couldn't have done it alone. Therefore, I would like to express my sincere gratitude to the people who played significant roles in my PhD journey.

First and foremost, I would like to express my deepest gratitude to my supervisor, Professor Nikolai Leonenko for his generously provided knowledge and expertise, and intellectual freedom. I highly valued the online meetings we held, which not only served as crucial checkpoints to keep me on track academically, but also provided me with plenty of encouragement. I am profoundly grateful for his immeasurable contributions to my development and his compassionate guidance.

Words cannot express my gratitude to Professor Nenad Šuvak for his encouragement and immense contribution to my research, even before the start of my PhD. Maintaining a personal connection in academic exchanges was truly appreciated. Also, special thanks to Professor Alla Sikorskii for providing the data and much needed explanations. Her experience working with data has helped to solve many stalemate positions we encountered.

I would like to thank my parents, Margita and Zdravko. Your love and belief in me has kept me going through this endeavour. I am also thankful to my partner, Lorenzo for his understanding and support, both emotional and practical. Many thanks to my friends from all around the world who were always a phone-call away when I needed them. I also thank my cat Sylvester for all the entertainment he provided, which I truly needed during these difficult times.

Lastly, my gratitude goes to the EPSRC who funded this research.

Dissemenation of work

Publications

- 2023: Leonenko, N. N., Salinger, Ž., Sikorskii, A., Šuvak, N., & Boivin, M.J., "Multimodal diffusion model for increments of electroencephalogram data", *Stochastic Environmental Research and Risk Assessment*, **37**(12), 4695 − 4706.
- 2023: Leonenko, N. N., Salinger, Ž., Sikorskii, A., Šuvak, N., & Boivin, M.J., "Generalized Gaussian time series model for increments of EEG data", *Statistics and Its Interface*, **16**(1), 17− 29.

Presentations

- 2022: Doctoral probability seminar, *Department of Mathematics, University of Zagreb*, Croatia
- 2022: Seminar of Croatian society for theoretical and mathematical biology, *Ruđer Bošković Institute*, Zagreb, Croatia
- 2021: Statistical seminar, *Department of Mathematics, J. J. Strossmayer University*, Osijek, Croatia
- 2021: 25^{th} Young Statisticians Meeting (YSM), Vorau, Austria
- 2021: 22^{nd} European Young Statisticians Meeting (EYSM), Athens, Greece

Abstract

Electroencephalography (EEG) is a useful diagnostic tool for many brain disorders, including coma. EEG signals are non-stationary, but it is possible to model EEG signal increments using stationary processes. In this thesis, EEG increments are viewed as discrete time observations from a diffusion process with marginal distributions which is independent of time. First, the basic theory needed for modelling of the diffusion processes is presented. Then, based on the histograms of EEG increments, the choice of a marginal distribution is the generalized Gaussian distribution (GGD) with a parametrization that comprises both light-tailed and heavy-tailed distributions. Some properties of the GGD are presented, along with the method of estimation of the tail index using the so-called empirical scaling function. The estimated parameters from models across EEG channels obtained from both subfamilies are explored as potential predictors of neurocognitive outcomes in children 6 months after recovering from cerebral malaria. To include a wider range of marginal distributions observed in histograms, a new strictly stationary strong mixing diffusion model with marginal multimodal (three-peak) distribution and exponentially decaying autocorrelation function is used for modelling of EEG increments. The marginal distribution is viewed as a mixture of three non-central generalized Gaussian distributions. Distribution parameters are estimated using the expectation-maximization (EM) algorithm, where the added shape parameter is estimated using the higher order statistics approach based on an analytical relationship between the shape parameter and the kurtosis. Similarly to the unimodal case, obtained estimates are then used for prediction of subsequent neurodevelopment and cognition of cerebral malaria survivors using the elastic net regression. All predictive models are compared to determine whether additional information obtained from multimodality of the marginal distributions can be used to improve the prediction. The results of analysis in this thesis show that stochastic modelling of EEG features can improve the explanation of variation in neurodevelopmental outcomes of children who were in coma due to cerebral malaria.

Contents

List of abbreviations

a.s. almost surely
AIC Akaike information criterion
BIC Bayes information criterion
CDF cummulative distribution function
ECM Expectation/Conditional Maximization
EEG electroencephalogram
EM Expectation/Maximization
GGD Generalized Gaussian distribution
GGDiff Generalized Gaussian diffusion
IID independent identically distributed
MixGGD Mixed Generalized Gaussian distribution
MixGGDiff Mixed Generalized Gaussian diffusion
PDF probability density function
RMSE Root mean squred error
RV random variable
SDE stochastic differential equation
w.r.t. with respect to

List of figures

List of tables

Chapter I

Introduction

1 Cerebral malaria and electroencephalogram (EEG)

Cerebral malaria is the most severe neurological complication of infection with a parasite *Plasmodium falciparum* where children in sub-Saharan Africa are the most affected with over half a million cases annually. Mortality rate is high and even with treatment, $15 - 20\%$ of children die while surviving patients sustain brain injury which manifests as long-term neurocognitive impairments [Idro et al., 2011]. Severe malaria can be caused by other species of *Plasmodium*, such as *P. vivax* and *P. knowlesi*. In addition to the presence of *P. falciparum* on a blood smear, main diagnostic difference between other complications of severe malaria and cerebral malaria is the onset of a coma. Wilson and Nordal [2013] define coma as "a pathological state of suspended consciousness and unresponsiveness to external or internal stimuli" which can last from an hour to several weeks and result in "regaining of consciousness with or without sequelae, death or [...] chronic vegetative state".

There are a number of novel neuroscience techniques for studying brain function such as functional magnetic resonance imaging (fMRI), positron emission tomography (PET), magnetoencephalography (MEG), transcranial magnetic stimulation (TMS), (for comparison of these techniques see [Hecht and Stout, 2015]) but electroencephalography is still widely used as a non-invasive way to monitor patients' brain functions. Electroencephalogram (EEG) registers electrical neural activity of the brain. EEG signal is used as a diagnostic tool for many brain disorders based on the visual inspection by clinical experts in the field [Sanei, 2013]. Signals are captured by multiple electrodes called channels located over the scalp and are usually presented in the time domain. Human brain is neither a fully deterministic nor a fully stochastic system [Klonowski, 2009], but EEG signals

are often considered as realisations of a stochastic process. Body's biological and physiological changes cause the changes in brain metabolism which result in EEG signals being non-linear. Also, the signals are generally non-stationary which means that their statistics such as mean, variance and higher order statistics may change over time. However, EEG signals can be considered quasi-stationary, i.e. stationary only within short time intervals because their statistical properties are steady within one short time segment. The duration and these quasi-stationary segments in EEG vary. For example, in their research Kaplan et al. [2005] report that the majority of these quasi-stationary segments had a duration of less than 1 second. Usually, under normal brain conditions and within such short intervals, the multichannel EEG process is described by multivariate Gaussian distribution. Statistical averages of the signals can be estimated by time averages since the signals are often considered ergodic. Although an accurate model for EEG signals, i.e. a model that can link the chemical processes within corresponding neurons generating the active potentials is hard to achieve, a number of such models has been introduced since 1950s. An overview of these models is given in Sanei [2013].

Even though EEG is underused in coma, combined with neurological examination, it can be useful in determining the prognosis in some conditions when the etiology of the coma is known [Young, 2000]. As mentioned, the sequelae of a coma can be severe and their extent can impact subsequent neurodevelopment and cognitive function. Thus, the evaluation of aforementioned functions while the subject is still in an unconscious state is of great importance. For this purpose EEG can be especially useful as it provides a language-independent clinical data for such evaluation.

Identification of factors that can predict the extent of neurocogntive impairment and other outcomes following cerebral malaria illness is an important problem [Birbeck et al., 2010; Patel et al., 2020]. While evidence-based rehabilitative interventions for survivors are available, resources to administer them are limited across sub-Saharan Africa. Therefore directing these interventions to those in most need, as determined by predictors of subsequent impairment, is key to the most efficient use of the available resources.

Statistical analyses of EEG data usually include classification and prediction of seizures using arrays of EEG features (Duncan et al. [2013]; Kirch et al. [2015]; Temko et al. [2011]), but few models for the underlying stochastic processes have been proposed [Piryatinska et al., 2009; Veretennikova et al., 2018]. However, Veretennikova et al. [2018] performed an analysis of EEG signals from children who were in coma due to cerebral malaria and noticed that the underlying stochastic processes across channels could be described by the stochastic process with stationary increments. Additionally, stochastic features from modelling of EEG

increments were used as predictors of neurodevelopment and cognition. For a review of other mathematical models for malaria epidemic see Mandal et al. [2011], while more recent results can be found in Wanduku [2019].

2 Thesis aim and structure

The aim of this thesis is to investigate stochastic models of EEG increments for the prediction of neurodevelopment and cognition of children who were in coma due to cerebral malaria.

The first step in achieving this aim is constructing the corresponding stochastic processes. EEG increments will be viewed as a time series $(X_n,\, n \in \mathbb{N})$, representing the model for discrete-time observations from a stationary diffusion process $\{X_t\}_{t\geq 0}$ with a specified probability density function. Once these diffusion processes are constructed, the methods for estimating their parameters need to be considered, and then the usefulness of the parameter estimates can be tested, answering whether the overall prediction of subsequent neurodevelopment or cognition can be improved with the addition of stochastic features.

The thesis is structured as follows:

Chapter II forms the theoretical part of the thesis. In Section II.1 the basic theory relevant for the stochastic differential equations and diffusion construction is presented. Section II.2 presents the unimodal marginal distribution used in modelling of EEG increments and includes the description of the generalized Gaussian distribution, its properties and the diffusion construction, with the method of estimation of its parameters, both for the light-tailed and the heavy-tailed subfamily. Section II.3 deals with the multimodal, marginal distribution which is a mixture of non-central generalized Gaussian distributions, for which again diffusion process was constructed and the algorithm for parameter estimation is presented.

Chapter III presents the empirical results of modelling the EEG increments using the stochastic models presented in Chapter II. Dataset description and analysis of the histograms of EEG increments are shown in Section III.1, a method used for the prediction of neurodevelopment is given in Section III.2, while Sections III.3 and III.4 include the results of fitting marginal distributions of constructed diffusion to EEG increments and introduce different models of combined non-EEG and stochastic features for the prediction of neurodevelopment.

Chapter IV discusses the results presented in the thesis and concludes with proposing future applications of presented diffusion models.

Chapter II

Marginal distributions and diffusion construction for modelling of EEG increments

The idea of modelling the EEG data is based on finding a suitable continuos-time model which would describe the underlying process that created the data. Hence, a continuos-time model would be used to describe the discrete-time observations which were obtained from EEG signals. The description of the system governing this data (i.e. brain function) is simplified by using a stochastic differential equation, and the realizations are then transformed into a stationary process, resulting in EEG increments. In order to characterize important probabilistic properties of EEG increments, including their dependence structure, the EEG increments are viewed as a time series $(X_n,\, n \in \mathbb{N})$, representing the model for discrete-time observations from the diffusion process $\{X_t\}_{t\geq 0}$ with a given marginal density f which would reflect the empirical data. First, in Section II.1 some basic definitions from probability and the theory of stochastic differential equations are presented. Next, a model based on the unimodal shape of the marginal distribution is given in Section II.2. In Section II.3 a more complex model covering a wider range of possible histogram shapes of EEG increments is given.

1 Basic definitions, theorems and properties

In the first part of this section, some basic definitions from probability theory are reviewed, without going into details. These definitions are used later on and thus needed for the definitions and theory of stochastic differential equations and its solutions. The second part of the section contains definitions related to stochastic differential equations and its solutions (diffusions). The definitions are mostly taken verbatim from Øksendal [2000], if not otherwise stated. The last

part of the section contains the main result from Bibby et al. [2005] which shows how to construct a diffusion process with a given marginal density.

1.1 Basic definitions

Definition 1.1 (σ-algebra).
If Ω is a given set, then σ-algebra $\mathcal{F}$ on Ω is a family $\mathcal{F}$ of subsets of Ω with the following properties:

(i) $\emptyset \in \mathcal{F}$,

(ii) if $F \in \mathcal{F}$ *then* $F^C \in \mathcal{F}$, *where* $F^C = \Omega \backslash F$ *is the complement of* F *in* Ω,

(iii) if $A_1, A_2, \cdots \in \mathcal{F}$, *then* $A \coloneqq \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

Definition 1.2 (σ-algebra generated by $\mathcal{U}$).
Given any family $\mathcal{U}$ of subsets Ω there is a smallest σ-algebra $\mathcal{H}_{\mathcal{U}}$ containing $\mathcal{U}$, namely

$$\mathcal{H}_{\mathcal{U}} = \bigcap \{\mathcal{H} \colon \mathcal{H} \text{ } \sigma\text{-algebra of } \Omega, \mathcal{U} \subset \mathcal{H}\}.$$

We call $\mathcal{H}_{\mathcal{U}}$ the σ-algebra generated by $\mathcal{U}$.

Definition 1.3 (Borel σ-algebra, Borel sets and Borel measure).
If $\mathcal{U}$ is the collection of all open subsets of a topological space Ω, then $\mathcal{B} = \mathcal{H}_{\mathcal{U}}$ is called Borel σ-algebra on Ω. The elements $B \in \mathcal{B}$ are called Borel sets. Any measure defined on Borel σ-algebra is called a Borel measure.

Definition 1.4 (Probability measure and probability space).
The pair $(\Omega, \mathcal{F})$ is called a measurable space. A probability measure $\mathbb{P}$ on a measurable space $(\Omega, \mathcal{F})$ is a function $\mathbb{P} \colon \mathcal{F} \to [0, 1]$ such that

(i) $\mathbb{P}(\emptyset) = 0$, $\mathbb{P}(\Omega) = 1$,

(ii) if $A_1, A_2, \cdots \in \mathcal{F}$ *and* $\{A_i\}_{i=1}^{\infty}$ *is disjoint, then*

$$\mathbb{P}\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a probability space.

Definition 1.5 ($\mathcal{F}$ measurable sets and $\mathcal{F}$ measurable function).
The subsets F of Ω which belong to $\mathcal{F}$ are called $\mathcal{F}$-measurable sets. If $(\Omega, \mathcal{F}, \mathbb{P})$ is a given probability space, then a function $Y \colon \Omega \to \mathbb{R}^n$ is called $\mathcal{F}$-measurable if

$$Y^{-1}(U) \coloneqq \{\omega \in \Omega, Y(\omega) \in U\} \in \mathcal{F}$$

for all Borel sets $U \in \mathbb{R}^n$.

Definition 1.6 (Random variable and its distribution).
A random variable (RV) X is an $\mathcal{F}$-measurable function $X\colon \Omega \to \mathbb{R}^n$. A probability measure m_X on $\mathbb{R}^n$ defined by

$$m_X(B) = \mathbb{P}(X^{-1}(B)), \quad B \in \mathcal{F}$$

is called the distribution of X.

Definition 1.7 (Expectation of a random variable). *Expectation of a random variable X w.r.t. to the probability measure m_X, if it exists, i.e. if X is absolutely integrable w.r.t. m_X, is defined by*

$$\mathbb{E}[X] = \int_{\mathbb{R}^n} x \, \mathrm{d}m_X(x). \tag{1.1}$$

Throughout this thesis the focus will be one-dimensional random variables having the density function corresponding to the measure m_X.

Definition 1.8 (Characteristic function and moment generating function). *For a random variable X with probability density function f, the characteristic function is defined by*

$$\phi(t) = \mathbb{E}\left[\mathrm{e}^{itX}\right] = \int_{-\infty}^{\infty} \mathrm{e}^{itx} f(x) \, \mathrm{d}x, \quad i = \sqrt{-1}. \tag{1.2}$$

A moment generating function is defined by

$$\phi(t) = \mathbb{E}\left[\mathrm{e}^{tX}\right] = \int_{-\infty}^{\infty} \mathrm{e}^{tx} f(x) \, \mathrm{d}x. \tag{1.3}$$

Since the relation between the distribution function and the characteristic function is one-to-one, knowing the characteristic function is synonymous to knowing the distribution function [Karlin and Taylor, 1975].

Definition 1.9 (Stochastic process and its state space).
A (stochastic) process $X = \{X_t\}_{t\in T}$ is a parametrized collection of random variables $\{X_t\}_{t\in T}$ defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and assuming values in $\mathbb{R}^n$. The range (possible values) of the random variables in a stochastic process is called the state space of the process.

The stochastic process may be regarded as a function of two variables $X(t,\omega)\colon T \times \Omega \to \mathbb{R}^n$ but for convenience, the notations $X_t(\omega)$ (or even X_t, omitting the variable ω completely), are used.

Definition 1.10 (Path of a stochastic process).
A path of a stochastic process $X = \{X_t\}_{t\in T}$ *is a function*

$$t \mapsto X_t(\omega), \quad t \in T$$

for every fixed $\omega \in \Omega$.

Definitions from Karlin and Taylor [1975] present important concepts of (strictly) stationary and wide-sense stationary processes, which are the main focus of the thesis.

Definition 1.11 ((Strictly) stationary process).
A stationary process is a stochastic process $X = \{X_t\}_{t\in T}$ *with the property that for any positive integer* k *and any* $t_1, \ldots, t_k \in T$ *and* $h \in T$*, the joint distribution of*

$$X_{t_1}, \ldots, X_{t_k}$$

is the same as the joint distribution of $X_{t_1+h}, \ldots, X_{t_k+h}$.

Usually, the term "stationary process" refers to the strictly stationary processes.

Definition 1.12 (Wide-sense (weakly or 2nd order) stationary process).

A wide-sense stationary process is a stochastic process $X = \{X_t\}_{t\in T}$ *having the following properties:*

(i) $\mathbb{E}[X_t]$ *exists and is constant for all* $t \in T$,
(ii) $\mathbb{E}[X_t^2 < \infty]$ *for all* $t \in T$ *and*
(iii) $\mathrm{Cov}(X_t, X_s) = \mathrm{Cov}(X_{t+h}, X_{s+h}), \quad$ *for all* $t, s \in T$ *and* $h > 0$
where Cov *is the covariance function defined as* $\mathbb{E}\Big[\Big(X_t - \mathbb{E}[X_t]\Big)\Big(X_s - \mathbb{E}[X_s]\Big)\Big]$.

When a time-average of a wide-sense stationary process is the same as its average over the probability space, the process is called (mean) ergodic. A formal definition from Papoulis and Pillai [2002] is given below.

Definition 1.13 (Ergodic process). *Let* $X = \{X_t\}_{t\in T}$ *be a wide-sense stationary process and* $\mathbb{E}[X_t] = \mu$. *A time average of a process on an arbitrary interval* $[-L, L]$ *is defined by*

$$\mu_L = \frac{1}{2L} \int_{-L}^{L} X(t)\,\mathrm{d}t.$$

We say that a process X *is mean-ergodic if*

$$\lim_{L\to\infty} \mu_L = \mu.$$

Definition 1.14 (Finite-dimensional distributions of a process).
The finite-dimensional distributions of the process $X = \{X_t\}_{t\in T}$ *are the measures* $m_{t_1,\dots,t_k}$ *defined on* $\mathbb{R}^{nk}$, $k = 1, 2, \dots$ *by*

$$m_{t_1,\dots,t_k}(F_1 \times F_2 \times \cdots \times F_k) = \mathbb{P}[X_{t_1} \in F_1, \cdots, X_{t_k} \in F_k], \quad t_i \in T$$

where $F_1, \dots, F_k$ *denote Borel sets in* $\mathbb{R}^n$.

An important example of a stochastic process which will be used throughout the thesis is called a Brownian motion.

Definition 1.15 (Brownian motion (Wiener process)).
A stochastic process $B = \{B_t\}_{t\geq 0}$ *which satisfies the properties*

(i) $B_0 = 0$ *a.s.,*

(ii) $t \mapsto B_t$ *is a.s. continuous,*

(iii) B *has independent increments, i.e.*

$$B_{t_1}, B_{t_2} - B_{t_1}, \dots, B_{t_k} - B_{t_{k-1}} \quad \text{are independent for all} \quad 0 \leq t_1 < t_2 < \cdots < t_k,$$

(iv) increments are normally distributed i.e. $B_t - B_s \sim \mathcal{N}(0, t-s)$ *for all* $0 \leq s < t$,

is called a Brownian motion or Wiener process B.

From the properties (i) and (iv) it follows

$$B_t - B_s \sim \mathcal{N}(0, t-s) \implies B_t - B_0 \sim \mathcal{N}(0, t-0) \implies B_t \sim \mathcal{N}(0, t).$$

Moreover, it could be shown that all finite dimensional marginal distributions are Gaussian. The Gaussianity of the whole process follows from (iii) and (iv).

Definition 1.16 (Filtration).
A filtration is a family $\mathcal{H} = \{\mathcal{H}_t\}_{t\geq 0}$ *of* σ*-algebras* $\mathcal{H}_t \subset \mathcal{F}$ *such that*

$$0 \leq s < t \implies \mathcal{H}_s \subset \mathcal{H}_t$$

i.e. $\mathcal{H} = \{\mathcal{H}_t\}_{t\geq 0}$ *is increasing.*

Definition 1.17 ($\mathcal{H}$-adapted process).
Let $\mathcal{H} = \{\mathcal{H}_t\}_{t\geq 0}$ *be a filtration. A process* $g(t,\omega)\colon [0,\infty\rangle \times \Omega \to \mathbb{R}^n$ *is called* $\mathcal{H}$*-adapted process if for each* $t \geq 0$ *the function* $\omega \to g(t,\omega)$ *is* $\mathcal{H}_t$*-measurable.*

Definition 1.18 (Martingale).
An n*-dimensional stochastic process* $M = \{M_t\}_{t\geq 0}$ *on* $(\Omega, \mathcal{F}, \mathbb{P})$ *is called a martingale with respect to filtration* $\mathcal{H} = \{\mathcal{H}_t\}_{t\geq 0}$ *and with respect to* $\mathbb{P}$ *if*

(i) M_t *is* $\mathcal{H}$*-adapted process,*

(ii) $\mathbb{E}[|M_t|] < \infty, \quad \forall t,$

(iii) $\mathbb{E}[M_t|\mathcal{H}_s] = M_s, \quad \forall t \geq s.$

1.2 Stochastic differential equations and its solutions

Stochastic differential equations (SDEs) are viewed as an extension of classical differential equations where some randomness is allowed in some of the coefficients [Øksendal, 2000]. A white noise term is added to the equation which introduces an uncertainty in the description of the system and results in dependence between observations [Bibby et al., 2005]. This inclusion of the uncertainty of the system can be represented by the equation of the form

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \gamma(X_t)\,\mathrm{d}V_t, \quad t \geq 0 \tag{1.4}$$

where b, γ are real continuos functions and $V = \{V_t\}_{t\in T}$ is some stochastic process. Functions b and γ are called the *infinitesimal parameters* of the process, where b is called the *drift parameter, infinitesimal mean* or *expected infinitesimal displacement* and γ is called the *diffusion parameter*, *infinitesimal variance* or *volatility* [Karlin and Taylor, 1981].

Within this thesis, the driving process V shall be *Browninan motion* $B = \{B_t\}_{t\geq 0}$ from Definition 1.15, thus (1.4) can be rewritten as

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \gamma(X_t)\,\mathrm{d}B_t, \quad t \geq 0. \tag{1.5}$$

Stochastic differential equations can also be defined using an Itô integral which is defined using a special class of functions.

Definition 1.19. *Let $S, T \subset [0,\infty\rangle \times \Omega$ and $\mathcal{V} \coloneqq \mathcal{V}(S,T)$ be the class of functions*

$$f(t,\omega)\colon [0,\infty\rangle \times \Omega \to \mathbb{R}$$

such that

(i) $(t,\omega) \to f(t,\omega)$ *is $\mathcal{B}\times\mathcal{F}$-measurable, where $\mathcal{B}$ is a Borel σ-algebra on $[0,\infty\rangle$;*

(ii) $f(t,\omega)$ *is $\mathcal{F}$-adapted;*

(iii) $\mathbb{E}\left[\int_S^T f(t,\omega)^2 dt\right] < \infty$.

The Itô integral can be defined for a larger class of integrands f than $\mathcal{V}$ by relaxing the conditions (ii) and (iii) in the following way:

(ii)' there exists a filtration $\mathcal{H} = \{\mathcal{H}_t\}_{t\geq 0}$ such that $B = \{B_t\}_{t\geq 0}$ is a martingale with respect to filtration $\mathcal{H}$ and f is $\mathcal{H}$-adapted;

(iii)' $\mathbb{P}\left[\int_S^T f(t,\omega)^2 dt < \infty\right] = 1$.

From Øksendal [2000], an important example where (ii)' holds, but (ii) doesn't is described as follows: suppose $(B_1, \ldots, B_n)$ is n-dimensional Brownian motion and $B_t(\omega) = B_k(t)$ is its k^{th} coordinate. If $\mathcal{F}_t^{(n)}$ is the σ-algebra generated by

$(B_1(s_1), \dots B_n(s_n))$ where $s_k \leq t$, then $B_k(t)$ is a martingale with respect to $\mathcal{F}_t^{(n)}$ since $B_k(s) - B_k(t)$ is independent of $\mathcal{F}_t^{(n)}$ for $s > t$. This defines integrals of the form

$$\int_0^t f(s,\omega) \dot{\mathrm{B}}_k(s)$$

for $\mathcal{F}_t^{(n)}$-adapted integrands $f(t,\omega)$.

Definition 1.20 (The Itô integral).
Let $f \in \mathcal{V}(S,T)$ where $S,T \subset [0,\infty\rangle \times \Omega$. Then the Itô integral of f from S to T is defined by

$$\int_S^T f(t,\omega)\,\mathrm{d}B_t = \lim_{n\to\infty} \int_S^T \varphi_n(t,\omega)\,\mathrm{d}B_t(\omega) \quad (\textit{limit in } L^2(\mathbb{P})), \tag{1.6}$$

where $\{\varphi_n\}$ is a sequence of simple functions[†] such that

$$\mathbb{E}\left[\int_S^T (f(t,\omega) - \varphi_n(t,\omega))^2\,\mathrm{d}t\right] \to 0 \quad \textit{as } n \to \infty,$$

where $L^2(\mathbb{P})$ refers to the convergence in mean square.

Definition 1.21 (One-dimensional Itô processes).
Let $B = \{B_t\}_{t\geq 0}$ be a one-dimensional Brownian motion on $(\Omega, \mathcal{F}, \mathbb{P})$. A (one-dimensional) Itô process is a stochastic process $X = \{X_t\}_{t\geq 0}$ on $(\Omega, \mathcal{F}, \mathbb{P})$ of the form

$$X = X_0 + \int_0^t u(y,\omega)\,\mathrm{d}y + \int_0^t v(y,\omega)\,\mathrm{d}B_y, \tag{1.7}$$

where v belongs to the class of processes satisfying the condition (i) from Definition 1.19 and conditions (ii)' and (iii)', so that

$$\mathbb{P}\left[\int_0^t v(y,\omega)^2\,\mathrm{d}y < \infty,\ \forall t \geq 0\right] = 1.$$

Also, it is assumed that u is $\mathcal{H}$-adapted, where $\mathcal{H}$ is from (ii)', and

$$\mathbb{P}\left[\int_0^t |u(y,\omega)|\,\mathrm{d}y < \infty,\ \forall t \geq 0\right] = 1.$$

[†]A function is called simple if it has only finitely many values [Cohn, 2013, p. 42]. A formal definition of these functions in relation to the construction of Itô integrals can be found in [Øksendal, 2000, p. 26] where they are called *elementary functions*.

Theorem 1.1 (The one-dimensional Itô formula).
Let $X = \{X_t\}_{t\geq 0}$ be an Itô process given by (1.7), written in the shorter differential form as

$$\mathrm{d}X_t = u(t, X_t)\,\mathrm{d}t + v(t, X_t)\,\mathrm{d}B_t,$$

where u and v are $\mathcal{H}$-adapted processes, with $\mathcal{H}$ satisfying the conditions in (ii)'. Let g be twice continuously differentiable on $[0, \infty\rangle \times \mathbb{R}$. Then $Y = \{Y_t\}_{t\geq 0}$ where

$$Y_t = g(t, X_t)$$

is again an Itô process and

$$\mathrm{d}Y_t = \frac{\partial}{\partial t}g(t, X_t)\,\mathrm{d}t + \frac{\partial}{\partial x}g(t, X_t)\,\mathrm{d}X_t + \frac{1}{2}\frac{\partial^2}{\partial x^2}g(t, X_t)(\mathrm{d}X_t)^2,$$

where $(\mathrm{d}X_t)^2 = (\mathrm{d}X_t)\cdot(\mathrm{d}X_t)$ is computed according to the rules

$$\mathrm{d}t \cdot \mathrm{d}t = \mathrm{d}t \cdot \mathrm{d}B_t = \mathrm{d}B_t \cdot \mathrm{d}t = 0, \quad \mathrm{d}B_t \cdot \mathrm{d}B_t = \mathrm{d}t. \tag{1.8}$$

Proof. See [Øksendal, 2000, Theorem 4.1.2]. ■

The term $(\mathrm{d}X_t)^2 = (\mathrm{d}X_t)\cdot(\mathrm{d}X_t)$ is in fact

$$(\mathrm{d}X_t)^2 = (u(t, X_t)\,\mathrm{d}t + v(t, X_t)\,\mathrm{d}B_t)^2.$$

In terms of size $\mathrm{d}t \ll \mathrm{d}B_t \ll 1$ and $\mathrm{d}B_t \approx \sqrt{\mathrm{d}t}$, which results in $(\mathrm{d}t)^2 \approx 0$, $\mathrm{d}t \cdot \mathrm{d}B_t \approx 0$ and $(\mathrm{d}B_t)^2 \approx (\sqrt{\mathrm{d}t})^2$. This gives the explanation of the rules defined in (1.8).

More about Itô calculus is given in [Øksendal, 2000, Chapters 3 and 4].

When discussing the solutions of the SDE (1.5), the question of existence and uniqueness of the solution arises. Next, the theorem which gives the conditions under which the solution(s) exist and is unique is presented below. The theorem is stated in a more general form where b and γ are allowed to be n-dimensional.

Theorem 1.2 (Existence and uniqueness theorem for SDEs).
Let $T > 0$, $t \in [0, T]$ and $g(t, x) \coloneqq g(x_t)$. Also, let $b\colon \mathbb{R}^n \to \mathbb{R}^n$, $\gamma\colon \mathbb{R}^n \to \mathbb{R}^{n\times m}$ be measurable functions satisfying

$$|b(x_t)| + |\gamma(x_t)| \leq C(1 + |x_t|), \quad x_t \in \mathbb{R}^n, \tag{1.9}$$

for some constant C, where $|\gamma|^2 = \sum |\gamma_{ij}|^2$ such that

$$|b(x_t) - b(y_t)| + |\gamma(x_t) - \gamma(y_t)| \leq D|x_t - y_t|, \quad x_t, y_t \in \mathbb{R}^n, \tag{1.10}$$

for some constant D. Let Z be a random variable which is independent of the σ-algebra $\mathcal{F}_\infty^{(m)}$ generated by $B_s, s \geq 0$, such that $\mathbb{E}[|Z|^2] < \infty$. Then the stochastic differential equation

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \gamma(X_t)\,\mathrm{d}B_t, \quad 0 \leq t \leq T,\ X_0 = Z. \tag{1.11}$$

has a unique t-continuos solution $X = \{X_t\}_{t\geq 0}$ with the property that X is adapted to the filtration $\{\mathcal{F}_t^Z\}_{t\geq 0}$ [†] generated by Z and B_s, $s \leq t$ and

$$\mathbb{E}\Big[\int_0^T |X_t|^2\,\mathrm{d}t\Big] < \infty. \tag{1.12}$$

Proof. For proof see [Øksendal, 2000, Theorem 5.2.1, p. 66]. ■

Linear growth condition (1.9) ensures that the solution X doesn't explode, i.e. that $|X|$ doesn't tend to infinity in finite time. The Lipschitz condition (1.10) guarantees uniqueness of the solution, i.e. if X_1 and X_2 are two continuous processes which satisfy the theorem conditions, then

$$X_1(t,\omega) = X_2(t,\omega), \quad \forall t \leq T \quad \text{a.s.} \tag{1.13}$$

Two possibilities for a solution exist. When the version of the Brownian motion B is given in advance and the solution constructed from it is $\mathcal{F}^Z$ adapted, then the solution is called a *strong solution*. However, if a filtration $\mathcal{H} = \{\mathcal{H}_t\}_{t\geq 0}$ such that processes $\widetilde{X}$ and $\widetilde{B}$ satisfy the SDE (1.11) (where $\widetilde{X}$ is $\mathcal{H}$-adapted and $\widetilde{B}$ is $\mathcal{H}$-Brownian motion, i.e. $\widetilde{B}$ is a Brownian motion and a martingale w.r.t. $\mathcal{H}$) and one needs to find a pair of process $\big((\widetilde{X}, \widetilde{B}), \mathcal{H}\big)$ on a probability space $(\Omega, \mathcal{H}, \mathbb{P})$ with only the functions b and γ given, then the solution is called a *weak solution*. As expected, a strong solution is also a weak solution, but not vice versa (in general).

Additionally, two forms of the uniqueness can be defined. A *strong* or *pathwise uniqueness* is defined by (1.13). A *weak uniqueness* means that any two solutions are unique in the probability law, i.e. they have the same finite-dimensional distributions. As stated in [Øksendal, 2000, Lemma 5.3.1, p. 71], if b and γ satisfy the conditions of Theorem 1.2 then a solution of (1.11) is weakly unique.

In summary, for the *strong solution*

- the initial value is given,
- the probability space is known and there is a Brownian motion B in that space,
- one is looking for the pathwise unique solutions,

and for the *weak solution*

[†] $\mathcal{F}_t^Z = \{\mathcal{F}_t^Z\}_{t\geq 0}$ is a notation for the filtration $\mathcal{F}_t = \{\mathcal{F}_t\}_{t\geq 0}$ generated by $Z = X_0$, to distinguish it from the filtration $\mathcal{F}_t = \{\mathcal{F}_t\}_{t\geq 0}$ generated by X.

- the probability law is given,
- there exists a probability space where (1.11) holds and one is free to choose the Brownian motion on that probability space,
- there is no point in looking for the pathwise uniqueness since they might be defined on different probability spaces,
- the notion of uniqueness refers to the weak uniqueness where solutions have the same finite-dimensional distribution.

From a modelling point and for mathematical reasons, it is convenient to look for weak solutions since there are SDEs that have no strong solution, but still have a (weakly) unique weak solution. Also, the weak solution concept is natural because it doesn't specify the explicit representation of a white noise [Øksendal, 2000].

Stochastic processes which are the solutions of a stochastic differential equation are called *(Itô) diffusions* (here just "diffusions").

Definition 1.22 (Diffusion).
A (time-homogeneous) diffusion is a stochastic process $X = \{X_t\}_{t\geq 0}$, where

$$X_t(\omega) \coloneqq X(t,\omega)\colon\ [0,\infty\rangle \times \Omega \to \mathbb{R}^n,$$

satisfying a stochastic differential equation of the form

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \gamma(X_t)\,\mathrm{d}B_t, \quad t \geq s, \quad X_s = x$$

where B is an m-dimensional Brownian motion and $b\colon \mathbb{R}^n \to \mathbb{R}^n$, $\gamma\colon \mathbb{R}^n \to \mathbb{R}^{n\times m}$ satisfy the conditions of Theorem 1.2. The (unique) solution is denoted by

$$X_t = X_t^{s,x},\ t \geq s.$$

For $\{X_t\}_{t\geq 0}$ to be time-homogenous means that

$$\{X_{s+h}^{s,x}\}_{h\geq 0} \quad \text{and} \quad \{X_h^{0,x}\}_{h\geq 0}$$

have the same $\mathbb{P}^0$-distributions (where $\mathbb{P}^0 = \mathbb{P}$ is the probability law of B_t starting at 0), i.e. a process is time-homogenous when it is insensitive in the distributional sense to the time-shifts. Thus, the definition implies that the process $\{X_h^{0,x}\}_{h\geq 0}$ shifted by $s \geq 0$ has the same probability properties as the process $\{X_{s+h}^{s,x}\}_{h\geq 0}$.

Definition 1.23 (Scale density, scale function and speed measure).
Let X be a diffusion which satisfies the SDE

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \gamma(X_t)\,\mathrm{d}B_t.$$

Then the scale density of the process is defined as

$$\mathfrak{s}(x) = \exp\left(-\int_{\cdot}^{x} \frac{2b(y)}{\gamma(y)}\,\mathrm{d}y \right),$$

where the notation $\int_{\cdot}^{x}$ *refers to an integral with an arbitrary lower limit (with respect to the domain of the integrand).*
The scale function of the process is then defined by

$$\mathfrak{S}(x) = \int_{\cdot}^{x} \mathfrak{s}(\eta)\,\mathrm{d}\eta = \int_{\cdot}^{x} \exp\left(-\int_{\cdot}^{\eta} \frac{2b(y)}{\gamma(y)}\,\mathrm{d}y \right) \mathrm{d}\eta.$$

The speed measure of the process is defined by

$$\mathfrak{m}(x) = \frac{1}{\gamma(x)\mathfrak{s}(x)}.$$

The property that the process is "memoryless", i.e. the future behaviour of the process after time t is the same as the behaviour obtained when the process started at X_t, is called the *Markov property*. If by $\mathbb{E}^x$ we denote expectation w.r.t. to the probability measure $\mathbb{Q}^x$ and $\mathbb{E}^y[f(X_h)]$ (which means $\mathbb{E}[f(X_h^y)]$) is the function evaluated at $y = X_t(\omega)$, then the Markov property for diffusions is given in the following theorem.

Theorem 1.3 (The Markov property for diffusion).
Let $f\colon \mathbb{R}^n \to \mathbb{R}$ *be a bounded Borel function. Then for* $t, h \geq 0$

$$\mathbb{E}^x\left[f(X_{t+h})|\mathcal{H}\right]_\omega = \mathbb{E}^{X_t(\omega)}\left[f(X_h)\right]$$

Proof. For proof see [Øksendal, 2000, Theorem 7.1.2, p. 111]. ■

Definition 1.24 ((Strict) stopping time).
Let $\{\mathcal{N}_t\}$ *be a filtration on* $(\Omega, \mathcal{F})$*. A function* $\tau\colon \Omega \to [0, \infty\rangle$ *is called a (strict) stopping time w.r.t* $\{\mathcal{N}_t\}$ *if*

$$\{\omega,\, \tau(\omega) \leq t\} \in \mathcal{N}_t, \quad \forall t \geq 0.$$

Theorem 1.4 (The strong Markov property for diffusion).
Let $f\colon \mathbb{R}^n \to \mathbb{R}$ *be a bounded Borel function,* τ *be a stopping time w.r.t.* $\{\mathcal{H}_t\}$*,* $\tau < \infty$ *a.s. Then for* $t, h \geq 0$

$$\mathbb{E}^x\left[f(X_{\tau+h})|\mathcal{H}_\tau\right]_\omega = \mathbb{E}^{X_\tau}\left[f(X_h)\right], \quad \forall h \geq 0.$$

Proof. For proof see [Øksendal, 2000, Theorem 7.4.4, p. 113]. ■

In other words, a diffusion can be defined as a continuos time stochastic process which possesses the strong Markov property and for which the sample paths $X_t(\omega)$ are almost always continuous functions of t [Karlin and Taylor, 1981].

In this section, the concept of stochastic differential equations was introduced within the framework of Markov and regular processes. More general concepts are outside of the scope of this thesis, but can be found in e.g. Heyde and Leonenko [2005], Taufer and Leonenko [2009] and Fournier [2009].

1.3 Diffusion construction

Throughout the thesis, the main focus is the construction of a diffusion with a given marginal density f which would be based on the data at hand, i.e. the distribution should fit the histogram of the data. The construction of such a diffusion with an exponentially decreasing autocorrelation function and a specified marginal distribution is presented in [Bibby et al., 2005]. The main theorem from the paper is given below.

Theorem 1.5 (Construction of diffusion with an exponential autocorrelation function and a specified marginal distribution).
Let f be a probability density with expectation μ and F beits distribution function. Suppose f is continuos, bounded and strictly positive on $\langle l, u\rangle$ and zero outside, where $-\infty \leq l < u \leq \infty$, and suppose it has a finite variance. Then

(i) the stochastic differential equation

$$\mathrm{d}X_t = -\theta(X_t - \mu)\,\mathrm{d}t + \sqrt{v(X_t)}\,\mathrm{d}B_t, \quad t \geq 0, \tag{1.14}$$

where $\theta > 0$, $\mu \in \langle l, u\rangle$ and v is a non-negative function defined on $\langle l, u\rangle$

$$v(x) = \frac{2\theta \int_l^x (\mu - y) f(y)\,\mathrm{d}y}{f(x)} = \frac{2\theta\mu F(x) - 2\theta \int_l^x y f(y)\,\mathrm{d}y}{f(x)}, \quad x \in \langle l, u\rangle, \tag{1.15}$$

has a unique Markovian weak solution. The diffusion coefficient is strictly positive for all $x \in \langle l, u\rangle$;

(ii) the diffusion process X that solves (1.14) *and* (1.15) *is ergodic with invariant density f;*

(iii) the function fv satisfies

$$\int_l^u v(x) f(x)\,\mathrm{d}x < \infty \tag{1.16}$$

and

$$\mathbb{E}\left[X_{s+t} | X_s = x\right] = x\mathrm{e}^{-\theta t} + \mu(1 - \mathrm{e}^{-\theta t}). \tag{1.17}$$

If $X_0 \sim f$, then X is stationary and the autocorrelation function of X is given by

$$\operatorname{Corr}(X_{s+t}, X_s) = \mathrm{e}^{-\theta t}, \quad s, t \geq 0; \tag{1.18}$$

(iv) if $-\infty < l$ or $u < \infty$, then the diffusion given by (1.14) *and* (1.15) *is the only ergodic diffusion with drift $-\theta(x-\mu)$ and invariant density f. If the state space is $\mathbb{R}$, it is the only ergodic diffusion with drift $-\theta(x-\mu)$ and invariant density f for which the condition* (1.16) *is satisfied.*

Lemma 1.6. *Suppose that $\mathbb{E}[f(x)] \leq \mu$ and that v is given by* (1.15)*. Then the function*

$$g(x) \coloneqq f(x)v(x) = 2\theta \int_l^x (\mu - y) f(y)\,\mathrm{d}y \tag{1.19}$$

is strictly positive for all $x \in \langle l, u\rangle$ and $\lim_{x\to l} g(x) = 0$. If $\mathbb{E}[f(x)] = \mu$, then $\lim_{x\to u} g(x) = 0$

Proof. See [Bibby et al., 2005, Lemma 2.2.] ■

Proof of Theorem 1.5.

(i) From Lemma 1.6 and the fact that f is continuous follows that the diffusion coefficient v is strictly positive for all $x \in \langle l, u\rangle$. Next for $x \in \langle l, u\rangle$ and some interior point $x^* \in \langle l, u\rangle$, the scale density is

$$\mathfrak{s}(x) = \exp\left(2\theta \int_x^{x^*} \frac{y-\mu}{v(y)}\,\mathrm{d}y\right) = \frac{g(x^*)}{g(x)} \tag{1.20}$$

and the scale function

$$\mathfrak{S}(x) = \int_x^{x^*} s(y)\,\mathrm{d}y = g(x^*) \int_x^{x^*} \frac{1}{g(y)}\,\mathrm{d}y.$$

Then $\mathfrak{S}$ is strictly increasing, twice continuously differentiable and maps $\langle l, u\rangle$ to $\mathbb{R}$. The function $\mathfrak{s}(\mathfrak{S}^{-1}(\cdot))\sqrt{v(\mathfrak{S}^{-1}(\cdot))}$ is continuos on $\mathbb{R}$, so the stochastic differential equation

$$\mathrm{d}Y_t = \mathfrak{s}(\mathfrak{S}^{-1}(Y_t))\sqrt{v(\mathfrak{S}^{-1}(Y_t))}\,\mathrm{d}B_t \tag{1.21}$$

satisfies the conditions of Theorem 2.2 of Engelbert and Schmidt [1985] so the SDE has a unique Markovian weak solution with state space $\mathbb{R}$. The process $\mathfrak{S}^{-1}(Y_t)$ solves the SDE (1.14) by Itô's formula and is the only such process. To show that it is the only solution, let X be a solution of the SDE (1.14). Then, again by Itô's formula, $\mathfrak{S}(X_t)$ solves (1.21).

(ii) To prove that X has invariant density f, it needs to be shown that the scale measure diverges at both end-points and that the speed measure has density proportional to f (see Skorokhod [1989]). If $\langle l, u\rangle = \mathbb{R}$, the scale measure diverges based on Lemma 1.6. If u is finite, from present theorem conditions, there exists $K > 0$ such that

$$g(x) = 2\theta \int_x^u (y-\mu) f(y)\,\mathrm{d}y \leq K(u-x)$$

from which follows that $\lim_{x\to u} S(x) = \infty$. Similarly, for l finite, $\lim_{x\to l} S(x) = -\infty$. From Karlin and Taylor [1981], the invariant density is proportional to the density of the speed measure, so using (1.19) and (1.20) the speed measure has the density

$$\frac{1}{v(x)\mathfrak{s}(x)} = \frac{f(x)}{g(x^*)}.$$

(iii) For the proof of (1.17) and (1.18), it needs to be shown that (1.16) holds. In the case of $x \in \langle l, u\rangle$, where $-\infty < l < u < \infty$, this follows from Lemma 1.6. For the infinite boundaries, vf needs to go sufficiently fast to zero, which is ensured by the condition that f has a finite variance. Hence, for $u = \infty$

$$\begin{aligned}\int_\mu^\infty g(x)\,\mathrm{d}x &= 2\theta \int_\mu^\infty \int_x^\infty (y-\mu) f(y)\,\mathrm{d}y\,\mathrm{d}x \\ \text{(Tonelli's theorem)} &= 2\theta \int_\mu^\infty \int_\mu^y \mathrm{d}x (y-\mu) f(y)\,\mathrm{d}y \\ &= 2\theta \int_\mu^\infty (y-\mu)^2 f(y)\,\mathrm{d}y < \infty.\end{aligned}$$

For $l = -\infty$, this can be verified in a similar way. Since (1.16) holds, from (1.14) follows (1.17), which then implies (1.18).

(iv) For an ergodic diffusion given by (1.14) with invariant density f and some $K > 0$, by using the general expression for the speed measure, it holds

$$f(x) = \frac{K}{v(x)} \exp\left(-2\theta \int_{x^*}^x \frac{y-\mu}{v(y)}\right) \mathrm{d}y.$$

Function $g = fv$ is differentiable so for some constant C follows

$$v(x) = \frac{2\theta \int_l^x (\mu - y) f(y)\,\mathrm{d}y + C}{f(x)}. \tag{1.22}$$

From Lemma 1.6 the integral goes to zero at l and u. Thus, it can be concluded that $C \geq 0$, since v needs to be positive for all $x \in \langle l, u\rangle$. If

$l = -\infty$ and $u = \infty$, the expression (1.22) defines an ergodic diffusion with invariant density f for all $C \geq 0$. However, for the scale measure $\frac{1}{fv}$ to diverge at a finite boundary, it is necessary that $C = 0$. Thus (1.22) holds only when $C = 0$. This proves the statement in Theorem 1.5 (iv).

■

Next theorem from Bibby et al. [2005] states how to determine the squared diffusion coefficient.

Theorem 1.7 (Squared diffusion coefficient).
Consider an invariant density for a diffusion process which belongs to an exponential family of the form

$$f(x;\xi) = a(\xi)b(x)\mathrm{e}^{\xi_1 x + \alpha(\xi)t(x)}, \tag{1.23}$$

where $\xi = (\xi_1, \ldots, \xi_p)$ and α, t may be vectors. Then the squared diffusion coefficient is given by

$$v(x;\xi) = -\frac{2\theta}{f(x;\xi)}\frac{\partial}{\partial \xi_1}F(x;\xi), \quad l < x < u. \tag{1.24}$$

Proof. For proof see [Bibby et al., 2005, Theorem 2.4]. ■

This review of definitions and theorems concludes the theoretical part needed for the understanding of the methods and results presented in the rest of the thesis.

2 Unimodal case: Generalized Gaussian distribution

As mentioned in Section 1.3, the choice of the marginal density of the diffusion should be based on the empirical data. Analysis of the dataset at hand showed that the histograms of EEG increments have a symmetric distribution and the majority of the histograms displayed a one-peak distribution with a maximum at zero, which resembled a standard normal distribution, but in some cases the tails were either heavier or lighter than the tails of normal distribution. A detailed analysis of the dataset will be presented in Section III.1. To reflect the diversity of the empirically observed distribution candidates for the unimodal case, probability density function of the distribution of the increments was chosen to be the *generalized Gaussian distribution (GGD)*[†] with a specific parametrization

[†]Another name for the GGD is Generalized error distribution, which is mostly used in economics.

adapted from Lutwak et al. [2004] with mean zero. Hence, omitting the parameter μ form the parametrization (i.e. using $\mu = 0$) reflects the situation observed on the EEG increments. The parametrization is then given as

$$f_{s,b}(x) = \begin{cases} \dfrac{1}{2(s\sigma^2)^{1/s}\Gamma\left(1+\frac{1}{s}\right)} \exp\left(-\dfrac{|x|^s}{s\sigma^2}\right), & b = 0, \\ \dfrac{bs}{2\sigma^2}\left(\dfrac{s\sigma^2}{b}\right)^{-1/s} \dfrac{\Gamma\left(1+\frac{1}{s}+\frac{\sigma^2}{b}\right)}{\Gamma\left(\frac{1}{s}\right)\Gamma\left(\frac{\sigma^2}{b}\right)} \left(1+\dfrac{b}{s\sigma^2}|x|^s\right)^{-\frac{\sigma^2}{b}-\frac{1}{s}-1}, & b > 0, \end{cases} \tag{2.1}$$

where the parameter b is used as an indicator for making the distinction between the light-tailed and heavy-tailed distributions within the family. The subfamily characterized by $b = 0$ resembles the usual GGD parametrization including, for $s = 2$, the zero-mean normal distribution with variance σ^2. For $b > 0$, distributions in the GGD subfamily admit heavy tails, e.g., for $s = 2$ this distribution is of the Student type.

Since the majority of the analysis deals with the subfamily where $b = 0$, some of the basic properties of this GGD subfamily are presented first for the general case where $\mu \neq 0$, with some remarks related to the subfamily where $b > 0$.

2.1 Properties of the generalized Gaussian distribution

Basic properties of the generalized Gaussian distribution which are presented in this section are derived from Dytso et al. [2018]. For completeness, all properties are stated, but proofs here are given only for some of the properties which are important for the analysis.

Dytso et al. [2018] use the following parametrization of GGD

$$f_{X_p}(x) = \frac{c_p}{\alpha} e^{-\frac{|x-\mu|^p}{2\alpha^p}}, \quad c_p = \frac{p}{2^{\frac{p+1}{p}}\Gamma(\frac{1}{p})}, \quad x \in \mathbb{R},\ p > 0. \tag{2.2}$$

So X_p follows the Generalized Gaussian distribution, denoted as $X_p \sim \mathcal{N}_p(\mu, \alpha^p)$ where μ, p, α are the location, shape and scale parameters, respectively. Gamma function, lower incomplete gamma function and upper incomplete gamma function are defined respectively as follows:

$$\begin{aligned} \Gamma(x) &= \int_0^\infty t^{x-1}e^{-t}dt, \\ \gamma(a,x) &= \int_0^x t^{a-1}e^{-t}dt, \\ \Gamma(a,x) &= \int_x^\infty t^{a-1}e^{-t}dt. \end{aligned}$$

In order to relate this parametrization to the previous notation in (2.1) for $b = 0$, following relations for shape and scale parameters are used:

$$p \coloneqq s, \qquad \alpha \coloneqq \left(\frac{s\sigma^2}{2}\right)^{\frac{1}{s}},$$

thus the density function becomes

$$f_{X_s}(x) = \frac{c_{s,\sigma^2}}{\alpha} e^{-\frac{|x-\mu|^s}{s\sigma^2}}, \quad c_{s,\sigma^2} = \frac{1}{2(s\sigma^2)^{\frac{1}{s}}\Gamma\left(1+\frac{1}{s}\right)}, \quad x \in \mathbb{R},\ s > 0. \tag{2.3}$$

and $X_s \sim \mathcal{N}_s\left(\mu, \frac{s\sigma^2}{2}\right)$, which is the same parametrization as (2.1) (for $b = 0$ and $\mu = 0$). In this case, μ is again the location parameter, s is the shape parameter and σ^2 is a parameter related to the scale parameter α by the expression $\left(\frac{s\sigma^2}{2}\right)^{\frac{1}{s}}$. The cumulative distribution function (CDF) is then given by

$$F_{X_s}(x) = \frac{1}{2} + \operatorname{sgn}(x-\mu)\frac{\gamma\left(\frac{1}{s}, \frac{|x-\mu|^s}{\frac{s\sigma^2}{2}}\right)}{2\Gamma\left(\frac{1}{s}\right)}, \quad x \in \mathbb{R} \tag{2.4}$$

As previously stated, for $s = 2$ this distribution is Gaussian with mean μ and variance σ^2. Other important examples from this family of distributions are the Laplace distribution with location parameter μ and scale σ^2 for $s = 1$ and the uniform distribution on $\left[\mu - \left(\frac{s\sigma^2}{2}\right)^{\frac{1}{s}}, \mu + \left(\frac{s\sigma^2}{2}\right)^{\frac{1}{s}}\right]$ for $s = \infty$.

Mellin transform, moments and the moment problem

The Mellin transform is used as a tool in characterizing products of independent random variables.

Definition 2.1 (Mellin transform).
The Mellin transform of a positive random variable X is defined as

$$m_X(r) = \mathbb{E}[X^{r-1}], \quad r \in \mathbb{C}.$$

For independent random variables, it follows that

$$m_{X\cdot Y}(r) = m_X(r) \cdot m_Y(r).$$

Proposition 2.1 (Mellin transform of $|X_s|$).
For any $s > 0$ and $X_s \sim \mathcal{N}_s(0, \frac{s\sigma^2}{2})$

$$\mathbb{E}\left[\left|X_s\right|^{r-1}\right] = \frac{2^{\frac{r-1}{s}}}{\Gamma\left(\frac{1}{s}\right)}\left(\frac{s\sigma^2}{2}\right)^{\frac{r-1}{s}}\Gamma\left(\frac{r}{s}\right), \quad \operatorname{Re}(r) > 0.$$

For any $s > 0$ and $k > -1$, the absolute moments are given by

$$\mathbb{E}[|X|^k] = (s\sigma^2)^{\frac{k}{s}}\frac{\Gamma\left(\frac{k+1}{s}\right)}{\Gamma\left(\frac{1}{s}\right)}. \tag{2.5}$$

Proof. See [Dytso et al., 2018, Proposition 1]. ■

Hence, the s-th absolute moment of X_s is given by

$$\mathbb{E}\Big[\Big|X_s\Big|^s\Big] = \frac{s\sigma^2}{\Gamma\left(\frac{1}{s}\right)}\Gamma\left(\frac{s+1}{s}\right) = \sigma^2,$$

which gives the interpretation of the shape parameter.

In particular, integer moments are given by

$$\mathbb{E}[X^n] = \frac{1+(-1)^n}{2}(s\sigma^2)^{\frac{n}{s}}\frac{\Gamma\left(\frac{1+n}{s}\right)}{\Gamma\left(\frac{1}{s}\right)}, \quad n \in \mathbb{N}, \tag{2.6}$$

and therefore the first four moments are

$$\begin{aligned}
\mathbb{E}[X] &= \mathbb{E}[X^3] = 0,\\
\mathbb{E}[X^2] &= \frac{(s\sigma^2)^{\frac{2}{s}}\,\Gamma\left(\frac{3}{s}\right)}{\Gamma\left(\frac{1}{s}\right)},\\
\mathbb{E}[X^4] &= \frac{(s\sigma^2)^{\frac{4}{s}}\,\Gamma\left(\frac{5}{s}\right)}{\Gamma\left(\frac{1}{s}\right)}.
\end{aligned}$$

The relation between k-th moments of two GGDs of a different order is given by the following corollary.

Corollary 2.2. *Let $X_q \sim \mathcal{N}_q(0,1)$ and $X_s \sim \mathcal{N}_s(0,1)$. Then for $q \geq s > 0$*

$$\mathbb{E}\Big[\Big|X_q\Big|^k\Big] \leq \mathbb{E}\Big[\Big|X_s\Big|^k\Big] \tag{2.7}$$

for any $k \in \mathbb{R}^+$. Moreover, for any $q > s$

$$\lim_{k\to\infty}\left(\frac{\mathbb{E}\Big[\Big|X_s\Big|^k\Big]}{\mathbb{E}\Big[\Big|X_q\Big|^k\Big]}\right)^{\frac{1}{k}} = \infty.$$

Proof. See [Dytso et al., 2018, Appendix X]. ■

Remark 2.1. *For the heavy-tailed GGD subfamily ($b > 0$), the tail of the density decreases like $|x|^{-1-s\left(\frac{\sigma^2}{b}+1\right)}$. The absolute moment of order $k > 0$ exists for $k < s\left(\frac{\sigma^2}{b}+1\right)$ and is given by the following expression:*

$$\mathbb{E}[|X|^k] = \left(\frac{b}{s\sigma^2}\right)^{\frac{k}{s}}\frac{\Gamma\left(\frac{1+k}{s}\right)\Gamma\left(1-\frac{k}{s}+\frac{\sigma^2}{b}\right)}{\Gamma\left(\frac{1}{s}\right)\Gamma\left(1+\frac{\sigma^2}{b}\right)}. \tag{2.8}$$

The integer moments are given by

$$\begin{aligned}
\mathbb{E}[X^n] = \frac{s\left(1+(-1)^n\right)}{2}&\left(\frac{b}{s\sigma^2}\right)^{1-\frac{n}{s}}\\
&\cdot\frac{\Gamma\left(\frac{1+n}{s}\right)\Gamma\left(1-\frac{n}{s}+\frac{\sigma^2}{b}\right)}{\Gamma\left(\frac{1}{s}\right)\Gamma\left(\frac{\sigma^2}{b}\right)}, \quad n < s\left(\frac{\sigma^2}{b}+1\right).
\end{aligned} \tag{2.9}$$

The first six moments are of the following form:

$$\mathbb{E}[X] = \mathbb{E}[X^3] = \mathbb{E}[X^5] = 0,$$

$$\mathbb{E}[X^2] = \frac{\Gamma\left(\frac{3}{s}\right)\left(\frac{b}{s\sigma^2}\right)^{-\frac{2}{s}}\Gamma\left(\frac{\sigma^2}{b}-\frac{2}{s}+1\right)}{\Gamma\left(\frac{1}{s}\right)\Gamma\left(\frac{\sigma^2+b}{b}\right)},$$

$$\mathbb{E}[X^4] = \frac{\Gamma\left(\frac{5}{s}\right)\left(\frac{b}{s\sigma^2}\right)^{-\frac{4}{s}}\Gamma\left(\frac{\sigma^2}{b}-\frac{4}{s}+1\right)}{\Gamma\left(\frac{1}{s}\right)\Gamma\left(\frac{\sigma^2+b}{b}\right)},$$

$$\mathbb{E}[X^6] = \frac{\Gamma\left(\frac{7}{s}\right)\left(\frac{b}{s\sigma^2}\right)^{-\frac{6}{s}}\Gamma\left(\frac{\sigma^2}{b}-\frac{6}{s}+1\right)}{\Gamma\left(\frac{1}{s}\right)\Gamma\left(\frac{\sigma^2+b}{b}\right)}.$$

Remark 2.2. *As already stated, in the heavy-tailed part of* (2.1) *where* $b > 0$ *for the case* $s = 2$ *Student-type distribution is obtained with the PDF*

$$f_{2,b}(x) = \frac{\Gamma\left(\frac{\sigma^2}{b}+\frac{3}{2}\right)}{\sqrt{\frac{2\pi\sigma^2}{b}}\Gamma\left(\frac{\sigma^2}{b}+1\right)}\left(1+\frac{b}{2\sigma^2}x^2\right)^{-\frac{\sigma^2}{b}-\frac{3}{2}}, \quad b > 0. \tag{2.10}$$

The absolute moment of order $k > 0$ *exists for* $k < 2\left(\frac{\sigma^2}{b}+1\right)$ *and is given by the following expression:*

$$\mathbb{E}[|X|^k] = \left(\frac{b}{2\sigma^2}\right)^{-\frac{n}{2}}\frac{\Gamma\left(\frac{1+k}{2}\right)\Gamma\left(1-\frac{k}{2}+\frac{\sigma^2}{b}\right)}{\sqrt{\pi}\,\Gamma\left(1+\frac{\sigma^2}{b}\right)}. \tag{2.11}$$

The integer moments are given by

$$\mathbb{E}[X^n] = (1+(-1)^n)\left(\frac{b}{2\sigma^2}\right)^{1-\frac{n}{2}} \cdot \frac{\Gamma\left(\frac{1+n}{2}\right)\Gamma\left(1-\frac{n}{2}+\frac{\sigma^2}{b}\right)}{\sqrt{2\pi}\Gamma\left(\frac{\sigma^2}{b}\right)}, \quad n < 2\left(\frac{\sigma^2}{b}+1\right). \tag{2.12}$$

and the first six moments are of the following form:

$$\begin{aligned}\mathbb{E}[X] = \mathbb{E}[X^3] = \mathbb{E}[X^5] &= 0,\\ \mathbb{E}[X^2] &= 1,\\ \mathbb{E}[X^4] &= \frac{3\sigma^2}{\sigma^2-b},\\ \mathbb{E}[X^6] &= \frac{15\sigma^4}{\sigma^4-3b\sigma^2+2b^2}.\end{aligned}$$

For example, the explicit expressions for the moments enable the direct calculation of method of moments estimators of a strictly stationary diffusion with Student

marginals. For more details on analysis of probabilistic properties, moment-based parameter estimation and testing of statistical hypothesis regarding such diffusion see Leonenko and Šuvak [2010].

The moment problem refers to the question whether a distribution can be uniquely defined by its moments. If a random variable is defined on $\mathbb{R}$, the problem is referred to as the Hamburg moment problem. For variables defined on $\mathbb{R}^+$, the problem is called the Stieltjes moment problem. For random variables which are uniquely defined by their moments, the moment problem is called *determinate*. If there exists another distribution which has the same moments, the moment problem is *indeterminate*. Next theorem gives the conditions under which the GGD distribution (for $b = 0$) is determinate and indeterminate.

Theorem 2.3 (Determinacy of GGD).
The GGD distribution is determinate for $s \in [1, \infty\rangle$ and indeterminate for $s \in \langle 0, 1\rangle$.

Proof. To show that the GGD is indeterminate for $s \in \langle 0, 1\rangle$, it is enough to check the classical Krein sufficient condition, which states that an absolutely continuos distribution with PDF f is indeterminate when

$$\int_{-\infty}^{\infty} -\frac{\ln(f(x))}{1+x^2}\,\mathrm{d}x < \infty.$$

Imputing (2.3) for f and calculating the integral gives

$$\int_{-\infty}^{\infty} -\frac{\ln(f(x))}{1+x^2}\,\mathrm{d}x = -\pi \log(c_{s,\sigma^2}) + 2\int_{0}^{\infty} \frac{(x-\mu)^s}{1+x^2}\,\mathrm{d}x.$$

Since the first term on the right-hand side of the equation is a constant, it is enough to show that the integral on the right-hand side is finite for $s \in \langle 0, 1\rangle$. If $s \in \langle 0, 1\rangle$

$$(x-\mu)^s < (x^2+1),$$

thus the integral is finite. For $s = 1$, the integral is infinite and doesn't satisfy the condition. The part of the proof which shows that the GGD is determinate for $s \in [1, \infty\rangle$ will be shown using the characteristic function later on. ■

Since the GGD is indeterminate for $s \in \langle 0, 1\rangle$, one can conclude that the method of moments performs poorly for this parameter range which was in fact shown by Varanasi and Aazhang [1989].

Stochastic ordering and positive definiteness

Based on the inequality (2.7), there might be some ordering in GGD family.

Definition 2.2 (Stochastic dominance).
A random variable X dominates another random variable Y in the sense of

- *the first order-stochastic dominance if*

$$F_X(x) \leq F_Y(x), \quad \forall x \in \mathbb{R}; \tag{2.13}$$

- *the second order-stochastic dominance if*

$$\int_{-\infty}^{x} \Big(F_Y(u) - F_X(u)\Big)\,\mathrm{d}u \geq 0, \quad \forall x. \tag{2.14}$$

Proposition 2.4. *Let $X_s \sim \mathcal{N}_s(0,1)$ and $X_q \sim \mathcal{N}_q(0,1)$. Then for $s \leq q$, X_q dominates X_s in the sense of the second-order stochastic dominance.*

Proof. See [Dytso et al., 2018, Appendix B]. ■

The first order-stochastic dominance doesn't hold since, for $s \leq q$

$$F_{X_q}(x) \leq F_{X_s}(x), \quad x \leq 0,$$
$$F_{X_q}(x) \geq F_{X_s}(x), \quad x > 0.$$

Proposition 2.5. *Let $X_s \sim \mathcal{N}_s(0,1)$ and $X_q \sim \mathcal{N}_q(0,1)$. Then for $s \leq q$ and for any nondecreasing concave function $g : \mathbb{R} \to \mathbb{R}$*

$$\mathbb{E}\left[g(X_q)\right] \geq \mathbb{E}\left[g(X_s)\right]. \tag{2.15}$$

Proof. See [Dytso et al., 2018, Proposition 4]. ■

Depending on the value range of parameter s, the GGD shows different properties. Next theorems gives the conditions under which the GGD is completely monotone and positive definite. The definitions are given first.

Definition 2.3 (Completely monotone and Bernstein functions).
A function $f\colon [0,\infty\rangle \to [0,\infty\rangle$ is said to be completely monotone if

$$(-1)^k \frac{\mathrm{d}^k f(x)}{\mathrm{d}x^k} \geq 0, \quad \forall x > 0 \text{ and } k \in \mathbb{N}^+. \tag{2.16}$$

A function $f\colon [0,\infty\rangle \to [0,\infty\rangle$ is said to be a Bernstein function if the derivative of f is a completely monotone function.

Since a composition of a completely monotone and a Bernstein function is completely monotone [Schilling et al., 2012], the following result is obtained.

Corollary 2.6. *For $s \in \langle 0, 1]$ the function $\mathrm{e}^{-\frac{x^s}{2}}$ is completely monotone.*

Definition 2.4 (Positive definite functions).
A function $f\colon \mathbb{R} \to \mathbb{C}$ is called positive definite if for every positive integer n and $x_1, x_2, \ldots, x_n \in \mathbb{R}$ the $n \times n$ matrix

$$A \coloneqq (a_{ij})_{i,j=1}^{n}, \quad a_{ij} = f(x_i - x_j)$$

is positive semi-definite.

Theorem 2.7. *The function $\mathrm{e}^{-\frac{x^s}{2}}$ is*

- *not positive definite for $s \in \langle 2, \infty\rangle$;*
- *positive definite for $s \in \langle 0, 2]$. Also, there exists a finite non-negative Borel measure[†] m_s on $\mathbb{R}^+$ such that for $x > 0$*

$$\mathrm{e}^{-\frac{x^s}{2}} = \int_0^\infty \mathrm{e}^{-\frac{u}{2}x^2} \,\mathrm{d}m_s(u). \tag{2.17}$$

Proof. See [Dytso et al., 2018, Appendix C]. ■

Based on the results stated in the previous theorem, the property of the positive-definiteness is reflected on the properties of the GGD in the parameter s, depending whether $s \leq 2$ or $s > 2$.

Another result which will be useful for analysing the properties of the characteristic function of the GGD is given by the next corollary.

Corollary 2.8. *For any $0 < q \leq s \leq 2$ let $r = \frac{2q}{s}$. Then for $x > 0$*

$$\mathrm{e}^{-\frac{x^q}{2}} = \int_0^\infty \mathrm{e}^{-\frac{u}{2}x^r} \,\mathrm{d}m_s(u). \tag{2.18}$$

Product decomposition of GGD random variables

There exists a decompositional representation of the GGD random variables which is a consequence of Theorem 2.7 and is given by the following proposition.

Proposition 2.9. *For any $0 < q \leq s \leq 2$ let $X_q \sim \mathcal{N}_q(0, 1)$ and $X_{\frac{2q}{s}} \sim \mathcal{N}_{\frac{2q}{s}}(0, 1)$. Then*

$$X_q \overset{d}{=} V_{s,q} X_{\frac{2q}{s}} \tag{2.19}$$

where $V_{s,q}$ is a positive random variable independent of $X_{\frac{2q}{s}}$ and has the following properties

[†]Such Borel measure can be constructed by assigning $\mathbb{R}^+$ numbers to Borel sets such that the measure is countly additive for disjoint unions and that equation (2.17) holds.

- *$V_{s,q}$ is an unbounded RV for $s < 2$ and $V_{s,q} = 1$ for $s = 2$;*
- *for $s < 2$ $V_{s,q}$ is a continuos RV with PDF given by*

$$f_{V_{s,q}}(y) = \frac{1}{2\pi} \frac{\Gamma\left(\frac{s}{2q}\right)}{\Gamma\left(\frac{1}{q}\right)} \int_{\mathbb{R}} y^{-iu-1} \frac{2^{\frac{iu}{q}} \Gamma\left(\frac{iu+1}{q}\right)}{2^{\frac{ius}{2q}} \Gamma\left(\frac{s(iu+1)}{2q}\right)} \,\mathrm{d}u, \quad y > 0. \tag{2.20}$$

Proof. To show that X_q is decomposable, denote $r = \frac{2q}{s}$ and let m_s be a finite non-negative Borel measure defined in Theorem 2.7. Since $X_q \sim \mathcal{N}_q(0,1)$, from

$$\begin{aligned}
1 = \mathbb{P}(X_q \in \mathbb{R}) &= \int_{\mathbb{R}} c_q \mathrm{e}^{-\frac{|x|^q}{2}} \,\mathrm{d}x \\
\text{(using the representation of } \mathrm{e}^{-\frac{|x|^q}{2}} \text{ from Cor. 2.8)} &= \int_{\mathbb{R}} c_q \int_0^\infty \mathrm{e}^{-\frac{u}{2}|x|^r} \,\mathrm{d}m_s(u) \,\mathrm{d}x \\
\text{(Tonelli's theorem)} &= c_q \int_0^\infty \int_{\mathbb{R}} \mathrm{e}^{-\frac{u}{2}|x|^r} \,\mathrm{d}x \,\mathrm{d}m_s(u) = c_q \int_0^\infty \frac{1}{c_r u^{\frac{1}{r}}} \,\mathrm{d}m_s(u) \\
\text{(defining } \mathrm{d}y(u) \coloneqq \tfrac{c_q}{c_r} \tfrac{1}{u^{\frac{1}{r}}} \mathrm{d}m_s(u)) &= \int_0^\infty \mathrm{d}y(u)
\end{aligned}$$

it can be concluded that y is a probability measure on $[0, \infty\rangle$.

Now, for any measurable set $S \subset \mathbb{R}$, it holds

$$\begin{aligned}
\mathbb{P}(X_q \in S) &= \int_S c_q \mathrm{e}^{-\frac{|x|^q}{2}} \,\mathrm{d}x \\
\text{(using the representation of } \mathrm{e}^{-\frac{|x|^q}{2}} \text{ from Cor. 2.8)} &= \int_S c_q \int_0^\infty \mathrm{e}^{-\frac{u}{2}|x|^r} \,\mathrm{d}m_s(u) \,\mathrm{d}x \\
&= \int_0^\infty \int_S c_q \frac{c_r u^{\frac{1}{r}}}{c_r u^{\frac{1}{r}}} \mathrm{e}^{-\frac{u}{2}|x|^r} \,\mathrm{d}x \,\mathrm{d}m_s(u) \\
\text{(}\mathrm{d}y(u) = \tfrac{c_q}{c_r} \tfrac{1}{u^{\frac{1}{r}}} \mathrm{d}m_s(u)\text{) is a probability measure)} &= \int_0^\infty \mathbb{P}\left(\frac{1}{T^{\frac{1}{r}}} X_r \in S \mid T = u \right) \frac{c_q}{c_r} \frac{1}{u^{\frac{1}{r}}} \,\mathrm{d}m_s(u) \\
(X_r \text{ is independent of } u) &= \mathbb{E}\left[\mathbb{P}\left(\frac{1}{T^{\frac{1}{r}}} X_r \in S \mid T \right) \right] \\
\text{(renaming } \tfrac{1}{T^{\frac{1}{r}}} \coloneqq V_{s,q}) &= \mathbb{P}\left(V_{s,q} X_r \in S\right) = \mathbb{P}\left(V_{s,q} X_{\frac{2q}{s}} \in S\right)
\end{aligned}$$

which proves (2.19). For the rest of the proof see [Dytso et al., 2018, Appendix D]. ∎

Based on the previous result, the GGD is a Gaussian mixture.

Definition 2.5 (Gaussian mixture).
A random variable X is called a (centred) Gaussian mixture if there exists a positive random variable V and a standard Gaussian random variable Z independent of V such that

$$X \overset{d}{=} VZ.$$

Corollary 2.10. *For $q \in \langle 0, 2]$, $X_q \sim \mathcal{N}_q(0,1)$ is a Gaussian mixture, i.e.*

$$X_q \stackrel{d}{=} V_{G,q} X_2,$$

where $V_{G,q} \coloneqq V_{q,q}$ is independent of X_2 and its PDF is defined by (2.20).

Proof. By choosing $s = q$ in (2.19), it follows that $X_q \stackrel{d}{=} V_{G,q} X_2$. ■

Additionally, when $q \leq 1$ and X_1 is a Laplace RV, it follows

$$X_q \stackrel{d}{=} V_{q,2q} X_1.$$

Thus, $V_{q,2q}$ will be denoted as $V_{L,q}$.

A different representation of the PDF given in (2.20) which may be easier to analyse is given by the next proposition.

Proposition 2.11. *For $0 < q \leq s \leq 2$ the PDF of a random variable $V_{s,q}$ has the following representations*

- *the power series representation*

$$f_{V_{s,q}}(y) = \frac{\Gamma\left(\frac{s}{2q}\right)}{\Gamma\left(\frac{1}{q}\right)} \sum_{k=1}^{\infty} a_k y^{kq}, \quad y > 0, \tag{2.21}$$

  where

$$a_k = \frac{q}{\pi} \frac{(-1)^{(k+1)}\, 2^{(kq+1)\left(\frac{s}{2q}-\frac{1}{q}\right)}\, \Gamma\left(\frac{kq}{2}+1\right) \sin\left(\frac{\pi kq}{2}\right)}{k!};$$

- *the integral representation*

$$f_{V_{s,q}}(y) = \frac{q^{2^{\frac{s}{2q}-\frac{1}{q}}}\, \Gamma\left(\frac{p}{2q}\right)}{\pi \Gamma\left(\frac{1}{q}\right)} \int_0^{\infty} \sin\left(a_s y^q x^{\frac{s}{2}}\right) \exp\left(-b_s y^q x^{\frac{s}{2}} - x\right) \mathrm{d}x, \tag{2.22}$$

  where

$$a_s = 2^{\frac{s}{2}-1} \sin\left(\frac{\pi s}{2}\right), \quad b_s = 2^{\frac{s}{2}-1} \cos\left(\frac{\pi s}{2}\right).$$

Proof. See [Dytso et al., 2018, Appendix E]. ■

For $s = q = 1$ the random variable $V_{G,1}$ is distributed according to the Rayleigh distribution, i.e.

$$f_{V_{G,1}}(y) = \frac{2^{-\frac{1}{2}}}{\sqrt{\pi}} \int_0^{\infty} \sin\left(\frac{y x^{\frac{1}{2}}}{\sqrt{2}}\right) \mathrm{e}^{-x}\, \mathrm{d}x = \frac{y}{4} \mathrm{e}^{-\frac{y^2}{8}}, \quad y \geq 0.$$

Similarly to Theorem 2.3, next proposition gives the conditions under which the distribution of $V_{G,q}$ is determinate.

Proposition 2.12. *The distribution of $V_{G,q}$ is determinate for $q \geq \frac{2}{5}$.*

Proof. See [Dytso et al., 2018, Proposition 8]. ■

Remark 2.3. *The product of two independent RVs with determinate distributions can have an indeterminate distribution. From Corollary 2.10 it is known that*

$$X_q \overset{d}{=} V_{G,q} X_2.$$

Based on Theorem 2.3, X_q has an indeterminate distribution on $q \in \left[\frac{2}{5}, 1\right]$, but from Proposition 2.12 it is known that $V_{G,q}$ is determinate on $q \in \left[\frac{2}{5}, 1\right]$.

Characteristic function

The characteristic function of the GGD can be presented in an integral form which is given by the following theorem.

Theorem 2.13 (Characteristic function of GGD).
The characteristic function of $X_s \sim \mathcal{N}_s(0,1)$ is given by

(i) for any $s > 0$

$$\phi_s(t) = 2c_{s,\sigma^2} \int_0^\infty \cos(tx) \mathrm{e}^{-\frac{x^s}{2}} \,\mathrm{d}x, \quad t \in \mathbb{R}; \tag{2.23}$$

(ii) for any $s \in \langle 0, 2]$

$$\phi_s(t) = \mathbb{E}\left[\mathrm{e}^{-\frac{t^2 V_{G,s}^2}{2}}\right], \quad t \in \mathbb{R}, \tag{2.24}$$

where the density of $V_{G,s}$ is defined in Proposition 2.9 and the PDF of $V_{G,s}^2$ is $\frac{1}{2\sqrt{y}} f_{V_{G,s}}(\sqrt{y})$.

Proof. (i) Since $\mathrm{e}^{-\frac{|x|^s}{2}}$ is an even function, its Fourier transform is equivalent to the cosine transform.

(ii) Characteristic function of X_2 is $\mathrm{e}^{-\frac{t^2}{2}}$. By the definition of characteristic function given in Definition 1.8 and the decomposition property given in Proposition 2.9, it follows

$$\phi_s(t) = \mathbb{E}\left[\mathrm{e}^{itX_s}\right] = \mathbb{E}\left[\mathrm{e}^{itV_{G,s}X_2}\right] = \mathbb{E}\left[\mathbb{E}\left[\mathrm{e}^{itV_{G,s}X_2} | V_{G,s}\right]\right]$$

where the last equality follows from the law of total expectation. Equality (2.24) follows from the fact that $V_{G,s}$ is independent of X_2.

■

Note that ϕ_s for $s \in \langle 0, 2]$ is a decreasing function for $t > 0$.

A special class of distributions which are closed under convolution of independent copies is called stable distribution (or stable random variable).

Definition 2.6 (Stable distribution). *Let X_1 and X_2 be independent copies of a random variable X. Then X is said to be stable if for all constants $a > 0$, $b > 0$ there exist $c > 0$ and $d \in \mathbb{R}$ such that*

$$aX_1 + bX_2 \stackrel{d}{=} cX + d \tag{2.25}$$

which is equivalent to

$$\phi_X(at)\phi_X(bt) = \phi_X(ct)\mathrm{e}^{itd}, \quad \forall t \in \mathbb{R}, \tag{2.26}$$

where ϕ_X is the characteristic function of X.

The characteristic function of a stable distribution has the following canonical representation

$$\phi_X(t) = \mathrm{e}^{-it\mu-|ct|^\alpha(1-i\beta\,\mathrm{sgn}(t)\Delta(t))}, \quad \text{where } \Delta(t) = \begin{cases} \tan\left(\frac{\pi\alpha}{2}\right), \ \alpha \neq 1, \\ -\frac{2}{\pi}\ln|t|, \ \alpha = 1. \end{cases} \tag{2.27}$$

Here $\mu \in \mathbb{R}^+$ is the shift parameter, $c \in \mathbb{R}^+$ is the scaling parameter, $\beta \in [-1, 1]$ is the skewness parameter and $\alpha \in \langle 0, 2]$ is the order parameter. When $\beta = 0$ (i.e symmetric stable distributions) the class is called α-stable distributions and the characteristic function is given by

$$\phi_X(t) = \mathrm{e}^{-|t|^\alpha}, \quad t \in \mathbb{R}. \tag{2.28}$$

For $s \in \langle 0, 2]$, the PDF of the GGD is equal to the characteristic function of an α-stable distribution (up to a normalizing constant), and vice versa, the PDF of an α-stable RV is equal to the characteristic function of the GGD, again up to a normalizing constant. This gives another integral representation of the characteristic function of the GGD. The representation is useful in numerical computations of ϕ_s since the integral is performed over a finite interval.

Proposition 2.14. *For $s \in \langle 0, 2] \setminus \{1\}$*

$$\phi_s(t) = 2\pi c_{s,\sigma^2}\frac{s|t|^{\frac{1}{s-1}}}{2|s-1|}\int_0^1 U_s(x)\mathrm{e}^{-|t|^{\frac{s}{s-1}}U_s(x)}\,\mathrm{d}x, \tag{2.29}$$

where

$$U_s(x) = \left(\frac{\sin\left(\frac{\pi x s}{2}\right)}{\cos\left(\frac{\pi x}{2}\right)}\right)^{\frac{s}{1-s}}\frac{\cos\left(\frac{\pi x(s-1)}{2}\right)}{\cos\left(\frac{\pi x}{2}\right)}.$$

Also, let $g_s(x) := U_s(x)\mathrm{e}^{-|t|^{\frac{s}{s-1}}U_s(x)}$, $x \in [0, 1]$ define the integrand in (2.29).

- *U_s is a non-negative function;*

- *for* $s \in \langle 0, 1\rangle$, U_s *is an increasing function with*

$$\lim_{x\to 0^+} U_s(x) = 0, \quad \lim_{x\to 1^-} U_s(x) = \infty;$$

- *for* $s \in \langle 1, 2]$, U_s *is a decreasing function with*

$$\lim_{x\to 0^+} U_s(x) = \infty, \quad \lim_{x\to 1^-} U_s(x) = 0;$$

- *for* $s \in \langle 0, 2] \setminus \{1\}$

$$\lim_{x\to 0^+} g_s(x) = 0, \quad \lim_{x\to 1^-} g_s(x) = 0;$$

- *the function* g_s *had a single maximum given by*

$$\max_{x\in[0,1]} g_s(x) = \frac{1}{\mathrm{e}^{|t|^{\frac{s}{s-1}}}}.$$

Proof. See [Dytso et al., 2018, Proposition 9]. ■

Since the moments of the GGD are known for every k (see Proposition 2.1), it would be useful if its characteristic function could be represented as a power series of the form

$$\sum_{k=0}^{\infty} \frac{(it)^k}{k!} \mathbb{E}\left[X^k\right].$$

Next proposition gives the conditions for the analyticity of the GGD's characteristic function.

Proposition 2.15. *Characteristic function* ϕ_s *is*

- *a real analytic function*[†] *for* $t \in \mathbb{R}$ *for* $s > 1$ *and* $|t| < \frac{1}{2}$ *for* $s = 1$*;*
- *not real analytic for* $s < 1$.

Proof. See [Dytso et al., 2018, Appendix G]. ■

Based on the previous proposition, it follows that for $s > 1$ the moment generating function of X_s exists for all $t \in \mathbb{R}$.

The distribution of zeros of the characteristic function is given by the following theorem.

Theorem 2.16. *The characteristic function* ϕ_s *has the following properties:*

- *for* $s > 2$, ϕ_s *has at least one zero (positive to negative crossing) and the number of zeros is at most countable,*
- *for* $s \in \langle 0, 2]$, ϕ_s *is a positive function,*
- *for* $s \in \langle 2, \infty\rangle$ *zeros of* ϕ_s *do not appear periodically,*
- *for* $s = \infty$ *zeros do appear periodically since* $\phi_\infty(t) = \frac{\sin(t)}{t}$.

[†]A real analytic function is infinitely differentiable and can be represented by a convergent power series. For a formal definition, see Rudin [1976].

Proof. See [Dytso et al., 2018, Appendix H]. ■

The asymptotic behaviour of ϕ_s as $t \to \infty$ can be extended to a more general function

$$t \mapsto \mathbb{E}\left[V_{G,s}^m \mathrm{e}^{-\frac{V_{G,s}^2 t^2}{2}}\right]$$

for some $m > 0$ which is shown in the following proposition. It also gives the exact tail behaviour of ϕ_s.

Proposition 2.17. *Let $m \in \mathbb{R}^+$. Then*

$$\lim_{t\to\infty} t^{m+s+1}\, \mathbb{E}\left[V_{G,s}^m \mathrm{e}^{-\frac{V_{G,s}^2 t^2}{2}}\right] = \frac{s}{2}\Gamma\left(\frac{m+s+1}{2} 2^{\frac{m+2s}{2}-\frac{s+1}{s}}\right) := A_m. \tag{2.30}$$

For $s \in \langle 0, 2\rangle$

$$\lim_{t\to\infty} \phi_s(t) t^{s+1} = A_0,$$

where A_0 is defined by (2.30)*. Also, for $q > 0$, $s < 2$ and some $a > 0$*

$$\lim_{t\to\infty} \frac{\phi_s(at)}{\phi_s(t)} = \begin{cases} 0, & q{>}s \\ \frac{1}{a^{q+1}}, & q{=}s \\ \infty, & q{<}s \end{cases}.$$

Proof. See [Dytso et al., 2018, Appendix I]. ■

Proposition 2.18. *For $n \in \mathbb{R}$, $\mathbb{E}\left[V_{G,s}^n\right]$ is finite if and only if $n + s > -1$.*

Proof. See [Dytso et al., 2018, Proposition 13]. ■

According to Propositions 2.1 and 2.9 for $n > -1$, it follows

$$\mathbb{E}\left[V_{G,s}^n\right] = \frac{\mathbb{E}\left[|X_s|^n\right]}{\mathbb{E}\left[|X_2|^n\right]} < \infty.$$

This means that $\mathbb{E}\left[V_{G,s}^n\right]$ is finite even if absolute moments of X_s and X_2 are infinite. For $n \leq -1$ it is unclear if $\mathbb{E}\left[V_{G,s}^n\right]$ is finite since both $\mathbb{E}\left[|X_s|^n\right]$ and $\mathbb{E}\left[|X_2|^n\right]$ are infinite.

Additive decomposition

To determine whether $X_q \sim \mathcal{N}_q\left(0, \left(\frac{s\sigma^2}{2}\right)^{\frac{q}{s}}\right)$ can be decomposed into a sum of independent RVs, definitions and properties of divisibility and decomposability are stated first.

Definition 2.7 (Infinite divisibility).
A characteristic function ϕ is said to be infinitely divisible if for every $n \in \mathbb{R}$ there exists a characteristic function ϕ_n such that

$$\phi(t) = (\phi_n(t))^n. \tag{2.31}$$

Theorem 2.19 (Properties of infinitely divisible distributions).
An infinitely divisible distribution satisfies the following properties

(i) an infinitely divisible characteristic function has no real zeros;

(ii) a symmetric distribution which has a completely monotone PDF on $\langle 0, \infty \rangle$ is infinitely divisible;

(iii) (Lévy-Khinchine canonical representation) the function ϕ is an infinitely divisible characteristic function if and only if it can be written as

$$\ln\left(\phi(t)\right) = ita + \int_{-\infty}^{\infty} \left(\mathrm{e}^{itx} - 1 - \frac{itx}{1+x^2}\right) \frac{1+x^2}{x^2}\,\mathrm{d}L(x), \tag{2.32}$$

where $a \in \mathbb{R}$ and L is a non-decreasing and bounded function such that

$$\lim_{x\to-\infty} L(x) = 0.$$

The function $\mathrm{d}L$ is called a Lévy measure. The integrand is defined for $x = 0$ by continuity to be equal to $-\frac{t^2}{2}$ and the representation is unique;

(iv) a non-degenerate infinitely divisible random variable X has a Gaussian distribution if and only if it satisfies

$$\limsup_{x\to\infty} \frac{-\ln \mathbb{P}\left[|X| \geq x\right]}{x \ln x} = \infty. \tag{2.33}$$

The distribution L is not bounded by one since in general the Lévy measure is not a probability measure.

Theorem 2.20. *The characteristic function ϕ_s of the GGD is infinitely divisible if and only if $s \in \langle 0, 1] \cup \{2\}$.*

Proof. From Corollary 2.6, the PDF is completely monotone on $\langle 0, \infty \rangle$. Since the distribution is symmetric and completely monotone on $\langle 0, \infty \rangle$, from Theorem 2.19 (ii) it follows that ϕ_s is infinitely divisible for $s \in \langle 0, 1]$.

To determine (2.33) from Theorem 2.19 (iv), by using the CDF in expression (2.4), it follows

$$\begin{aligned}
\limsup_{x\to\infty} \frac{-\ln \mathbb{P}\left[|X| \geq x\right]}{x \ln x} &= \limsup_{x\to\infty} \frac{-\ln\left(\frac{\Gamma\left(\frac{1}{s}, \frac{x^s}{2}\right)}{\Gamma\left(\frac{1}{s}\right)}\right)}{x \ln x} \\
\text{(using the limit } \lim_{x\to\infty} \tfrac{\Gamma(a,x)}{x^{a-1}\mathrm{e}^{-x}}\text{)} \quad &= \limsup_{x\to\infty} \frac{-\ln\left(x^{\frac{1}{s}-1}\mathrm{e}^{-\frac{x^s}{2}}\right)}{x \ln x} \\
&= \limsup_{x\to\infty} \frac{x^s}{2x \ln x} = \begin{cases} 0, & s \leq 1 \\ \infty, & s > 1 \end{cases}.
\end{aligned}$$

The GGD is Gaussian only for $s = 2$, hence from Theorem 2.19 (iv) ϕ_s for $s \geq 1$ is infinitely divisible only if $s = 2$. Also, from Theorem 2.16 ϕ_s has at least one zero for $s > 1$, which from 2.19 (i) means it is not infinitely divisible. ∎

A new representation of ϕ_s for $s \in \langle 0, 1]$ can be obtained by showing that the Lévy measure in (2.32) is an absolutely continuous measure which is stated in the following proposition.

Proposition 2.21. *For $s \in \langle 0, 1]$, the Lévy measure is absolutely continuous with density f_L and ϕ_s can be expressed as follows*

$$\phi_s(t) = \exp\left(-\int_{\infty}^{\infty} \left(1 - \cos(tx) \frac{1+x^2}{x^2} \right) f_L(x)\, \mathrm{d}x \right). \tag{2.34}$$

For $x \neq 0$

$$(1+x^2) f_L(x) = -\frac{x}{\pi} \int_0^{\infty} (\ln(\phi_s(t))' \sin(tx)\, \mathrm{d}t. \tag{2.35}$$

Proof. See [Dytso et al., 2018, Appendix J]. ■

If a RV can be decomposed into a sum of two independent RVs where one of the RVs belongs to the same family of distributions, the RV (i.e. its distribution) is called self-decomposable. A formal definition is given below.

Definition 2.8 (Self-decomposable characteristic function).
A characteristic function ϕ is said to be self-decomposable if for every $\alpha \geq 1$ there exists a characteristic function ψ_α such that

$$\phi(\alpha t) = \phi(t)\psi_\alpha(t). \tag{2.36}$$

In the case of the GGD random variable, $X_q \sim \mathcal{N}_q\left(0, \left(\frac{s\sigma^2}{2}\right)^{\frac{q}{s}}\right)$ is self-decomposable if for every $\alpha \geq 1$ there exists a random variable $\widetilde{X_\alpha}$ such that

$$\alpha X_s \overset{d}{=} \widetilde{X_\alpha} + Z_s, \tag{2.37}$$

where $Z_s \sim \mathcal{N}_s(0, 1)$ is independent of $\widetilde{X_\alpha}$. The generalization of self-decomposability is given as

$$\alpha X_q \overset{d}{=} \widetilde{X_\alpha} + Z_s,$$

where again for every $\alpha \geq 1$ there exists a RV $\widetilde{X_\alpha}$ independent of $Z_s \sim \mathcal{N}_s(0, 1)$ and $X_q \sim \mathcal{N}_q(0, 1)$. This existence of a RV $\widetilde{X_\alpha}$ which is independent of Z_s and X_q is equivalent to showing that

$$\phi_{(q,s,\alpha)}(t) = \frac{\phi_q(\alpha t)}{\phi_s(t)}$$

is a valid characteristic function, which requires checking if $\phi_{(q,s,\alpha)}$ is a positive definite function, but the next theorem gives the conditions under which $\phi_{(q,s,\alpha)}$ is a characteristic function.

Theorem 2.22. *For $(s,q) \in \mathbb{R}_+^2$ let*

$$\begin{aligned}\mathbb{S} &= \mathbb{S}_1 \cup \mathbb{S}_2,\\ \mathbb{S}_1 &= \{(s,q)\colon 2 < q < s\},\\ \mathbb{S}_2 &= \{(s,q)\colon q = s \in \langle 0,1] \cup \{2\}\}.\end{aligned}$$

Then the function $\phi_{(q,s,\alpha)}$ has the following properties

- *for $(s,q) \in \mathbb{S}_2$, $\phi_{(q,s,\alpha)}$ is a characteristic function (i.e. X_s is self-decomposable);*
- *for $(s,q) \in \mathbb{R}_+^2 \setminus \mathbb{S}$, $\phi_{(q,s,\alpha)}$ is not a characteristic function for any $\alpha \geq 1$;*
- *for $(s,q) \in \mathbb{S}_1$ and almost all $\alpha \geq 1$, $\phi_{(q,s,\alpha)}$ is not a characteristic function*
- *for $2 < q \leq s < \infty$, $\phi_{(q,s,\alpha)}$ is not a characteristic function for all $\alpha > 1$.*

Proof. See [Dytso et al., 2018, Appendix K]. ■

From the previous theorem it can be concluded that both Gaussian ($s = 2$) and Laplace ($s = 1$) RVs are self-decomposable.

Returning to Proposition 2.9, the question of decomposability for the case $s > 2$ can be answered by using the characteristic function. Since $|X_s| \overset{d}{=} V \cdot |X_q|$ for some positive random variable V independent of $X_q \sim \mathcal{N}_q(0,1)$ and using Proposition 2.1, decomposability reduces to determining whether

$$\phi_{\ln V}(t) = \mathbb{E}\left[V^{it}\right] = \frac{\mathbb{E}\left[|X_s|^{it}\right]}{\mathbb{E}\left[|X_q|^{it}\right]} = \frac{2^{\frac{it}{s}}\Gamma\left(\frac{it+1}{s}\right)\Gamma\left(\frac{1}{q}\right)}{2^{\frac{it}{q}}\Gamma\left(\frac{it+1}{q}\right)\Gamma\left(\frac{1}{s}\right)}, \quad t \in \mathbb{R}$$

is a proper characteristic function, the answer to which is given by the next proposition.

Proposition 2.23. *The function $\phi_{\ln V}$*

- *is not a valid characteristic function for $s > q$, thus the decomposition $X_s \overset{d}{=} V \cdot X_q$ doesn't exist;*
- *is an integrable function for $s < q$. Also, $\phi_{\ln V}$ is a valid characteristic function when the PDF of V is given by*

$$f_V(y) = \frac{1}{2\pi}\frac{\Gamma\left(\frac{1}{q}\right)}{\Gamma\left(\frac{1}{s}\right)}\int_{\mathbb{R}} y^{-it-1}\frac{2^{\frac{it}{s}}\Gamma\left(\frac{it+1}{s}\right)}{2^{\frac{it}{q}}\Gamma\left(\frac{it+1}{q}\right)}\,\mathrm{d}t, y > 0.$$

Proof. See [Dytso et al., 2018, Appendix L]. ■

Entropy

Entropy is a measure of the uncertainty of a random variable [Cover and Thomas, 2005]. Formally, for continuous variables, the definition is given as follows.

Definition 2.9 (Differential entropy).
The (differential) entropy of a continuous random variable X with density f defined on S is given by

$$H(X) = -\int_S f(x) \ln f(x) \, \mathrm{d}x.$$

Generalization of the entropy is given by the Rényi entropy.

Definition 2.10 (Rényi and Shannon entropy).
The Rényi entropy H_r of order r is defined as

$$H_r(X) = \frac{1}{1-r} \ln \left(\int f^r(x \, \mathrm{d}x) \right)$$

for $0 < r < \infty$, $r \neq 1$. For $r \to 1$, the Shannon entropy function is obtained

$$H(X) = H_1(X) = -\int f(x) \ln f(x) \, \mathrm{d}x.$$

The problem of finding the distribution which maximizes entropy the $H(f)$ over all probability densities is stated by the following theorem.

Theorem 2.24 (Maximum entropy distribution).
Let f satisfy the following conditions

(i) $f > 0$, *with equality outside the support set S;*

(ii) $\int_S f(x) \, \mathrm{d}x = 1$;

(iii) $\int_S f(x) r_i(x) \, \mathrm{d}x = \beta_i$ *for* $1 \leq i \leq m$ *("moment constraints").*

Let

$$f^* = f_\lambda(x) = \exp\left(\lambda_0 + \sum_{i=1}^{m} \lambda_i r_i(x) \right), \; x \in S,$$

where $\lambda_0, \ldots, \lambda_m$ are chosen such that f^* satisfies the mentioned conditions. Then f^* uniquely maximizes $H(f)$ over all probability densities f which satisfy the mentioned conditions.

Proof. See [Cover and Thomas, 2005, Theorem 12.1.1] ■

In Lutwak et al. [2004], the GGD distribution was studied in the framework of affine moments to prove some sharp moment-entropy inequalities using a parametrization similar to (2.1). Namely, it is known that the GGD satisfies the maximum entropy principle which is stated here in order to emphasize the importance of this family of distributions to the entropy-based problems (see e.g. Lutwak et al. [2007]).

Theorem 2.25. *Let X be a real random variable with the PDF g such that for some $s > 0$*

$$E\left[|X|^s\right] < \infty,$$

and let H_r be the corresponding Shannon ($r = 1$) or Rényi ($\frac{1}{1+s} < r < 1$) entropy.

(i) If $\sup(y) = \mathbb{R}$ and $r = 1$, then

$$H_1(y) \leq H_1(f_{s,0}) = \frac{1}{s} - \ln\left(\frac{1}{2(s\sigma^2)^{\frac{1}{s}}\Gamma\left(1+\frac{1}{s}\right)}\right),$$

where $f_{s,0}$ is the PDF of the GGD (2.1) for $b = 0$ and

$$H_1(y) = -\int_{\mathbb{R}} y(x)\ln\left(y(x)\right)dx,$$

is the corresponding Shannon entropy. The equality holds if and only if $y = f_{s,0}$ a.s.

(ii) If $\sup(y) = \mathbb{R}$ and $\frac{1}{1+s} < r < 1$, then

$$H_r(y) \leq H_y(f_{s,b}) = \frac{1}{1-r}\ln\left(1+\frac{b}{s}\right) - \ln\left(\frac{bs}{2\sigma^2}\left(\frac{s\sigma^2}{b}\right)^{-\frac{1}{s}}\frac{\Gamma\left(1+\frac{1}{s}+\frac{\sigma^2}{b}\right)}{\Gamma\left(\frac{1}{s}\right)\Gamma\left(\frac{\sigma^2}{b}\right)}\right),$$

where $f_{s,b}$ is the PDF of the GGD (2.1) for $b > 0$ and

$$H_r(y) = \frac{1}{1-r}\ln\left(\int_{\mathbb{R}} y^r(x)\,dx\right),$$

is the corresponding Rényi entropy. The equality holds if and only if $y = f_{s,b}$ a.s.

For more details on the Rényi and Shannon entropy and their relations to the GGD (2.1) and its special cases see Lutwak et al. [2004], Heyde and Leonenko [2005] and Johnson and Vignat [2007]. Also, more details on the light-tailed GGD subfamily can be found in Nadarajah [2005].

2.2 Generalized Gaussian diffusion

As mentioned before, EEG increments are viewed as a time series $(X_n,\, n \in \mathbb{N})$, representing the model for discrete-time observations from the diffusion process $\{X_t\}_{t\geq 0}$ with a given marginal density f which would reflect the empirical data. Recall from Theorem 1.5 that if the PDF (which ideally fits the histogram of the data) is continuous, bounded, and strictly positive on the whole $\mathbb{R}$, a diffusion

can be constructed with that specified marginal density. Since the PDF (2.1) satisfies the conditions of Theorem 1.5, the SDE

$$dX_t = -\theta X_t\, dt + \sqrt{v(X_t)}\ \mathrm{d}B_t, \quad \theta > 0, \quad t \geq 0, \tag{2.38}$$

driven by the standard Brownian motion $\{B_t\}_{t\geq 0}$, where the drift reflects the mean reversion of the process to zero and the diffusion coefficient v is obtained as

$$v(x) = \frac{2\theta}{f(x)} \int_{-\infty}^{x} (-y) f(y)\ \mathrm{d}y$$

admits the unique weak ergodic solution and defines the diffusion with marginal distribution (2.1) which will be called the *Generalized Gaussian diffusion* (GGDiff).

The important properties of the GGDiff are here emphasized as follows:

- If X_0 has the PDF $f_{s,b}$, the GGDiff is a strictly stationary process;
- If $\mathbb{E}[X_t^2] < \infty$, the autocorrelation function of the GGDiff is given by $\mathrm{Corr}(X_s, X_t) = \mathrm{e}^{-\theta|t-s|}$;
- Mixing is a measure of dependence defined by

  $$\alpha(A, B) = \sup|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|,$$

  where

  $$A \in \mathcal{F}_s = \sigma\{X_u,\, u \leq s\}, \quad B \in \mathcal{F}^{s+t} = \sigma\{X_u,\, u \geq s+t\}.$$

  The dependence coefficient is given by $\alpha(s) = \sup_{s\geq 0} \alpha(\mathcal{F}_s, \mathcal{F}^{s+t})$. A process is called a strongly mixing process if $\alpha(s) \to 0$ as $s \to \infty$. According to Doukhan [1994][Proposition 3, p. 115], the GGDiff is then a strong mixing process with exponentially decaying mixing coefficient.

The time series $(X_n,\, n \in \mathbb{N})$ inherits the strong stationarity, autocorrelation structure, and the strong mixing property of the GGDiff (2.38), making it a reasonable model for EEG increments and enabling statistical analysis of the EEG data.

The diffusion (2.38) has the linear drift governed by the autocorrelation parameter θ. The diffusion coefficient is defined as follows. For the light-tailed case

($b = 0$) in (2.1), the diffusion coefficient is given by

$$v(x) = \begin{cases} 2\theta \exp\left(\frac{(-x)^s}{s\sigma^2}\right)\left(\frac{(s\sigma^2)^{2/s}}{s}\Gamma\left(\frac{2}{s}\right) \right. \\ \qquad \left. - \int_x^0 y\mathrm{e}^{-\frac{y^s}{s\sigma^2}}\,\mathrm{d}y\right), \quad x < 0 \\ 2\theta \exp\left(\frac{x^s}{s\sigma^2}\right)\left(\frac{(s\sigma^2)^{2/s}}{s}\Gamma\left(\frac{2}{s}\right) \right. \\ \qquad \left. + \int_0^x y\mathrm{e}^{-\frac{y^s}{s\sigma^2}}\,\mathrm{d}y\right), \quad x \geq 0 \end{cases} \tag{2.39}$$

$$= 2\theta \exp\left(\frac{(x \cdot \operatorname{sgn}(x))^s}{s\sigma^2}\right)\left(\frac{(s\sigma^2)^{2/s}}{s}\Gamma\left(\frac{2}{s}\right) \right.$$
$$\left. + \sigma^{\frac{4}{s}} s^{\frac{2}{s}-1}\left(\gamma\left(\frac{2}{s}, \frac{x^s}{s\sigma^2}\right) - \gamma\left(\frac{2}{s}, \frac{1}{s\sigma^2}\right)\right)\right),$$

where γ is the lower incomplete gamma function. For $b > 0$ the diffusion coefficient takes the following form:

$$v(x) = \begin{cases} 2\theta\left(1 + \frac{b}{s\sigma^2}(-x)^s\right)^{\frac{\sigma^2}{b}+\frac{1}{s}+1}\left(\frac{1}{s}\left(\frac{\sigma^2 s}{b}\right)^{2/s} \right. \\ \quad \cdot B\left(\frac{2}{s}, 1 - \frac{1}{s} + \frac{\sigma^2}{b}\right) \\ \quad \left. - \int_x^0 y\left(1 + \frac{b}{s\sigma^2}y^s\right)^{-\frac{\sigma^2}{b}-\frac{1}{s}-1}\,\mathrm{d}y\right), \qquad x < 0 \\ 2\theta\left(1 + \frac{b}{s\sigma^2}x^s\right)^{\frac{\sigma^2}{b}+\frac{1}{s}+1}\left(\frac{1}{s}\left(\frac{\sigma^2 s}{b}\right)^{2/s} \right. \\ \quad \cdot B\left(\frac{2}{s}, 1 - \frac{1}{s} + \frac{\sigma^2}{b}\right) \\ \quad \left. + \int_0^x y\left(1 + \frac{b}{s\sigma^2}y^s\right)^{-\frac{\sigma^2}{b}-\frac{1}{s}-1}\,\mathrm{d}y\right), \qquad x \geq 0 \end{cases} \tag{2.40}$$

$$= 2\theta\left(1 + \frac{b}{s\sigma^2}\left(x \cdot \operatorname{sgn}(x)\right)^s\right)^{\frac{\sigma^2}{b}+\frac{1}{s}+1}$$
$$\cdot \left(\frac{1}{s}\left(\frac{s\sigma^2}{b}\right)^{\frac{2}{s}} B\left(\frac{2}{s}, 1 - \frac{1}{s} + \frac{\sigma^2}{b}\right) + s^{\frac{2}{s}-1}\left(\frac{\sigma^2}{b}\right)^{\frac{2}{s}} \right.$$
$$\left.\left(\beta\left(\frac{bx^s}{s\sigma^2}; \frac{1}{s}, -\frac{\sigma^2}{b} - \frac{1}{s}\right) - \beta\left(\frac{b}{s\sigma^2}; \frac{1}{s}, -\frac{\sigma^2}{b} - \frac{1}{s}\right)\right)\right),$$

where $B(\cdot, \cdot)$ is the standard beta function and

$$\beta(x; b, s, \sigma) = \int_{\frac{b}{s\sigma^2}}^{\frac{bx^s}{s\sigma^2}} t^{\frac{1}{s}-1}(1+t)^{-\frac{\sigma^2}{b}-\frac{1}{s}-1}\,\mathrm{d}t.$$

The existence of the strong solution of the SDE (2.38) for $b > 0$ and $s = 2$ follows from the analysis of generally parametrized Student diffusion in Leonenko and Šuvak [2010]. However, due to the nature of the diffusion parameter for $b = 0$ and general $b > 0$, in this case existence of a unique strong solution of the GGDiff (2.38) is verified just up to the explosion time $T(X_0)$. In practical application to the EEG data, this explosion time may correspond to the end of coma.

Remark 2.4. *There are many schemes for diffusion discretization, see, for example, Iacus [2009] or Kloeden and Platen [1992] for a detailed exposition on Euler and Milstein schemes for numerical solutions to SDEs. However, the main problem with the time series obtained by these discretization schemes, comprising some form of autoregressive structure, is the lack of the strict stationarity. The detailed exposition on this matter is given in [Ozaki, 1985, p. 55-56], where it is stated that the stationary time series can be obtained for diffusions with linear drift and unit volatility. For general non-linear diffusions the transformation of the diffusion to the diffusion with the unit volatility and the local linearization of the drift is proposed. Furthermore, it is shown that the discretization of this transformed diffusion under some technical assumptions [Ozaki, 1985, Theorems* 3.1 *and* 3.2*] yields the non-explosive and ergodic time-series that converges to the unit-volatility diffusion. For example regarding the diffusion with marginal Student distribution see [Ozaki, 1985, Example* 6*, p.* $69 - 70$*].*

Remark 2.5. *According to [Heyde and Leonenko, 2005, Theorem* 3.2*] there exists the strictly stationary Ornstein-Uhlenbeck type process* $\{X_t\}_{t\in\mathbb{R}}$

$$\begin{aligned} X_t &= \mathrm{e}^{-\lambda t}X_0 + \mathrm{e}^{-\lambda t}\int_0^t \mathrm{e}^{\lambda s}\,\mathrm{d}Y(\lambda s) \\ &= \int_{-\infty}^t \mathrm{e}^{-\lambda(t-s)}\,\mathrm{d}Y(\lambda s), \quad \lambda > 0, \end{aligned}$$

with the marginal Student $T(\nu, \delta, \mu)$ *distribution with the PDF*

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\delta\sqrt{\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1 + \left(\frac{x-\mu}{\delta}\right)^2\right)^{-\frac{\nu+1}{2}},$$

$$\nu > 0, \quad \delta > 0, \quad \mu \in \mathbb{R}, \quad x \in \mathbb{R},$$

governed by the so-called background driving Lévy process[†] $\{Y_t\}_{t\in\mathbb{R}}$*. Due to the self-decomposability of the Student distribution, there also exists the strictly stationary solution of the autoregressive equation*

$$X_n = cX_{n-1} + \varepsilon_n, \quad n \in \mathbb{N}, \tag{2.41}$$

[†]Stochastic process having stationary independent increments and continuos in probability (i.e $\lim_{s\to t} \mathbb{P}\{|X(t) - X(s)| > \varepsilon\} = 0$, $t \geq 0$, $\forall \varepsilon > 0$).

where $c \in \langle 0,1\rangle$ *and* $\{\varepsilon_n\}_{n\in\mathbb{N}}$ *is a sequence of IID random variables (the innovation process) independent of the process* $\{X_n\}_{n\in\mathbb{N}}$*. Also,*

$$X_0 \overset{d}{=} cX_0 + \varepsilon_1 \sim T(\nu,\delta,\mu),$$

and so, if the relation (2.41) *holds for every* $c \in \langle 0,1\rangle$*, it follows that* $T(\nu,\delta,\mu)$ *can be observed as a marginal distribution of the autoregressive time series* $(X_n,\, n \in \mathbb{N}_0)$*. Furthermore, according to [Heyde and Leonenko, 2005, Remark* 3.1*] and [Jurek, 2001, Proposition* 2*],*

$$\varepsilon_1 \overset{d}{=} \int_0^{-\ln(c)} \mathrm{e}^{-s}\,\mathrm{d}Y(s) \overset{d}{=} Y(1),$$

with the distribution having the cumulate transform[†]

$$\begin{aligned}\kappa_{Y(1)}(0) &= 0,\\ \kappa_{Y(1)}(z) &= \ln \mathbb{E}\left[\mathrm{e}^{izY(1)}\right]\\ &= iz\mu - \delta|z|\frac{K_{\nu/2-1}(\delta|z|)}{K_{\nu/2}(\delta|z|)}, \quad z \in \mathbb{R}\setminus\{0\},\end{aligned}$$

where

$$K_\lambda(x) = \frac{1}{2}\int_0^\infty u^{\lambda-1}\exp\left\{\frac{1}{2}x\left(u+\frac{1}{u}\right)\right\}\mathrm{d}u, \quad x>0,$$

is the modified Bessel function of the third kind with the index $\lambda \in \mathbb{R}$*. For fixed* $\lambda > 0$ *this function is positive, decreasing and, as* $x \to 0^+$*,*

$$K_\lambda(x) \sim \Gamma(\lambda)2^{\lambda-1}x^{-\lambda}.$$

Furthermore, the autoregressive time series $(X_n,\, n \in \mathbb{N}_0)$ *has the following important properties:*

- $X_n \sim T(\nu,\delta,\mu)$*,* $\forall n \in \mathbb{N}_0$*,*
- $\mathbb{E}[X_n] = \mu$ *for* $\nu > 1$*,*
- $\mathrm{Corr}\,(X_n, X_{n+\tau}) = c^{|\tau|}$ *for* $\nu > 2$*,* $c \in \langle 0,1\rangle$ *and* $\tau = 0, \pm1, \pm2, \ldots$*.*

Since the heavy-tailed subfamily of the GGD (2.1) *for* $s = 2$ *comes down to the Student-type distribution, in this case the* $AR(1)$ *process constructed in such a way could be used as a time-series model for EEG signals. This* $AR(1)$ *time series is used for modelling the EEG signals in Veretennikova et al. [2018].*
Furthermore, in the light-tailed ($b = 0$*) case of the GGD* (2.1) *for* $s \in \langle 0,1]\cup\{2\}$ *the strictly-stationary* $AR(1)$ *time series with the GGD* (2.1) *marginals could be constructed since for these values of the parameter* s *this distribution is infinitely-divisible and self-decomposable. However, due to lack of these properties for*

[†]Cumulate-generating function is the natural logarithm of the characteristic function

$s \in \langle 1, 2)$ in $b = 0$ case, strictly stationary $AR(1)$ time-series with the GGD (2.1) marginals with arbitrary parameter values cannot be constructed. Therefore, the model presented here, based on a completely different type of the underlying stochastic process, is more general.

2.3 Estimation of the parameters of the GGDiff distribution

In order to estimate the parameters of the GGD, estimation was divided into two parts, since the GGD distribution (2.1) is comprised of two subfamilies with different sets of parameters.

2.3.1 GGD subfamily with $b = 0$: Quasi-likelihood estimation

For the light-tailed case ($b = 0$), the two-dimensional parameter $\zeta = (s, \sigma^2)$ of the marginal distribution of the GGDiff $X = \{X_t\}_{t \geq 0}$ was estimated by the quasi-likelihood method.

For $\Delta > 0$ and a Markovian diffusion X, let

$$p_X(\Delta, x | x_0; \zeta) = \frac{\mathrm{d}}{\mathrm{d}x} P(X_{t+\Delta} \in \mathrm{d}x | X_t = x_0)$$

be the conditional PDF of $X_{t+\Delta}$ given $\{X_t = x_0\}$. Due to the Markovian structure of the GGDiff X, the corresponding likelihood function based on the time series of observations $(X_{\Delta n},\, n \in \mathbb{N})$ is

$$L_n(\zeta) = \prod_{i=1}^{n} p_X\left(\Delta, X_{i\Delta} | X_{(i-1)\Delta}; \zeta\right),$$

and the log-likelihood function is

$$l_n(\zeta) = \sum_{i=1}^{n} \ln\left(p_X\left(\Delta, X_{i\Delta} | X_{(i-1)\Delta}; \zeta\right)\right).$$

The transition density $p_X\left(\Delta, X_{i\Delta} | X_{(i-1)\Delta}; \zeta\right)$ is rarely known in the explicit form, which is also the case here for the GGDiff. However, it is known that the transition density converges to the ergodic density (see [Levin, D., Peres, Y., Wilmer, 2017, Chapter 1]). Therefore, for the purpose of the estimation of parameter ζ, the existing exponentially decaying autocorrelation structure of the diffusion was disregarded and the quasi-likelihood function was simply defined as

$$L_n(\zeta) = \prod_{i=1}^{n} f_{s,b}(X_{\Delta i}),$$

where $f_{s,b}$ is the GGDiff density given by (2.1) for $b = 0$. Then the quasi log-likelihood function is

$$l_n(\zeta) = \sum_{i=1}^{n} \ln \left(\frac{1}{2(s\sigma^2)^{\frac{1}{s}}\Gamma\left(1+\frac{1}{s}\right)} \mathrm{e}^{-\frac{|X_i|^s}{s\sigma^2}} \right), \tag{2.42}$$

with the obvious simplification $\Delta = 1$ in the construction of the time-series of observations. The estimate $\hat{\zeta} = (\hat{s}, \widehat{\sigma^2})$ of the parameter $\zeta = (s, \sigma^2)$ is then obtained by maximising (2.42), which can be performed using existing non-linear optimization methods. For more details on maximum likelihood estimation for diffusion processes see Bishwal [2007] and Chenxu [2013].

Even though the estimation was simplified by disregarding the autocorrelation structure of the diffusion, the method worked well in practice. Examples of the estimated values and their fit to the histograms of EEG increments will be presented in Section 3.1.

2.3.2 GGD subfamily with $b > 0$: Tail index estimation

In the case of the GGD for $b > 0$, there is an additional parameter that needed to be estimated. However, due to the more complicated form of the PDF of the GGD subfamily and the fact that all three parameters that needed to be estimated are arguments of the Γ function, usual estimation methods didn't produce satisfactory results. Still, some of the histograms of EEG increments displayed heavy-tailed distributions which couldn't be properly captured using the light-tailed subfamily of the GGD. Despite the inability to obtain all parameter estimates in the heavy-tailed case, it was possible to estimate just the tail index.

As mentioned in Section 2.1, the asymptotic behaviour of the tail index follows the behaviour of the function $|x|^{-1-s\left(\frac{\sigma^2}{b}+1\right)}$, so the tail index is of the form $\alpha = s\left(\frac{\sigma^2}{b}+1\right)$. There are many methods for tail index estimation, see e.g., Embrechts et al. [1997] for a concise overview. The approach used here was introduced by Grahovac et al. [2015] and is based on the so-called empirical scaling function. The method relies on the graphical inspection since the shape of the scaling function is strongly influenced by the tail index. The method is summarized below.

Let $X_1, X_2, \ldots, X_n$ be a zero mean sample coming from a stationary heavy-tailed sequence with strong mixing property with an exponentially decaying rate. The partition function is a special kind of the sample moment statistic based on consecutive blocks of data. The partition function of the sample is then defined by the following expression:

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

where $q \in \mathbb{R}$ and $1 \leq t \leq n$. As the partition function S_q can be considered an estimator of $\mathbb{E}\left[|X(t)|^q\right]$, it is expected that $\ln(S_q)$ will be linear in $\ln t$. In a simple linear regression of $\ln(S_q)$ on $\ln t$ based on some points $1 \leq t_i \leq n$, $i = 1, \ldots, N$, the slopes with varying q will be called the empirical scaling function. Thus, for a fixed value of q and using $t_i = n^{s_i}$, $s_i \in \langle 0, 1 \rangle$, $i = 1, \ldots, N$, the empirical scaling function is defined as

$$\widehat{\tau}_{N,n}(q) = \frac{\sum_{i=1}^{N} s_i \frac{\ln S_q(n, n^{s_i})}{\ln n} - \frac{1}{N} \sum_{i=1}^{N} s_i \sum_{j=1}^{N} \frac{\ln S_q(n, n^{s_j})}{\ln n}}{\sum_{i=1}^{N} (s_i)^2 - \frac{1}{N} \left(\sum_{i=1}^{N} s_i \right)^2}. \tag{2.43}$$

By repeating this for different values of q, a plot of the empirical scaling function $\hat{\tau}_{N,n}$ is obtained.

Then estimation of the tail index is based on the asymptotic behaviour of the empirical scaling function $\widehat{\tau}_{N,n}$ as $n, N \to \infty$. For each $q > 0$, the limit in probability is

$$\tau_{\alpha}^{\infty}(q) = \begin{cases} \frac{q}{\alpha}, & \text{if } q \leq \alpha \text{ and } \alpha \leq 2, \\ 1, & \text{if } q > \alpha \text{ and } \alpha \leq 2, \\ \frac{q}{2}, & \text{if } 0 < q \leq \alpha \text{ and } \alpha > 2, \\ \frac{q}{2} + \frac{2(\alpha - q)^2 (2\alpha + 4q - 3\alpha q)}{\alpha^3 (2 - q)^2}, & \text{if } q > \alpha \text{ and } \alpha > 2, \end{cases} \tag{2.44}$$

where α is the tail index (see [Grahovac et al., 2015, Theorem 2]).

Clearly, the shape of the empirical scaling function strongly depends on the value of the tail index, and the sample-based empirical counterpart $\widehat{\tau}_{N,n}$ can be used to estimate the index. The asymptotic shape of τ_{α}^{∞} is shown in Figure II.2.1. Heavy-tailed samples are characterized by the (approximately) broken line shape of the empirical scaling function, with the break occurring at α. The limiting case as $\alpha \to \infty$ corresponds to the light-tailed distributions with the straight line scaling function $q/2$ (dashed in Figure II.2.1).

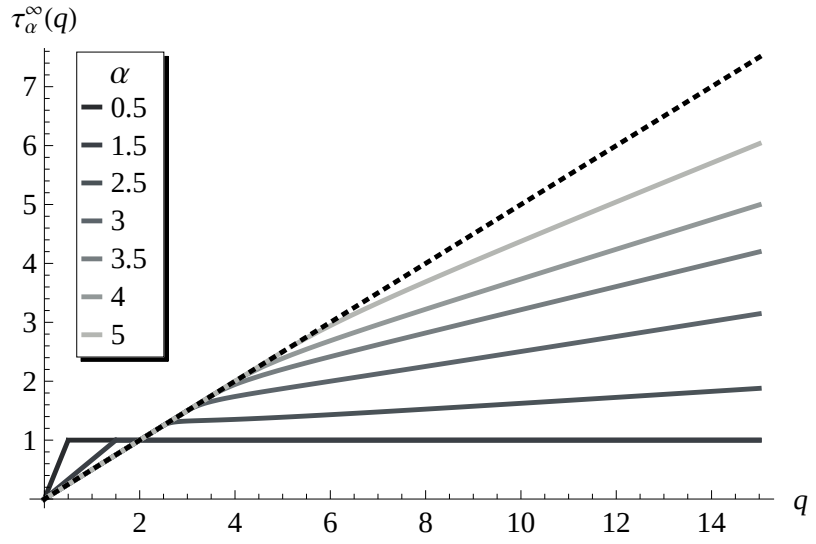


Figure II.2.1 The asymptotic form of the scaling function.

Estimation of the tail index can be done by fitting the empirical scaling function to its asymptotic form. As τ_α^∞ is complex, the estimation is done separately for $\alpha \leq 2$ and $\alpha > 2$ and it is first necessary to graphically determine in which of these two cases the break in the graph occurs. After determining whether $\alpha \leq 2$ or $\alpha > 2$, the tail index estimate is then given by

$$\hat{\alpha} = \underset{\alpha \in \langle 0, \infty \rangle}{\arg\min} \sum_{i=1}^{M} (\widehat{\tau}_{N,n}(q_i) - \tau_\alpha^\infty(q_i))^2. \tag{2.45}$$

for arbitrary points $q_i \in \langle 0, q_{max} \rangle)$, $i = 1, \ldots, M$, where τ_α^∞ takes the form given in (2.44) based on the determined range for α. Examples of the obtained values of the tail index parameter will be shown in Section 3.2.

The important thing to note is that this estimation method does not depend on the particular form of the underlying distribution.

Remark 2.6. *For the special Student case ($s = 2$) of the $b > 0$ GGD* (2.1) *introduced in Remark 2.2, in order to estimate the tail index, using the parametrization of the Student distribution* (2.10)*, it is enough to estimate the parameter σ^2/b. By using the method of moments, the estimator of the parameter σ^2/b is based on the forth empirical moment of the set of discrete observations $(X_1, \ldots, X_n)$:*

$$m_4 = \frac{1}{n} \sum_{k=1}^{n} X_k^4.$$

Then the estimator for σ^2/b is given by

$$\widehat{\left(\frac{\sigma^2}{b}\right)} = \frac{m_4}{m_4 - 3}.$$

For the method of moments estimation of the parameters of the Student distribution see Leonenko and Šuvak [2010].

3 Multimodal case: Mixture of generalized Gaussian distributions

3.1 Mixture of non-central generalized Gaussian distributions

As will be obvious in Section III.1, not all of the histograms of EEG increments were unimodal. Some of the histograms displayed multimodal distributions, with two to six peaks. Hence, there was a need for a more flexible distribution which could capture this behaviour. As a natural extension of the unimodal case where the generalized Gaussian distribution was the best candidate, the distribution which was chosen as a proper fit for a multimodal case was a mixture of three non-central generalized Gaussian distributions which will be called *mixed generalized Gaussian distribution (MixGGD)*. Although some histograms of EEG increments displayed more than three peaks, the mixture distribution here used was comprised of only three parameters for reasons which will be clear in Section 4.1.

The probability density function of the MixGGD is of the following form:

$$f_\zeta(x) = \sum_{k=1}^{3} w_i p_k(x|\zeta) = \sum_{k=1}^{3} \frac{w_k}{2(s_k\sigma_k^2)^{1/s_k}\Gamma\left(1+\frac{1}{s_k}\right)} \mathrm{e}^{-\frac{|x-\mu_k|^{s_k}}{s_k\sigma_k^2}} \tag{3.1}$$

where $\zeta = (\zeta_1, \zeta_2, \zeta_3)$, $\zeta_k = (\mu_k, s_k, \sigma_k^2, w_k)$ are its parameters. Note that $\mu_k \in \mathbb{R}$, $s_k > 0$, $\sigma_k > 0$ and $0 \leq w_k \leq 1$ such that $\sum_{k=1}^{3} w_k = 1$.

For the specific values of some parameters, the MixGGD family contains the following well-known and widely applied distributions:

- for $s_1 = s_2 = s_3 = 1$ the MixGGD becomes the mixture of three Laplace distributions,
- for $s_1 = s_2 = s_3 = 2$ the MixGGD becomes the mixture of three Gaussian distributions,
- for $s_1 = 1$ and $s_2 = s_3 = 2$ the MixGGD becomes the mixture of one Laplace and two Gaussian distributions (similarly for other combinations of values 1 and 2 of parameters s_1, s_2, s_3),
- for $s_1 = 1$ and $s_2 = s_3 > 0$ the MixGGD becomes the mixture of one Laplace and two generalized Gaussian distributions (similarly for other combinations of same values of parameters s_1, s_2, s_3),
- for $s_1 = 2$ and $s_2 = s_3 > 0$ the MixGGD becomes the mixture of one Gaussian and two generalized Gaussian distributions (similarly for other combinations of same values of parameters s_1, s_2, s_3).

Different combinations of s_k and w_k with fixed values of μ_k and σ_k^2 for a two- and three-component MixGGD are shown in Figures II.3.2 and II.3.3. For the sake of consistency, the leftmost peak is related to the first component in the mixture ($k = 1$), the middle peak to the second component ($k = 2$) and the rightmost peak to the third component ($k = 3$) of the MixGGD. Observe that a three-component MixGGD becomes a two-component MixGGD when the weight of the middle peak is set to 0, i.e. $w_2 = 0$ and by setting $w_1 = w_3 = 0$ a three-component MixGGD is reduced to a (unimodal) generalized Gaussian distribution. Hence, by appropriately setting the weights, the MixGGD can cover cases of histograms where both one, two and three peaks are observed.

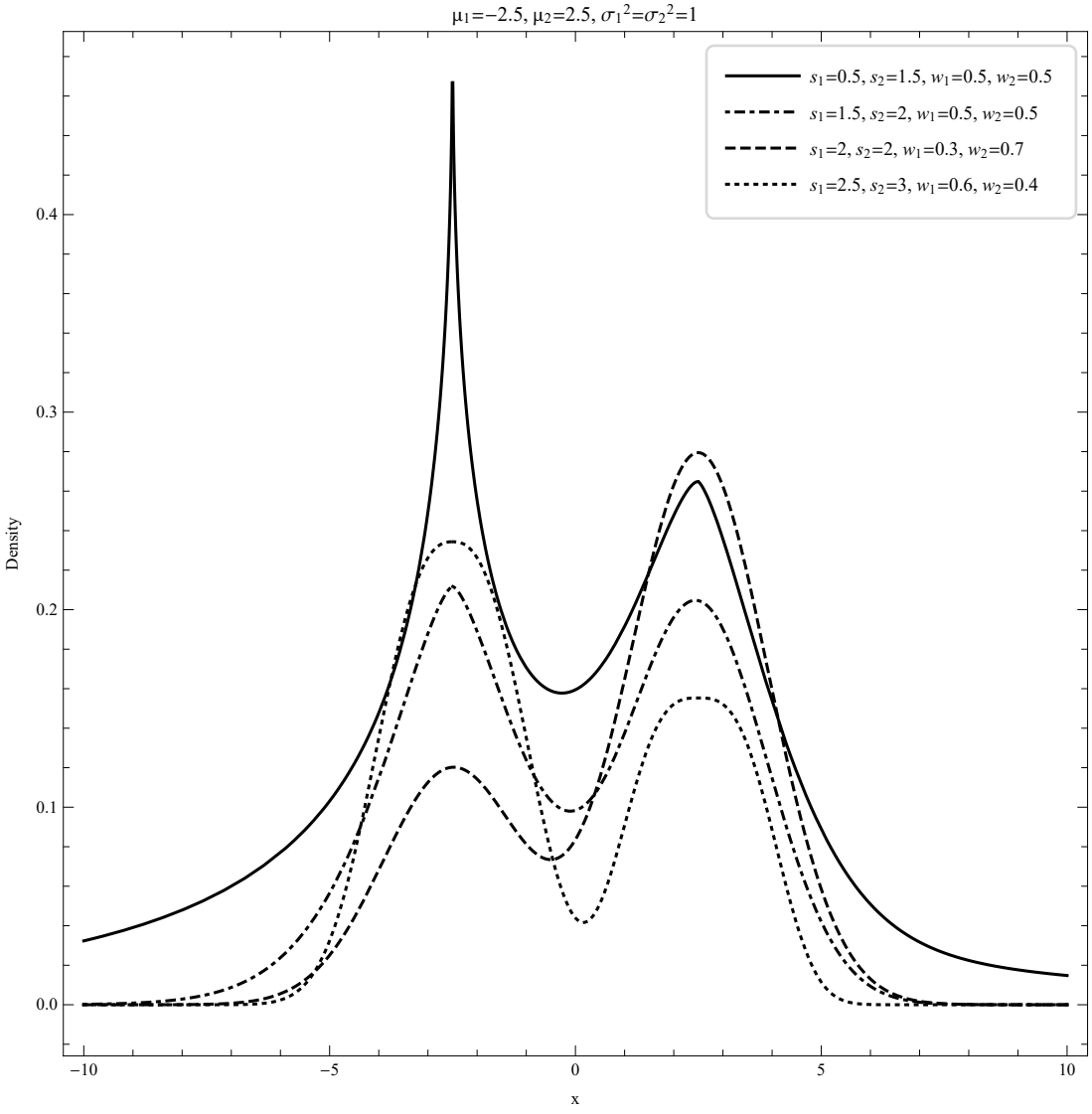


Figure II.3.2 The two-component MixGGD with different combinations of s_k and w_k.

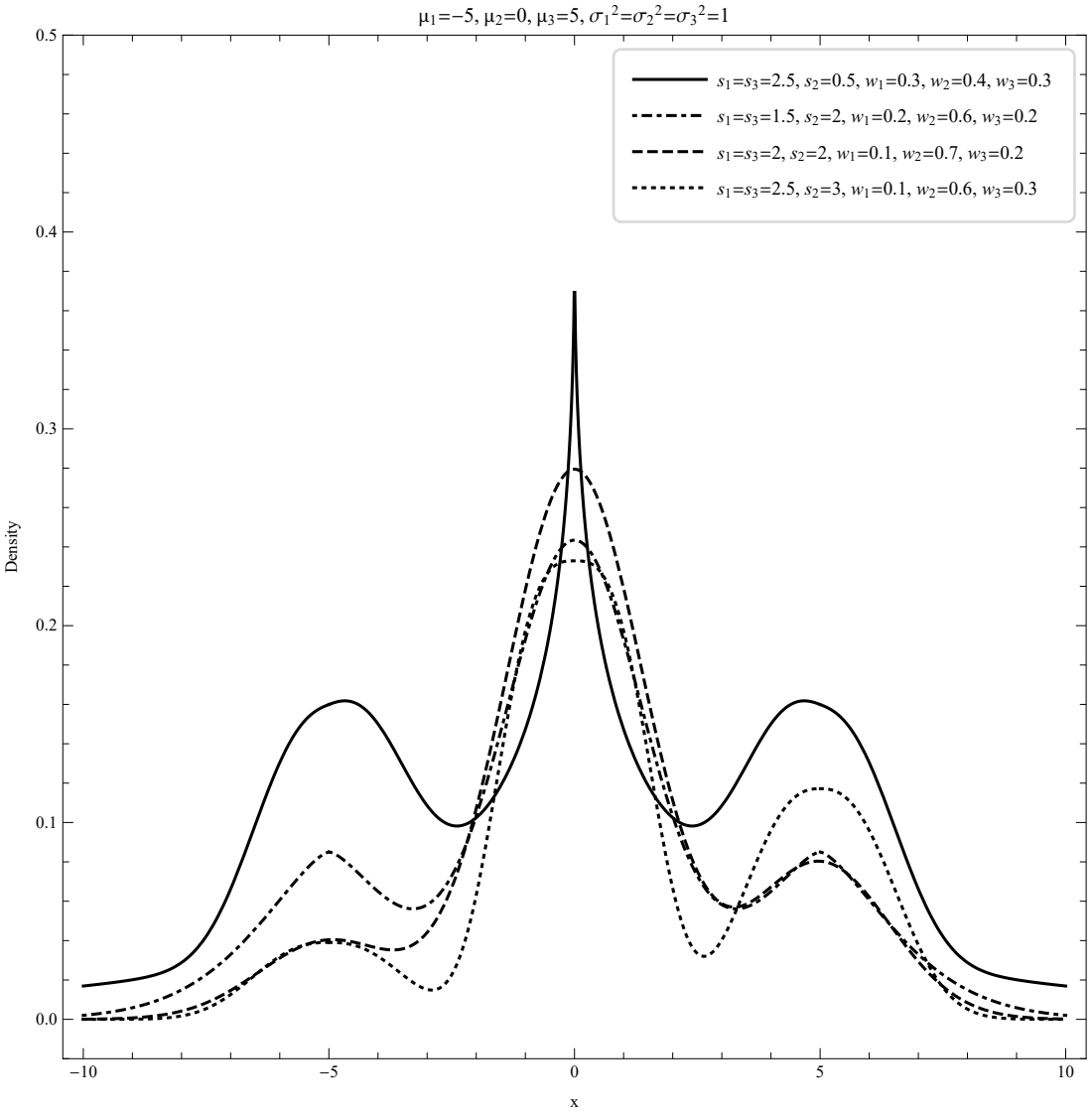


Figure II.3.3 The tree-component MixGGD with different combinations of s_k and w_k.

According to the characteristic function form of the generalized Gaussian distribution given in Pogány and Nadarajah [2010], the characteristic function of the MixGGD (3.1) is given by

$$\varphi_\zeta(t) = \sum_{k=1}^{3} w_k \frac{\sqrt{\pi} \mathrm{e}^{it\mu_k}}{\Gamma\left(\frac{1}{s_k}\right)} \sum_{j=0}^{\infty} \frac{1}{j!} \frac{\Gamma\left(\frac{1}{s_k} + \frac{2j}{s_k}\right)}{\Gamma\left(\frac{1}{2} + j\right)} \left(\frac{-\left(s_k \sigma_k^2\right)^{-1/s_k}}{4} \right)^j. \tag{3.2}$$

Further, some important properties related to moments of the MixGGD are mentioned. All moments exist, without any restrictions on parameter values. The integer moment of order $n \in \mathbb{N}$ is given by

$$\mathbb{E}[X^n] = \sum_{k=1}^{3} w_k \mu_k^n \left(\sum_{i=0}^{n} \binom{n}{i} \left(\frac{\left(s_k \sigma_k^2\right)^{1/s_k}}{\mu_k} \right)^i \frac{\left(1 + (-1)^i\right) \Gamma\left(\frac{i+1}{s_k}\right)}{2\Gamma\left(\frac{1}{s_k}\right)} \right). \tag{3.3}$$

In particular, the first five moments are

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{k=1}^{3} w_k \mu_k =: \mu, \\
\mathbb{E}[X^2] &= \sum_{k=1}^{3} w_k \left(\mu_k^2 + \frac{s_k^{2/s_k} \sigma_k^{4/s_k} \Gamma\left(\frac{3}{s_k}\right)}{\Gamma\left(\frac{1}{s_k}\right)} \right), \\
\mathbb{E}[X^3] &= \sum_{k=1}^{3} w_k \left(\mu_k^3 + \frac{3\mu_k s_k^{2/s_k} \sigma_k^{4/s_k} \Gamma\left(\frac{3}{s_k}\right)}{\Gamma\left(\frac{1}{s_k}\right)} \right), \\
\mathbb{E}[X^4] &= \sum_{k=1}^{3} w_k \left(\mu_k^4 + \frac{6\mu_k^2 s_k^{2/s_k} \sigma_k^{4/s_k} \Gamma\left(\frac{3}{s_k}\right)}{\Gamma\left(\frac{1}{s_k}\right)} + \frac{s_k^{4/s_k} \sigma^{8/s_k} \Gamma\left(\frac{5}{s_k}\right)}{\Gamma\left(\frac{1}{s_k}\right)} \right), \\
\mathbb{E}[X^5] &= \sum_{k=1}^{3} w_k \left(\mu_k^5 + \frac{10\mu_k^3 s_k^{2/s_k} \sigma^{4/s_k} \Gamma\left(\frac{3}{s_k}\right)}{\Gamma\left(\frac{1}{s_k}\right)} + \frac{5\mu_k s_k^{4/s_k} \sigma^{8/s_k} \Gamma\left(\frac{5}{s_k}\right)}{\Gamma\left(\frac{1}{s_k}\right)} \right),
\end{aligned}$$

while the central moments are given by

$$\begin{aligned}
\mathbb{E}[(X - \mathbb{E}[X])^n] = \sum_{k=1}^{3} w_k \Bigg(\sum_{i=0}^{n} \binom{n}{i} (\mu_k - \mu)^{n-i} \left(s_k \sigma_k^2\right)^{i/s_k} \\
\cdot \left(1 + (-1)^i\right) \frac{\Gamma\left(\frac{i+1}{s_k}\right)}{2\Gamma\left(\frac{1}{s_k}\right)} \Bigg).
\end{aligned} \tag{3.4}$$

The skewness and kurtosis are given by

$$\mathfrak{s} = \frac{\sum_{k=1}^{3} w_k \left((\mu_k - \mu)^3 + 3(\mu_k - \mu) \frac{\left(s_k \sigma_k^2\right)^{2/s_k} \Gamma\left(\frac{3}{s_k}\right)}{\Gamma\left(\frac{1}{s_k}\right)} \right)}{\left(\sum_{k=1}^{3} w_k \left((\mu_k - \mu)^2 + \frac{\left(s_k \sigma_k^2\right)^{2/s_k} \Gamma\left(\frac{3}{s_k}\right)}{\Gamma\left(\frac{1}{s_k}\right)} \right) \right)^{3/2}},$$

$$\kappa = \frac{\sum_{k=1}^{3} w_k \left((\mu_k - \mu)^3 + 6(\mu_k - \mu)^2 \frac{\left(s_k \sigma_k^2\right)^{2/s_k} \Gamma\left(\frac{3}{s_k}\right)}{\Gamma\left(\frac{1}{s_k}\right)} + \frac{\left(s_k \sigma_k^2\right)^{4/s_k} \Gamma\left(\frac{5}{s_k}\right)}{\Gamma\left(\frac{1}{s_k}\right)} \right)}{\left(\sum_{k=1}^{3} w_k \left((\mu_k - \mu)^2 + \frac{\left(s_k \sigma_k^2\right)^{2/s_k} \Gamma\left(\frac{3}{s_k}\right)}{\Gamma\left(\frac{1}{s_k}\right)} \right) \right)^{2}}.$$

Some properties of a similar parametrization of MixGGD, where $\left(s_k \sigma_k^2\right)^{1/s_k}$ is treated as one parameter, and its relation to the multivariate generalized error

distribution are studied in Wen et al. [2022]. In particular, Wen et al. [2022] consider the MixGGD with two components and present the explicit expressions for the distribution function, hazard rate function, characteristic function and some numerical characteristics (moments, central moments, skewness and kurtosis) as well as the study of moment estimation, maximum likelihood estimation and the Expectation/Conditional Maximization ECM algorithm for estimation of parameters of this special case of MixGGD.

3.2 Mixed Generalized Gaussian diffusion

Similarly to the unimodal case, important probabilistic properties of EEG increments will be described by fitting these discrete-time observations to the continuous-time diffusion process $X = \{X_t\}_{t\geq 0}$ with the marginal PDF f_ζ given by (3.1).

Since the PDF (3.1) is continuous, bounded, strictly positive on $\mathbb{R}$, has expectation $\mu = \sum_{k=1}^{3} w_k\mu_k$ and finite variance, according to Theorem 1.5, the stochastic differential equation

$$dX_t = -\theta\,(X_t - \mu)\,\mathrm{d}t + \sqrt{v(X_t)}\,\mathrm{d}B_t, \qquad t \geq 0,$$

has a unique Markovian weak solution $X = \{X_t\}_{t\geq 0}$ that is ergodic with the invariant density (3.1), where

- $\{B_t\}_{t\geq 0}$ is the driving Brownian motion independent of X_0,
- $\theta > 0$ is the parameter describing the speed of reversion of diffusion X towards expectation μ,
- diffusion coefficient v is given by

$$\begin{aligned} v(x) &= \frac{2\theta}{f_\zeta(x)} \int_{-\infty}^{x} (\mu - y) f_\zeta(y)\, dy \\ &= 2\theta \left(\sum_{k=1}^{3} \frac{w_k}{c_k} \mathrm{e}^{-\frac{|x-\mu_k|^{s_k}}{s_k\sigma_k^2}} \right)^{-1} \left(\sum_{k=1}^{3} c_k \int_{-\infty}^{x} (\mu - y) \mathrm{e}^{-\frac{|y-\mu_k|^{s_k}}{s_k\sigma_k^2}}\, dy \right) \end{aligned} \tag{3.5}$$

  and it is strictly positive, i.e. $v(x) > 0$ for all $x \in \mathbb{R}$.

The diffusion X given by the SDE will be called the *mixed generalized Gaussian diffusion (MixGGDiff)*. According to Theorem 1.5, some important properties of MixGGDiff are:

- The function $g = f_\zeta v$ satisfies

$$\int_{-\infty}^{\infty} g(x)\, dx < \infty$$

  and

$$\mathbb{E}[X_{s+t}|X_s = x] = x\mathrm{e}^{-\theta t} + \mu(1 - \mathrm{e}^{-\theta t});$$

- If $X_0 \sim f_\zeta$, MixGGDiff is a strictly stationary process with $\mathbb{E}[X_t] = \mu$ for all $t \geq 0$;
- The autocorrelation function of MixGGDiff is given by $\mathrm{Corr}(X_{s+t}, X_s) = \mathrm{e}^{-\theta t}$, $s, t \geq 0$.
- Since the drift coefficient is linear, it satisfies both Lipschitz and the linear growth conditions. However, non-differentiability of the diffusion coefficient v implies that the existence of a unique strong solution could be verified just up to the explosion time $T(X_0)$, similarly as in the unimodal case for $b = 0$. In practical application to the EEG data, this explosion time $T(X_0)$ can be interpreted as the end of coma.

Remark 3.1. *Other mixture distributions are also used in neurology, e.g. Gaussian mixture is used for estimating and comparing the shapes of distributions of neuroimaging data related to aging effects in brain white matter Kim et al. [2014]. Mixture diffusion processes are extensively used in finance, since due to their multimodality and heavier tails they often better describe the properties of financial data. For example, Alexander and Narayanan [2001] studies the option pricing under Gaussian mixture distributed returns. Furthermore, in Brigo and Mercurio [2002*b*] and Brigo and Mercurio [2002*a*] diffusions with mixture of log-normal densities are used for modeling of market smile phenomenon, while in Brigo [2008] the overview of general properties of mixture diffusion SDEs under assumption of existence of strong solution and constant starting value is given.*

3.3 Estimation of the parameters of the MixGGDiff distribution

Due to the added shape parameter s_k, GGD is more flexible and can approximate a large class of statistical distributions. However, it also implies there is an additional parameter to estimate, compared to a Gaussian distribution. Thus, a K-component MixGGD requires the estimation of $4K$ parameters.

Method of estimation of $\zeta_k = (\mu_k, s_k, \sigma_k^2, w_k)$ in MixGGD (3.1) was adopted from Mohamed and Jaïdane-Saïdane [2009] and is based on the expectation maximization (EM) algorithm first proposed by Dempster et al. [1977]. EM algorithm is comprised of two steps: the step to find the expectation (E-step) and the maximization step (M-step). The M-step can be computationally complicated because maximum likelihood estimation requires multidimensional numerical iteration. Thus, Meng and Rubin [1993] proposed a special class of generalized EM algorithms, called Expectation/Conditional Maximization (ECM) algorithm which replaces a complicated M-step with several computationally simpler conditional maximization steps. The algorithm maximizes the complete log-likelihood

function

$$l(\zeta) = \sum_{k=1}^{K} \sum_{i=1}^{N} h_{k,i} \ln \left(w_k p_k(x_i|\mu_k, s_k, \sigma_k^2) \right), \tag{3.6}$$

where

$$h_{k,i} = p(k|x_i) = \frac{w_k p(x_i|\zeta_k)}{\sum_{j=1}^{K} w_j p(x_i|\zeta_j)}$$

represents the conditional expectation of p_k given the observation x_i, i.e. the posterior probability that x_i belongs to the k^{th} component.

Steps of the algorithm are as follows:

- *initialization* of the model parameter ζ;
- *expectation (E-step)* where the conditional probability $h_{k,i}$ is calculated using

$$h_{k,i}^{(n+1)} = \frac{w_k^{(n)} p(x_i|\mu_k^{(n)}, s_k^{(n)}, \sigma_k^{2(n)})}{\sum_{j=1}^{K} w_j^{(n)} p(x_i|\mu_j^{(n)}, s_j^{(n)}, \sigma_j^{2(n)})} \tag{3.7}$$

  and the computation of $l(\zeta)$ is calculated based on the estimate $\zeta^{(n)}$;
- *maximization (M-step)* deals with the numerical maximization of the log-likelihood function (3.6) and the parameter estimate at iteration $(n+1)$ is estimated as

$$\widehat{\zeta}^{(n+1)} = \arg\max_{\zeta} l(\zeta^{(n)})$$

Parameters w_k, μ_k, σ_k^2 are estimated using a set of iterative equations

$$\widehat{w_k}^{(n+1)} = \frac{1}{N} \sum_{i=1}^{N} h_{k,i}^{(n)},$$

$$\widehat{\mu_k}^{(n+1)} = \frac{\sum_{i=1}^{N} h_{k,i}^{(n)} x_i}{\sum_{i=1}^{N} h_{k,i}^{(n)}},$$

$$\widehat{\sigma_k^2}^{(n+1)} = \frac{\sum_{i=1}^{N} h_{k,i}^{(n)} \left(x_i - \mu_k^{(n)} \right)^2}{\sum_{i=1}^{N} h_{k,i}^{(n)}}.$$

For the added shape parameters s_k, the approach based on the use of higher order statistics from Mohamed and Jaïdane-Saïdane [2009] was used. Specifically, values of the kurtosis for each mixture component can be used to derive an approximation of shape parameters s_k. Tesei and Regazzoni [1998] define an analytical relationship between the shape parameter s_k and the kurtosis κ_k as

$$\kappa_k = \frac{\Gamma\left(\frac{5}{s_k}\right) \Gamma\left(\frac{1}{s_k}\right)}{\left(\Gamma\left(\frac{3}{s_k}\right)\right)^2}. \tag{3.8}$$

It is impossible to express s_k in terms of κ_k due to the Γ function definition, but an approximation can be found by applying the least squared method on a

generic second-order monotonic analytical expression

$$\kappa_k = \frac{1.865 s_k^2 + \alpha_1 s_k + \alpha_2}{s_k^2 + \delta_1 s_k + \delta_2}$$

which gives

$$\kappa_k \approx \frac{5 + 1.865(s_k + 0.12)^2}{(s_k + 0.12)^2}. \tag{3.9}$$

Kurtosis in iteration $(n+1)$ is estimated using the iterative equation

$$\kappa_k^{(n+1)} = \frac{\sum_{i=1}^{N} h_{k,i}^{(n)} \left(x_i - \mu_k^{(n)}\right)^2}{\left(\sigma_k^{(n)}\right)^4 \sum_{i=1}^{N} h_{k,i}^{(n)}}$$

with the same weights as for $\mu_k^{(n+1)}$ and $\sigma_k^{2(n+1)}$. Shape parameter s_k in the iteration $(n+1)$ is then approximated by inverting the monotonic expression (3.9) resulting in

$$\widehat{s_k}^{(n+1)} \approx \sqrt{\frac{5}{\kappa_k^{(n)} - 1.865}} - 0.12 \tag{3.10}$$

which is defined for $\kappa_k > 1.865$ and gives a good approximation of shape parameter s_k as a function of κ_k when κ_k lies in the range of $\langle 1.865, 30]$ [Tesei and Regazzoni, 1996], which corresponds to s_k range of $[0.302, \infty\rangle$. For s in range $\langle 0.3, 10]$, comparison of the true value of the kurtosis and the approximated value of the kurtosis as a function of the shape parameter s based on (3.9) is shown in Figure II.3.4

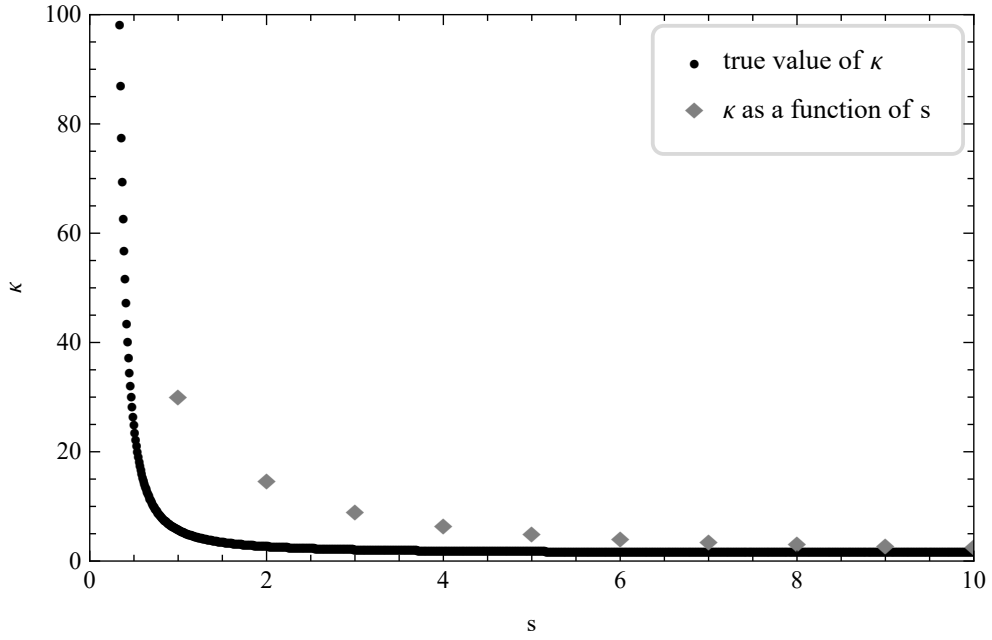


Figure II.3.4 True value of the kurtosis and the approximated value of the kurtosis as a function of the shape parameter s.

Remark 3.2. *Another approach in estimating the shape parameters s_k based on the numerical optimization of the log-likelihood function can be found in Mohamed and Jaïdane-Saïdane [2009]. However, the approach is much more complex due to the system of equations being strongly nonlinear, sensitivity of initial parameter values and computational time.*

Chapter III

Prediction of neurodevelopment using EEG increments

To illustrate an application of the constructed diffusion processes presented in Chapter II to real data and test whether the information obtained from parameters of their marginal distributions can contribute to the prediction of neurodevelopment in children who were in a coma due to cerebral malaria, analysis of EEG data collected during an observational study was performed. Information about the data collection and variables included in the study is described in Section III.1. Additionally, Section III.1 also includes the summary of analysis of shapes of histograms of EEG increments, showing examples of histograms with different number of peaks. The method used in the identification of important predictors for neurodevelopment is described in Section III.2. Section III.3 includes examples of estimating the parameters of the (unimodal) GGD and the tail-index and their contribution to prediction of neurodevelopment, while Section III.4 presents the same for the MixGGD.

1 Description and analysis of the dataset

Focus of the analysis were the increments of EEG signal, but in addition to EEG signals, dataset also included non-EEG data which is known to be of predictive value for neurodevelopment. Hence, the question was whether the use of parameters from stochastic models can improve the prediction, i.e. to determine their usefulness over and above socio-demographic characteristics, clinical factors, and child neurodevelopment or cognitive score at the time of hospitalization for cerebral malaria after coma.

1.1 Description of the dataset

Data used in the analysis were collected during the observational study of the pathogenesis of severe malaria (cerebral malaria (CM) and severe malarial anaemia) in surviving children. Data collection was performed between 2008 and 2015 at Mulago National Referral and Teaching Hospital which is a public tertiary care facility in Kampala, Uganda. The observational study was approved by the Institutional Review Boards of the Makerere University School of Medicine and the University of Minnesota. The sample consisted of 78 children between 18 months and 12 years of age. The children were recruited for projects studying CM pathogenesis and clinical epidemiology and were admitted with a clinical diagnosis of CM with a Blantyre coma score ≤ 2, *P. falciparum* present on the peripheral blood smear and no other known cause of coma [Postels et al., 2018]. As cerebral malaria results in a coma, EEG signals were recorded during coma for the children who were diagnosed with cerebral malaria. Recording of the EEG was performed within 12 hours of admission using the International $10-20$ system, meaning that the cap placed on the scalp for recording of the signal consisted of 19 electrodes called *channels* (see Figure III.1.1). Sampling rate of the recordings was 500 Hz and the average record duration was 30 minutes. In the case of a patient becoming medically unstable, the recording of EEG was terminated before this time mark. Since the Cz electrode is placed at the centre of midline sagittal plane, it was chosen as the reference electrode. Data from some channels for some children contained a substantial number of zero observations, potentially due to a poor connection between the electrode and the scalp. This consideration was included in the interpretation of the results. Paediatric neurologist performed the research interpretations after the hospital discharge. Artefacts due to breathing, muscle movement and heartbeat had already been removed from this dataset prior to the statistical analysis. The study also included community control children who were recruited from household compound area of children with cerebral malaria or severe malarial anaemia to control for socioeconomic variables that affect neurodevelopment and cognition [Bangirana et al., 2016].

In order to asses the effects of coma, age-appropriate assessments of neurodevelopment and cognition were performed using the Mullen Scales of Early Learning [Shank, 2011] for children 5 years of age or younger. For children over the age of five, cognitive assessments was performed using Kaufman Assessment Battery for Children, second edition [Kaufman, 2004]. In the absence of Ugandan norms, all measures were standardized using the available United States norms. A single measure of neurodevelopment and cognition regardless of age was then obtained by computing the z-scores using the mean and standard deviation of the community control children, as discussed in Veretennikova et al. [2018]. For all children, assessment was performed at three points in time - after the dis-

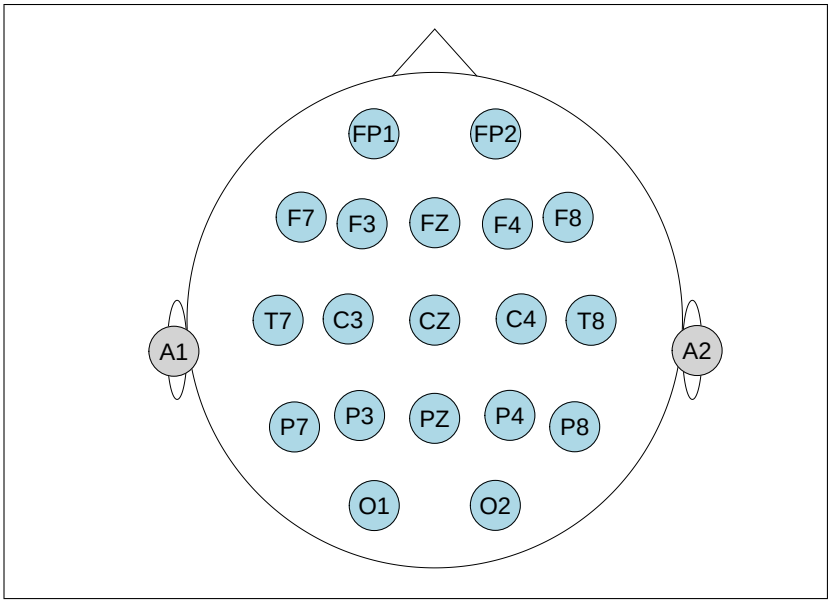


Figure III.1.1 The International $10-20$ system for EEG signal recording.

charge from the hospital (for community control children, this was the point of enrolment), 6 months after the discharge (or enrolment) and 12 months after the discharge (or enrolment). In addition to performing the neurological exams, children were also classified as either "alive with sequelae" or "normal" based on the parents' assessment of their child's motor, cognitive or behavioural abnormalities.

Other non-EEG data that were collected in the study included demographic and anthropometric characteristics (age, sex, height-for-age and weight-for-age-z-score using The World Health Organization reference norms [WHO, 2009]). Socioeconomic status was measured using a checklist of material possessions, housing quality, cooking resources and water accessibility. Quality and quantity of stimulation to which the child is exposed in the home environment was assessed using Home Observation for the Measurement of the Environment (HOME) measure [Bradley and Caldwell, 1979], where higher HOME scores indicate higher quality of home environment. Severity of coma caused by cerebral malaria was assessed using the Blantyre coma scale [Taylor, 2009]. At the point of hospitalization of children with cerebral malaria, biomarker panels from plasma and cerebrospinal fluid were collected. Preprocessing of the data included location and scale transformation for all potential predictors included in feature matrices.

More details about the data collection can be found in Postels et al. [2018] and Veretennikova et al. [2018].

1.2 Analysis of histograms of EEG increments

Data analysis was performed using `R` for Windows. The software was regularly updated to reflect the changes in the packages and functions which happen with

the release of a new version of the software. For the initial part of the analysis, `R` 3.6.3 version was used. Histograms of the increments of EEG signals were investigated for every child and every channel separately, creating a child-channel pair, resulting in a total of 1482 histograms (78 children times 19 channels). Visual inspection of the histograms of EEG increments for each child-channel pair was performed separately to try to classify the shape of the underlying distribution. Classification of the child-channel pairs was performed based on the number of peaks observed on the histograms of the increments of EEG signal and it showed that the channels displayed from one to a maximum of 6 peaks. The frequency of the observed number of peaks is shown in Table III.1.1. As it can be concluded from the Table III.1.1, approximately 95% of the channels consisted of a maximum of three peaks, with the majority of them displaying the unimodal distribution on histograms. Examples of histograms classified as "1-peak" can be seen in Figure III.1.2. Although some histogram look quite similar, tail index estimation suggested that they can differ in their tail behaviour, some displaying heavy tails, while others were light-tailed. Behaviour of the tail will be evident in Section 3.2. Additionally, last figure in Figure III.1.2 shows an example of an almost degenerate distribution, with a single narrow peak. The analysis of the data on such peaks showed that this is due to a large number of zero observations in EEG signals, and *not* increments. As mentioned before, this can happen when there is a poor connection between the electrode and the scalp. However, these channels were also included in the analysis and prediction of neurodevelopment since they model a situation that happens in practice when doing electroencephalography.

Table III.1.1 Classification of EEG channels based on the shape of histograms of increments.

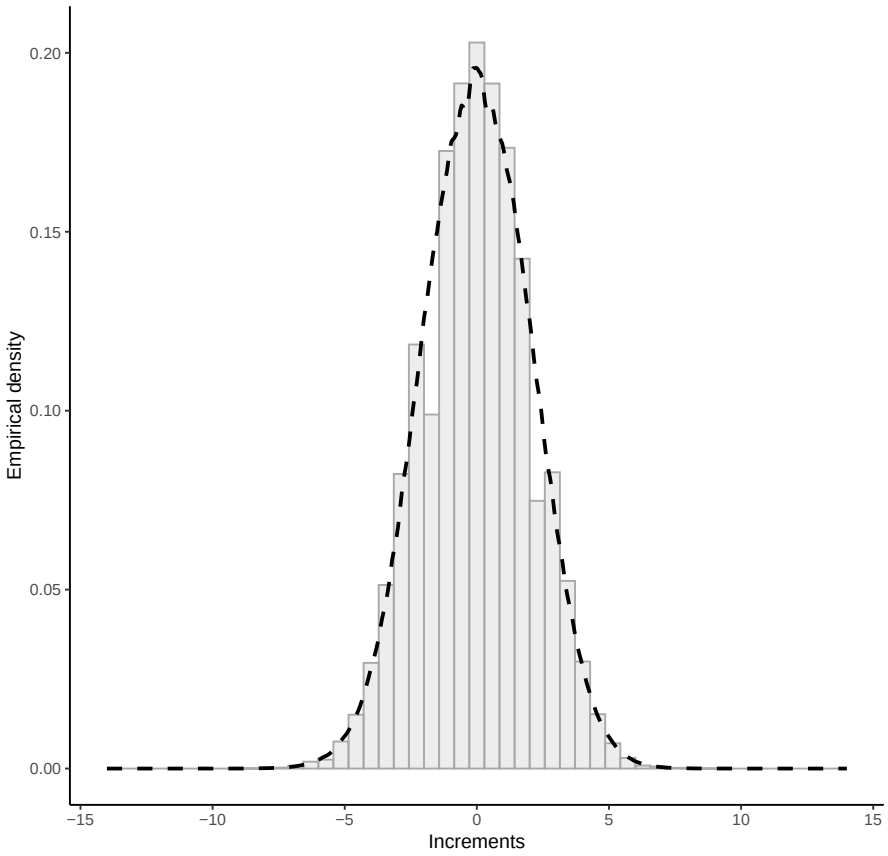
Histogram shape	Number of channels	% of channels
1-peak	1100	74.22
2-peak	144	9.72
3-peak	169	11.40
4-peak	25	1.69
5-peak	40	2.70
6-peak	4	0.27
total	1482	100

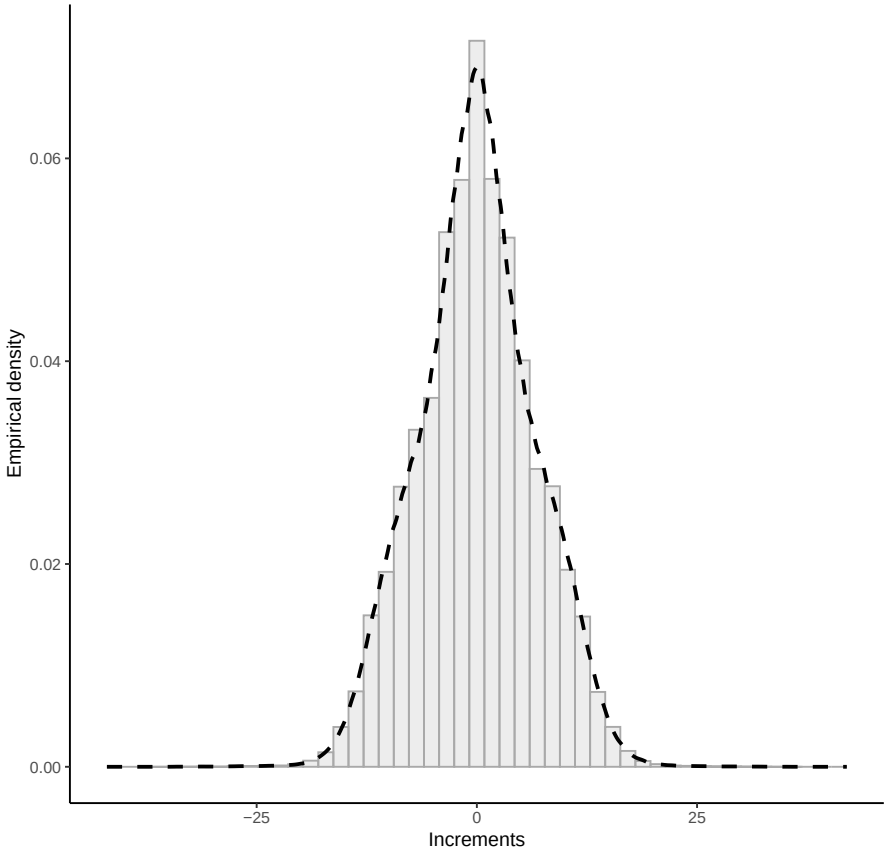
Histograms of channels that displayed a 2-peak distribution comprised approximately 10% of the channels. The examples of histograms displaying a 2-peak

distribution are shown in Figure III.1.3.

Three-peak channels were the most frequent after unimodal channels and they comprised approximately 11% of the channels. Some examples of histograms of EEG increments displaying a 3-peak distribution can be seen in Figure III.1.4 Least represented were 4-, 5- and 6-peak channels comprising less than 5% of the channels. As it will be seen in Section 4.1, this was the reason to use a mixture of generalized Gaussian distributions consisting of only up to three components. Examples of 4-, 5- and 6-peak histograms can be seen in Figures III.1.5 - III.1.7

This visual analysis of the histograms of EEG increments showed that the suitable marginal distributions for the diffusion process which would model the EEG increments were either unimodal distribution with a flexible shape parameter, including a possibility of heavy tails, or a multimodal distribution with the flexibility of the shape of each components. Now it is clear why the chosen marginal distributions of the diffusion process were the (unimodal) GGD presented in Section II.2 with the parametrization (2.1) and the MixGGD presented in Section II.3.


0.20
0.15
0.10
0.05
0.00
Empirical density
−15
−10
−5
0
5
10
15
Increments


0.06
0.04
0.02
0.00
Empirical density
−25
0
25
Increments

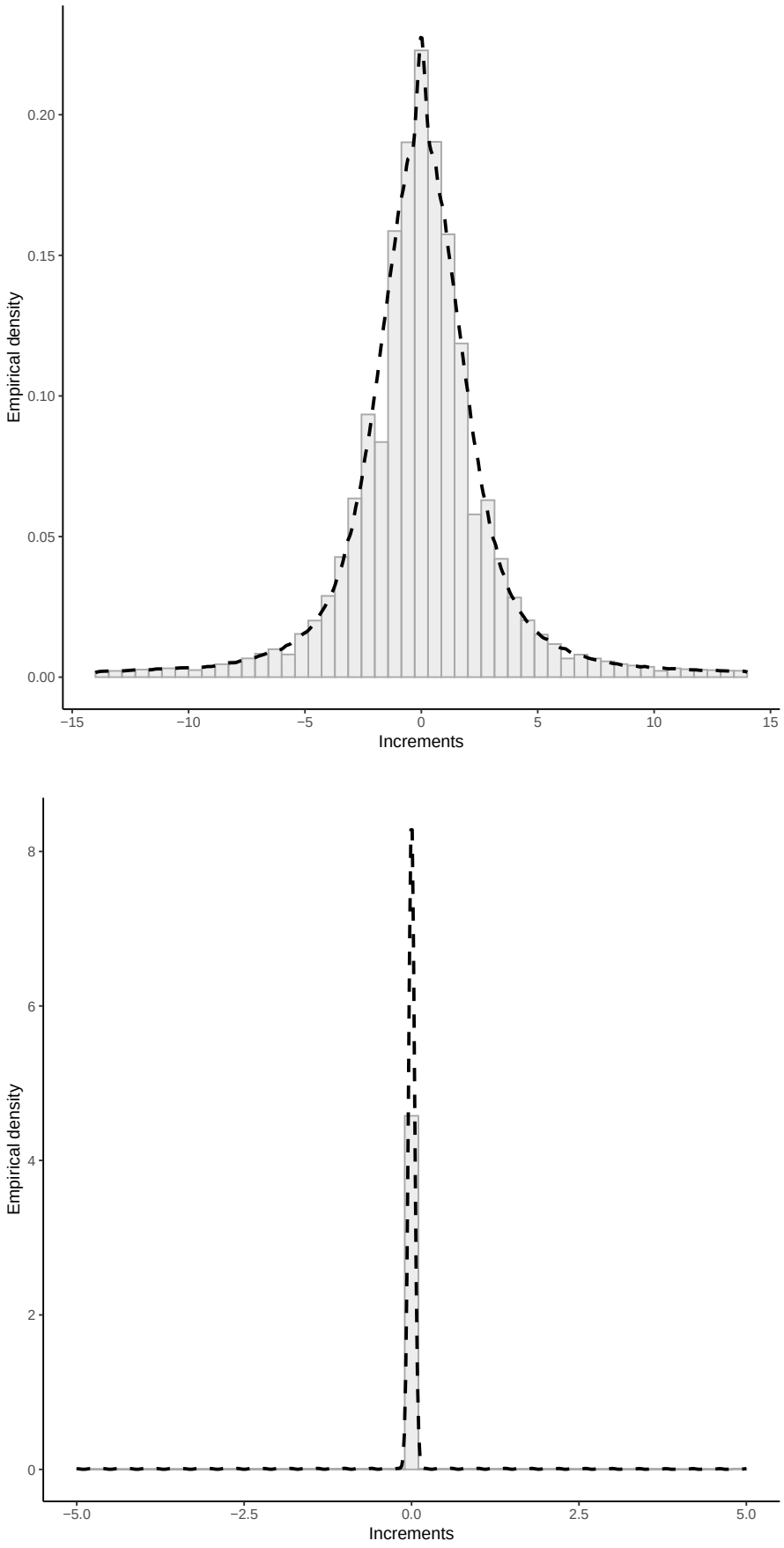
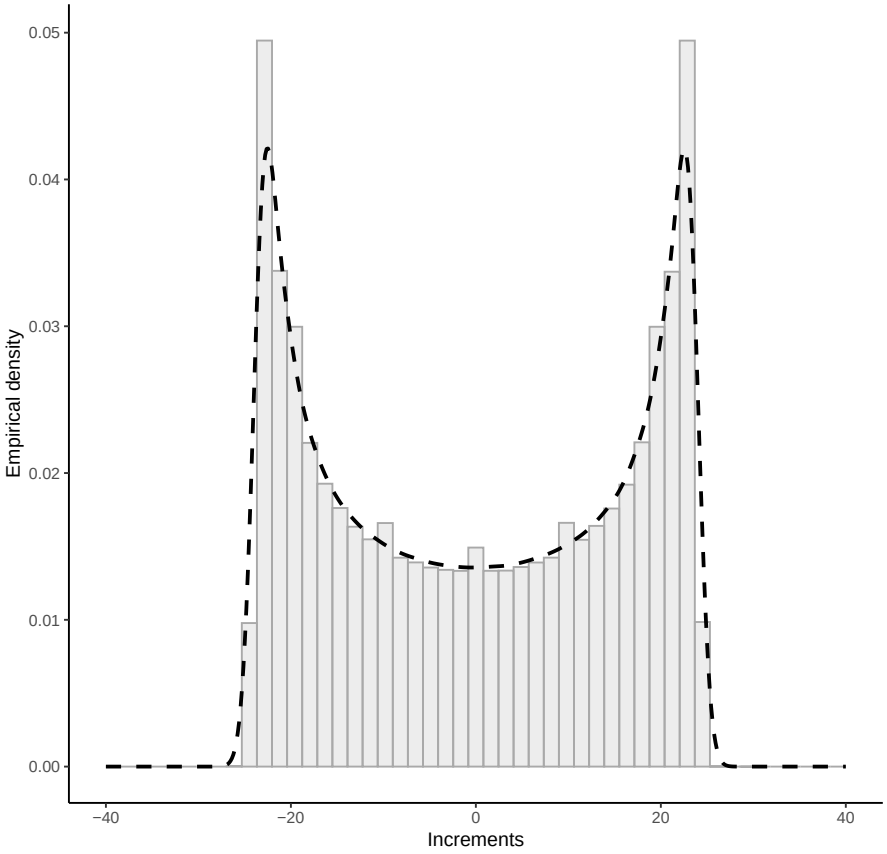
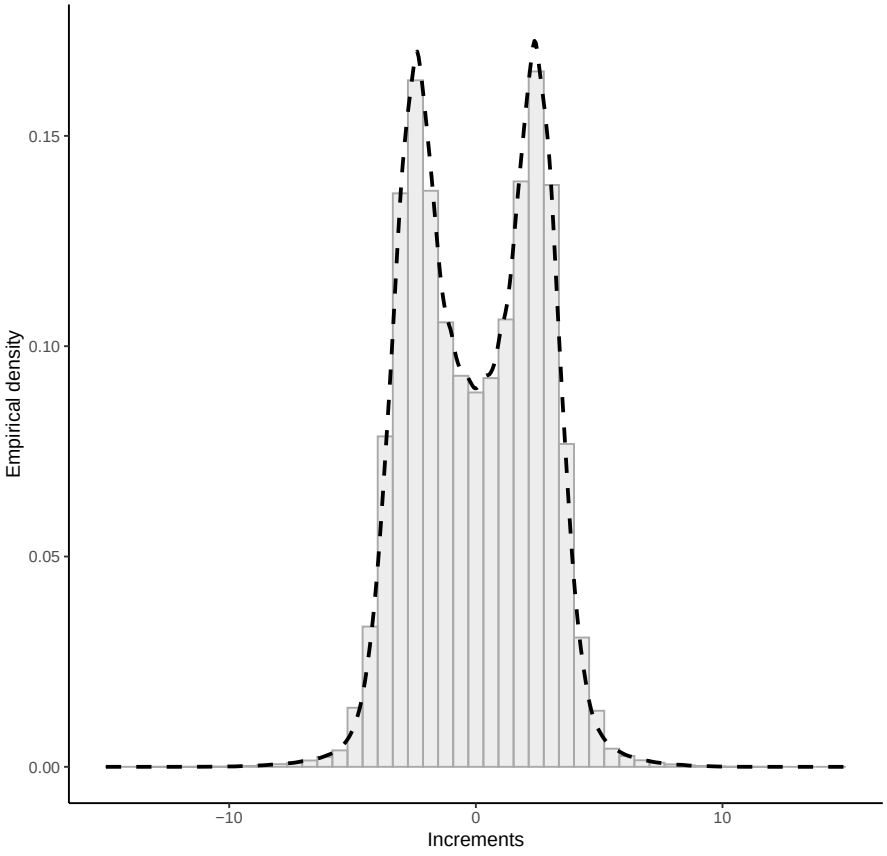


Figure III.1.2 Histograms of EEG increments displaying a 1-peak distribution


0.05
0.04
0.03
0.02
0.01
0.00
Empirical density
−40
−20
0
20
40
Increments


0.15
0.10
0.05
0.00
Empirical density
−10
0
10
Increments

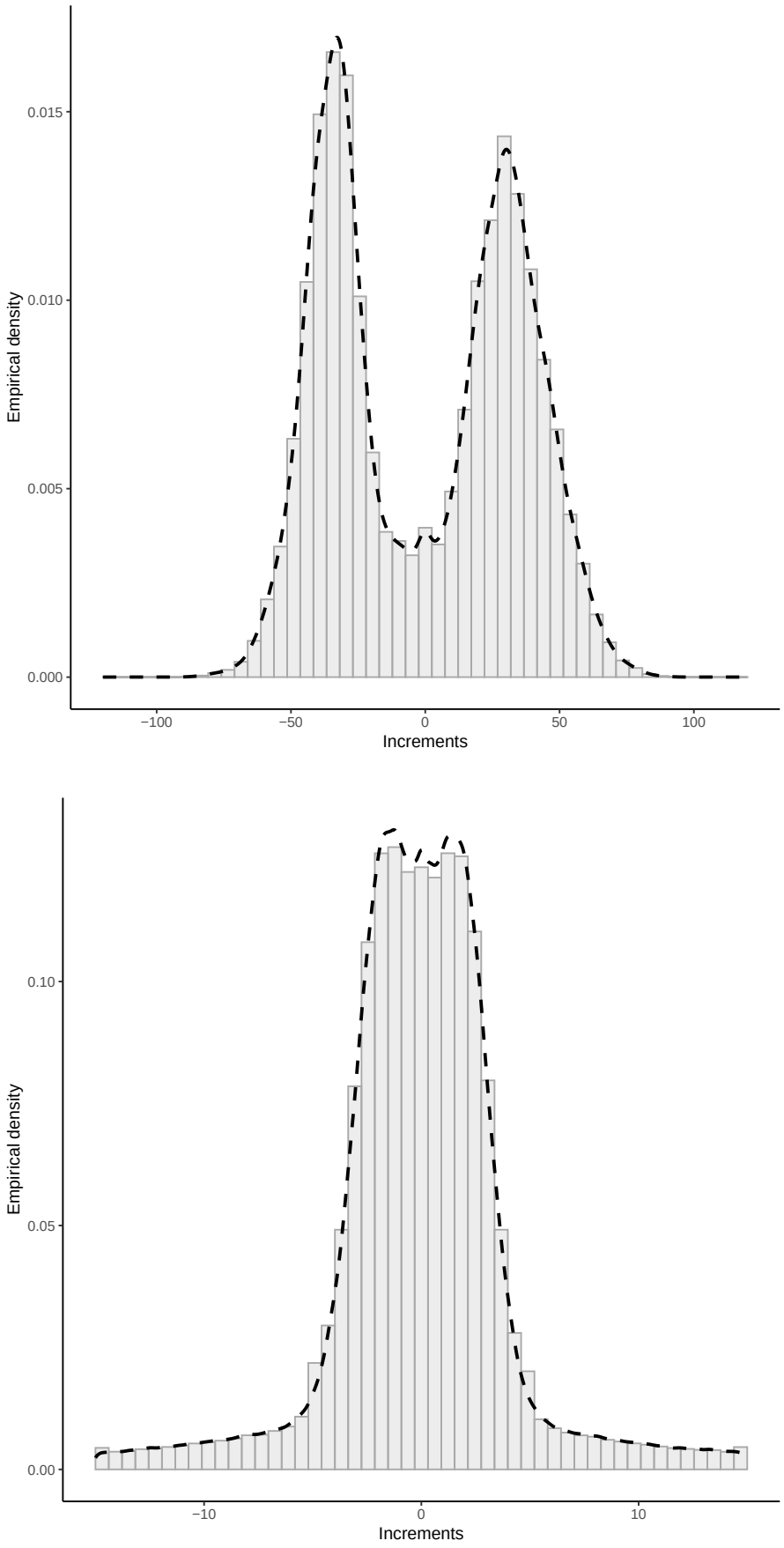
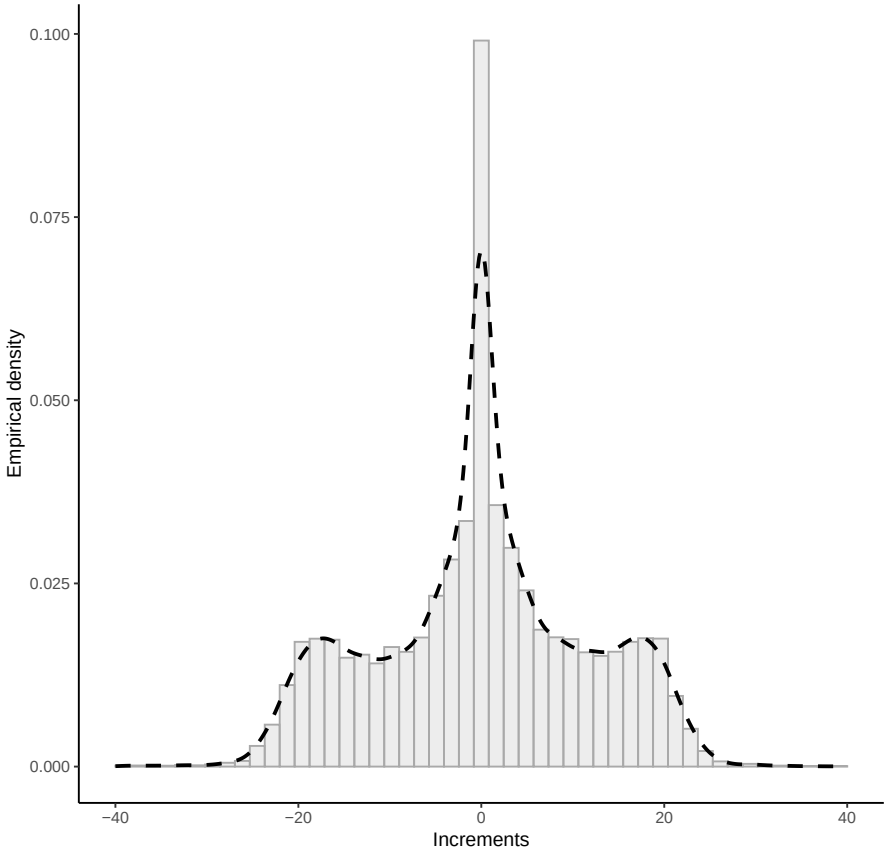
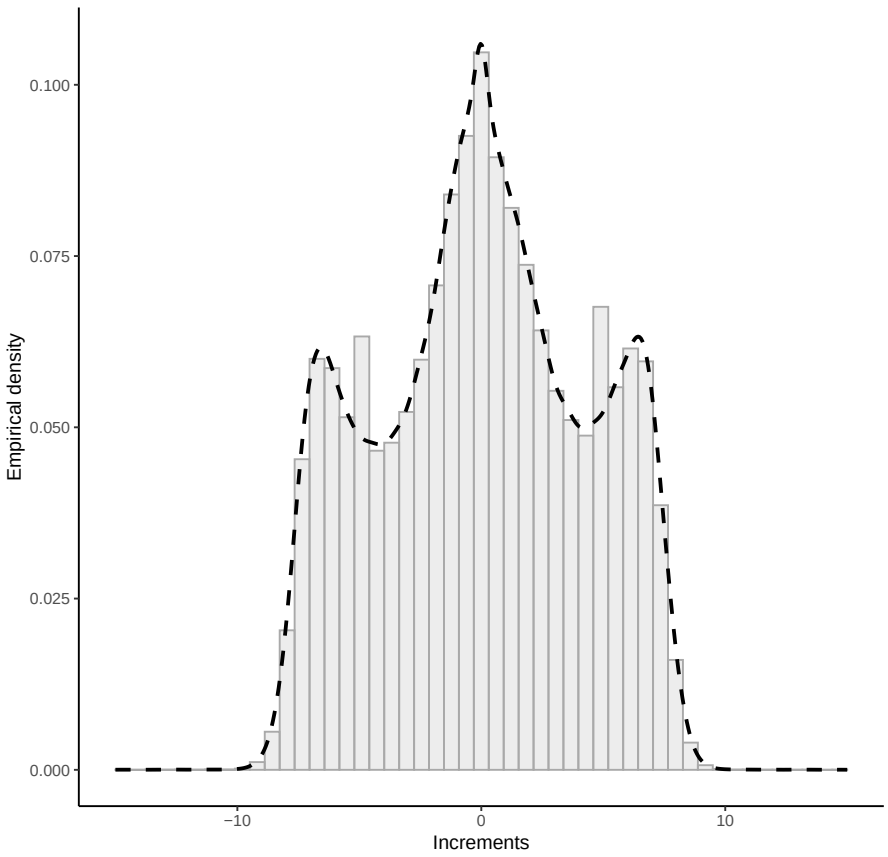


Figure III.1.3 Histograms of EEG increments displaying a 2-peak distribution


0.100
0.075
0.050
0.025
0.000
Empirical density
−40
−20
0
20
40
Increments


0.100
0.075
0.050
0.025
0.000
Empirical density
−10
0
10
Increments

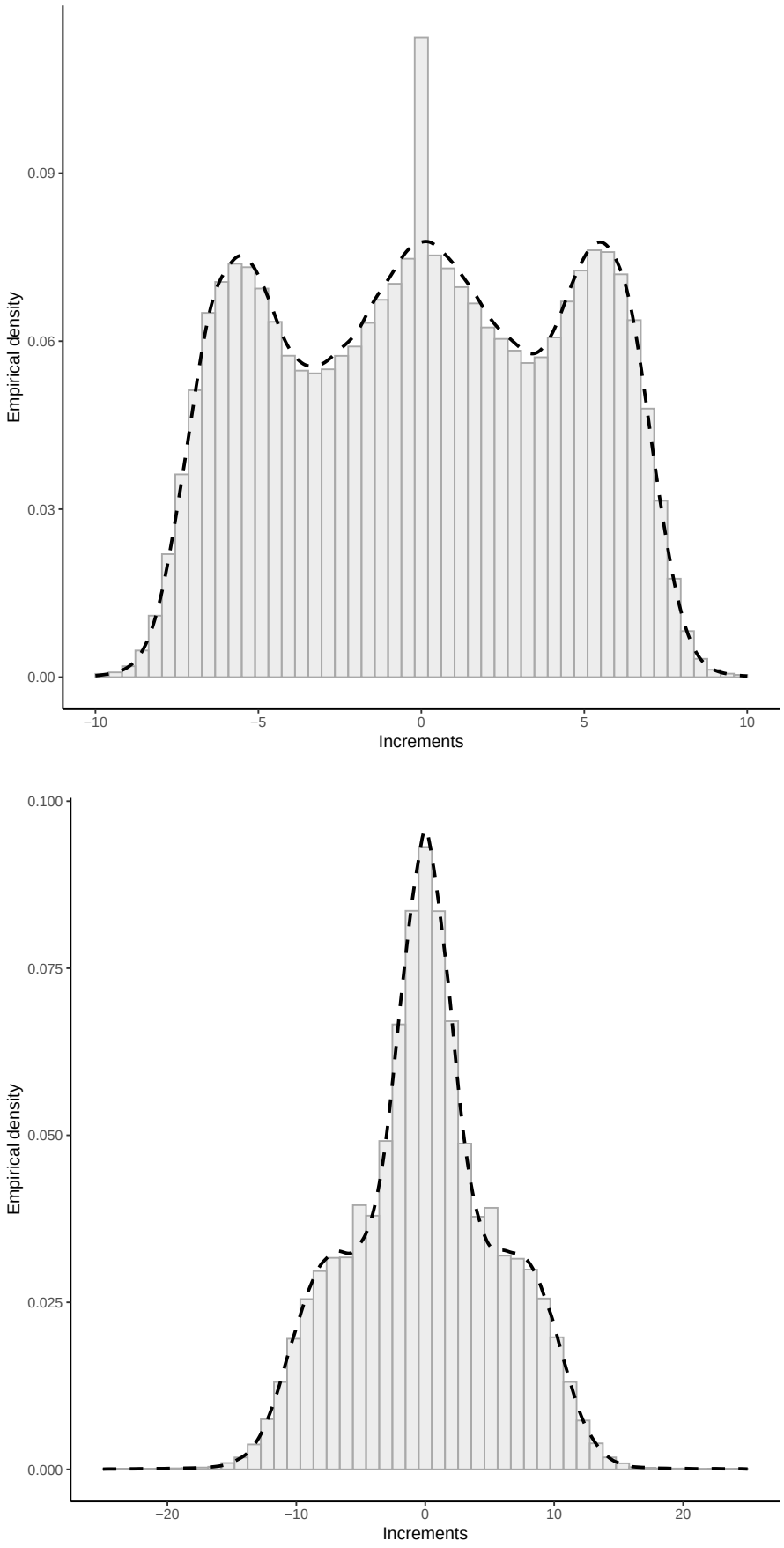
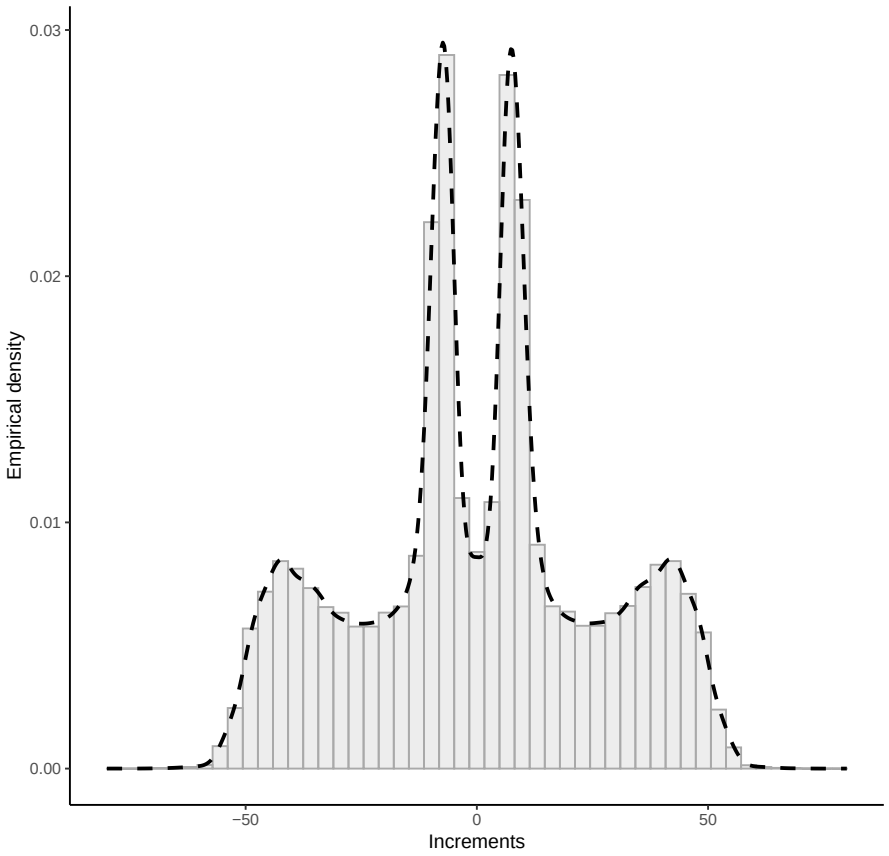
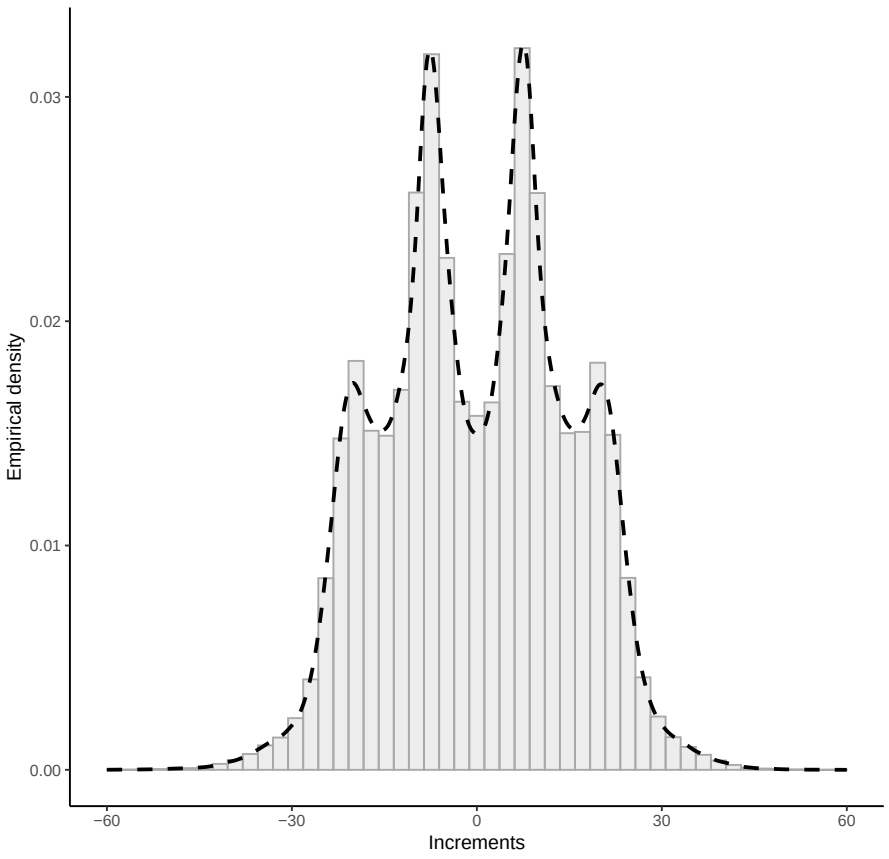


Figure III.1.4 Histograms of EEG increments displaying a 3-peak distribution


0.03
0.02
0.01
0.00
Empirical density
−50
0
50
Increments


0.03
0.02
0.01
0.00
Empirical density
−60
−30
0
30
60
Increments

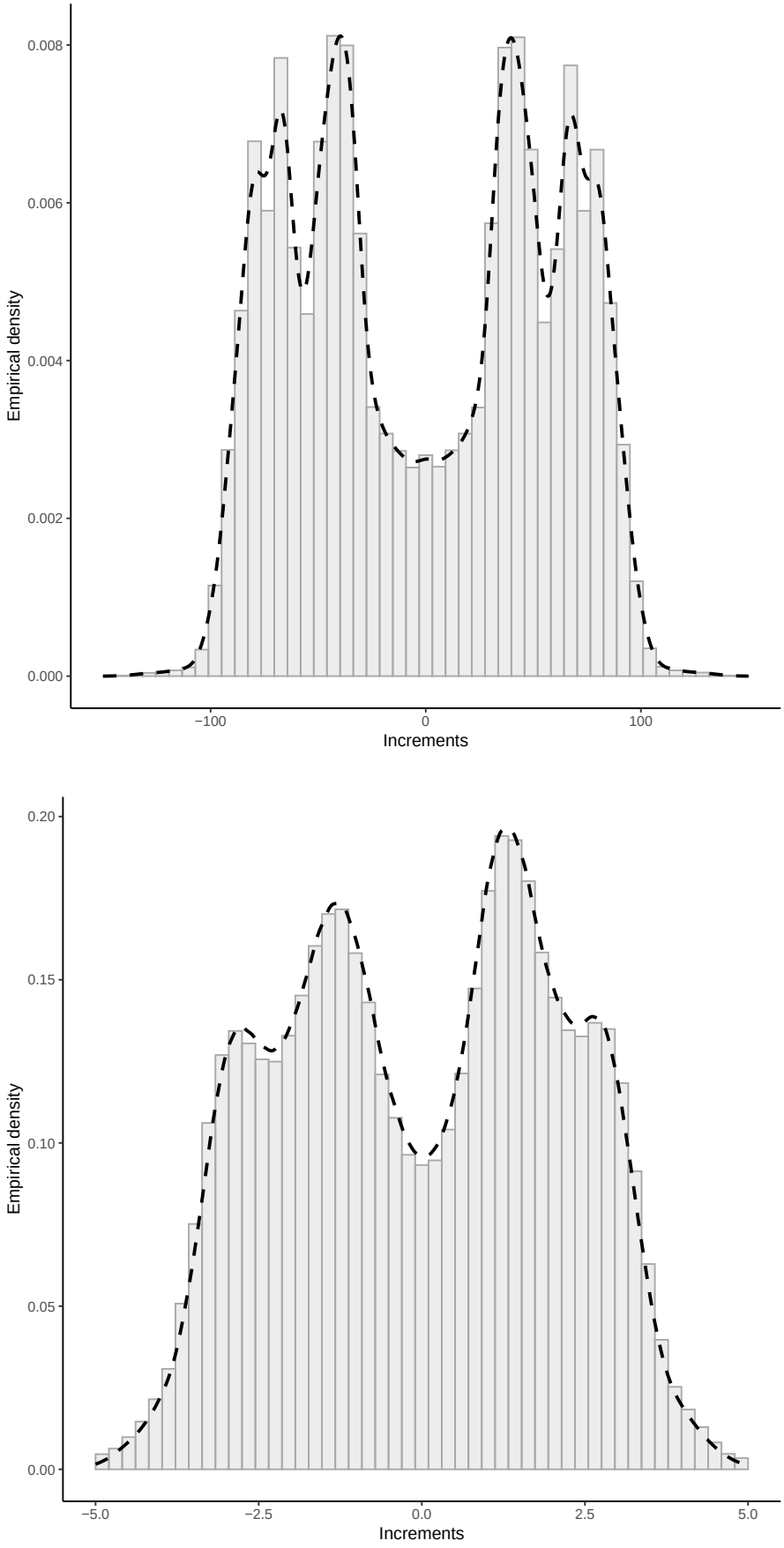


Figure III.1.5 Histograms of EEG increments displaying a 4-peak distribution

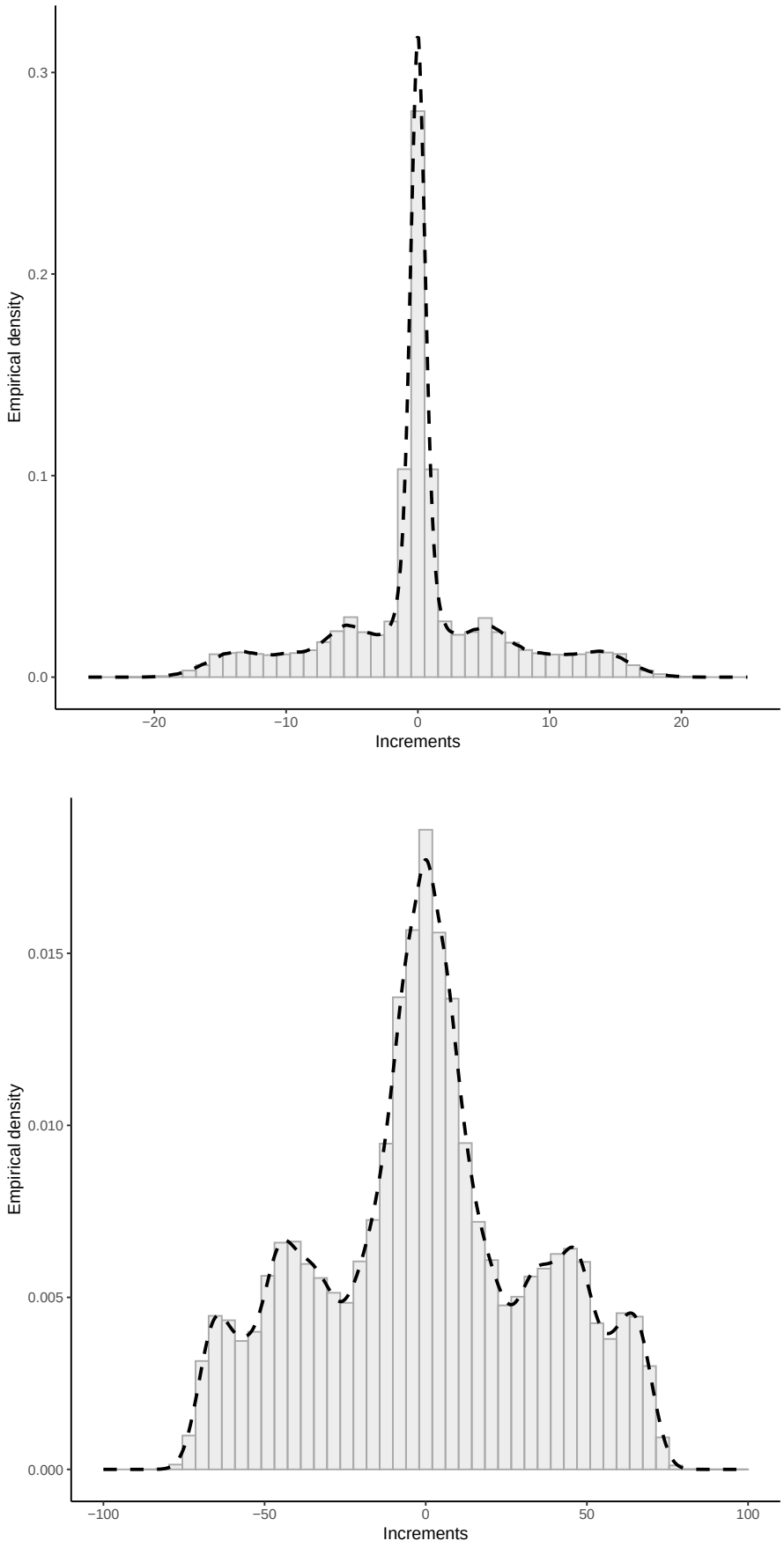


Figure III.1.6 Histograms of EEG increments displaying a 5-peak distribution

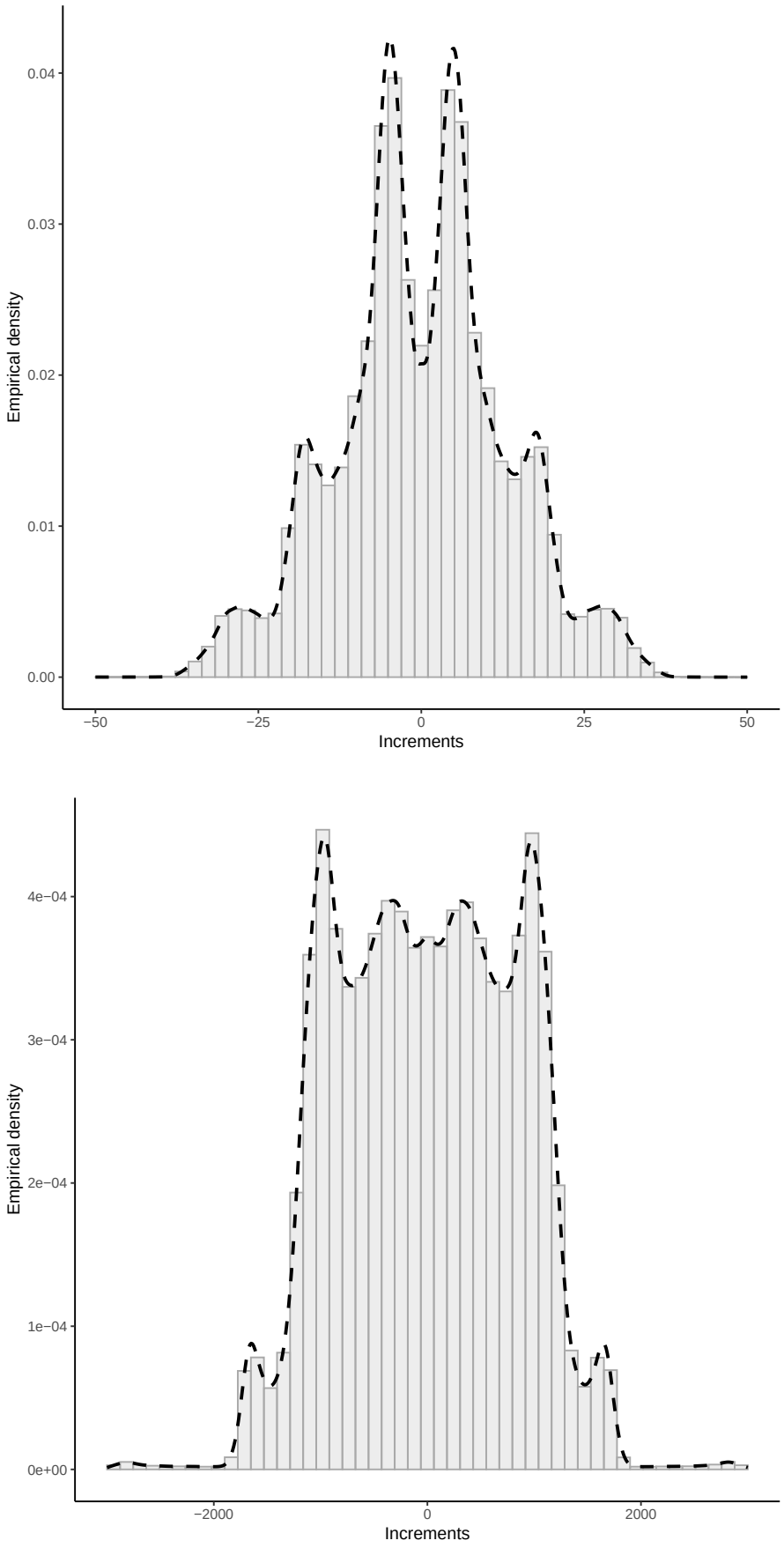


Figure III.1.7 Histograms of EEG increments displaying a 6-peak distribution

2 Prediction method: Elastic net regression

The purpose of the analysis of the EEG increments was whether these EEG features should be retained as a potentially important information in considering future cognition in children who suffered from cerebral malaria. There are several possible prediction approaches such as random forests or neural networks (NNs), both of which can deal with a vast amount of data and are highly flexible. However, they are prone to overfitting or are difficult to interpret (NNs), which presents an issue when it comes to justification of the procedures to the professionals who are not familiar with prediction methods. Thus, a method which is suited to high-dimensional problems and balances the advantages and disadvantages of L_1 (LASSO) and L_2 (ridge) regression was chosen. To identify important predictors of neurodevelopment and cognition 6 months after coma from cerebral malaria, elastic net regression was used. The method was introduced by Zou and Hastie [2005] as a way of controlling for correlations among predictors and dealing 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 defined by

$$L(\alpha, \lambda, \boldsymbol{\beta}) = |\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}|_2 + \lambda \left(\frac{1-\alpha}{2} |\boldsymbol{\beta}|_2 + \alpha |\boldsymbol{\beta}|_1 \right), \tag{2.1}$$

where

$$|\boldsymbol{\beta}|_2 = \sum_{j=1}^{p} \beta_j^2, \qquad |\boldsymbol{\beta}|_1 = \sum_{j=1}^{p} |\beta_j|$$

$\boldsymbol{y} = (y_1, \ldots, y_n)^\tau$ is the response, $\boldsymbol{X} = (\boldsymbol{x}_1 | \ldots | \boldsymbol{x}_p)$ is the model matrix and $\boldsymbol{x}_j = (x_{1j}, \ldots, x_{nj})^\tau$, $j = 1, \ldots, p$ are the predictors [Zou and Hastie, 2005].

Hyperparameter α can be seen as a mixing parameter between ridge ($\alpha = 0$) and LASSO ($\alpha = 1$) regression. Tuning of hyperparameters α and λ was performed using `caret` package [Kuhn, 2020] with leave-one-out cross validation. The tuning grid was constructed from values of $\lambda \in \{10^{-5}, 10^{-4}, \ldots, 10^3\}$ and 7 equidistant points from the interval $[0.0001, 1]$ for α. The pair which had the lowest root mean squared error (RMSE) was chosen for the final model.

The response variable was the standardized neurodevelopment or cognitive score taken 6 months after the discharge from the hospital, and the scores were in the range of $[-1.99, 1.5]$. Predictor variables were taken from the non-EEG features described in Section 1.1 and the EEG features obtained from analysis of increments. The model including just the non-EEG features was chosen as the baseline model to which all other models were compared (in terms of their RMSE). In feature matrices, the neurodevelopmental or cognitive score immediately after discharge from the hospital was included. In reporting of the results, this score is referred to as the baseline neurodevelopmental score. The rationale for the inclusion of this score was that it can be obtained relatively easily compared to

the estimation of the EEG parameters for stochastic models. Fitting of stochastic models would be warranted if the EEG parameters were shown to be important over and above other measures that can be obtained easily or as part of routine clinical care for cerebral malaria.

Some features contained missing values, especially in the non-EEG dataset. Since the dataset comprised both continuous and categorical data, imputation methods suitable to mixed-type data were used. Based on the comparison of different imputation methods presented in Robin et al. [2021], `missForest` package for `R` was selected. The algorithm is based on an iterative imputation scheme by training a random forest on observed values in a first step, followed by predicting the missing values and then proceeding iteratively. Main function of the package is the function `missForest` which takes the argument in the form of a matrix (dataframe) with missing values, where the columns correspond to the variables and the rows to the observations. The function also takes optional arguments, such as the maximum number of iterations to be performed, number of trees to grow in each forest, complete data matrix which can be supplied to test the performance etc., and the function can also be run parallel. The function `missForest` internally calls function `varClass` which returns the variable types of a provided dataframe. Imputation scheme is run iteratively - the difference between the previous and the new imputed data matrix is assessed after each iteration for the continuous and categorical parts and when both differences increase for the first time, the algorithm stops. In addition, when a complete data matrix is supplied, `missForest` can internally call the function `mixError` which calculates the imputation error in the case of mixed-type data (a normalized RMSE or proportion of falsely classified entries based on the type of the data) ([Stekhoven and Buhlmann, 2012]).

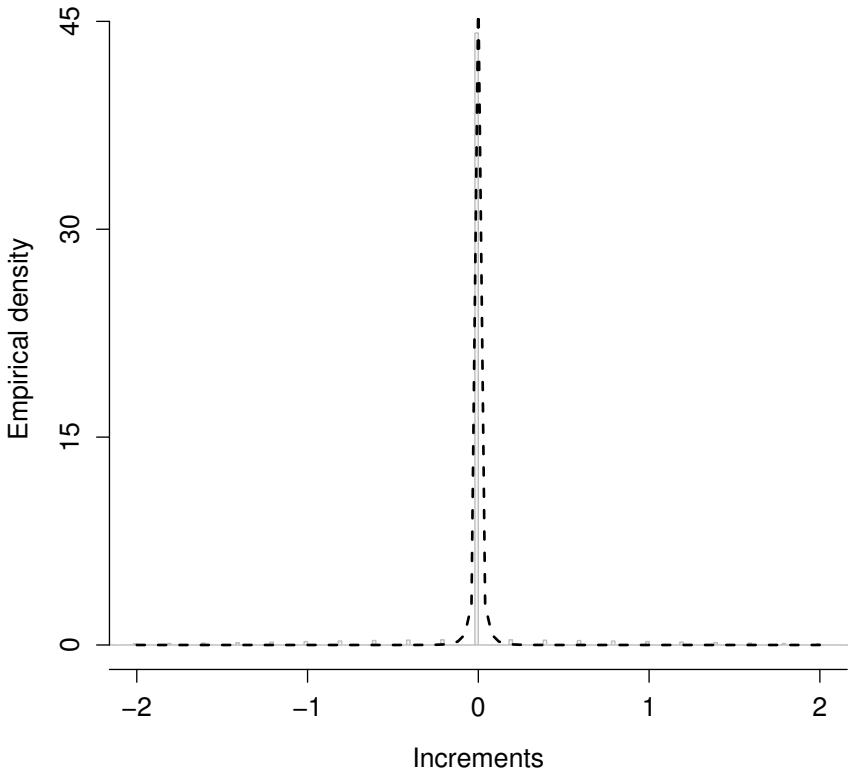
3 Prediction using Generalized Gaussian distribution

Prediction of neurodevelopment for children who were in coma due to cerebral malaria was first investigated for the estimates obtained by fitting the GGD to EEG increments, which would represent the simplest model containing stochastic features. In order to obtain the estimates of GGD parameters, all channels were viewed as unimodal having the GGD distribution. For the heavy-tailed subfamily of then GGD, parameter estimation didn't prove to be successful, thus the parameters that presented a suitable fit for heavy-tailed channels couldn't be obtained. Instead, tail-index estimation was performed on all the channels, with estimates representing the second collection of features included in prediction. Results from this section were published in Leonenko et al. [2023*a*].

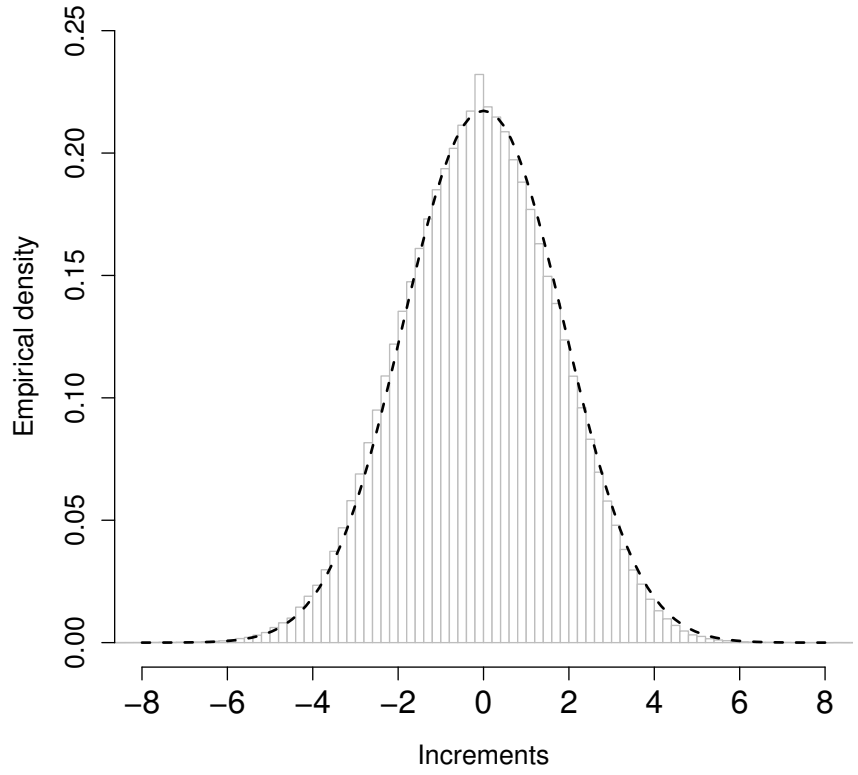
3.1 Fitting of GGD to EEG increments

Histograms of the increments of EEG signals were investigated for every child and every channel separately resulting in a total of 1482 histograms. Histograms were approximately symmetrical with means close to zero but also displayed higher peaks than the normal distribution, which as mentioned, further justifies the choice of the GGD (2.1) as the marginal distribution for modelling. Estimation of the parameter $\zeta = (s, \sigma^2)$ of the light-tailed GGD(s, σ^2) was obtained using the quasi-likelihood approach presented in Section 2.3.1. Non-linear optimization was performed using `maxLik` package [Henningsen and Toomet, 2011] in `R` version 4.0.4 for Windows. Due to the fact that the shape parameter s appears inside the Γ function in (2.1), constrains for the minimum value of s were used in the optimization to prevent computational problems. Optimization was performed on the entire dataset of EEG increments values and no sampling was used.

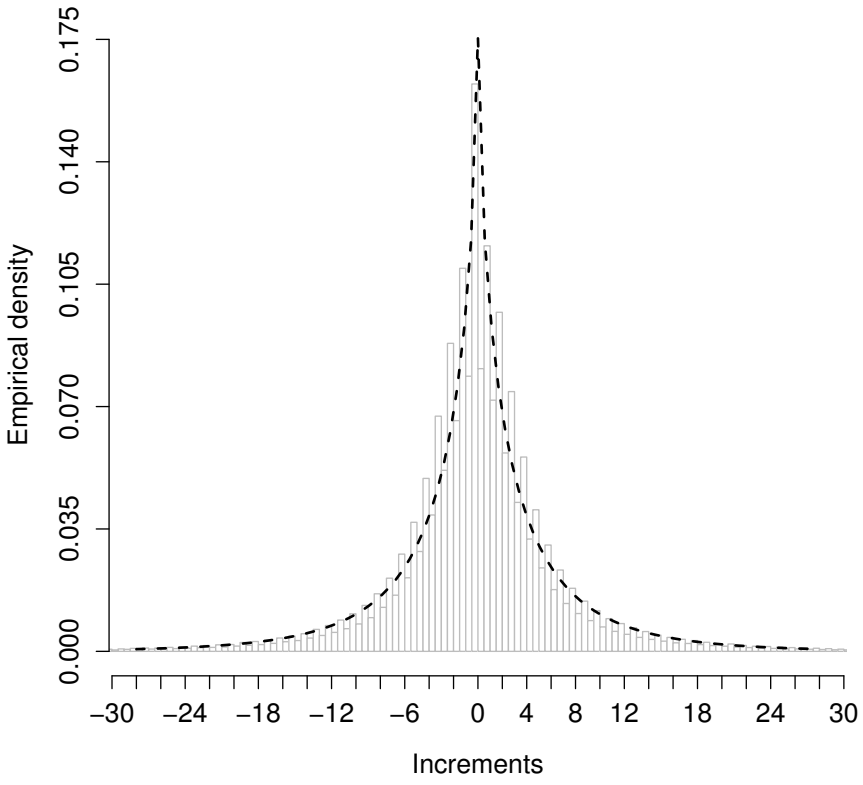
Values of $\hat{s}$ were in the range of $[0.02, 8.23]$. Extremely low values of $\hat{s}$ (close to 0.02) and $\widehat{\sigma^2}$ (close to 0.1) appeared on the channels containing very large number of zero observations and manifested in histograms with high and narrow peaks (Figure III.3.8a). Values of $\hat{s}$ near 2 result in a GGD fit that resembles a zero-mean normal distribution, which can be seen in Figure III.3.8b. Figure III.3.8c and Figure III.3.8d show examples of histograms where obtained estimates for the shape parameter were $\hat{s} < 2$ and $\hat{s} > 2$, respectively. Dashed line in the plots represents the fitted GGD.



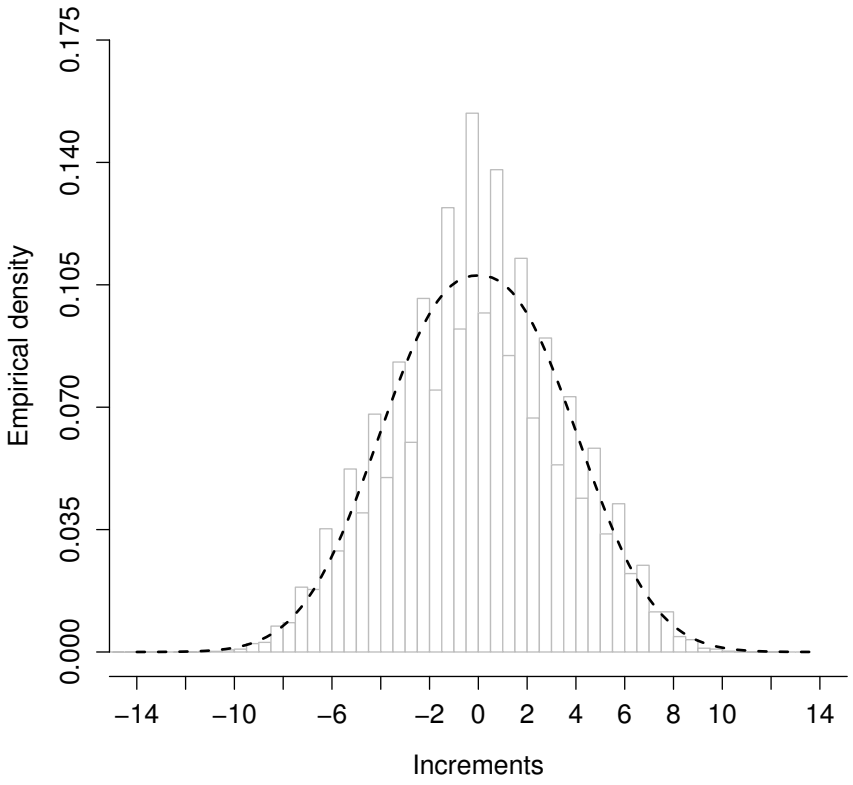
(a) GGD($\hat{s} = 0.02$, $\widehat{\sigma^2} = 0.12$)



(b) GGD($\hat{s} = 2.09$, $\widehat{\sigma^2} = 3.52$)



(c) GGD($\hat{s} = 0.68$, $\widehat{\sigma^2} = 2.50$)



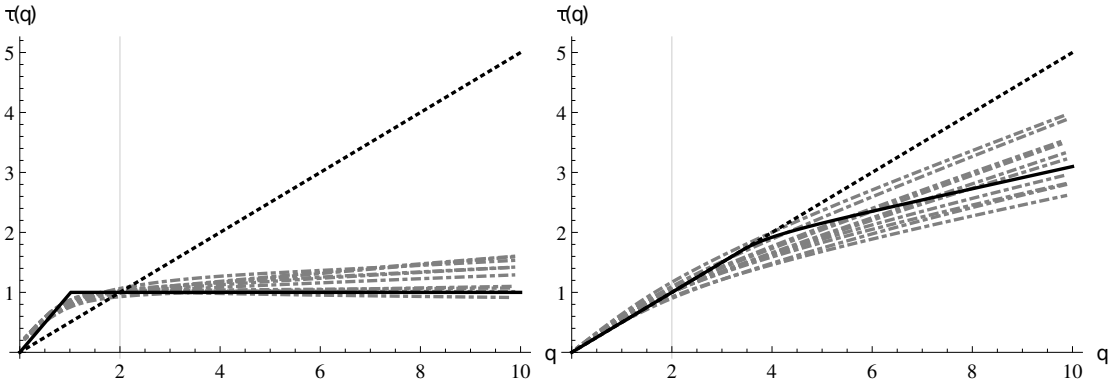
(d) GGD($\hat{s} = 2.32$, $\widehat{\sigma^2} = 20.10$)

Figure III.3.8 Fitting of the light-tailed GGD to EEG increments.

3.2 Estimation of tail index on EEG increments

Tail index estimation of EEG increments was performed using the graphical method presented in Section 2.3.2. The estimation was divided into two cases, $\alpha \leq 2$ and $\alpha > 2$, due to complexity of the expression τ_α^∞ in (2.44). This means that the corresponding part of τ_α^∞ based on the true value of α was used as a model function for the estimate $\hat{\alpha}$ in (2.45). Hence, before obtaining the value of the estimate $\hat{\alpha}$, it was necessary to visually inspect the plot of empirical scaling function and determine where the break point happens.

To obtain the numerical value of $\hat{\alpha}$, the empirical scaling function was fitted to the asymptotic form τ_α^∞ using ordinary least squares method. For the calculation of $\hat{\tau}_{N,n}$ values for $s_i \in \langle 0, 1\rangle$ were chosen to be equidistant points ($N = 23$) in the interval $[0.1, 0.9]$, while for q_j a total of 40 equidistant points were taken from interval $[0.11, 10]$. A random sample of EEG increments of size 10000 was chosen for every channel, and the sampling was repeated 10 times. The empirical scaling function was plotted for each of these 10 samples (shown by dot-dashed lines in Figure III.3.9), obtaining estimates $\hat{\alpha}_i$, $i = 1, \ldots, 10$. The final value of the tail index estimate $\hat{\alpha}$ was chosen to be the median of values $\hat{\alpha}_i$, represented by a solid line in Figure III.3.9. The bilinear shape (i.e a broken line) is clearly visible on both plots and identifies the breakpoint which determines whether the data are from the distribution with infinite (Figure III.3.9a) or finite (FigureIII.3.9b) variance. Analysis was performed using `Mathematica` version 11.3 for Windows.



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

Figure III.3.9 The function τ and the tail index estimates of EEG increments.

3.3 Generalized Gaussian distribution models

Models with different combinations of features were investigated for the prediction of neurodevelopment. As mentioned before, model containing only the non-EEG features was selected as the baseline model.

3.3.1 Non-EEG features model

The feature matrix for this model included a total of 54 non-EEG features with 78 observations. Maximum number of missing values per feature was 23, and most of the empty entries occurred in biomarker panels from cerebrospinal fluid. Categorical variables such as *sex* (2 levels) and *bcs* (Blantyre coma score, 6 levels) were coded into dummy variables for the inclusion in the elastic net regression. For the variable *bcs*, a score of 0 (poor results) was chosen as the reference level. Only the categorical variables were recoded into dummy variables (after the imputation of the missing values), other variables were included as continuous predictors in the feature matrix.

3.3.2 Combined non-EEG and GGD features model

The feature matrix for this model was a combination of aforementioned 54 non-EEG features from Section 3.3.1 and additional 38 features (for 19 channels) which were estimates of s and σ^2 parameters obtained by fitting the light-tailed GGD to EEG increments (see Section 3.1). Thus, for each channel, two predictors were included in the feature matrix, one including the continuous values of $\widehat{(s)}$ and one including continuous values of $\widehat{(\sigma^2)}$. There were no missing values in the GGD features subset.

3.3.3 Combined non-EEG and tail index features model

The feature matrix for this model was a combination of the same 54 non-EEG features and additional 19 features (for 19 channels) which were median values of estimates $\hat{\alpha}$ of the tail index from the estimation in Section 3.2). Missing values occurred for two cases and were imputed using `missForest` package. Median values of the tail index estimates in the feature matrix were in the range of $[0.01, 8.76]$. To reduce the noise of this variable within the model, the tail index for every channel was classified into 3 levels based on distributional tertiles. Thus, values of 3.1 and 4.5 were chosen as cut-offs. Since this resulted in creating categorical variables, they were recoded into dummy variables with values below 3.1 acting as reference level. Both versions of the model (with continuous values of tail index and categorized tail index values) were examined in the analysis.

3.4 Results and comparison

After running over a grid of different combinations of tuning parameters, best tuning parameters across all models were $\alpha = 0.833$ and $\lambda = 0.1$. These values produced the lowest RMSE. Comparison of models based on the leave-one-out cross validation RMSE is given in Table III.4.5. Since the `caret` package doesn't

provide standard errors of the coefficients in the elastic net regression model, standard errors of predictors' coefficients were obtained by bootstrapping. For this purpose `R` package `boot`[Canty and Ripley, 2021] was used with the number of bootstrap replicates set to 1000. Ordinary bootstrapping procedure was used due to the convenience of the function `boot` in package `boot`, but other options for nonparametric resampling such as "parametric", "balance", "permutation" or "antithetic" are also possible. The function takes the data, the statistic of interest and number of replicates as mandatory arguments, with an extensive list of other arbitrary arguments (for more details see Canty and Ripley [2021]).

Table III.3.2 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

Table III.3.3 Predictors selected by elastic net models.

Predictor	Model coefficient (SE)			
	Non-EEG features model	Non-EEG and GGD features model	Non-EEG and continuous tail index features model	Non-EEG and categorical tail index features model
Baseline ND	4.9848×10^{-1} (8.3221×10^{-2})	5.0301×10^{-1} (8.2076×10^{-2})	4.9848×10^{-1} (8.0763×10^{-2})	4.9860×10^{-1} (8.3407×10^{-2})
Blantyre coma score	-1.1657×10^{-1} (8.3327×10^{-2})	-1.0944×10^{-1} (8.4044×10^{-2})	-1.1657×10^{-1} (8.1889×10^{-2})	-1.2040×10^{-1} (8.1088×10^{-2})
Hemoglobin level	2.1215×10^{-2} (1.6438×10^{-2})	2.3149×10^{-2} (1.7477×10^{-2})	2.1215×10^{-2} (1.7165×10^{-2})	2.3891×10^{-2} (1.7051×10^{-2})
White blood cell count	-7.2054×10^{-3} (6.3639×10^{-3})	-7.1759×10^{-3} (5.7969×10^{-3})	-7.2054×10^{-3} (6.1838×10^{-3})	-8.2922×10^{-3} (6.2979×10^{-3})
Interleukin (IL)-10 csf[1] level	6.9793×10^{-3} (5.0404×10^{-3})	7.2020×10^{-3} (4.7077×10^{-3})	6.9793×10^{-3} (4.7335×10^{-3})	7.9912×10^{-3} (4.8946×10^{-3})
Age	1.4945×10^{-3} (1.5223×10^{-2})	7.2271×10^{-4} (1.2848×10^{-2})	1.4945×10^{-3} (1.4590×10^{-2})	–
IL-1α receptor level in csf	8.9480×10^{-4} (6.0332×10^{-4})	9.6438×10^{-4} (6.3683×10^{-4})	8.9480×10^{-4} (6.1714×10^{-4})	7.6489×10^{-4} (5.4993×10^{-4})
HOME score	7.1907×10^{-4} (7.5640×10^{-3})	1.5538×10^{-4} (7.9678×10^{-3})	7.4088×10^{-3} (1.4590×10^{-2})	–
IL-6 csf level	4.4212×10^{-4} (8.5593×10^{-4})	5.0712×10^{-4} (8.4274×10^{-4})	4.4212×10^{-4} (8.2988×10^{-4})	5.8932×10^{-4} (7.4943×10^{-4})
Von Willebrand factor	3.9401×10^{-5} (3.1056×10^{-4})	4.5009×10^{-5} (3.1598×10^{-4})	3.9401×10^{-5} (3.0665×10^{-4})	–
IL-8 csf level	3.0274×10^{-5} (3.4552×10^{-5})	2.7567×10^{-5} (3.2663×10^{-5})	3.0274×10^{-5} (3.3118×10^{-5})	1.9938×10^{-5} (3.1701×10^{-5})
IL-1α receptor level in plasma	-6.1415×10^{-6} (6.6809×10^{-6})	-5.5330×10^{-6} (6.1356×10^{-6})	-6.1415×10^{-6} (6.1917×10^{-6})	-6.9700×10^{-7} (4.9735×10^{-6})
T3_CZ σ^2 estimate	N/A	2.5271×10^{-8} 1.3371×10^{-8}	N/A	N/A
T6_CZ2	N/A	N/A	N/A	-1.6478×10^{-1} 1.0155×10^{-1}

[1] cerebrospinal fluid

Models displayed similar RMSE values but generally, the addition of EEG features resulted in the RMSE reduction. The lowest RMSE of 0.5549 was obtained for the model containing non EEG features and categorical tail index features. This model also had the smallest number of non-zero coefficients. The list of predictors selected by each model is shown in Table III.3.3. The table shows the values of predictors' coefficients along with their standard errors obtained by bootstrapping. Predictors that weren't included as features in models are marked by N/A and those that weren't selected by a certain model are marked by "–".

4 Prediction using mixed generalized Gaussian distribution

After investigating the GGD as the marginal distribution of the diffusion processes of EEG increments and using their estimates in prediction of neurodevelopment, the question arose whether a more complex model which would account for the multimodality in histograms would improve the prediction of neurodevelopment. For this reason, estimation was performed to obtain the parameters of the MixGGD and include them in feature matrices. Results from this section were published in Leonenko et al. [2023*b*].

4.1 Fitting of MixGGD to EEG increments

Parameter estimation was performed without randomly sampling the values of increments, however, due to the sensitivity of the estimating equations, outliers had to be removed prior to the estimation. Outliers were defined as data points more than 1.5 of the interquartile range below the first or above the third quartile.

Although some channels' histograms displayed a distribution of up to six peaks, a GGD mixture of maximum three components ($K = 3$) was used as the model in the analysis because 95% of the child-channel pairs had 3 or fewer peaks. Four- and six-peak channels were reclassified as having two peaks, and five-peak channels as having three peaks which can be seen as merging peaks which are in proximity of each other in the histograms. The reason for this is that less than 5% of the total number of channels manifested in more than three peaks, and fitting a six-component GGD mixture doubles the number of parameters needed to be estimated and further complicates the analysis. Additionally, the parameter estimates were intended to be used as predictors in modelling of neurodevelopment. If a six-component mixture were to be fitted to all channels, most of the parameter estimates would be missing by default in the feature matrices since the majority of channels' histograms (approximately 95%) displayed only up to three peaks.

Analysis was performed using `R` version 4.1.2 for Windows. The ECM algorithm for estimation of parameters of the MixGGD wasn't available in `R` from any of the existing packages so the algorithm had to be implemented using the ECM method described in Section 3.3. The algorithm was implemented for the case of up to three peaks and additionally, improved with the possibility of forcing a symmetrical mixture (in terms of location, scale and shape) where for the case of two-component mixture the conditions $\widehat{\mu_1} = \widehat{\mu_3}$ and $\widehat{\mu_2} = 0$, $\widehat{\sigma_1^2} = \widehat{\sigma_3^2}$ and $\widehat{\sigma_2^2} = 0$, $\widehat{s_1} = \widehat{s_3}$ and $\widehat{s_2} = 0$ were applied. Similarly, for the three-component mixture that condition corresponded to $\widehat{\mu_1} = \widehat{\mu_3}$, $\widehat{\sigma_1^2} = \widehat{\sigma_3^2}$, $\widehat{s_1} = \widehat{s_3}$; only the weight

parameters weren't forced to be equal. The main function implemented in the algorithm was the `ECM_algorithm` function which takes a vector of data and the number of components as mandatory arguments. Recall that the number of components was determined beforehand from the visual inspection of histograms of the EEG increments. Other arbitrary arguments are the number of iterations and tolerance which were used as stopping criteria for the algorithm. The function `ECM_algorithm` internally calls functions `e.step` and `m.step`, which are respectively the E- and M- step of the ECM algorithm. The function `e.step` takes as arguments the vector of data and initial (or subsequent) values of parameters estimates of the MixGGD and returns the log-likelihood value in (3.6) calculated based on the current value of parameter estimates $\hat{\zeta}$, along with the conditional probability defined in (3.7). This value is then passed to the function `m.step` which performs the numerical maximization of the log-likelihood function (3.6) using the previously obtained value of $\hat{\zeta}$ and returns the new value of parameter estimate. This process runs iteratively until a stopping criterion is met. In addition to these functions, a plotting function for the histograms and their empirical density functions was implemented utilizing the `ggplot2` package from Wickham [2016].

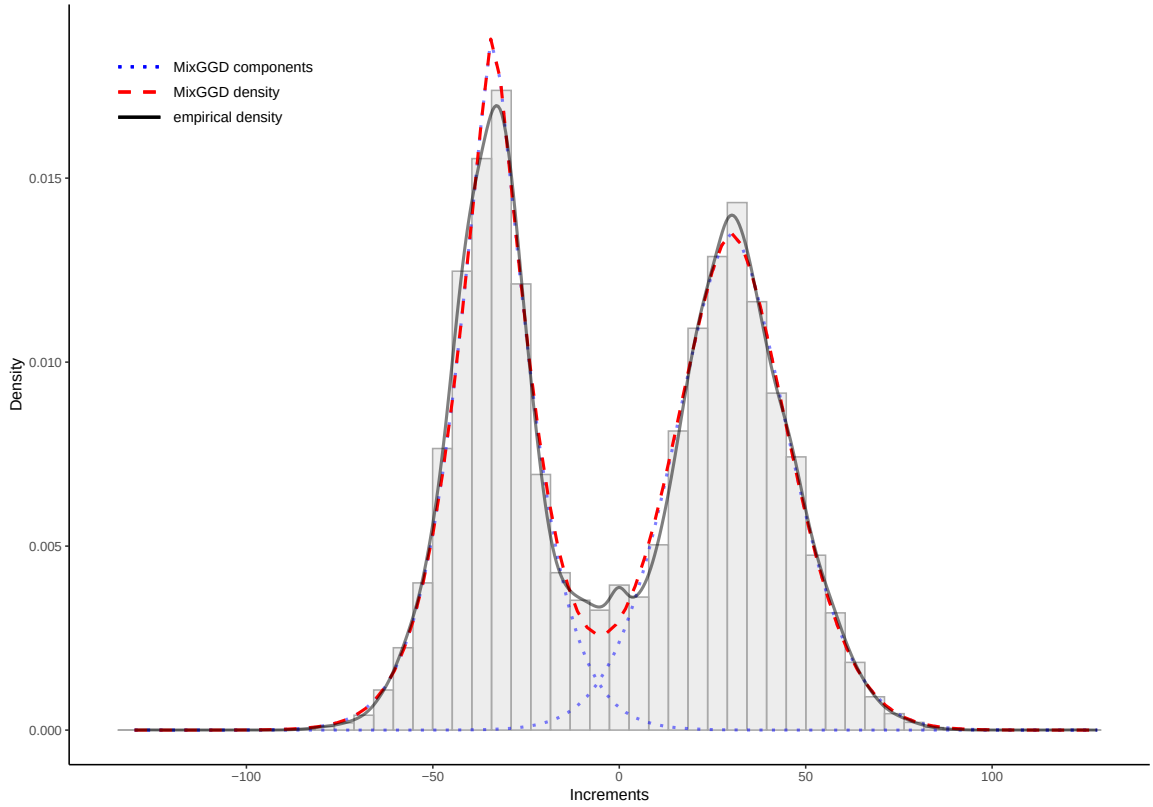
To initialize the algorithm, k-means clustering available from base `R` package was performed, and then the initial values of parameters were calculated for each of the derived clusters. Since the channels had been previously classified based on their histogram shape, clustering was performed using the available information about the number of peaks. Initial values of parameters μ_k, σ_k^2, w_k, along with the initial value of the kurtosis κ_k, for $k = 1, 2, 3$ were estimated on each of the clusters. The initial value of the shape parameter s_k was then calculated using the expression (3.10). These values of $\widehat{\mu_k}$, $\widehat{\sigma_k^2}$, $\widehat{s_k}$ and $\widehat{w_k}$ were then used as the starting points of the algorithm. Stopping criterion for the algorithm was a tolerance of 10^{-4} or 1000 iterations. In cases where the number of peaks wasn't conclusive, the ECM algorithm with all the possible number of peaks was tested and the performance compared based on the value of AIC, BIC and/or log-likelihood value. In some cases, the obtained estimates didn't present a good fit so an improvement was tried by forcing the ECM algorithm for a symmetrical mixture; this resulted with an improvement in 4 channels. This could be explained by the fact that the disturbances on the individual channels during EEG recording result in "pulling" the distribution to either side of the zero, thus forcing the MixGGD to be symmetrical on such channels could be justified in order to compensate for these effects. The algorithm converged (was considered successful) in 1263 (85.3%) of the cases (channels) and estimates couldn't be obtained in a total of 219 (14.8%) cases. Of these 219 cases where the algorithm was unsuccessful, 77 cases corresponded to unimodal channels with a "degener-

ate" distribution, and the rest was on multimodal channels. The reason behind the failure of the algorithm in these cases was determined to be the value of the kurtosis, which was either lower than 1.865 or greater than 349.087 (these values of the kurtosis result in negative values of the shape parameter s_k) so the shape parameter couldn't be approximated using the expression (3.10).

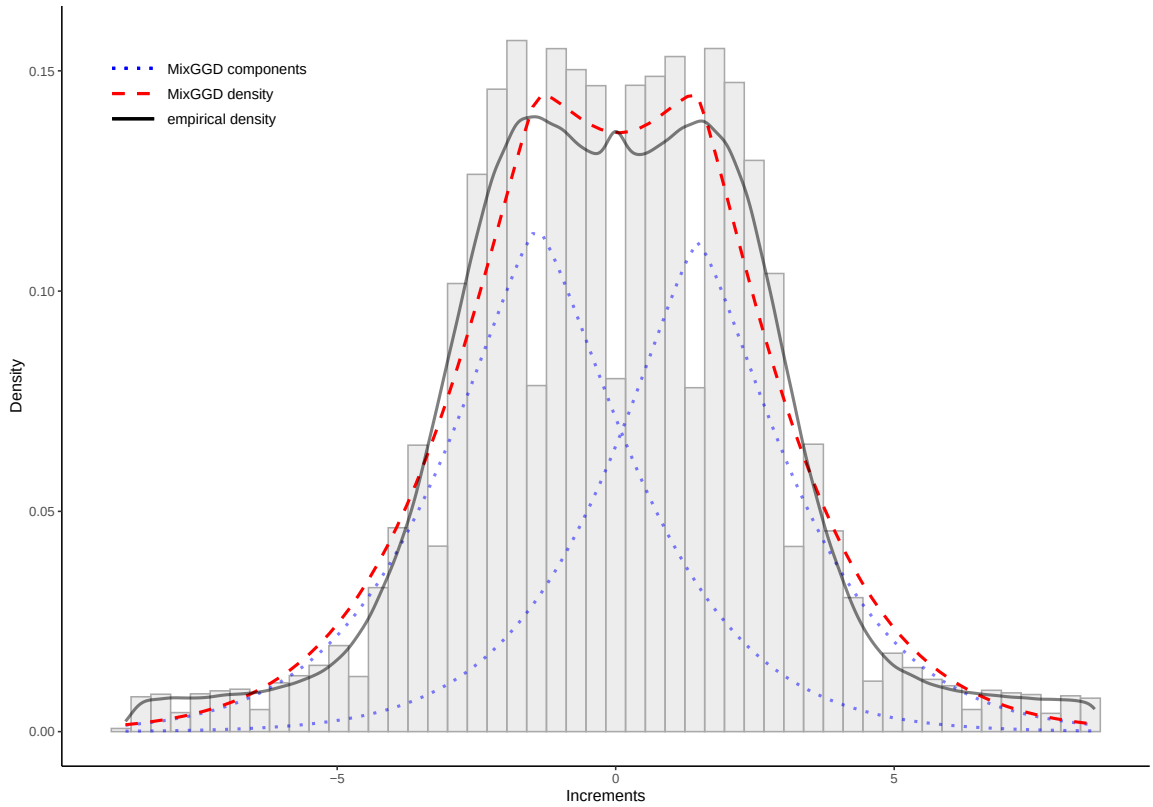
Examples of the obtained fit for the two- and three-component MixGGD can be seen in Figures III.4.10 and III.4.11, respectively. Examples of the fit where a two- or three-component MixGGD was forced on histograms with more than three peaks can be seen in Figure III.4.12. In each of the figures, (blue) dotted lines represent the fit of the individual components of MixGGD, (red) dashed line represents the mixture of these components, and the (gray) solid line represents the empirical density. Values of parameter estimates are given in Table III.4.4.

Table III.4.4 The MixGGD parameter estimates.

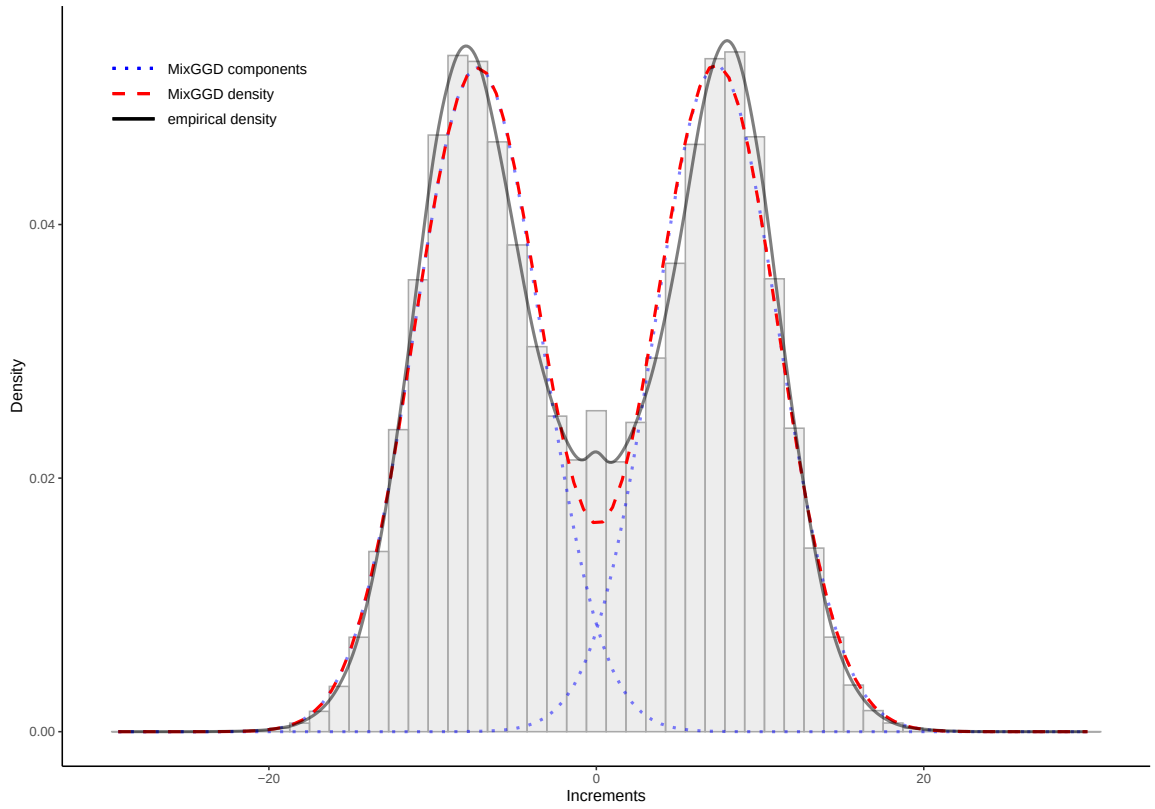
	MixGGD parameters											
	$\widehat{\mu_1}$	$\widehat{s_1}$	$\widehat{\sigma_1^2}$	$\widehat{w_1}$	$\widehat{\mu_2}$	$\widehat{s_2}$	$\widehat{\sigma_2^2}$	$\widehat{w_2}$	$\widehat{\mu_3}$	$\widehat{s_3}$	$\widehat{\sigma_3^2}$	$\widehat{w_3}$
Fig. III.4.10a	-33.86	1.34	166.18	0.47					30.16	1.74	283.97	0.53
Fig. III.4.10b	-1.41	1.34	5.54	0.52					1.48	1.32	5.29	0.48
Fig. III.4.10c	-7.26	2.05	14.30	0.50					7.33	2.00	14.18	0.50
Fig. III.4.10d	-1.51	1.14	1.00	0.49					1.47	1.11	0.97	0.51
Fig. III.4.11a	-36.89	3.77	112.89	0.23	0.02	0.65	50.48	0.54	37.13	3.88	112.00	0.23
Fig. III.4.11b	-7.05	2.74	16.39	0.28	-0.05	1.67	11.55	0.44	6.99	2.70	16.41	0.28
Fig. III.4.11c	-5.82	2.36	1.32	0.19	0.11	2.22	9.69	0.63	5.74	2.43	1.18	0.18
Fig. III.4.11d	-3.66	1.80	5.88	0.34	-0.23	1.28	2.52	0.24	3.06	1.79	6.78	0.42
Fig. III.4.12a	-1.81	2.51	1.23	0.49					1.74	2.24	1.15	0.51
Fig. III.4.12b	-673.06	1.81	232133.93	0.48					613.24	1.98	260249.63	0.52
Fig. III.4.12c	-44.84	3.76	227.45	0.26	-0.11	2.00	128.94	0.48	45.27	3.90	218.94	0.26
Fig. III.4.12d	-40.14	1.07	502.34	0.40	-0.09	1.32	194.85	0.20	40.54	1.10	501.41	0.40


MixGGD components
MixGGD density
empirical density
Density
0.015
0.010
0.005
0.000
−100
−50
0
50
100
Increments

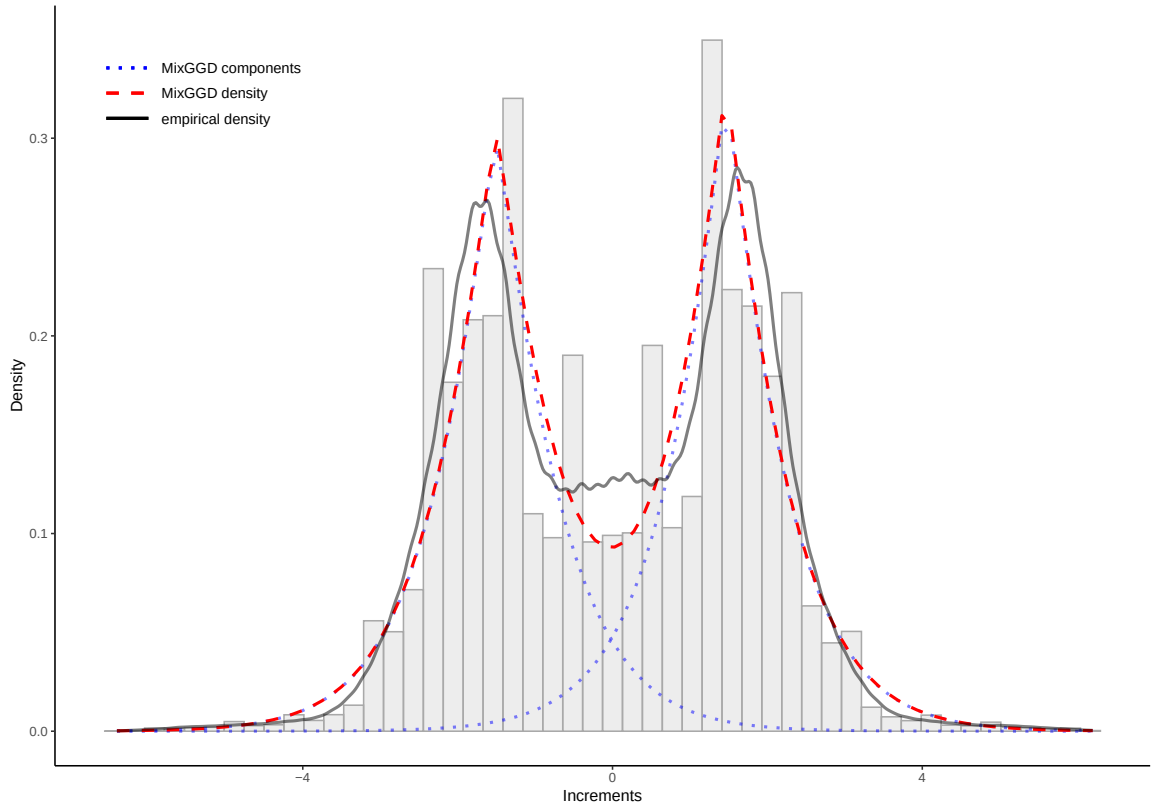
(a)


MixGGD components
MixGGD density
empirical density
Density
0.15
0.10
0.05
0.00
−5
0
5
Increments

(b)

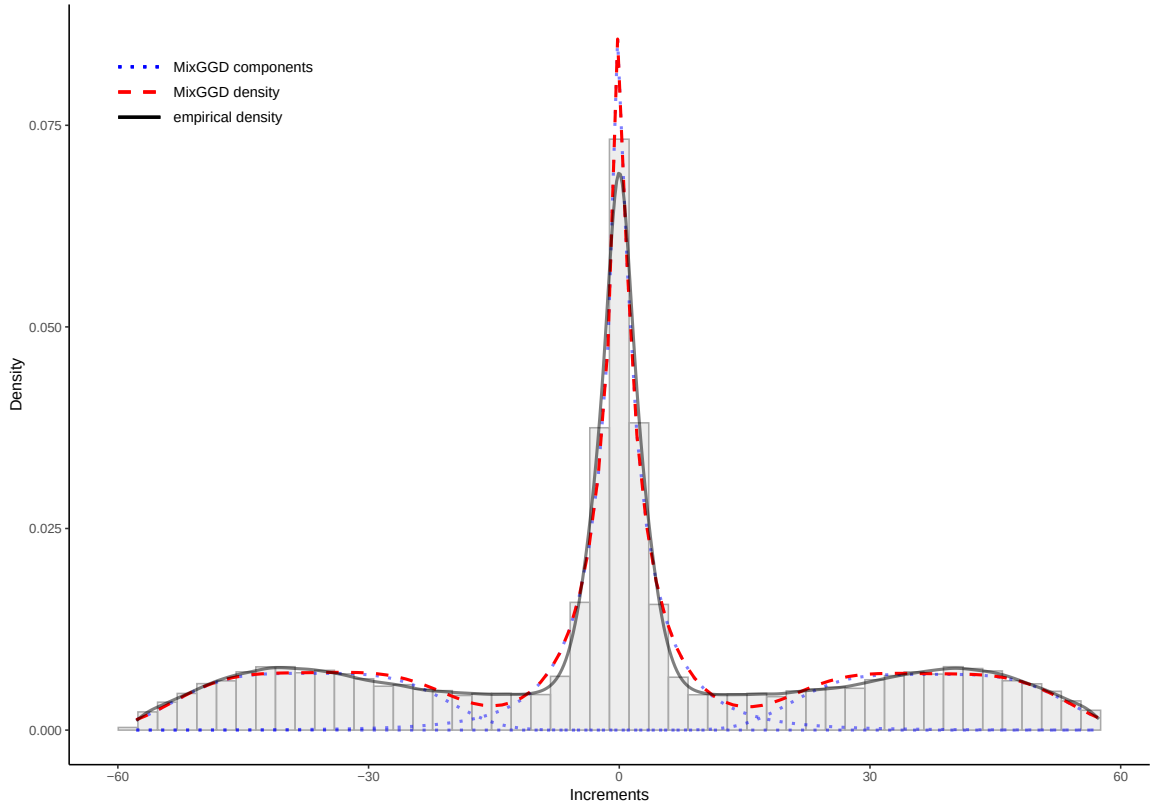


(c)

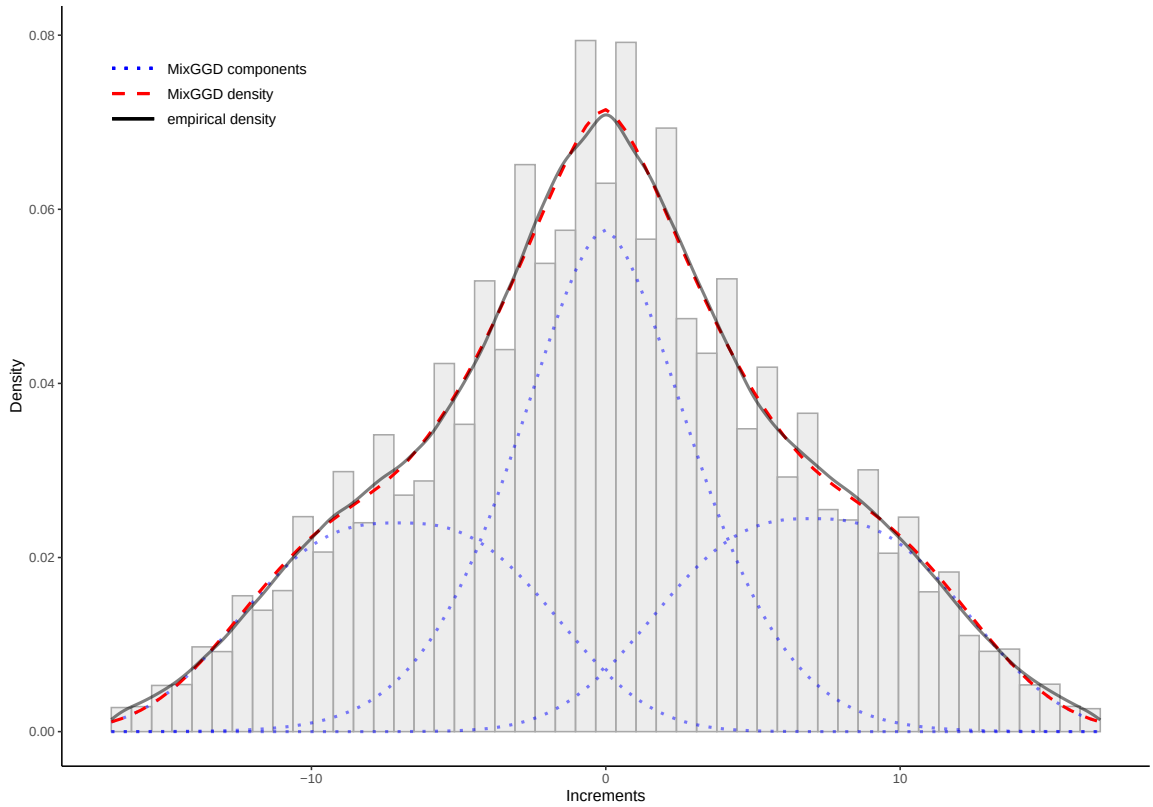


(d)

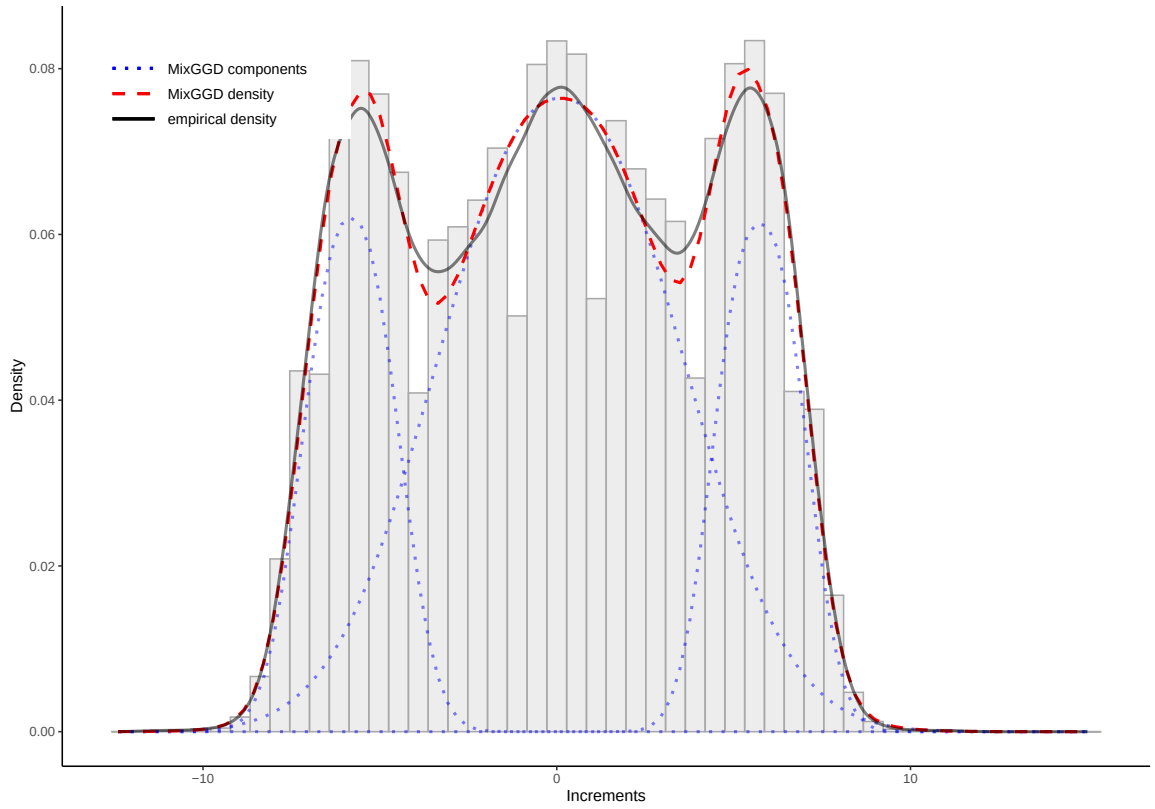
Figure III.4.10 Fitting a two-component MixGGD to EEG increments.


MixGGD components
MixGGD density
empirical density
0.075
0.050
0.025
0.000
Density
−60
−30
0
30
60
Increments

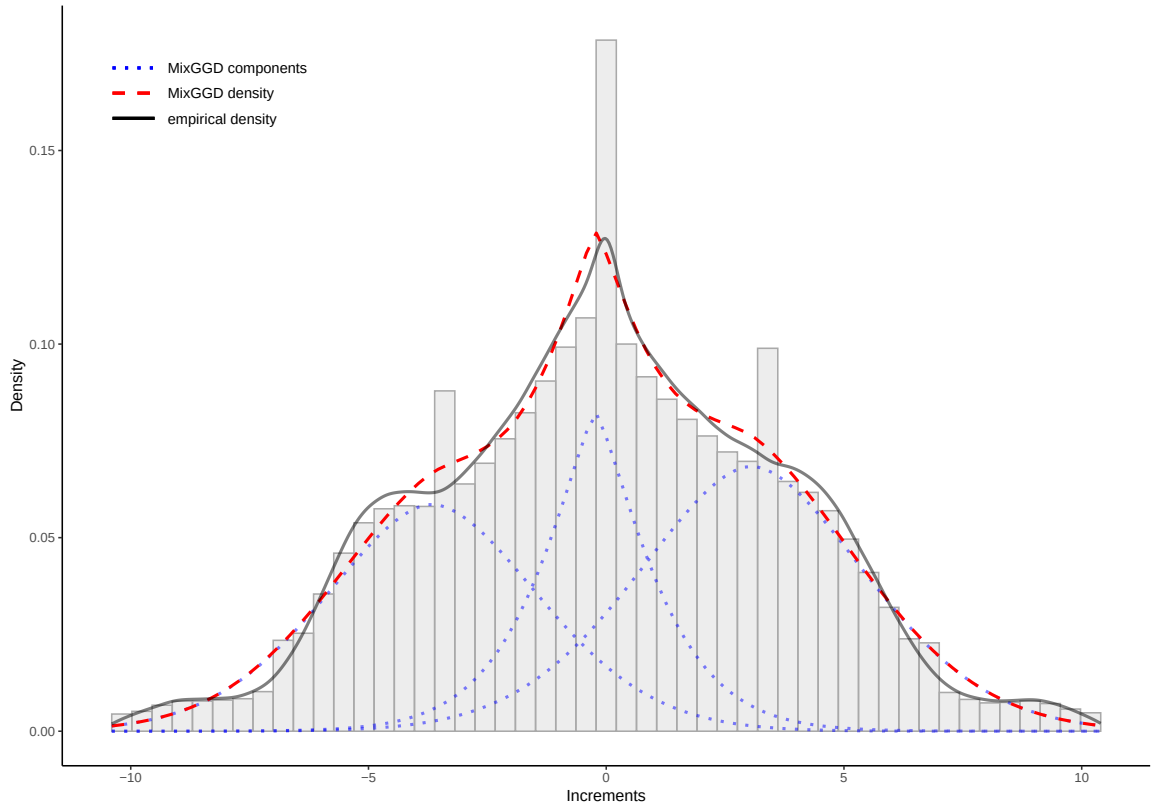
(a)


MixGGD components
MixGGD density
empirical density
0.08
0.06
0.04
0.02
0.00
Density
−10
0
10
Increments

(b)

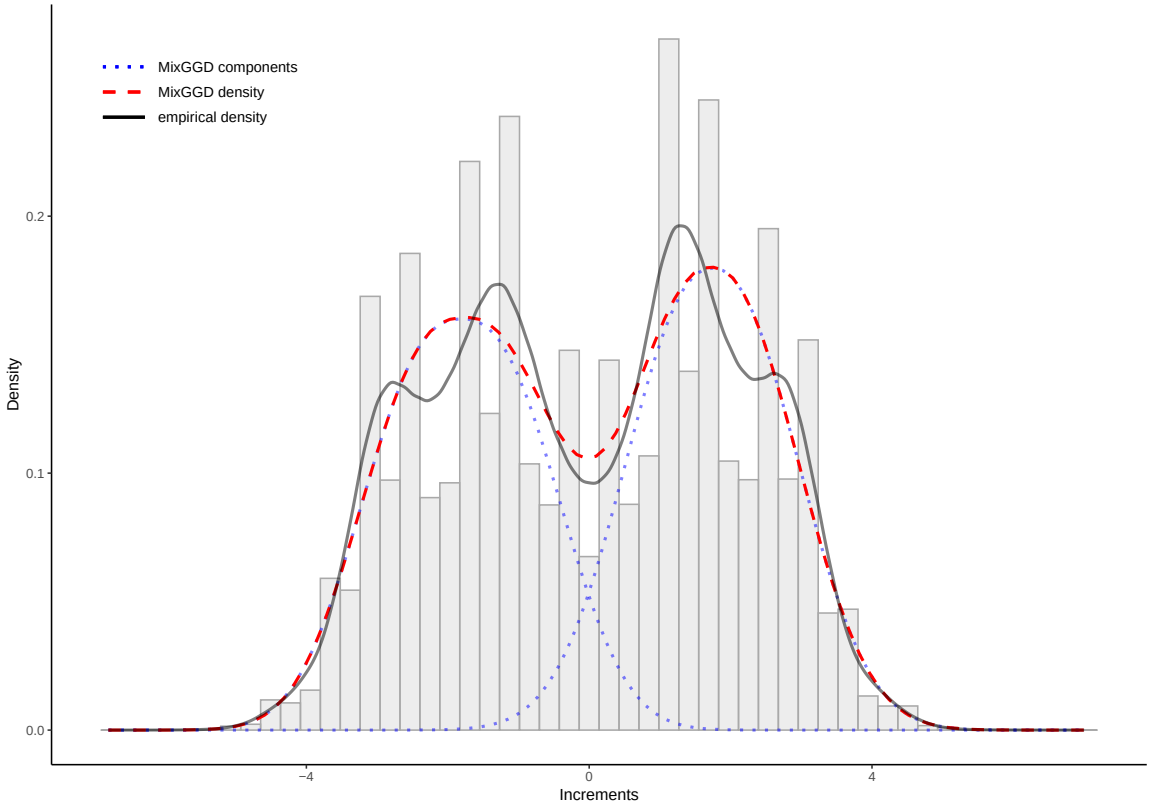


(c)

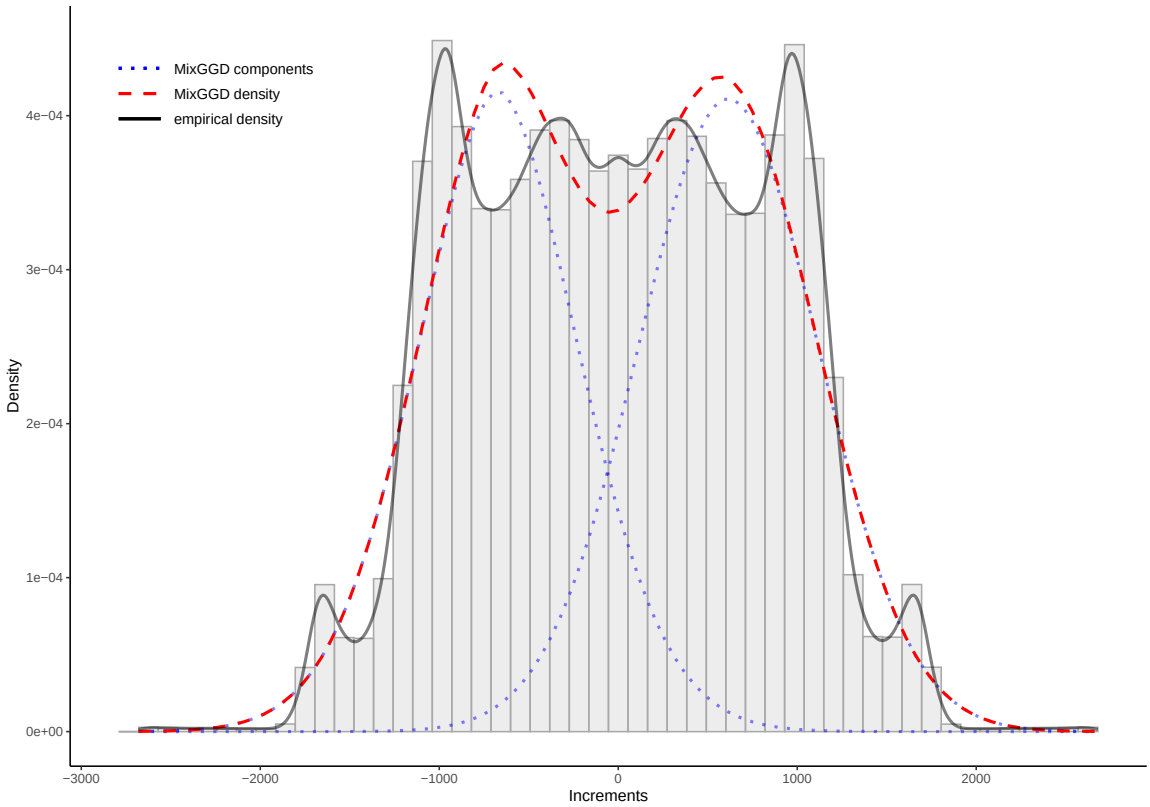


(d)

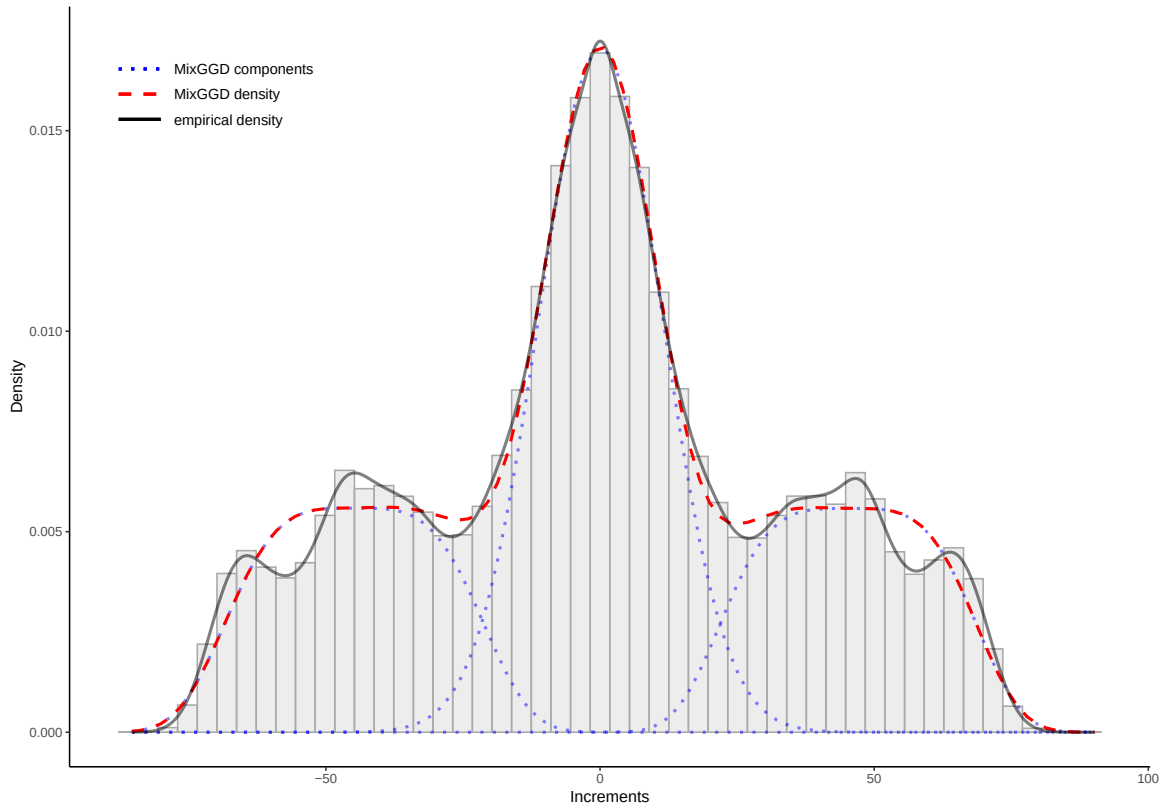
Figure III.4.11 Fitting a three-component MixGGD to EEG increments.


MixGGD components
MixGGD density
empirical density
Density
0.2
0.1
0.0
−4
0
4
Increments

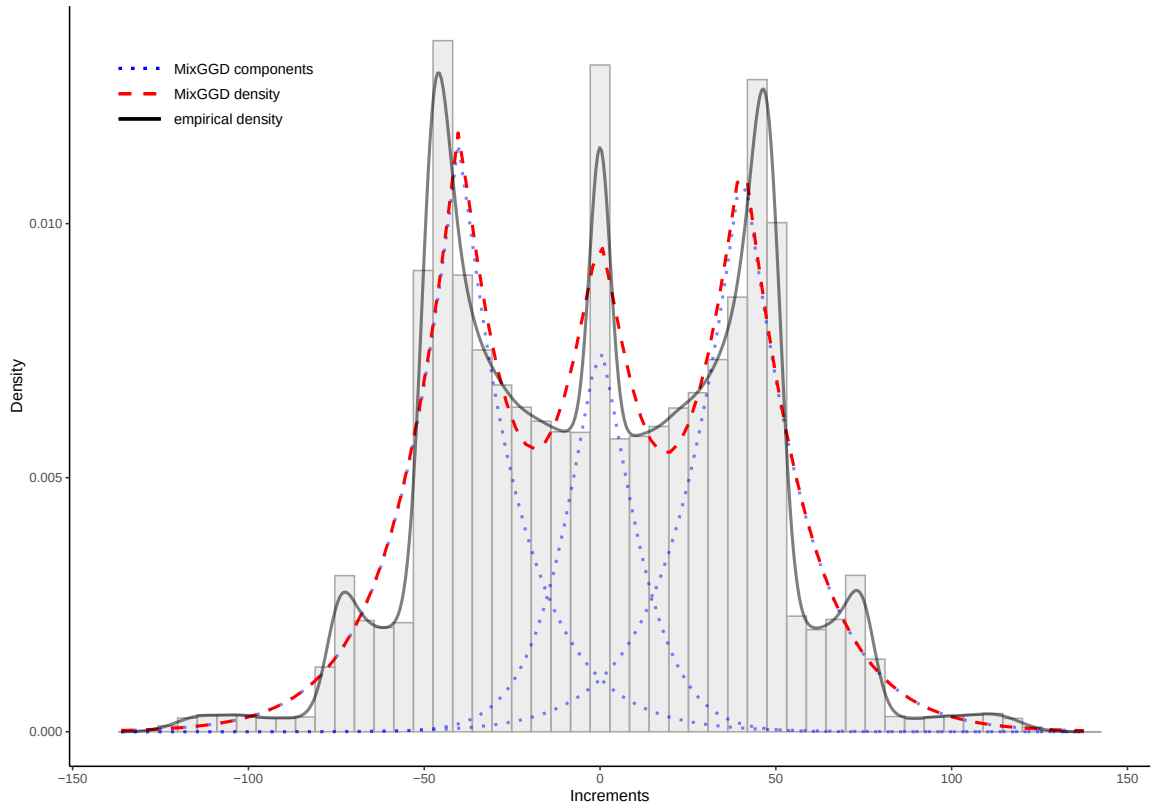
(a)


MixGGD components
MixGGD density
empirical density
Density
4e−04
3e−04
2e−04
1e−04
0e+00
−3000
−2000
−1000
0
1000
2000
Increments

(b)



(c)



(d)

Figure III.4.12 Fitting a MixGGD to EEG increments with more than three peaks.

4.2 Mixed generalized Gaussian distribution models

Similarly to Section 3.3, MixGGD parameters were investigated as potential predictors of age-appropriate scores of neurodevelopment and cognition at 6 months after the discharge from the hospital for cerebral malaria survivors. The 54 non-EEG features formed the same baseline model as in Section 3.3.1. To these features, the EEG parameters reflecting multimodality were added. First, the exact number of peaks (from 1 to 6) on each channel was added, treated as a categorical variable to reflect potentially non-linear relationship to the outcome. In the second group of models, the MixGGD parameters were treated as continuos or categorical variables.

As described in Section III.2, elastic net regression was used for the selection of predictors. Again, the response variable was the standardized neurodevelopment or cognitive score taken 6 months after the discharge from the hospital. Fitting of stochastic models with multimodal distributions would be warranted if the EEG parameters were shown to be important over and above other measures that can be obtained easily or as part of routine clinical care for cerebral malaria.

Features containing missing values from the non-EEG dataset had already been imputed for the analysis of unimodal estimate predictors. The same values (hence, the same feature matrix) were used in this analysis. Missing values from the EEG dataset (i.e. missing values of $\widehat{\mu_k}$, $\widehat{\sigma_k^2}$, $\widehat{s_k}$ and $\widehat{w_k}$) were imputed by hand using a "random-by-peak" method. This consisted of imputing random values from successful channels, i.e. for each missing case (channel), randomly choosing one channel where the estimation was successful (with respect to the number of peaks) and using those parameter estimates in place of missing values.

4.2.1 Combined non-EEG and number of peaks features models

The feature matrix for these models was a combination of aforementioned 54 non-EEG features and additional predictors containing the information about the number of peaks on each channel. First model (Model A) contained predictor variables with the exact number of peaks (from 1 to 6) on each channel. Second model (Model B) contained a five-category predictor variable with the number of peaks (0 for "degenerate" channels, 1, 2, 3 or more than 3 peaks). Third model (Model C) included a three-category predictor variable ("degenerate", unimodal and multimodal) and the fourth model (Model D) included a two-category predictor variable (unimodal vs. multimodal). In case of the categorical predictor variables, recoding into dummy variables was performed automatically within `caret` package [Kuhn, 2020] with the first category acting as the reference level.

4.2.2 Combined non-EEG and MixGGD features models

The feature matrix for these models consisted of aforementioned 54 non-EEG features and parameter estimates $\widehat{\mu_k}$, $\widehat{\sigma_k^2}$, $\widehat{s_k}$ and $\widehat{w_k}$ obtained by fitting a three-component MixGGD to EEG increments. Out of these parameter estimates, the first model (Model E) included predictors as a continuous variable of all parameter estimates, hence for each channel a total of 12 features were included as a result of fitting a three-component MixGGD. This meant that the channels displaying one or two peaks didn't have the values of parameter estimates of some components, as these components weren't naturally present in the channel. Those values were imputed by using the value 0 both as an indicator and as a natural selection for a parameter estimate of a missing component, as this value (with the exception of location parameter μ_k) wouldn't occur for any of the parameters in the MixGGD. Other models that were tested in the analysis used categorical variables in some form of cut off values, either for the shape parameters of the components or for all the parameter estimates. Predictors for the feature matrices of these models were formed as follows:

- Model F: three-category for $\widehat{s_k}$, $k = 1, 2, 3$ where values of 0 and 2 for the shape parameter s_k of each component were chosen as cut-offs,
- Model G: two-category for $\widehat{s_k}$, $k = 1, 2, 3$ where 0 was chosen as a cut-off value for s_k, defining a "peak not present/present in channel" category,
- Model H: two-category for $\widehat{s_1}, \widehat{s_3}$ and three-category for $\widehat{s_2}$ where "peak not present/present in channel" category was used for the first and third peak, and the cut-off values for the second peak were chosen to be 0, again indicating that the middle peak wasn't present in channel, and 2.5 which was chosen as a natural cut-off value based on the frequency of values of $\widehat{s_2}$ in all channels,
- Model I: two-category for all estimates, except $\widehat{\mu_k}$, $k = 1, 2, 3$ where 0 was selected as a cut-off value for $\widehat{\sigma_k^2}$, $\widehat{s_k}$ and $\widehat{w_k}$, $k = 1, 2, 3$ to indicate "peak not present/present in channel",
- Model J: two-category for all estimates, selecting 0 as a cut-off value for all $\widehat{\mu_k}$, $\widehat{\sigma_k^2}$, $\widehat{s_k}$ and $\widehat{w_k}$, $k = 1, 2, 3$.

4.3 Results and comparison

Comparison of the best models based on the leave-one-out cross validation RMSE is given in Table III.4.5

Models displayed similar RMSE values with the addition of categorizing the number of peaks in channels showing only negligible reduction in the RMSE, but with a higher number of non-zero coefficients compared to non-EEG features model. Note that estimating the actual parameters related to multimodality

Table III.4.5 Model characteristics.

Model features included (number of features, including dummy)	RMSE	Number of non-zero coefficients	Number of non-zero coefficients from EEG features subset
non-EEG features model (58)	0.5670	12	N/A
Model A (77)	0.5659	12	1
Model B (134)	**0.5530**	17	2
Model C (96)	0.5642	13	3
Model D (77)	0.5641	13	3
Model I (286)	0.5716	10	1
Model F (305)	0.5812	14	6
Model G (286)	0.5695	14	6
Model H (286)	0.5707	15	7
Model I (286)	0.5693	21	13
Model J (286)	0.5654	34	24

did not improve upon the model where the number of peaks was accounted for without the peak parameter estimates. It should be noted that models containing the number of peaks as predictors performed better in terms of the RMSE than the models containing MixGGD parameter estimates as predictors.

Chapter IV

Discussion and conclusion

The EEG findings on admission to the hospital have been used to predict mortality and morbidity following illness [Postels et al., 2018]. Separately, selected plasma biomarkers have been shown to be associated with cognitive impairment in paediatric severe malaria [Ouma et al., 2020]. Data on plasma and cerebrospinal fluid biomarkers were combined together with parameters of stochastic models for the EEG data to determine their usefulness for prediction of neurodevelopment and cognition 6 months following cerebral malaria illness. Previous analysis of these data showed that stochastic modelling of EEG features can better explain the variation in neurodevelopmental and cognitive outcomes of children who were affected by cerebral malaria [Veretennikova et al., 2018]. An overview of cerebral malaria and the use of electroencephalography is presented in Chapter I.

In Chapter II basic theory needed for the modelling of EEG increments was presented. Modelling was based on the construction of a stationary diffusion with a given marginal density. Based on the analysis of the histograms of increments, two distributions were chosen. The first approach in modelling of EEG increments was based on a unimodal generalized Gaussian distribution consisting of a light-tailed and a heavy-tailed subfamily. In Section 2.3.2 properties of the generalized Gaussian distribution were presented, along with the method of estimating the tail index for the heavy-tailed GGD subfamily. As some histograms displayed multimodal distributions, the second choice of the distribution was a mixture of non-central generalized Gaussian distributions, which is presented in Section II.3.

Chapter III focused on implementation of the presented theory to real world data. Firstly in Section III.3, the light-tailed GGD subfamily was fitted to EEG increments to create the features for prediction. The second stochastic model was based on a heavy-tailed distribution with the tail index estimated using the empirical scaling function. Additionally, the model containing non-EEG features (such as socio-demographic and anthropometric data and biomarker panels) was examined to see whether including the information from EEG signals can help in explaining of variation in neurodevelopmental and cognitive scores 6 months post-

coma. Results showed that the baseline neurodevelopmental score (taken right after coma) was the most important predictor of neurodevelopment at point 6 months after coma, which was expected. The rationale behind this comes from the fact that the baseline neurodevelopmental score is a direct measure of the outcome variable, i.e., neurodevelopmental score 6 months after coma, and thus measures the same attribute, but at a different time point. Other non-EEG features retained in the model generally overlap with the non-EEG features found to be important predictors in the analysis of Veretennikova et al. [2018] and mostly contain biomarker panels from cerebrospinal fluid and/or plasma. Also, negative values of the model coefficient for some features such as white blood cell count and IL-1α receptor level in plasma are in accordance with the intuitive assumption of increased inflammatory response negatively affecting child's development. However, this model was improved by the addition of EEG features.

The addition of EEG features from fitting of the GGD and estimation of the tail index resulted in an improved RMSE for both light-tailed and heavy-tailed stochastic models. Features from the non-EEG dataset which were retained in the elastic net as the important predictors for the neurodevelopment and cognition were also kept in the combined non-EEG and GGD features model. The additional feature that was selected in this case was the GGD parameter estimate σ^2 for the channel T3 (temporal electrode placed on the left side of the head). The model also had a slightly lower RMSE compared to the pure non-EEG model, meaning that the addition of stochastic features can improve the prediction of neurodevelopment and cognition.

Further improvement in the explanation of variation in neurodevelopment and cognition was achieved by introducing tail index estimates as features into the elastic net model. The tail index treated as a continuous feature failed to improve the model, and none of the EEG features were chosen by the model to be significant. However, categorization of the tail index based on distributional tertiles brought improvements and resulted in the model with the lowest RMSE. This means there is a threshold for the effect of the tail index value on neurodevelopmental or cognitive score. EEG feature retained in the predictive model was the tail index estimate on the channel T6 (temporal electrode placed on the right side of the head) as the dummy variable with level 2. Since level 1 (values of the tail index lower than 3.1) was chosen as the reference level, the interpretation behind it is that the tail index value above 3.1 and below 4.5 on the channel T6 has an increased negative influence on neurodevelopment and cognition compared to the tail index values of less than 3.1 on the same channel. The feature manifested in the 2$^{\text{nd}}$ highest ranked value of the coefficient in the final model which could indicate it's importance over majority of other non-EEG features.

Since the inclusion of parameter estimates from a unimodal distribution im-

proved the prediction of neurodevelopment, in Section III.4 the multimodal distribution was considered as a refinement for the previously used unimodal distribution. For this purpose, a multimodal diffusion model with the mixed generalized Gaussian distribution (MixGGD) as a marginal probability density function was constructed. The model was applied to the EEG data to determine the extent to which modelling multimodality may improve the results obtained from stochastic models with unimodal marginal distributions. As a compromise between precision in using the exact number of peaks and keeping the number of parameters which need to be estimated within reason, the MixGGD was limited to a total of three components, as there were less than 5% of channels which manifested in more than three peaks. Parameter estimation was performed using the Expectation/Conditional Minimization (ECM) algorithm, and the added shape parameter s was estimated using the approach of higher order statistics, where the value of the kurtosis is used to derive an approximation of the shape parameter s.

Several models with different sets of predictors were investigated to determine significant predictors of neurodevelopment at point 6 months after coma. Model containing non-EEG features (such as socio-demographic and anthropometric data and biomarker panels) was used as a benchmark to see whether including the information from EEG signals can help in explaining of variation in neurodevelopmental and cognitive scores 6 months post-coma. To test this, a total of ten combined (non-EEG and EEG features) models were investigated. One model consisted of non-EEG features and predictors with observed number of peaks in channels as a categorical variable with five levels. Another model consisted of non-EEG features and MixGGD features as continuos variables. Last two models consisted of non-EEG features and MixGGD features as categorical variables with a cut-off value for parameter estimates (except the location parameter) or all parameter estimates. Categorization of predictors was used to reduce the noise of the predictors and account for potential non-linearity in their relation to the neurocognitive outcome score.

The addition of EEG features resulted in an improvement of the RMSE in seven out of ten models, but the improvement was negligible. Also, none of the models which showed an improvement had a smaller number of non-zero coefficients, which means that an improvement in the RMSE is at the cost of a more complicated model, i.e. a model with a larger number of predictors. However, all of the models included at least one predictor from non-EEG dataset. The model with the lowest RMSE had the addition of the number of peaks as a categorical predictor with 5 levels (0 for "degenerate" channels, 1, 2, 3 or more than 3 peaks), where two predictors from this set of predictors were chosen as significant. Models that didn't show an improvement in the RMSE were the model with the MixGGD features as continuos predictors, thde model with MixGGD

features where the estimate of the shape parameter s_k was a categorical predictor with 3 levels and the model with MixGGD features where estimates of s_1 and s_3 were categorical with 2 levels and estimate of s_2 was categorical with 3 levels.

In general, the group of models which had the addition of features relating to the number of peaks on each of the channels showed an improvement in the RMSE compared both to the benchmark model and the combined non-EEG and MixGGD features models. This could mean that the information gained only from the visual investigation of EEG increments might bring an improvement in prediction of neurodevelopment, without the need for a more detailed classification of the underlying distribution and/or estimating the parameters of such distributions. However, investigating other marginal distributions appropriate for modelling of EEG signal increments should be performed to see whether an improvement in prediction can be made using a more adequate distribution.

In summary, the addition of stochastic EEG modelling can improve the prediction of children's brain function 6 months following coma. Further improvement can be made by investigating other marginal distributions appropriate for modelling of EEG signal increments (e.g., the family of generalized Pearson distributions considered in Cobb et al. [1983]). One modelling approach could be to use the convolutional neural networks (CNN). For an overview of the CNN techniques in analysis of EEG signal see Rajwal and Aggarwal [2023]. Another possible approach could be to incorporate spatial information in the form of random fields (see for e.g. Chung [2020], Gonzalez-Castillo et al. [2023], Ovalle-Muñoz and Ruiz-Medina [2022]). Additionally, time series models similar to the one introduced here can be investigated in other diseases that affect the brain and in electrical activity of other types of cells such as cells impacted by cancer and its treatment.

References

Alexander, C. and Narayanan, S. [2001], 'Option Pricing with Normal Mixture Returns: Modelling Excess Kurtosis and Uncertanity in Volatility'.

Bangirana, P., Opoka, R. O., Boivin, M. J., Idro, R., Hodges, J. S. and John, C. C. [2016], 'Neurocognitive domains affected by cerebral malaria and severe malarial anemia in children', *Learn. Individ. Differ.* **46**, 38–44.

Bibby, B. M., Skovgaard, I. M. and Sørensen, M. [2005], 'Diffusion-type models with given marginal distribution and autocorrelation function', *Bernoulli* **11**(2), 191–220.

Birbeck, G. L., Molyneux, M. E., Kaplan, P. W., Seydel, K. B., Chimalizeni, Y. F., Kawaza, K. and Taylor, T. E. [2010], 'Blantyre Malaria Project Epilepsy Study (BMPES) of neurological outcomes in retinopathy-positive paediatric cerebral malaria survivors: A prospective cohort study', *Lancet Neurol.* **9**(12), 1173–1181.

Bishwal, J. P. N. [2007], *Parameter Estimation in Stochastic Differential Equations*, Springer-Verlag, Berlin, Heidelberg.

Bradley, R. H. and Caldwell, B. M. [1979], 'Home observation for measurement of the environment'.

Brigo, D. [2008], 'The General Mixture Diffusion SDE and its Relationship with an Uncertain-volatility Option Model with Volatility-asset Decorrelation', *SSRN Electron. J.* .

Brigo, D. and Mercurio, F. [2002*a*], *Displaced and Mixture Diffusions for Analytically-Tractable Smile Models*, Springer, pp. 151–174.

Brigo, D. and Mercurio, F. [2002*b*], 'Lognormal-mixture dynamics and calibration to market volatility smiles', *Int. J. Theor. Appl. Financ.* **05**(04), 427–446.

Canty, A. and Ripley, B. D. [2021], *boot: Bootstrap R (S-Plus) Functions*.

Chenxu, L. [2013], 'Maximum-likelihood estimation for diffusion processes via closed-form density expansions', *Ann. Stat.* **41**(3), 1350–1380.

Chung, M. K. [2020], 'Introduction to Random Fields', pp. 425–458.

Cobb, L., Koppstein, P. and Chen, N. H. [1983], 'Estimation and Moment Recursion Relations for Multimodal Distributions of the Exponential Family', *J. Am. Stat. Assoc.* **78**(381), 124–130.

Cohn, D. L. [2013], *Measure Theory*, 2 edn, Springer New York, New York, NY.

Cover, T. M. and Thomas, J. A. [2005], *Differential Entropy*.

Dempster, A. P., Laird, N. M. and Rubin, D. B. [1977], 'Maximum Likelihood from Incomplete Data via the EM Algorithm on JSTOR', *J. R. Stat. Soc. Ser. B* **39**(1), 1–22.

Doukhan, P. [1994], *Mixing - properties and examples*, Springer.

Duncan, D., Talmon, R., Zaveri, H. P. and Coifman, R. R. [2013], 'Identifying preseizure state in intracranial EEG data using diffusion kernels', *Math. Biosci. Eng.* **10**(3), 579–590.

Dytso, A., Bustin, R., Poor, H. V. and Shamai, S. [2018], 'Analytical properties of generalized Gaussian distributions', *J. Stat. Distrib. Appl.* **5**(1), 6.

Embrechts, P., Klüppelberg, C. and Mikosch, T. [1997], *Modelling Extremal Events: for Insurance and Finance*, Vol. 33, Springer-Verlag, Berlin, Heidelberg, Berlin.

Engelbert, H. J. and Schmidt, W. [1985], 'On one-dimensional stochastic differential equations with generalized drift', *Stoch. Differ. Syst. Filter. Control* pp. 143–155.

Fournier, N. [2009], 'Simulation and approximation of Lévy-driven stochastic differential equations', *ESAIM - Probab. Stat.* **15**(2), 402–416.

Gonzalez-Castillo, J., Fernandez, I. S., Lam, K. C., Handwerker, D. A., Pereira, F. and Bandettini, P. A. [2023], 'Manifold learning for fMRI time-varying functional connectivity', *Front. Hum. Neurosci.* **17**(July).

Grahovac, D., Jia, M., Leonenko, N. and Taufer, E. [2015], 'Asymptotic properties of the partition function and applications in tail index inference of heavy-tailed data', *Statistics (Ber).* **49**(6), 1221–1242.

Hecht, E. and Stout, D. [2015], Techniques for Studying Brain Structure and Function, *in* E. Bruner, ed., 'Hum. Paleoneurology.', Springer International Publishing, Cham, pp. 209–224.

Henningsen, A. and Toomet, O. [2011], 'MaxLik: A package for maximum likelihood estimation in R', *Comput. Stat.* **26**(3), 443–458.

Heyde, C. C. and Leonenko, N. N. [2005], 'Student processes', *Adv. Appl. Probab.* **37**(2), 342–365.

Iacus, S. M. [2009], *Simulation and Inference for Stochastic Differential Equations: with R Examples*, Vol. 1, Springer.

Idro, R., Marsh, K., John, C. C. and Newton, C. R. J. [2011], 'Cerebral Malaria: Mechanisms Of Brain Injury And Strategies For Improved Neuro-Cognitive Outcome', *Pediatr. Res.* **68**(4), 267–274.

Johnson, O. and Vignat, C. [2007], 'Some results concerning maximum Rényi entropy distributions', *Ann. l'Institut Henri Poincare Probab. Stat.* **43**(3), 339–351.

Jurek, Z. J. [2001], 'Remarks on the selfdecomposability and new examples',

Demonstr. Math. **34**(2), 241–250.

Kaplan, A. Y., Fingelkurts, A. A., Fingelkurts, A. A., Borisov, S. V. and Darkhovsky, B. S. [2005], 'Nonstationary nature of the brain activity as revealed by EEG/MEG: methodological, practical and conceptual challenges', *Signal Processing* **85**(11), 2190–2212.

Karlin, S. and Taylor, H. M. [1975], *A first course in stochastic processes*, 2 edn, Academic Press.

Karlin, S. and Taylor, H. M. [1981], *A second course in stochastic processes*, 1 edn, Academic Press.

Kaufman, A. S. [2004], *Manual for the Kaufman Assessment Battery for Children*, Circle Pines, MN: AGS Publishing.

Kim, N., Heo, M., Fleysher, R., Branch, C. A. and Lipton, M. L. [2014], 'A Gaussian Mixture Model Approach for Estimating and Comparing the Shapes of Distributions of Neuroimaging Data: Diffusion-Measured Aging Effects in Brain White Matter', *Front. Public Heal.* **2**.

Kirch, C., Muhsal, B. and Ombao, H. [2015], 'Detection of Changes in Multivariate Time Series With Application to EEG Data', *J. Am. Stat. Assoc.* **110**(511), 1197–1216.

Kloeden, P. E. and Platen, E. [1992], *Numerical Solution of Stochastic Differential Equations*, Springer-Verlag, Berlin, Heidelberg.

Klonowski, W. [2009], 'Everything you wanted to ask about EEG but were afraid to get the right answer', *Nonlinear Biomed. Phys.* **3**(1), 2.

Kuhn, M. [2020], *caret: Classification and Regression Training*.
URL: *https://cran.r-project.org/package=caret*

Leonenko, N. N., Salinger, Ž., Sikoriskii, A., Šuvak, N. and Boivin, M. J. [2023*a*], 'Generalized Gaussian time series model for increments of EEG data', *Stat. Interface* **16**(1), 17–29.

Leonenko, N. N., Salinger, Ž., Sikoriskii, A., Šuvak, N. and Boivin, M. J. [2023*b*], 'Multimodal diffusion model for increments of electroencephalogram data', *Stoch. Environ. Res. Risk Assess.* **37**(12), 4695–4706.

Leonenko, N. N. and Šuvak, N. [2010], 'Statistical inference for student diffusion process', *Stoch. Anal. Appl.* **28**(6), 972–1002.

Levin, D., Peres, Y., Wilmer, E. [2017], *Markov Chains and Mixing Times*, 2 edn, American Mathematical Society.

Lutwak, E., Yang, D. and Zhang, G. [2004], 'Moment-entropy inequalities', *Ann. Probab.* **32**(1B), 757–774.

Lutwak, E., Yang, D. and Zhang, G. [2007], 'Moment-entropy inequalities for a random vector', *IEEE Trans. Inf. Theory* **53**(4), 1603–1607.

Mandal, S., Sarkar, R. R. and Sinha, S. [2011], 'Mathematical models of malaria - a review', *Malar. J. (part Springer Nature)* **10**(202), 124–130.

Meng, X.-L. and Rubin, D. B. [1993], 'Maximum Likelihood Estimation via the ECM Algorithm: A General Framework', *Biometrika* **80**(2), 267.

Mohamed, O. M. M. and Jaïdane-Saïdane, M. [2009], On the parameters estimation of the Generalized Gaussian mixture model, *in* 'Eur. Signal Process. Conf.', number Eusipco, pp. 2273–2277.

Nadarajah, S. [2005], 'A generalized normal distribution', *J. Appl. Stat.* **32**(7), 685–694.

Øksendal, B. [2000], *Stochastic Differential Equations*, 5 edn, Springer.

Ouma, B. J., Ssenkusu, J. M., Shabani, E., Datta, D., Opoka, R. O., Idro, R., Bangirana, P., Park, G., Joloba, M. L., Kain, K. C., John, C. C. and Conroy, A. L. [2020], 'Endothelial activation, acute kidney injury, and cognitive impairment in pediatric severe malaria', *Crit. Care Med.* **48**(9), 734–743.

Ovalle-Muñoz, D. P. and Ruiz-Medina, M. D. [2022], 'LRD spectral analysis of multifractional functional time series on manifolds'.

Ozaki, T. [1985], 2 Non-linear time series models and dynamical systems, *in* 'Time Ser. Time Domain', Vol. 5 of *Handbook of Statistics*, Elsevier, pp. 25–83.

Papoulis, A. and Pillai, U. S. [2002], *Probability, random variables, and stochastic proceses*, 4 edn, McGraw-Hill.

Patel, A. A., Jannati, A., Dhamne, S. C., Sapuwa, M., Kalanga, E., Mazumdar, M., Birbeck, G. L. and Rotenberg, A. [2020], 'EEG markers predictive of epilepsy risk in pediatric cerebral malaria – A feasibility study', *Epilepsy Behav.* **113**.

Piryatinska, A., Terdik, G., Woyczynski, W. A., Loparo, K. A., Scher, M. S. and Zlotnik, A. [2009], 'Automated detection of neonate EEG sleep stages', *Comput. Methods Programs Biomed.* **95**(1), 31–46.

Pogány, T. K. and Nadarajah, S. [2010], 'On the characteristic function of the generalized normal distribution', *Comptes Rendus Math.* **348**(3), 203–206.

Postels, D. G., Wu, X., Li, C., Kaplan, P. W., Seydel, K. B., Taylor, T. E., Kousa, Y. A., Idro, R., Opoka, R., John, C. C. and Birbeck, G. L. [2018], 'Admission EEG findings in diverse paediatric cerebral malaria populations predict outcomes', *Malar. J.* **17**(1), 208.

R Core Team [2019], *A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

Rajwal, S. and Aggarwal, S. [2023], 'Convolutional Neural Network-Based EEG Signal Analysis: A Systematic Review', *Arch. Comput. Methods Eng.* **30**(6), 3585–3615.

Robin, G., Mayer, I. and Sportisse, A. [2021], How to impute missing values?, Technical report.
URL: *https://rmisstastic.netlify.app/how-to/impute/missimp*

Rudin, W. [1976], *Principles o mathematical analysis*, 3 edn, McGraw-Hill.

Sanei, S. [2013], *Adaptive processing of brain signals*, 1 edn, John Wiley & Sons.

Schilling, R. L., Song, R. and Vondracek, Z. [2012], 'Bernstein Functions', *Bernstein Funct. .*

Shank, L. [2011], Mullen Scales of Early Learning, *in* 'Encycl. Clin. Neuropsychol.', Springer New York, New York, NY, pp. 1669–1671.

Skorokhod, A. V. [1989], *Asymptotic Methods in the Theory of Stochastic Differential Equations*, American Mathematical Society.

Stekhoven, D. J. and Buhlmann, P. [2012], 'MissForest–non-parametric missing value imputation for mixed-type data', *Bioinformatics* **28**(1), 112–118.

Taufer, E. and Leonenko, N. [2009], 'Simulation of Lévy-driven Ornstein-Uhlenbeck processes with given marginal distribution', *Comput. Stat. Data Anal.* **53**(6), 2427–2437.

Taylor, T. E. [2009], 'Caring for children with cerebral malaria: insights gleaned from 20 years on a research ward in Malawi', *Trans. R. Soc. Trop. Med. Hyg.* **103**(Supplement_1), S6–S10.

Temko, A., Thomas, E., Marnane, W., Lightbody, G. and Boylan, G. [2011], 'EEG-based neonatal seizure detection with Support Vector Machines', *Clin. Neurophysiol.* **122**(3), 464–473.

Tesei, A. and Regazzoni, C. S. [1996], Use of fourth-order statistics for non-gaussian noise modelling: the generalized gaussian pdf in terms of kurtosis, *in* '1996 8th Eur. Signal Process. Conf. (EUSIPCO 1996)', IEEE.

Tesei, A. and Regazzoni, C. S. [1998], 'HOS-based generalized noise pdf models for signal detection optimization', *Signal Processing* **65**(2), 267–281.

Varanasi, M. K. and Aazhang, B. [1989], 'Parametric generalized Gaussian density estimation', *J. Acoust. Soc. Am.* **86**(4), 1404–1415.

Veretennikova, M. A., Sikorskii, A. and Boivin, M. J. [2018], 'Parameters of stochastic models for electroencephalogram data as biomarkers for child's neurodevelopment after cerebral malaria', *J. Stat. dstributions Appl.* **5**(1), 1–12.

Wanduku, D. [2019], 'The stochastic extinction and stability conditions for non-linear malaria epidemics', *Math. Biosci. Eng.* **16**(5), 3771–3806.

Wen, L., Qiu, Y., Wang, M., Yin, J. and Chen, P. [2022], 'Numerical characteristics and parameter estimation of finite mixed generalized normal distribution', *Commun. Stat. - Simul. Comput.* **51**(7), 3596–3620.

WHO [2009], 'Wold Health Organization Growth Standards.'.
URL: *https://www.who.int/growthref/en/.*

Wickham, H. [2016], *ggplot2: Elegant Graphics for Data Analysis*, Springer-Verlag New York.
URL: *https://ggplot2.tidyverse.org*

Wilson, J. A. and Nordal, H. J. [2013], 'EEG in connection with coma', *Tidsskr. den Nor. lægeforening Tidsskr. Prakt. Med. ny række* **133**(1), 53–57.

Wolfram Research, Inc. [n.d.], 'Mathematica, Version 11.3'.

Young, G. B. [2000], 'The EEG in coma', *J. Clin. Neurophysiol.* **17**(5), 473–485.

Zou, H. and Hastie, T. [2005], 'Regularization and variable selection via the elastic net', *J. R. Stat. Soc. Ser. B (statistical Methodol.* **67**(2), 301–320.